Functional characterization of Ysh1p, the yeast endonuclease involved in 3' end processing and in transcription termination of RNA polymerase II transcripts

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Monika Garas

aus Polen

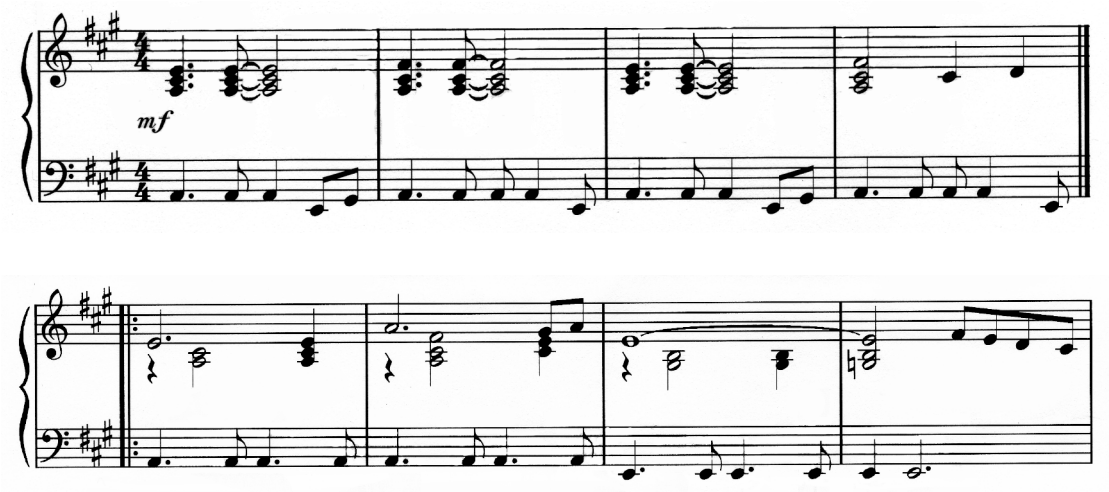
Basel, 2008

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel auf Antrag von der Herrn Professoren Walter Keller und Witold Filipowicz

Basel, den 22. Mai 2007

Prof. Dr. Hans-Peter Hauri

Dekan der Philosophisch-
Naturwissenschaftlichen Fakultät



Dla mojej Mamy i mojego Taty

Dedicated to my Mum and Dad

Acknowledgements

I wish to thank...

Walter Keller for giving me the opportunity and freedom to carry out this work in his laboratory, and for his empathy and support in scientific and non-scientific matters.

Witold Filiowicz for being the co-referee of this Ph.D. thesis.

My lab colleagues, both former and present, for sharing advice, experience and a good laughter. Bertrand Paguet, Salome Roeck and Andrea Kyburz for sharing good and bad times on the way to our Ph.D.s. Stepanka Vanacova for her ideas, discussions and encouragements. Bernhard Dichtl for his supervision in the initial phase of the projects, and helpful advices and discussions.

Moim kochanym rodzicom, braciom i całej rodzinie, za ich nieskończoną miłość i wsparcie, za ich ispirację i wiarę w moje nieograniczone możliwości. My parents, brothers, and family, for their endless love and support, for being my inspiration and believing only 'the sky is the limit'.

All my wonderful friends, from here and there, now and then, for all the joy, sorrow, fascinations and music we shared on the way... I cherish having you in my life. Special thank you to D.W.K., M.P.L., D.K., T.L., S.V., K.M., A.S., B.K. Friends in need are friends indeed.

Table of contents

Summary

Eukaryotic RNA polymerase II (RNAP II) is involved in the synthesis of two major classes of transcripts: messenger RNAs (mRNAs) and small nuclear and small nucleolar RNAs. In order to be biologically functional, primary transcripts of RNAP II require extensive processing and modifications. Biogenesis of mature mRNAs involves capping at the 5' end, splicing out of the introns and poly(A) tail addition at the 3' end. Only correctly processed mRNAs can be exported to the cytoplasm where they act as templates for protein translation.

Eukaryotic pre-mRNA 3' end formation is initiated by endonucleolytic cleavage at the poly(A) site, followed by polyadenylation of the upstream cleavage product. In contrast, small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) precursors are cleaved at their 3' ends, but in their mature form they are not polyadenylated. The seemingly simple reactions of 3' end cleavage and polyadenylation are nevertheless performed by surprisingly complex protein machineries. In yeast, the pre-mRNA 3' end processing apparatus consists of cleavage and polyadenylation factor (CPF), cleavage factor IA (CF IA) and cleavage factor IB (CF IB; reviewed in Zhao et al., 1999). The complexity of the 3' end processing machinery is in part due to the necessity of precise RNA sequence recognition and also to the regulation in a wider transcriptional context. Both the exact mechanism of 3' end processing, and many of the factors involved in these reactions exhibit a high level of similarity between metazoans and yeast. Cleavage and polyadenylation factors are co-transcriptionally recruited to the carboxy-terminal domain (CTD) of RNAP II and together with the *cis*-acting 3' end processing signals are required for transcription termination on mRNA genes (reviewed in Buratowski, 2005; Proudfoot, 2004).

The original aim of this thesis was the identification and characterization of the yeast endonuclease involved in pre-mRNA 3' end processing. Whereas it has long been known that the poly(A) tails of mRNA are synthesized by poly(A) polymerase, the endonucleolytic activity involved in 3' end cleavage remained enigmatic for many years. Therefore, in the beginning of this work we assigned putative endonucleolytic activity to the yeast CPF subunit Ysh1p/Brr5p and to its archaean homologue, *M. jannaschii* MJ1236, based on highly conserved metallo-β-lactamase and β-CASP domains present in these factors.

Very little has been known about Ysh1/Brr5 protein and its role within the 3' end processing machinery. We found that the conserved metallo-β-lactamase motif present in Ysh1p/Brr5p is essential for yeast viability *in vivo*, as any mutation within its conserved β-lactamase signature HXHXDH is detrimental to the cell. Although this fact underscored the functional importance of the metallo-β-lactamase motif in Ysh1p/Brr5p, it hampered further attempts to analyze the effects of such mutations. Moreover, biochemical assignment of a potential enzymatic activity to this factor *in vitro* was virtually impossible, as recombinant Ysh1p/Brr5p alone neither bound to RNA nor exhibited any nucleolytic activity. Consistently, specific cross-linking of the yeast 3' end processing factors to the poly(A) site did not identify Ysh1p/Brr5p as the factor present at the cleavage site. Therefore, to better understand

the role of Ysh1p/Brr5p in pre-mRNA 3' end formation, we generated a series of conditional mutants of *YSH1*. Analysis of several temperature- and cold- sensitive *ysh1* alleles revealed several important features of Ysh1p/Brr5p in different aspects of RNA processing and their coupling to RNAP II transcription termination and splicing. Firstly, we showed that Ysh1p/Brr5p is generally required for 3' end cleavage and polyadenylation as well as for poly(A) site selection of *ACT1* pre-mRNA. Interestingly, RNAP II transcription termination defects on a plasmid-borne CYC1 gene were observed in *ysh1* mutant strains. Northern blot analysis of steady-state RNA extracted from *ysh1-12* mutant cells detected read-through transcripts on several endogenous mRNA genes, confirming the general requirement of Ysh1p/Brr5p for transcription termination. Secondly, a significant proportion of RNAP II molecules failed to terminate transcription properly on *SNR3* snoRNA gene locus in *ysh1-12* mutant and extended transcripts produced from several snoRNA genes accumulated in this strain, pointing towards the involvement of Ysh1p/Brr5p in snoRNA 3' end formation. Furthermore, we showed that Ysh1p/Brr5p is involved in the regulation of *NRD1* mRNA levels. Interestingly, mutations in *ysh1-12* strain resulted in splicing defects on mRNA and snoRNA genes, thus suggesting a function for Ysh1p/Brr5p in coupling of pre-mRNA 3' end formation and splicing reactions in *S. cerevisiae*.

In addition, we analyzed functions of Syc1p, a new yeast 3' end processing subunit, which exhibits a high level of homology to the C-terminus of Ysh1p/Brr5p. Syc1p has possible regulatory functions in pre-mRNA 3' end formation and possibly links the processing machinery to other nuclear events.

Last not least, we carried out *in vitro* analyses of the recombinant *M. jannaschii* protein MJ1236, which is homologous to the β-lactamase and β-CASP domains of Ysh1p/Brr5p. Intriguingly, MJ1236 possesses also a KH-RNA binding domain, thus further suggesting a potential function of this factor in RNA metabolism. Heterogeneous expression and assaying of MJ1236 revealed its endonucleolytic activity on CYC1, ADH1 and GAL7 RNA substrates *in vitro*. This finding strongly implied the same type of hydrolyzing activity for its *S. cerevisiae* homologue Ysh1p/Brr5p.

Only recently the pre-mRNA 3' end endonucleolytic activity has been assigned to CPSF73, subunit of the mammalian 3' end processing machinery, as based on its crystal structure and *in vitro* activity (Mandel et al., 2006). Because of its high level of homology to CPSF73, Ysh1p/Brr5p is now generally believed to be the 3' end processing endonuclease in *S. cerevisiae*.

This thesis is a record of a fascinating yet sometimes frustrating quest towards identification of the yeast pre-mRNA 3' end processing endonuclease and understanding its functions in a wider transcriptional context.

CHAPTER 1

GENERAL INTRODUCTION

1.1 Biochemistry of pre-mRNA 3' end cleavage and polyadenylation

Cis –acting signals involved in pre-mRNA 3' end formation in *S. cerevisiae*

Mammalian *cis*-acting polyadenylation signals

Plant *cis*-acting polyadenylation signals

Comparison of eukaryotic polyadenylation signals

Trans-acting 3' end processing factors of *S. cerevisiae*

Mammalian 3' end processing factors

1.2 3' end formation of small nuclear and small nucleolar RNA transcripts

Cis-acting sequences required for snRNA and snoRNA 3' end formation and transcription termination in *S. cerevisiae*

Yeast *trans*-acting factors involved in 3' end formation and in transcription termination of snRNAs and snoRNAs

1.3 Evolutionary comparison of 3' end processing of different RNA transcripts

1.4 Coupling of the processing events to RNA polymerase II transcription

The CTD of RNAP II orchestrates the coupling of pre-mRNA processing to transcription

Capping is tightly coupled to RNAP II transcription

Splicing and transcription

Transcription and 3' end processing are interdependent events

Transcription termination and 3' end processing

Coupling of snRNA and snoRNA 3' end formation to transcription termination

Histone 3' end processing and termination

Coupling mRNA export with transcription, splicing and pre-mRNA processing

1.1 Biochemistry of pre-mRNA 3' end cleavage and polyadenylation

Eukaryotic pre-mRNA transcripts require 3' end processing in order to become functional. Generally, this involves an endonucleolytic cleavage at a specifically determined site, followed by addition of a poly(A) tail. The 3' end formation event is tightly coupled to transcription and promotes transcription termination and export of the mRNA from the nucleus. Moreover, the acquisition of a poly(A) tail provides binding sites for poly(A) binding proteins, and thus enhances the translation and stability of mRNA. 3' end formation is an essential step in pre-mRNA maturation, as any defects in this process can be detrimental to cell viability, growth and development. Formation of mRNA 3' ends serves as a key regulatory step in the expression of many genes, and in some cases aberrant polyadenylation leads to disease (Higgs et al., 1983; Lin et al., 1998), thus indicating the crucial role of 3' end processing in normal cell physiology.

Although cleavage and polyadenylation are seemingly simple catalytic reactions, a complex multisubunit machinery is required to perform them (reviewed in: Keller and Minvielle-Sebastia, 1997; Wahle and Rüegsegger, 1999; Zhao et al., 1999). In large part, this complexity is due to the need of specific poly(A) site recognition and to the regulation on transcriptional level. In *S. cerevisiae*, the 3' end processing machinery consists of cleavage and polyadenylation factor CPF, cleavage factor IA (CF IA) and cleavage factor IB (CF IB), and it recognizes specific *cis*-acting signals on the pre-mRNA transcript that define the correct poly(A) site position. Specific endonucleolytic cleavage is then catalyzed by the 3' end processing endonuclease Ysh1p, which in turn generates a 3'OH end for the poly(A) polymerase Pap1p that synthesizes the poly(A) tail. Moreover, poly(A) binding proteins Pab1p and Nab2p are necessary to achieve the correct length of the poly(A) tail. The yeast 3' end formation steps are shown schematically in Figure 1.1.

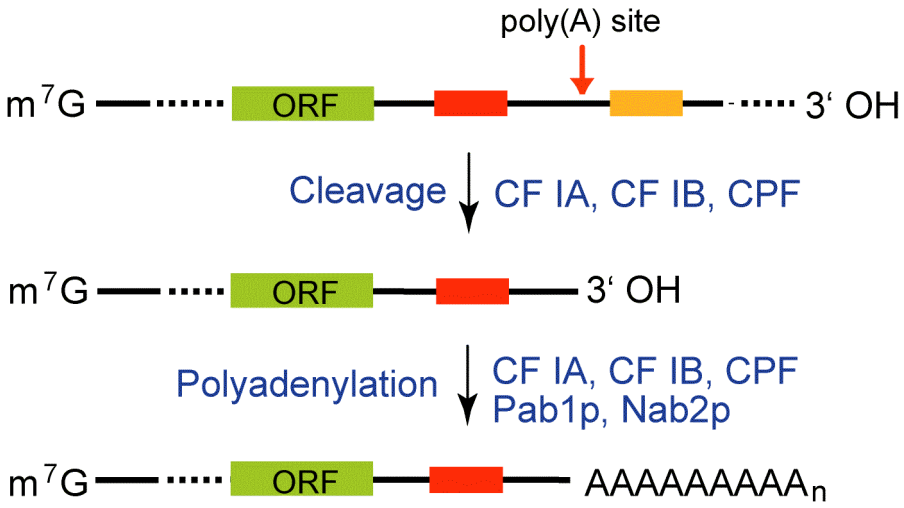


Figure 1.1 The yeast pre-mRNA 3’ end processing steps

The pre-mRNA transcript is cleaved endonucleolytically at the poly(A) site, followed by the addition of adenylate residues to the 3' end of the upstream fragment to form a poly(A) tail. The factors responsible for each step of the reaction are indicated on the right.

Cis–acting signals involved in pre-mRNA 3’ end formation in *S. cerevisiae*

Cis-acting sequences on the pre-mRNA precursors ultimately determine the processing efficiency in a given cellular environment. Polyadenylation signals directing pre-mRNA 3’ end processing in *Saccharomyces cerevisiae* are degenerated and less conserved than those of higher eukaryotes. The exact mechanism of poly(A) site recognition in yeast is rather poorly understood. Yeast polyadenylation signals were initially defined through an analysis of a mutant *CYC1* allele, *cyc1-512,* that has a severe 3′ end formation defect *in vivo* (Zaret and Sherman, 1982). The 3′ UTR region of *cyc1-512* allele carries a deletion of elements that were also required for processing of synthetic RNA *in vitro* (Butler and Platt, 1988). These elements were therefore termed efficiency element (EE) and positioning element (PE). Equivalent elements in many other yeast pre-mRNAs were subsequently described (reviewed in Zhao et al., 1999). Efficiency elements (EE) are found upstream of the cleavage site and often contain alternating UA dinucleotides or U-rich stretches (Graber et al., 1999b; Guo and Sherman, 1996). EE functions in activating the positioning element (PE). PE, the positioning element, directs endonucleolytic cleavage to a position approximately 20 nucleotides downstream of this sequence. Typically, PE is an adenine-rich sequence; approximately 13% of all analyzed yeast expressed sequence tags (EST) contain the AAUAAA hexamer (Graber et al., 1999b). However, the efficiency and positioning elements are not only degenerate but also redundant, and most yeast polyadenylation signals are more complex than that. Deletion or mutation of EE and PE motifs in several yeast genes has only slight or no effect on 3’ end processing of corresponding pre-mRNAs. Notably, the U-rich elements

encompassing the poly(A) site of the *CYC1* and *ADH1* pre-mRNAs are essential for correct 3′ end formation *in vivo*. Consistent with the *in vivo* results, mutation of sequences upstream and downstream of the poly(A) site affected 3′ end formation *in vitro* (Dichtl and Keller, 2001). These elements act in concert with the poly(A) site to produce multiple recognition sites for the processing machinery, since combinations of mutations within these elements were most effective in cleavage inhibition. Moreover, introduction of a U-rich element downstream of the *GAL7* poly(A) site strongly enhanced cleavage, underscoring the importance of downstream sequences in general (Dichtl and Keller, 2001). In contrast, the efficiency and positioning elements were dispensable for poly(A) site recognition within the short CYC1 substrate *in vitro*. Consistently, analysis of the polyadenylation signals present in the *URA4* gene of the fission yeast *Schizosaccharomyces pombe* revealed that three elements were important for 3' end processing: two site-determining elements upstream of the poly(A) sites and an efficiency element downstream of the poly(A) sites (Humphrey et al., 1994). The poly(A) site itself, which is in most cases $Py(A)_n$, also acts as a polyadenylation signal (Heidmann et al., 1992). Due to the convergent transcription of closely packed genes in *S. cerevisiae*, some polyadenylation signals function in both orientations (Aranda et al., 1998; Irniger et al., 1991).

Several proteins involved in pre-mRNA cleavage are believed to mediate poly(A) site recognition. RNA-binding analyses with the CPF subunit Ydh1p/Cft2p showed that interactions of this factor with the poly(A) site region determine the cleavage site (Dichtl and Keller, 2001). These results underscore the important function of CPF in the recognition of the poorly conserved yeast poly(A) site sequences.

Mammalian polyadenylation signals

The mammalian 3′ end formation machinery selects the poly(A) site through recognition of three mayor signals: the highly conserved hexanucleotide AAUAAA found 10 to 30 nucleotides upstream of the cleavage site, a moderately conserved U-rich or GU-rich element located downstream of the cleavage site, and the cleavage site itself which becomes the site of poly(A) tail synthesis. Additional sequences outside of this core element recruit regulatory factors or maintain the core signal in an open and accessible structure (reviewed in Zhao et al., 1999).

The conserved hexanucleotide sequence AAUAAA is present in a majority of polyadenylated mRNAs of animal cells and was initially discovered by a comparison of nucleotide sequences preceding the poly(A) sites in several mRNAs (Manley et al., 1985). The AAUAAA motif and its variants are essential for both cleavage and poly(A) addition. The downstream sequence element (DSE) is found within approximately 30 nucleotides downstream of the poly(A) site and can affect the cleavage site position and the efficiency of cleavage. DSE is more diffuse and poorly conserved and is present in two types, as a U-rich element or a GU-rich element. For most of the mammalian genes, the poly(A) site is defined by a CA dinucleotide and adenosine is found at the cleavage site of 70% of vertebrate mRNAs (Sheets et al., 1990). However, the overall sequence surrounding the cleavage site is

not conserved. Moreover, additional auxiliary sequence elements can modulate the efficiency of 3' processing in a positive or negative fashion, including the enhancer sequence of the U-rich upstream sequence element (USE). Interactions of the 3' end processing machinery with the polyadenylation signals on pre-mRNAs are mediated by CPSF-160 (Murthy and Manley, 1995) and CstF-64 proteins (MacDonald et al., 1994). Interactions of CPSF-160 with CstF-77 and other CstF subunits bridge CPSF and CstF factors across the poly(A) site to form a stable ternary complex on the pre-mRNA (Murthy and Manley, 1995).

Plant polyadenylation signals

The formation of pre-mRNA 3' ends in plants is rather poorly understood. In general, plant polyadenylation signals are more similar to the yeast ones rather than to those of animals, with both common content and arrangement of the signal elements. Upstream efficiency elements, upstream U-rich elements (UUE) and A-rich PE elements are similar between plants and yeast (Rothnie, 1996). Interestingly, the optimal yeast UE (UAUAUA) occurs both in *Oryza sativa* and *Arabidopsis thaliana,* with a positional distribution that peaks between 20 and 30 nucleotides upstream of the poly(A) site (Graber et al., 1999a). Moreover, U-rich sequences dominate the last 100 nucleotides upstream of the cleavage site in plants. Similar to other organisms, cleavage often occurs at a Py(A) dinucleotide. There are multiple cleavage sites in many genes and the use of a particular site is determined predominantly by the position of UE.

Comparison of eukaryotic polyadenylation signals

The mechanisms of pre-mRNA cleavage site recognition by the yeast, mammalian, and plant 3′ end processing complexes are highly conserved. An extensive analysis of polyadenylation signals from various organisms suggests that the composition and spatial organization of different elements are conserved rather than identical sequences (Graber et al., 1999a; Graber et al., 1999b; van Helden et al., 2000). The overall efficiency of poly(A) site selection is a function of all sequence *cis*-elements acting in concordance, as no single sequence element is universally required for processing (Graber et al., 1999a). Therefore, conservation of the hexanucleotide AAUAAA changes between species and is especially weak in yeast and plants. Consequently, the *cis*-acting elements in yeast resemble more the polyadenylation signal of plants than that of mammalian pre-mRNAs. For instance, the plant and yeast poly(A) sites are both flanked with upstream and downstream U-rich sequence elements (UUE, DUE), in contrast to the mammalian situation, where no U-rich sequences preceding poly(A) sites are found. Cleavage occurs preferentially at CA in mammals and at Py(A) in yeast and plants. Moreover, efficiency elements in both yeast and plants are similar. The U-rich sequences in mammals (DSE) are positioned more distant to the poly(A) site compared to the U-rich sequences (DUE) that follow the poly(A) site of yeast and plants. In summary, a common minimal polyadenylation signal in eukaryotes consists of an A-rich sequence

element, a U-rich element and a $Py(A)_n$ poly(A) site. Schematic comparison of the different 3' end processing signals in mammals, yeast and plants is presented in Figure 1.2.

Importantly, the recognition of yeast poly(A) signals is mediated by Yhh1p and Rna15p, whereas in mammals this step is achieved by their respective homologues CPSF160 and CstF-64 (Dichtl and Keller, 2001; MacDonald et al., 1994; Murthy and Manley, 1995). Thus, the topological arrangement not only of the *cis*-acting elements, but also of the *trans*-acting 3' end processing factors across the poly(A) site are conserved between yeast and mammals.

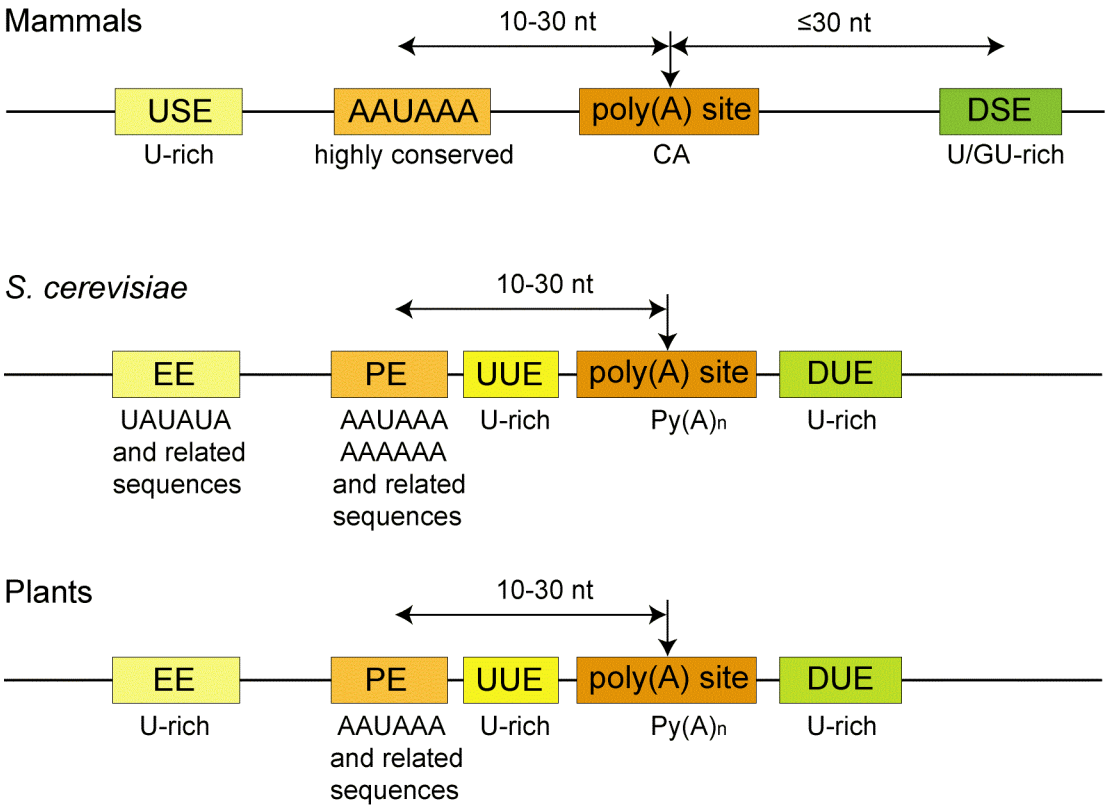


Figure 1.2 The pre-mRNA 3' end processing signals in mammals, *S. cerevisiae* and plants

Scheme representing the pre-mRNA 3' end processing signals in the respective kingdoms. USE, upstream sequence element; EE, efficiency element; PE, the A-rich positioning element including the highly conserved hexanucleotide AAUAAA of animals; UUE, upstream U-rich element; DUE, downstream U-rich element; DSE, U-rich downstream element. The figure was adapted from Graber et al., 1999a.

Trans-acting 3' end processing factors of *S. cerevisiae*

While the *cis*-acting polyadenylation signals used by mammals and yeast differ from each other, the *trans*-acting factors involved in 3' end cleavage and polyadenylation exhibit notable conservation (summarized in Table 1.1). Initial identification of the distinct activities required

for an *in vitro* reconstitution of 3' end cleavage and polyadenylation resulted from the biochemical fractionation of whole yeast cell extracts (Chen and Moore, 1992; Kessler et al., 1996, Minvielle-Sebastia et al., 1994; Preker et al., 1994). Specific cleavage and polyadenylation events occur when cleavage and polyadenylation factor (CPF), cleavage and polyadenylation factor IA (CF IA), and cleavage and polyadenylation factor IB (CF IB) are present. In addition, the specific polyadenylation step requires the poly(A) binding protein Pab1p. CPF is formed by factors that were initially believed to separate into two distinct complexes: cleavage factor (CF II) and polyadenylation factor (PF I; Chen and Moore, 1992). However, a multi-protein complex combining PF I/Pap1p and CF II activities was later isolated from yeast extracts by one-step affinity purification and therefore designated cleavage and polyadenylation factor CPF (Ohnacker et al., 2000).

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

Table 1.1 Yeast 3' end processing factors are evolutionarily conserved

Size (in kDa), protein sequence features and mammalian homologues are indicated.

CPF

The yeast cleavage and polyadenylation factor CPF, consisting of at least 15 polypeptides, contains the enzymes responsible for the two mayor activities of the 3' end processing machinery: the endonuclease Ysh1p/Brr5p, and the poly(A) polymerase Pap1p. Other CPF

subunits are involved in RNA recognition, poly(A) site selection, and in regulation of the 3' end processing at the transcriptional level.

Yhh1p/Cft1p was identified by its sequence similarity to the mammalian AAUAAA interacting protein CPSF160. It was subsequently implied in cleavage and polyadenylation of pre-mRNAs, based on the loss of both activities in extracts that were depleted with a specific antiserum (Stumpf and Domdey, 1996). Yhh1p is an RNA-binding protein that participates in poly(A) site recognition. The RNA-binding domain of Yhh1p is composed of predicted β-propeller repeats, which also occur in proteins involved in pre-mRNA splicing and UV-damage recognition. Mutant *yhh1* strains were defective in transcription termination, implicating a function for Yhh1p in this process (Dichtl et al., 2002b). Yhh1p/Cft1p bound specifically to the phosphorylated C-terminal domain (CTD) of RNAP II *in vitro* and in a two-hybrid test *in vivo*. Direct interactions of Yhh1p with both the pre-mRNA transcript and the CTD suggest that Yhh1p functions in the coupling of transcription termination and 3′ end formation (Dichtl et al., 2002b).

Ydh1p/Cft2p is an essential 105 kDa component of CPF. The sequence of Ydh1p has significant homology to its mammalian homologue CPSF100 (24% identity and 43% similarity). Ydh1p was shown to be required for 3' end cleavage and polyadenylation *in vitro* (Kyburz et al., 2003). Moreover, Ydh1p interacts with the poly(A) site region, and consistently mutant *ydh1* strains are deficient in the recognition of the *ACT1* cleavage site *in vivo* (Kyburz et al., 2003). Ydh1p interacts with the CTD of RNAP II, several other subunits of CPF and with Pcf11p, a component of CF IA. Therefore, Ydh1p contributes to the assembly of the polyadenylation machinery on the RNA substrate and mediates the dynamic association of CPF with RNAP II (Kyburz et al., 2003). Crystal structure of Ydh1p was recently revealed (Mandel et al., 2006). The protein conformation of Ydh1p shares sequence conservation and a similar domain architecture with Ysh1p but lacks the putative zinc-binding residues. The overall structure of Ydh1p is remarkably similar to that of human CPSF73, despite the low degree of sequence conservation between them. The most important difference between the two structures is that motifs for zinc binding in the metallo-β-lactamase domain are missing in Ydh1p, and therefore this protein cannot bind zinc and is unlikely to possess nuclease activity. The β-CASP domain of Ydh1p is much larger than that of CPSF73, and contains a highly flexible segment.

Ysh1p/Brr5p and its mammalian homologue CPSF73 are thought to possess the endonucleolytic activity involved in the 3' end cleavage of pre-mRNAs and also other RNAP II transcripts. Ysh1p/Brr5p (referred to as Ysh1p in the remainder of this work) was initially cloned and characterized in two independent reports (Chanfreau et al., 1996; Jenny et al., 1996). It was shown to be essential for cell viability and required for pre-mRNA 3' end formation both *in vitro* and *in vivo*. Depletion of Ysh1p resulted in inhibition of both cleavage and polyadenylation (Jenny et al., 1996).

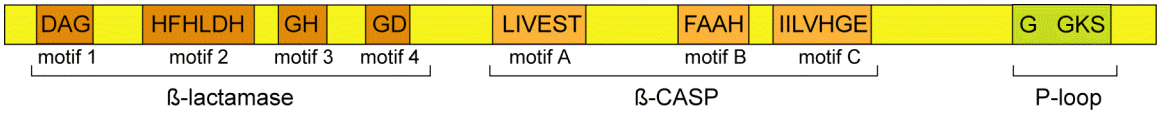


Figure 1.3 Conserved protein sequence motifs of Ysh1p

See text for explanations.

For years, the prediction for Ysh1p to be the putative 3' end processing endonuclease was based mainly on the protein sequence analysis which revealed a highly conserved β-lactamase fold commonly found in metal-dependent hydrolyzing enzymes (Aravind, 1999). The typical metallo-β-lactamase fold comprises of five highly conserved sequence motifs, which consist mostly of histidine and aspartic acid residues that participate in zinc coordination and hydrolysis reactions. Motif 2 is typical for the entire superfamily and is marked by the highly conserved HXHXDH signature. The third and fifth motifs each contain a conserved histidine; motif 4 contains an acidic residue or a cysteine. The conserved sequences C-terminal to motif 4, and including motif 5, constitute a hallmark of this family specifically acting on nucleic acids and were named the β-CASP motif, after metallo-β-lactamase-associated CPSF Artemis SNM1/PSO2 (Callebaut et al., 2002). A typical β-CASP domain contains 3 conserved motifs (A-C; Figure 1.3) that include two histidines and an acidic residue. In general, β-CASP proteins are defined by the presence of seven sequence motifs (four of the β-lactamase, and three of the β-CASP), all of which include an acidic or histidine residue known or predicted to participate in metal binding and hydrolysis. Specific features of these motifs allow distinguishing between enzymes involved in RNA metabolism from those acting on DNA substrates. This domain is highly conserved among Ysh1p/CPSF73 homologues and possibly has important functions in regulating their activity. Moreover, the C-terminal part of Ysh1p contains a P-loop motif ($GX_{4-8}GKS/T$) commonly found in proteins binding ATP or GTP (Figure 1.3). This motif is not present in the metazoan homologues of Ysh1p and was recently reported to be dispensable for yeast cell viability *in vivo* (Zhelkowsky et al., 2006).

A recent crystal structure of CPSF73, the mammalian homologue of Ysh1p, revealed its domain architecture. The amino-terminal residues form a domain similar to the structure of canonical metallo-β-lactamases, with a sandwich-layered αβ-βα architecture (Figure 1.4). The second domain, covering residues containing the β-CASP motif, forms a cassette inserted into the metallo-β-lactamase domain (Mandel et al., 2006). The β-CASP domain seems to be a new example of the nucleotide-binding fold (NBF) but is unlikely to bind nucleotides as it lacks the Walker A motif for binding nucleotide phosphate groups. The active site of CPSF73 contains two zinc atoms which are each bound in an octahedral environment, with a hydroxide ion as one of the bridging ligands (Figure 1.5). In metallo-β-lactamases, the hydroxide that bridges the two zinc ions is the nucleophile for the hydrolysis reaction, and the

substrate is directly liganded to the zinc atoms (Garrity et al., 2005). Indeed, in CPSF73 the hydroxide is located at the perfect position for an inline nucleophilic attack on the phosphate group to initiate the nuclease reaction (Mandel et al., 2006).

Interestingly, the active site in CPSF73 is located deep in the interface between the metallo-β-lactamase and the β-CASP domains, in contrast to canonical metallo-β-lactamases and RNase Z (the endonuclease cleaving the 3' ends of pre-tRNAs) that have an open zinc-binding site. The β-CASP domain of CPSF73 severely restricts access to the active site and the scissile phosphate group probably cannot reach the zinc ions in the structure. This indicates that a conformational change in the enzyme might be required for pre-mRNA binding (Mandel et al., 2006). Identification of Ysh1p/CPSF73 as members of the metallo-ß-lactamase fold superfamily is the most persuasive line of reasoning implicating those factors as the 3' end processing endonucleases. Moreover, Mandel et al., 2006, provided evidence for an unspecific *in vitro* RNA endonuclease activity of *E. coli* expressed CPSF73, when the protein was preincubated with Ca^{2+} ions prior to assaying. Most importantly, the nuclease activity of recombinant CPSF73 was abolished in a D75K/H76A mutant version of the protein where two of the zinc ligands in β-lactamase motif 2 were destroyed.

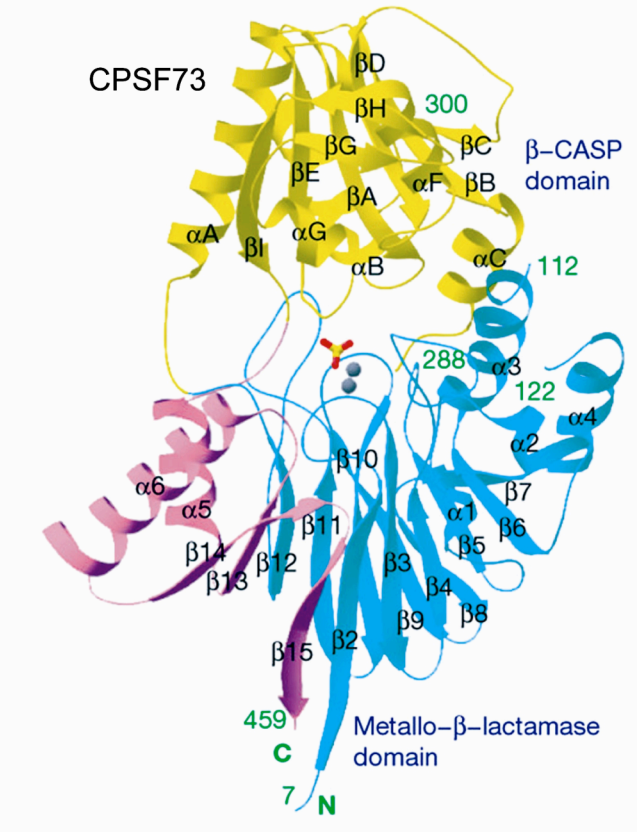


Figure 1.4 Crystal structure of the human 3' end processing endonuclease CPSF73

The β-strands and α-helices are labelled, and the two zinc atoms in the active site are shown as grey spheres. Scheme modified from Mandel et al., 2006.

The Ysh1p sequence shows similarity to both CPSF73 and CPSF100 subunits of the mammalian 3' end processing machinery. It is 53% identical and 73% similar in the first 500 amino acids to CPSF73, and 23% identical over the entire length of CPSF100 (Chanfreau et al., 1996; Jenny et al., 1996). The yeast homologue of CPSF100, Ydh1p, is also related to the Ysh1p by protein sequence showing 21% identity and 45% similarity. Protein sequence analysis revealed conserved β-lactamase and β-CASP motifs present in Ydh1p. However, several conserved amino acid residues in the β-lactamase motif 2 (HXHXDH signature) and also in the β-CASP motif are altered such that the protein is assumed not to be capable of any hydrolytic catalysis (Callebaut et al., 2002). This is also true for its mammalian homologue CPSF100, although to a lesser extend. The similarities that the two subunits Ysh1p/CPSF73 and Ydh1p/CPSF100 share within the metallo-β-lactamase/β-CASP domain suggest that they might have evolved from a common ancestor.

Characterization of the functions of the Ysh1p endonuclease in the 3' end processing of RNAP II transcripts is the main topic of this thesis.

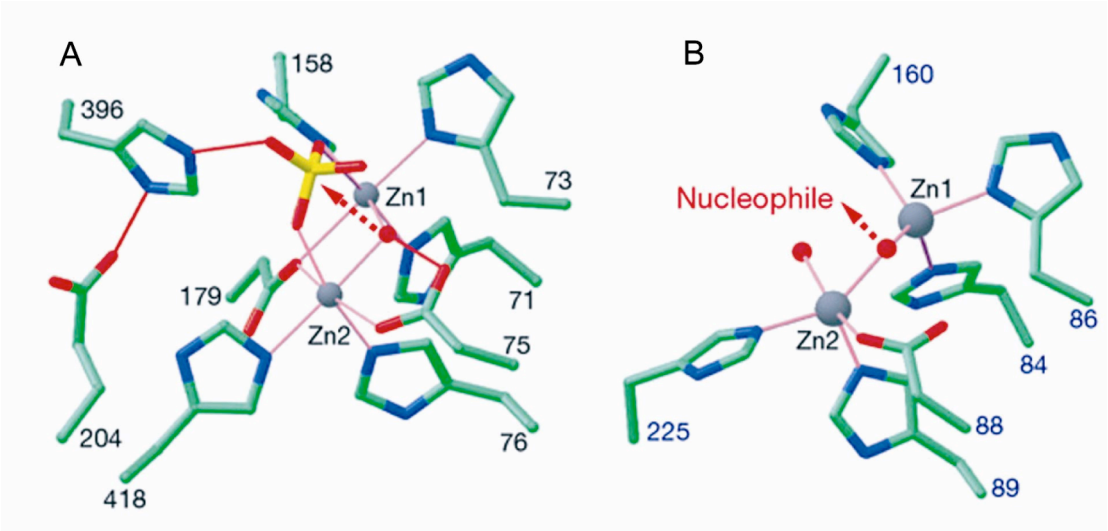


Figure 1.5 The active site of CPSF73

Comparison of the catalytic site of human CPSF73 **(A)** to the catalytic site of a typical metallo-β-lactamase (L1) **(B)**. The motifs are labelled, and the bridging hydroxide ion is shown as a red sphere. Ligand interactions are indicated by thin magenta lines, and hydrogen-bonding interactions by thin red lines. The arrow indicates the nucleophilic attack from the hydroxide ion. Scheme modified from Mandel et al., 2006.

Poly(A) polymerase (Pap1p) is the enzyme responsible for the addition of a polyadenine track at the 3' end of mRNA. Pap1p was the first factor of the yeast 3' end processing machinery to be purified and its gene, *PAP1*, was the first to be identified (Lingner et al., 1991; Patel and Butler, 1992). Pap1p shares a high level of similarity with its mammalian counterpart, being 47% identical within the first 400 amino acids. Poly(A)

polymerases have a modular organization with a catalytic domain near the N-terminus and an RNA-binding region that overlaps with a nuclear localization signal (NLS) near the C-terminus. The catalytic domain exhibits substantial homology with other nucleotidyl transferases (Bard et al., 2000; Martin et al., 2000). Yeast Pap1 is a template-independent RNA polymerase which is not required for efficient cleavage of pre-mRNAs *in vitro*. However, a mutation in Pap1p conferring temperature-sensitive growth can influence the choice of the poly(A) site in the *ACT1* pre-mRNA (Mandart and Parker, 1995).

Fip1p is encoded by an essential gene, *FIP1*, that was identified in a yeast 2-hybrid screen as a factor interacting with yeast poly(A) polymerase (*PAP1*; Preker et al., 1995). A temperature sensitive allele of *FIP1* showed a shortening of poly(A) tails and a decrease in the steady-state level of actin mRNAs in *vivo*. Mutant *fip1* extracts exhibited normal cleavage activity, but failed to polyadenylate the upstream cleavage product *in vitro.* Fip1p is the only protein known to interact directly with poly(A) polymerase (Pap1p). Its association with Pap1p inhibits the extension of an oligo(A) primer by limiting access of the RNA substrate to the C-terminal RNA binding domain (C-RBD) of Pap1p (Zhelkovsky et al., 1998). Fip1p tethers Pap1p to CPF and to RNA through its interactions with Pfs2p and Yth1p as well as to CF I through its interaction with Rna14p (Barabino et al., 2000; Ohnacker et al., 2000; Preker et al., 1995), thereby conferring specificity to poly(A) polymerase for pre-mRNA substrates. In summary, Fip1p is believed to have a crucial regulatory function in the polyadenylation reaction by controlling the activity of poly(A) tail synthesis through multiple interactions within the polyadenylation complex.

MPE1 is an essential gene encoding for Mpe1p, an integral subunit of CPF, which interacts genetically with *PCF11*. Mpe1p is an evolutionarily conserved protein, a homolog of which is encoded by the human genome (hMpe1). Protein sequence of Mpe1p contains a putative RNA-binding zinc knuckle motif. Extracts from a conditional mutant, *mpe1-1*, or from a wild-type extract immuno-depleted of Mpe1p are defective in 3' end processing, indicating that Mpe1p is required for the specific cleavage and polyadenylation of pre-mRNA (Vo et al., 2001).

Pfs2p is a 53 kDa protein that is encoded by an essential gene *PFS2* (Preker et al., 1995; 1997). The sequence of Pfs2p contains seven WD repeats. Extracts from *pfs2* mutant strains show cleavage and polyadenylation defects *in vitro* (Ohnacker et al., 2000). Furthermore, Pfs2p directly interacts with Ysh1p, Fip1p and Rna14p *in vitro*. One-step affinity purification of protein A-tagged Pfs2p results in the CPF complex pulled-down. Therefore, Pfs2p plays an essential role in the 3' end formation by bridging different processing factors.

The CPF subunit Yth1p is the yeast homologue of the mammalian CPSF30 (40% identity; Barabino et al., 1997). Yth1p is an RNA-binding protein that participates in the recognition of the cleavage site. Yth1p was also shown to be essential for both cleavage and polyadenylation of pre-mRNA (Barabino et al., 2000). The protein contains two distinct domains that have separate functions in 3’ end processing. The C-terminal part is required in polyadenylation to tether Fip1p and poly(A) polymerase to the rest of CPF. The second

domain, the highly conserved zinc finger motif, is essential for both cleavage and polyadenylation, and for interactions with the pre-mRNA and other CPF subunits (Barabino et al., 2000).

Pta1p, an essential 88 kDa protein, was originally implicated in tRNA splicing (O'Connor and Peebles, 1992). Extracts prepared from *pta1* mutant strains are impaired in both the cleavage and the poly(A) addition steps. Pta1p and Fip1p are tightly associated in a complex required for polyadenylation (Preker et al., 1997). Several subunits of the CPF complex were recently implied to form a distinct subcomplex, APT, in which Pti1p, Swd2p, Ssu72p, Glc7p, Syc1p and Ref2p subunits are associated with the Pta1p subunit of CPF (see below; Nedea et al., 2003). Importantly, Pta1p bridges the APT subcomplex to CPF. Chromatin immunoprecipitation assay identified Pta1p and other APT subunits on small nucleolar RNA (snoRNA) genes and primarily near the polyadenylation signals of the constitutively expressed *PYK1* and *PMA1* genes (Nedea et al., 2003). Interestingly, *pta1* mutants exhibited read-through from the *SNR33* gene locus into the downstream gene, suggesting that Pta1p is required for snoRNA 3' end formation (Nedea et al., 2003). Pta1p interacts physically and genetically with Glc7p and thus recruits Glc7p to the APT subcomplex. The activity of Pta1p is regulated via phosphorylation. The dephosphorylated state of Pta1p is essential for polyadenylation and is achieved by the Glc7p phosphatase activity (He and Moore, 2005). The depletion of Glc7p causes shortened poly(A) tails *in vivo* and accumulation of phosphorylated Pta1p.

APT, a subcomplex within CPF

Systematic tagging of subunits and purification of the CPF complex revealed its modular architecture. Six polypeptides associated with the CPF form a distinct complex, APT, in which Pti1p, Swd2p, Glc7p (a type I protein phosphatase), Ssu72p (a TFIIB and RNAP II-associated factor), Ref2p and Syc1p are associated with the Pta1p subunit (Nedea et al., 2003). APT subunits were shown to localise primarily near the polyadenylation signals of constitutively expressed mRNA genes and on snoRNA genes. Various studies revealed that most of the APT subunits are important for preventing transcriptional read-through at terminators of independently transcribed snoRNA genes (Dheur et al., 2003; Dichtl et al., 2004; Ganem et al., 2003; Morlando et al., 2002; Nedea et al., 2003; S. Roeck, personal communication).

PTI1, the essential gene encoding for the Pti1p subunit of the APT subcomplex was originally identified by a yeast 2-hybrid screen as strongly interacting with *PTA1* (W. Hübner, unpublished). Pti1p is subjected to phosphorylation upon growth arrest and displays significant sequence similarities to the CF IA subunit Rna15p and its mammalian homologue CstF-64. Moreover, Pti1p is required for correct snoRNA 3' end maturation, where it probably acts by uncoupling cleavage and polyadenylation (Dheur et al., 2003).

Another component of APT subcomplex is Ssu72p, a CTD phosphatase with specificity for the phosphorylated serine 5 (S5-P). Ssu72 catalyzes CTD S5-P dephosphorylation in association with Pta1p, although its essential role in 3' end processing is

independent of its catalytic activity (Krishnamurthy et al., 2004). Moreover, Ssu72p bridges CPF subunits Pta1p and Ydh1p/Cft2p, the general transcription factor TFIIB, and RNAP II via Rpb2p (Dichtl et al., 2002a). Mutant *ssu72-2* cells are defective in RNAP II transcription elongation and termination (Dichtl et al., 2002a). Interestingly, *ssu72* mutations also disrupt Nrd1-dependent termination of snoRNA precursors (Steinmetz and Brow, 2003). In summary, Ssu72p has a dual role in transcription, one as a CTD S5-P phosphatase that regenerates the initiation-competent, hypophosphorylated form of RNAP II and the other as a factor necessary for cleavage of pre-mRNA and efficient transcription termination.

The non-essential 20 kDa protein Syc1p was originally identified by extensive affinity-purification of the 3' end processing machinery (Gavin et al., 2002). It was later shown to associate with APT subcomplex (Nedea et al., 2003). Protein sequence analysis of Syc1p revealed a striking homology to the C-terminal domain of the Ysh1p endonuclease (Zhelkovsky et al., 2006). The characterization of Syc1p is one of the topics of this thesis (Chapter 3).

The WD-40 repeat protein Swd2p is a component of two functionally distinct multiprotein complexes: the cleavage and polyadenylation factor (CPF) that is involved in pre-mRNA and snoRNA 3' end formation and the SET1 complex (SET1C) that methylates histone 3 lysine 4. The role of Swd2p as component of CPF is functionally independent from the one in the SET1 complex. Swd2 protein is *in silico* predicted to form a seven-bladed beta-propeller structure. *swd2* mutant strains are defective in 3' end formation of specific mRNAs and snoRNAs (Dichtl et al., 2004). Overexpression of Ref2p, the protein previously implicated in Swd2p recruitment to APT, can rescue the growth and termination defects of the *swd2* mutant strains, indicating a functional interaction between the two proteins (Cheng et al., 2004).

Glc7p is a catalytic subunit of a type 1 serine/threonine protein phosphatase involved in many processes including glycogen metabolism, sporulation, and mitosis. Recent studies have found that Glc7p phosphatase associates with APT and is recruited to this complex by its physical interactions with Ref2p (Nedea et al., 2003). Conditional depletion of Glc7p causes defects in poly(A) addition but not in cleavage at the poly(A) site, and results in shortened poly(A) tails *in vivo*. The physiological substrate for Glc7p within the 3' end processing machinery is Pta1p (He and Moore, 2005). Downregulation of Glc7p, or its dissociation from CPF in the absence of CPF subunits Ref2p or Swd2p, results in snoRNA termination defects (Nedea et al., 2008).

Ref2p is encoded by the non-essential gene *REF2* that was originally identified by a genetic strategy predicted to detect decreases in the use of a *CYC1* poly(A) site interposed within the intron of an *ACT1-HIS4* fusion reporter gene (Russnak et al., 1995). Ref2p is required for efficient use of weak poly(A) sites and binds to RNA with specific affinity for pyrimidine bases (Russnak et al., 1995). This factor is also a negative regulator of poly(A) synthesis acting in opposition to Fir1p (Mangus et al., 2004). Extensive affinity-purification by Nedea et al., 2003, revealed that Ref2p is a part of APT subcomplex. Moreover, Ref2p genetically interacts with Glc7p and recruits it to APT/CPF. Interestingly, chromatin

immunoprecipitation localizes Ref2p on both mRNA and snoRNA genes (Nedea et al., 2003). Ref2p is also required for correct snoRNA 3' end maturation (Dheur et al., 2003).

CF IA

CF I (cleavage factor I) was originally identified as an activity needed for both the cleavage and polyadenylation steps of 3' end processing; further purification separated it into two components, CF IA and CF IB (Chen and Moore, 1992; Kessler et al., 1996). CF IA consists of four polypeptides, Rna14p, Rna15p, Pcf11p, and Clp1p.

The first indication of Rna15p and Rna14p involvement in pre-mRNA 3' end processing came from the observation that temperature-sensitive mutations in the *RNA14* and *RNA15* genes caused dramatic poly(A) tail shortening *in vivo* (Minvielle-Sebastia et al., 1991). The *rna14* and *rna15* mutants are defective in both cleavage and poly(A) addition *in vitro* (Minvielle-Sebastia et al., 1991). Moreover, these mutations are synergistically lethal with mutations in the *PAP1* gene. The 76 kDa Rna14 protein exhibits sequence homology to the mammalian CstF-77 (24% identity; Takagaki and Manley, 1994). Rna14p binds to the phosphorylated form of the CTD (Sadowski, 2003) and is also required for transcription termination on mRNA genes (Birse et al., 1998). The *RNA15* gene encodes for an essential component of CF IA. It contains an RNA recognition motif (RRM)-type RBD in its amino-terminal region and can be UV cross-linked to substrate RNA (Minvielle-Sebastia et al., 1991). Rna15p in complex with Nab4p/Hrp1p, Rna14p and Rna15p specifically interacts with the A-rich element; however, Rna15p alone is unable to recognize a particular sequence in the absence of other proteins (Gross and Moore, 2001). Recognition of the A-rich element depends on the tethering of Rna15p through an Rna14p bridge to Nab4p/Hrp1 bound to the UA-rich motif. Rna15p is essential for correct transcription termination on mRNA genes (Birse et al., 1998); it was also implied in correct snoRNA 3' end formation (Fatica et al., 2000). The vertebrate orthologue of Rna15p is CstF-64.

PCF11 is an essential gene encoding for a 70 kDa protein. This gene was isolated in a yeast 2-hybrid screen with Rna14p and Rna15p as baits (Amrani et al., 1997a). Temperature sensitive mutations in *PCF11* are synergistically lethal with mutant alleles of *RNA14* and *RNA15*. Pcf11p is required for 3' end cleavage and polyadenylation and for RNAP II transcription termination. Pcf11 protein contains a conserved CTD-interaction domain (CID), which is essential for cell viability. Through the CID Pcf11p binds to the phosphorylated CTD of RNAP II and thus bridges the CTD to the nascent transcript and causes dissociation of both RNAP II and the nascent transcript from the DNA. Possibly, Pcf11p transduces conformational changes between the CTD and the nascent transcript to cause termination (Zhang et al., 2005). CTD binding and 3' end processing activities of Pcf11p can be functionally uncoupled from each other and are provided by distinct Pcf11p fragments in *trans* (Sadowski et al., 2003). Interestingly, Pcf11p was also implied in transcription termination on snoRNA genes (Kim et al., 2006).

Clp1p, the fourth polypeptide of CF IA, was identified by peptide sequencing of the 50 kDa band in highly purified CF IA (Minvielle-Sebastia et al., 1997). Clp1p is evolutionarily conserved and contains P-loop motifs (Walker A and B motifs), which have been implicated in ATP/GTP binding. *CLP1* interacted genetically with *PCF11* in a yeast 2-hybrid screen, and mutant *clp1-12* strain was deficient in both cleavage and polyadenylation steps of 3' end processing (T. Wiederkehr and W. Hübner, personal communication).

CF IB

Purified CF IB consists of a single 73 kDa polypeptide named Nab4p/Hrp1p (Kessler et al., 1997). Originally, *NAB4* gene was identified as a suppressor of the mutant *npl3-1* allele, a gene encoding Npl3p protein involved in mRNA export (Henry et al., 1996). Synergistic lethality assays and yeast 2-hybrid screen showed that Nab4p/Hrp1p interacted *in vivo* with Rna14p and Rna15p but not with Pap1p (Kessler et al., 1997). The protein sequence analysis of Nab4p/Hrp1p revealed two RRMs in its middle region, both containing RNP1 and RNP2 motifs (Henry et al., 1996). SELEX analysis showed that the protein interacts with the UA-rich efficiency element (Valentini et al., 1999). Nab4p/Hrp1p is not essential for cleavage of pre-RNAs, but it regulates cleavage site utilization (Minvielle-Sebastia et al., 1998). Recombinant Nab4p/Hrp1p functionally replaces the yeast CF IB in reconstituted cleavage and polyadenylation assays (Kessler et al., 1997). More recently, Nab4p/Hrp1p was identified as a factor activating nonsense-mediated decay (NMD), the cytoplasmic pathway that monitors premature translation termination and degrades aberrant mRNAs. Nab4p/Hrp1p interacts with Upf1p, a component of the surveillance complex and binds specifically to a DSE-containing RNA, marking the transcript for rapid decay (Gonzalez et al., 2000). The closest counterpart to Nab4p/Hrp1p in the mammalian system, at least in terms of having an amino-terminal RBD and a function in cleavage, is CF I_m-68.

Poly(A) binding proteins

Regulation of the poly(A) tail length during mRNA 3' end formation requires a specific poly(A) binding protein in addition to the cleavage and polyadenylation machinery. The major poly(A) binding protein in yeast is Pab1p. It is a multifunctional protein that mediates many cellular functions associated with the poly(A) tail of messenger RNAs. Mutation or deletion of *PAB1* gene does not affect the cleavage step of 3' end formation; however, it generates mRNAs with abnormally long poly(A) tracts *in vitro*. Moreover, *in vitro* complementation with recombinant Pab1p not only restores the length of the poly(A) tails to normal, but also triggers a poly(A) shortening activity (Minvielle-Sebastia et al., 1997). Pab1p interacted with Rna15p in yeast 2-hybrid assays and in co-immunoprecipitation experiments (Amrani et al., 1997b). Furthermore, Pab1p was implicated in the regulation of translation initiation and in cytoplasmic mRNA stability. The inviability of a *PAB1* deletion strain was suppressed by a mutation in the 5'-3' exoribonuclease *RRP6*, a component of the nuclear exosome (Brune et al., 2005). Pab1p contains a nonessential leucine-rich nuclear export signal and shuttles

between the nucleus and the cytoplasm. The proper loading of Pab1p onto mRNAs and final trimming of the tail allows release from transcription sites and couples pre-mRNA processing to export (Dunn et al., 2005).

A recent report suggested that polyadenylation was more efficient and physiologically more relevant in the presence of Nab2p as opposed to Pab1p, and that possibly, Nab2p is both necessary and sufficient to regulate poly(A) tail length during the *de novo* synthesis in yeast (Dheur et al., 2005). Nab2p was first identified as a poly(A) binding protein by *in vivo* UV light-induced cross-linking of the proteins intimately associated with RNA (Anderson et al., 1993). The *NAB2* gene is located on chromosome VII within a cluster of ribonucleoprotein genes. Functional Nab2p is one of the major proteins associated with nuclear polyadenylated RNA *in vivo*, and it is required for export of poly(A) RNA from the nucleus. Export of Nab2p from the nucleus is dependent upon ongoing RNA polymerase II transcription (Green et al., 2002). Interestingly, the level of *NAB2* mRNA is controlled by the nuclear exosome subunit Rrp6p. *Cis*- and *trans*-acting mutations that inhibit the degradation activity of the nuclear exosome subunit Rrp6p result in elevated levels of *NAB2* mRNA. Control of *NAB2* mRNA levels occurs post-transcriptionally and requires a sequence of 26 consecutive adenosines (A_{26}) in the *NAB2* 3' untranslated region, which represses *NAB2* 3' end formation and sensitizes the transcript to degradation by Rrp6p (Roth et al., 2005).

Mammalian 3’ end processing factors

Extensive biochemical analysis of mammalian *trans*-acting factors revealed an astonishing complexity of proteins involved in 3’ end processing of pre-mRNA. Importantly, most of the protein factors required for pre-mRNA 3’ end processing are conserved between yeast and mammals. The mammalian cleavage and polyadenylation machinery consists of cleavage and polyadenylation specificity factor (CPSF), cleavage factors I_m and II_m (CF I_m and CF II_m), cleavage stimulatory factor (CstF), poly(A) polymerase (PAP), poly(A) binding protein (PABPN1), and the CTD of RNA polymerase II (Barabino and Keller, 1999; Colgan and Manley, 1997; Keller and Minvielle-Sebastia, 1997; Wahle and Rüegsegger, 1999; Zhao et al., 1999).

CPSF purified from HeLa cells or calf thymus contains four subunits: CPSF160, CPSF100, CPSF73, and CPSF30 (Bienroth et al., 1991; Jenny et al., 1994; Murthy and Manley, 1995). The yeast homologues of these subunits are Yhh1p, Ydh1p, Ysh1p, and Yth1p, respectively. CPSF recognizes the essential AAUAAA hexamer and is required for both the cleavage and polyadenylation reactions. The CPSF73 subunit is the 3’ end processing endonuclease (Mandel et al., 2006).

CstF is necessary for 3’ end cleavage, but not for polyadenylation of pre-mRNAs; however, it can stimulate poly(A) addition on substrates with a CstF binding site upstream of the AAUAAA sequence. Purification of CstF from HeLa cells showed that it consists of three polypeptides of 77, 64, and 50 kDa (CstF77, CstF64, and CstF50, respectively; Takagaki et al., 1990).

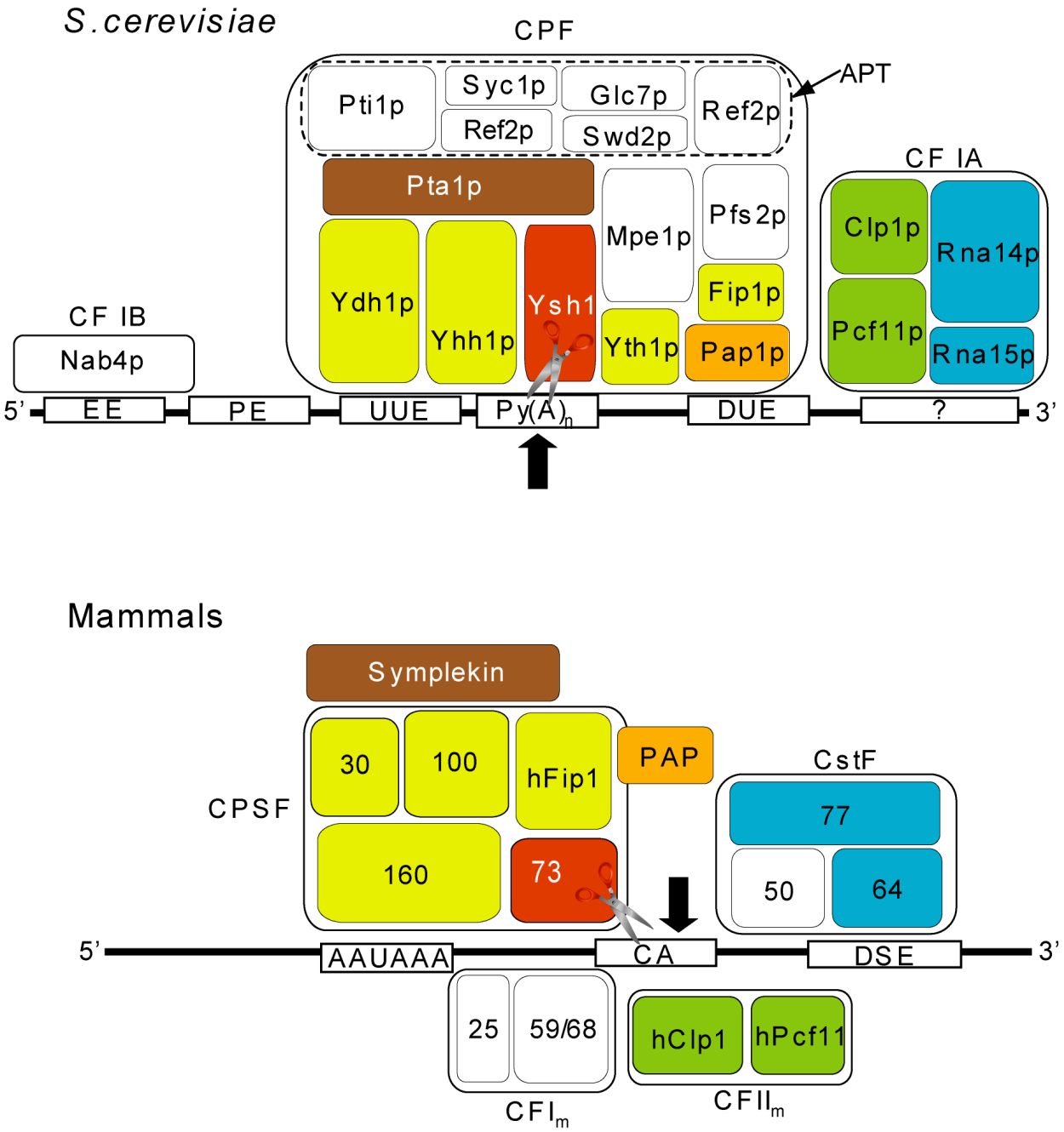


Figure 1.6 Eukaryotic 3' end processing machineries

Schematic representation of the 3' end processing complexes in *S.cerevisiae* and mammals and their arrangement on the pre-mRNA. Homologous subunits are indicated by equivalent colours.

The RNA binding domain of CstF64 recognizes GU- and U-rich sequences of the DSE region of pre-mRNA (MacDonald et al., 1994; Takagaki and Manley, 1997). The CstF77 and CstF64 subunits show a significant level of homology to the yeast CF IA factors Rna14p and Rna15p, respectively.

Cleavage factors I_m and II_m (CF I_m and CF II_m) are required only for the 3' end cleavage. Three polypeptides of 68, 59, and 25 kDa, and possibly a fourth one of 72 kDa copurify with CF I_m activity. The CF I_m subunits were shown to bind to pre-mRNA *in vitro* and to increase the stability of the CPSF-RNA complex (Rügsegger et al., 1996). Possibly, CF I_m interacts with CPSF and contributes to the overall stability of the 3' end processing machinery. Subunits of CF I_m have no orthologues in any of the known yeast 3' end processing factors. Purification of CF II_m from HeLa cell nuclear extract separated this activity into two components, CF IIA_m, essential for the 3' end cleavage, and CF IIB_m, stimulatory for the cleavage reaction. CF IIA_m contains the human homologues of two yeast 3' end processing factors, Pcf11p and Clp1p (thus named hPcf11 and hClp1, respectively; de Vries et al., 2000). The hClp1 protein is evolutionarily conserved and carries P-loop motifs commonly found in ATP/GTP binding proteins. Immunodepletion of hClp1 specifically inhibits the cleavage activity of 3' end processing (de Vries et al., 2000). Apart from its function in pre-mRNA 3' end processing, hClp1 was found to associate with the human tRNA splicing endonuclease complex (Paushkin et al., 2004), and to phosphorylate the 5' end of the 3' exon during human tRNA splicing, allowing the subsequent ligation of both exon halves by an unknown tRNA ligase (Weitzer and Martinez, 2007). hPcf11 also exhibits a high level of conservation and contains a CTD-interacting domain and two zinc finger motifs, but its function in 3' end formation remains to be investigated.

The mammalian poly(A) polymerase, PAP, plays a key role in the 3' end formation of mRNA in metazoan cells. Cloning and expression of the bovine PAP cDNAs have identified at least two isoforms of PAP which are generated by alternative splicing and differ only at their C termini; both forms are enzymatically active (Wahle et al., 1991). The N-terminal part of PAP is highly conserved in eukaryotes and contains a catalytic domain with homology to a family of nucleotidyltransferases including many DNA and RNA polymerases (Martin and Keller, 1996). PAP is recruited to the processing complex by interaction with CPSF160 (Murthy and Manley, 1995). However, in the absence of other 3' end processing factors, PAP has only a very low level of activity and shows no specificity for RNA substrates (Wahle and Keller, 1992). CPSF and PAP together are sufficient for the polyadenylation of pre-cleaved RNA substrates, although the reaction is slow and distributive. Rapid and processive elongation, and control of poly(A) tail length requires poly(A) binding protein PABPN1 that binds specifically to poly(A) (Bienroth et al., 1993).

Finally, RNAP II with its C-terminal domain (CTD) is required for the cleavage reaction (Hirose and Manley, 1998). CPSF and CstF could be retained on CTD affinity columns and copurified with RNAP II in a high molecular mass complex (McCracken et al., 1997b), pointing to the coupling of transcription and pre-mRNA 3' end processing.

1.2 3’ end formation of small nuclear and small nucleolar RNA transcripts

Small nuclear RNA (snRNA) is a class of small RNA molecules transcribed by RNAP II or RNAP III that are found within the eukaryotic nucleus. SnRNAs associate with specific proteins to form small nuclear ribonucleoproteins (snRNPs) involved in nuclear processes such as splicing, regulation of transcription factors, and maintaining the telomeres. Small nucleolar RNAs (snoRNAs) is a term used for a complex family of small RNA molecules localized in the nucleolus that participate in processing and modifications of ribosomal RNA (rRNA; Maxwell and Fournier, 1995), and also of snRNAs (Ganot et al., 1999; Jady and Kiss, 2001).

Formation of snRNA and snoRNA 3' ends differs between vertebrates and yeast. In vertebrates, snRNA genes utilize specialized promoters and 3' end formation elements that are distinct from those that direct mRNA synthesis (Hernandez, 2001; Medlin et al., 2003), whereas most snoRNAs are processed from pre-mRNA introns (Filipowicz and Pogacic, 2002). In contrast, all yeast snRNAs and most yeast snoRNA genes are transcribed by RNA polymerase II from promoters that appear similar to those of mRNAs. Similarly, nascent snRNA and snoRNA transcripts require capping at the 5’ end and processing at the 3’ end. However, unlike mRNAs, snRNAs and snoRNAs are not polyadenylated in their mature form.

Relatively little is known about the exact mechanism of 3’ end formation of snRNA and independently transcribed snoRNA in yeast, and the molecular mechanism responsible for uncoupling cleavage from polyadenylation at these genes remains unclear. In general, 3’ ends of snoRNAs are produced by either endonucleolytic cleavage or termination and subsequent exonucleolytic trimming by the nuclear exosome (Allmang et al., 1999; Butler, 2002; Chanfreau et al., 1998a; Chanfreau et al., 1998b; van Hoof et al., 2000). Processing of different prototypes of independently transcribed snoRNA units is summarized in Figure 1.7. Many dicistronic or polycistronic snoRNA units possess Rnt1p (the yeast RNase III) cleavage sites both at their 5’ and 3’ ends. In Rnt1p-dependent snoRNA 3’ end processing, the Rnt1p cleavage releases individual snoRNAs from the precursor molecule, often liberating it also from the 5’ end cap structure (Figure 1.7 A). The 5′ end of the snoRNAs encoded as monocistronic units either retains the 5’ end cap structure, or it is produced *de novo* by 5′–3′ trimming from an Rnt1p-processing site (Figure 1.7 B). However, many poly- and monocistronic units do not possess Rnt1p-processing sites in their 3′ ends. This type of snoRNA 3’ end formation is therefore termed Rnt1p-independent (Figure 1.7 B). In case of Rnt1p-independent 3’ end formation, it remains unclear how the entry site for the exonucleolytic trimming from the 3′ end is produced, and, more generally, how the 3′ end of snoRNA precursors is formed. It has been proposed that 3′ end formation of independently transcribed snoRNA relies on sequences that uniquely require the cleavage activity and not the polyadenylation activity of the canonical 3′ end processing machinery (Fatica et al., 2000;

Morlando et al., 2002). Both snoRNA 3' end formation and transcription termination events appear to be strictly interconnected. It is virtually impossible to clearly distinguish between the *cis*- and *trans*-acting elements involved in pre-snoRNA 3' end formation from those acting in transcription termination on snoRNA genes. Therefore, putative sequence elements and protein factors directing snoRNA 3' end processing and transcription termination will be described below.

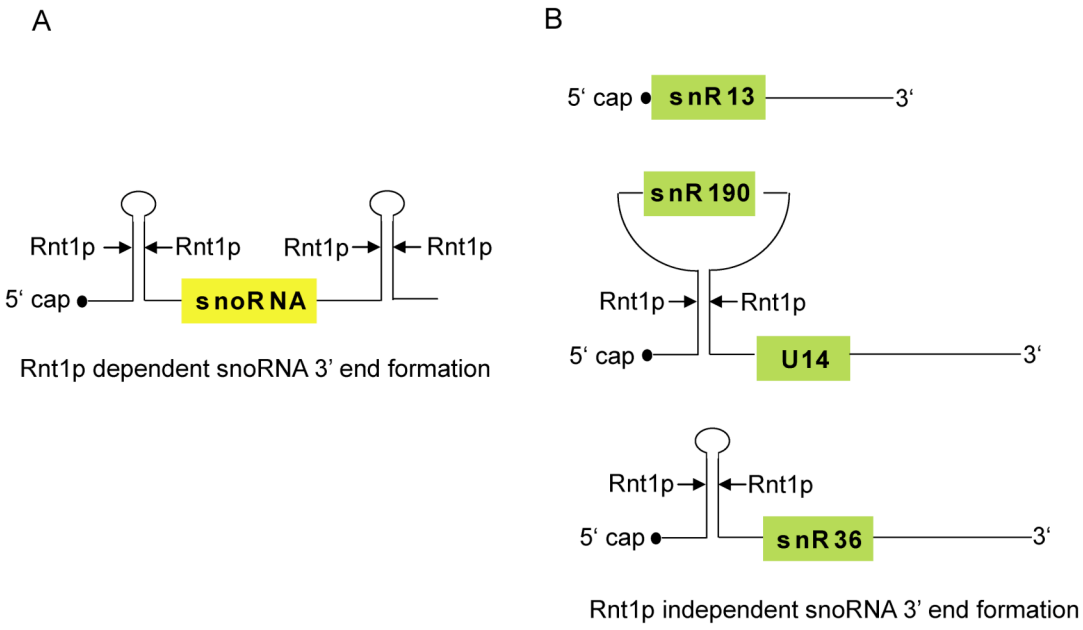


Figure 1.7 3' end formation of independently transcribed snoRNA transcripts

(**A**) and (**B**) Rnt1p-dependent and Rnt1p-independent snoRNA 3' end formation, respectively. See text for explanations.

Cis-acting sequences required for snRNA and snoRNA 3' end formation and transcription termination in *S. cerevisiae*

In a naïve consideration of snoRNA 3' end formation and transcription termination events, a clear distinction between the poly(A) and non-poly(A) pathways should be provided by the RNA signal on the nascent transcript. However, the requirement for certain factors of the canonical 3' end cleavage and polyadenylation machinery in snoRNA 3' end formation suggests that there are parallels between the poly(A) and the non-poly(A) 3' end formation and termination mechanisms. Moreover, similarly to pre-mRNA cleavage and polyadenylation signals, critical sequence elements directing pre-snoRNA 3' end formation are redundant.

A significant role in defining non-poly(A) terminators is played by GUA[AG] and UCUU sites that are specifically recognized by Nrd1 and Nab3 proteins, respectively (Carroll et al., 2004). Consistently, the CUU repeat was previously described to be involved in controlling 3' end maturation of snoRNAs and to respond to Sen1p factor (Rasmussen and Culbertson, 1998). These elements are clearly distinct from the *cis*-acting signals that direct 3' end formation of pre-mRNAs. In several snoRNA genes examined, both Nrd1- and Nab3- binding elements are present, but their number, spacing, and relative orientation of the GUA[AG] and UCUU sequences differs, thus indicating a large degree of genetic variability. Moreover, single base changes in these elements have relatively little effect on the ability to terminate, which is likely due to the redundancy of both the GUA[AG] and UCUU motifs (Carroll et al., 2004). In addition, the organization of these sequences is not strongly conserved among even closely related yeast species. Whereas both Nrd1- and Nab3-binding elements are essential for snoRNA 3' end formation and termination, it is not clear whether other *cis*-acting sequences are also essential/required for this function.

Interestingly, snoRNA terminators contain also sequences similar to mRNA polyadenylation sites (Steinmetz et al., 2006a). Extensive mapping of the *SNR13* and *SNR65* snoRNA gene terminators uncovered their bipartite structure, consisting of two distinct regions. Whereas region I contains sequences expected to bind Nrd1p, region II contains conserved features similar to pre-mRNA cleavage/polyadenylation sites. Each region separately has only weak terminator activity, but both together are required for efficient termination and point mutations in both of them cause a strong transcriptional read-through (Steinmetz and Brow, 2003; Steinmetz et al., 2006a). Strikingly, placing the *SNR13* or the *SNR65* terminator into the 3' untranslated region (UTR) of a protein-coding gene directed polyadenylation of the mRNA (Steinmetz et al., 2006a). Polyadenylation occurred at the downstream boundary of region II, within a few nucleotides of sites that were predicted computationally to resemble cleavage/polyadenylation sites of protein coding genes (Graber et al., 2002). This suggests that these polyadenylated 3' ends may be produced by the same 3' end processing machinery that acts on mRNAs. Bipartite terminators containing cleavage and polyadenylation signals may be a common characteristic of yeast snoRNA genes. The architecture of RNAP II-dependent 3' end processing is apparently modulated depending on the specific transcript. *Cis*-acting elements may therefore determine the type of 3' end

processing of the transcripts – whether they are to be cleaved and polyadenylated or only to be cleaved. Notably, 3' downstream sequences of snRNA and snoRNA precursors can be cleaved by the 3' end processing machinery *in vitro* (Morlando et al., 2002), thus suggesting the existence of putative poly(A) sites in their terminator regions that are recognizable by the canonical 3' end processing machinery.

Yeast *trans*-acting factors involved in 3' end formation and in transcription termination of snRNAs and snoRNAs

Several factors were reported to be essential for poly(A) independent transcription termination pathway on snoRNA genes, including sequence-specific RNA-binding proteins Nab3 and Nrd1, Sen1 helicase, Rpb3, Rpb11 and the CTD of RNAP II, the CTD kinase Ctk1, the CTD phosphatase Ssu72, and the Paf1 complex (Steinmetz & Brow, 1996; Conrad et al., 2000; Steinmetz et al., 2001; Ursic et al., 2004; Sheldon et al., 2005; Steinmetz et al., 2006a).

The yeast *NRD1* gene encodes an essential RNA-binding protein Nrd1p that associates with the CTD of RNAP II in yeast 2-hybrid assays via its CTD-interaction domain (CID; Steinmetz and Brow, 1998). Nrd1p regulates its own expression through premature termination directed by sequences in the nascent transcript, and levels of *NRD1* mRNA are increased in temperature sensitive *nrd1, nab3* and *sen1* strains that are known to impair correct termination of *NRD1* (Steinmetz et al., 2001; Arigo et al., 2006). Moreover, mutations in *NRD1* share some phenotypes with exosome mutants, including increased read-through transcription from several mRNA and sn/snoRNA genes. The Nrd1 complex physically interacts with the nuclear exosome and possibly attracts the exosome to the site of snoRNA 3' end processing (Vasiljeva and Buratowski, 2006).

Nab3p, a yeast hnRNP, is a high-copy suppressor of some *nrd1* temperature-sensitive alleles; it interacts with Nrd1p in a yeast 2-hybrid assay, and co-immunoprecipitates with Nrd1p. Temperature-sensitive alleles of *NAB3* are suppressed by deletion of *CTK1*, a kinase that has been shown to phosphorylate the CTD and to increase elongation efficiency *in vitro.* Nrd1p and Nab3p together form a heterodimer that recognizes *cis*-elements in the 3' regions of snoRNA and some snRNA precursors and directs their 3' end formation and termination (Conrad et al., 2000; Steinmetz and Brow, 1996; Steinmetz and Brow, 1998; Steinmetz et al., 2001). Nrd1p, Nab3p, the RNAP II CTD and the Ctk1 kinase interact genetically and physically with each other (Conrad et al., 2000). Both Nrd1p and Nab3p contain a single conserved RNA recognition motif (RRM), a domain common to proteins involved in RNA processing. As mentioned above, mutations of the Nrd1 and Nab3 binding sites result in RNAP II read-through into downstream genes, confirming the involvement of Nrd1/Nab3 in termination, and furthermore indicating that polyadenylation-independent 3′ end formation is also coupled to termination (Carroll et al., 2004; Steinmetz and Brow, 1996; Steinmetz and Brow, 1998).

Sen1 protein is another RNAP II termination factor that functions together with Nrd1 and Nab3 proteins on non-coding RNA genes (Steinmetz and Brow, 1996; Steinmetz et al.,

2001). Yeast Sen1p is an essential superfamily 1 helicase that is most similar to Upf1/SMG-2 and SDE3/Armitage (Steinmetz et al., 2006b). The Sen1p orthologue from *S. pombe* was shown to possess RNA-RNA, RNA-DNA, and DNA-DNA helicase activity *in vitro* (Kim et al., 1999). Mutations in Sen1p caused transcriptional read-through not only of the snoRNA gene terminators but also of the poly(A) site/terminators of the protein-coding *CYC1, NAB4,* and *IMD2* genes, thus confirming that Sen1p functions in termination of at least some short protein coding genes as well (Steinmetz et al., 2006b). Interestingly, Sen1p directly interacts with Glc7p, the yeast protein phosphatase 1 associated with CPF; Glc7p can dephosphorylate Sen1p *in vitro* (Nedea et al., 2008).

SnoRNA 3' end formation was shown to require the cleavage activity, but not the polyadenylation activity, of the canonical 3' end processing machinery (Fatica et al., 2000). Several pre-mRNA 3'end processing factors were consequently implicated in the snoRNA termination pathway. Recent reports indicated that several CPF factors forming the APT subcomplex (Ssu72p, Pti1p, Ref2p, Swd2p) are required for 3' end formation of snoRNAs (Dheur et al., 2003; Dichtl et al., 2004; Ganem et al., 2003; Nedea et al., 2003). In particular, Pti1p and Ref2p, which are required to prevent transcriptional read-through into downstream genes, were suggested to function in uncoupling cleavage and polyadenylation (Dheur et al., 2003). APT subcomplex was therefore proposed to mediate polyadenylation-independent 3' end formation on specific subpopulations of RNAP II transcripts, including snoRNAs and snRNAs. Components of the canonical 3' end processing apparatus associate with the snoRNA genes. A strong recruitment of several subunits of CFIA (Rna14p, Rna15p, and Pcf11p), CFIB (Hrp1p), and CPF (Ydh1p) to *SNR33* and *SNR13* snoRNA genes was observed by chromatin IP cross-linking (Kim et al., 2006). Similarly, Ref2p, Pta1p, Pti1p, and Ssu72p subunits of APT were reported to associate with *SNR13* gene (Nedea et al., 2003). CF IA subunits Rna14p, Rna15p and Pcf11p were required for the processing of snoRNA 3' ends (Fatica et al., 2000). In summary, the canonical 3' end processing machinery is recruited to snoRNA genes and is also required for the 3' end formation of snoRNA transcripts. However, it remains unclear whether endonucleolytic cleavage at the 3' ends is essential or dispensable for snoRNA 3' end formation.

Additionally, the yeast Paf1 complex (Paf1C) was shown to play a role in snoRNA 3' end formation. Primarily, Paf1C interacts with RNAP II and mediates histone modifications during elongation. Deletion of Paf1p subunit causes accumulation of snoRNA transcripts that are extended at their 3′ ends. Paf1C associates with and facilitates Nrd1p recruitment to *SNR47* gene, thus suggesting a direct involvement in snoRNA 3′ end formation (Sheldon et al., 2005).

1.3 Evolutionary comparison of 3' end processing of different RNA transcripts

RNA metabolism is one of the central parts of cell physiology. This term relates to a complex set of processes involving RNA-protein and RNA-RNA interactions, including transcription, processing and modification of RNA transcripts, translation, and RNA degradation. More than 80 orthologous groups of proteins are involved in RNA metabolism (Anantharaman et al., 2002). They are present in most species from the three primary kingdoms of eukarya, archaea and bacteria, and they are traceable to their last universal common ancestor (LUCA; Woese and Fox, 1977). RNA metabolism is therefore believed to be the most evolutionarily conserved system among all functional systems in the cell (Anantharaman et al., 2002), consistent with the hypothesis of a primordial 'RNA world' in which RNA-related functions had a dominant role (Gilbert, 1986). Although in the course of evolution RNA metabolism evolved similarly to other biological systems, it showed a strong tendency towards conservation of its ancient components (Anantharaman et al., 2002). Nevertheless, the emergence and evolution of eukaryotes resulted also in multiple innovations that led to the development of new functional systems within the RNA metabolism, such as splicing and alternative splicing, RNA interference, and other forms of post-transcriptional regulation.

All cells synthesize a vast array of RNAs. As most of the RNA transcripts require extensive processing and modifications to become physiologically competent, the eukaryotic and possibly also archaean cells developed sophisticated mechanisms involved in RNA maturation. For the majority of eukaryotic RNAs, irrespective of the type of transcript, formation of 3' ends requires an endonucleolytic cleavage. 3′ ends of mRNAs, tRNAs, microRNAs and certain small nuclear RNAs (snRNAs) are all cleaved in a specific yet sophisticated manner. Remarkably, a significant part of the RNA precursors is cleaved at the 3' end by identical or similar enzymes that carry a highly conserved β-lactamase signature. The β-lactamase family of enzymes is ubiquitously distributed in the three biological kingdoms, thus pointing to the functional importance and suggesting that they evolved from a common ancestor (Daiyasu et al., 2001).

As described already in part 1.1 of this chapter, 3' end maturation of eukaryotic pre-mRNAs involves endonucleolytic cleavage followed by polyadenylation. The yeast 3' end processing endonuclease Ysh1p and its mammalian homologue CPSF73 are members of a subfamily of metallo-β-lactamase enzymes, which use a distinct protein structure that coordinates two zinc ions. Similarly, ELAC2/tRNase Z, the endonuclease that cleaves 3′ extensions from pre-tRNAs in mammalian cells, like its close relatives in yeast, plants and archaea, also belongs to the metallo-β-lactamase class of enzymes (Schiffer et al., 2002). Moreover, ELAC2 and CPSF73 are similar in protein sequence (Callebaut et al., 2002).

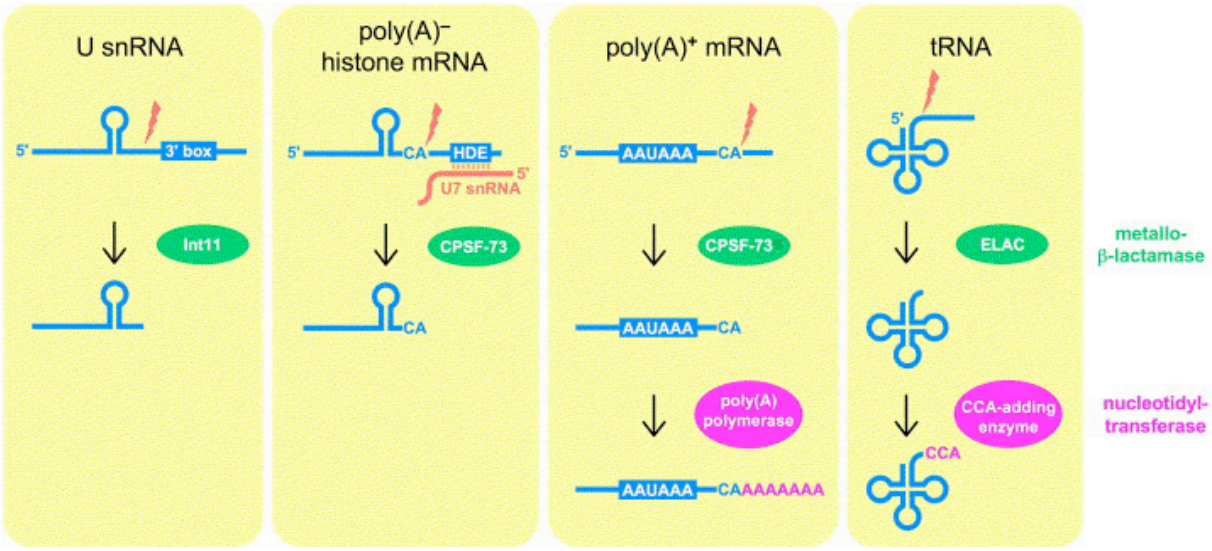


Figure 1.8 Similar 3' end processing of various types of metazoan RNA transcripts
Scheme taken from Weiner, 2005; see text for explanations.

Strikingly, the two steps of mRNA and tRNA 3′ termini formation are very similar, both involving an endonucleolytic cleavage performed by metallo-β-lactamase-related enzymes, and followed by the addition of non-templated RNA sequences by two related nucleotidyl transferases. The poly(A) tail on mRNAs is added by the poly(A) polymerase, and the CCA extension at the 3' end of tRNAs is added by the CCA-adding enzyme (Aravind, 1999).

The 3′ ends of metazoan histone mRNAs are also generated by the endonucleolytic cleavage by CPSF73 (Dominski et al., 2005); however, the cleavage site is defined by a mechanism different from that utilized for other pre-mRNAs and no poly(A) tail is added afterwards. Interestingly, the 3' ends of mammalian U-type small nuclear RNAs (U snRNAs) are also formed by an endonucleolytic cleavage performed by Int11, which is a component of the Integrator complex and a close homologue of CPSF73 (Baillat et al., 2005), suggesting an analogy between 3' end formation of U snRNAs and mRNAs. The similar processing of various metazoan RNA transcripts is summarized in Figure 1.8 (Weiner, 2005).

Interestingly, all the metallo-β-lactamases involved in 3' end processing seem to act as dimers (Aravind, 1999); either homodimers in the case of tRNase Z/ELAC2, or heterodimers of active and inactive subunits such as the yeast Ysh1p/Ydh1p and mammalian CPSF73/CPSF100 of the canonical 3' end processing apparatus, or Int11/Int9 of the Integrator complex. The catalytically inactive forms possibly have a modulatory function, helping to regulate the enzymatic activity of the active counterpart. Apparently, such gene duplications resulting in inactive-active enzyme pairs were important as one of the copies could be freed from the selective pressure on the active site and could therefore develop alternative regulatory functions (Aravind, 1999).

1.4 Coupling of the pre-mRNA processing events to RNA polymerase II transcription

Eukaryotic RNAP II transcripts require extensive processing in order to become functional. Soon after RNAP II transcription initiation the nascent RNA is capped at its 5' end. Intron sequences emerging on the transcript are removed by pre-mRNA splicing. Transcription is terminated upon reaching the end of the gene, and the newly synthesized RNA is cleaved and polyadenylated at its 3' end. Processed mRNAs are packed into export-competent mRNPs in order to be transported to the cytoplasm. For years, each of the steps in the pathway from gene to protein was studied separately, with little consideration given to the possible interconnections between them. However, more recent studies revealed functional links between various factors that carry out different steps of the gene expression pathway, suggesting that consecutive steps in the pathway are interdependent and/or are influenced by one another. Each stage of gene expression is in fact a subdivision of a continuous process, with each phase physically and functionally connected to the next. Protein factors involved in each individual step are often functionally and physically connected. Such organization of events allows for a series of quality control mechanisms at multiple stages (reviewed in Orphanides and Reinberg, 2002). Figure 1.9 schematically illustrates the process of gene expression from the activation of transcriptional regulators, through transcription coupled to processing, to the synthesis of a functional protein, with most steps functionally connected.

The CTD of RNAP II orchestrates the coupling of pre-mRNA processing to transcription

RNAP II plays a key role in the coordination of the transcription and processing steps required for mRNA biogenesis. RNAP II is a large multimeric complex (Cramer et al., 2001), the twelve subunits of which are remarkably conserved throughout eukaryotes. The enzymatic activity is located in the largest subunit of RNAP II, Rpb1p. Moreover, Rpb1p contains a carboxy-terminal domain (CTD) composed of 52 repeats of the heptad consensus peptide YSPTSPS. CTD and its phosphorylated residues integrate different nuclear events by binding proteins involved in mRNA biogenesis. The CTD-binding proteins recognize a specific CTD phosphorylation pattern, which changes during the transcription cycle due to the action of CTD-modifying enzymes. Recognition of the CTD phosphorylation patterns by the protein factors is either direct, i.e. by contacting phosphorylated residues or indirect, without contact to the phosphate (Meinhart et al., 2005). For example, the phosphorylated CTD binds tightly to a conserved CTD-interacting domain (CID) present in Pcf11p and Nrd1p, essential and evolutionarily conserved factors required for polyadenylation-dependent and -independent 3' end processing (Meinhart and Cramer, 2004), thus promoting a tight coupling of transcription with 3' end processing.

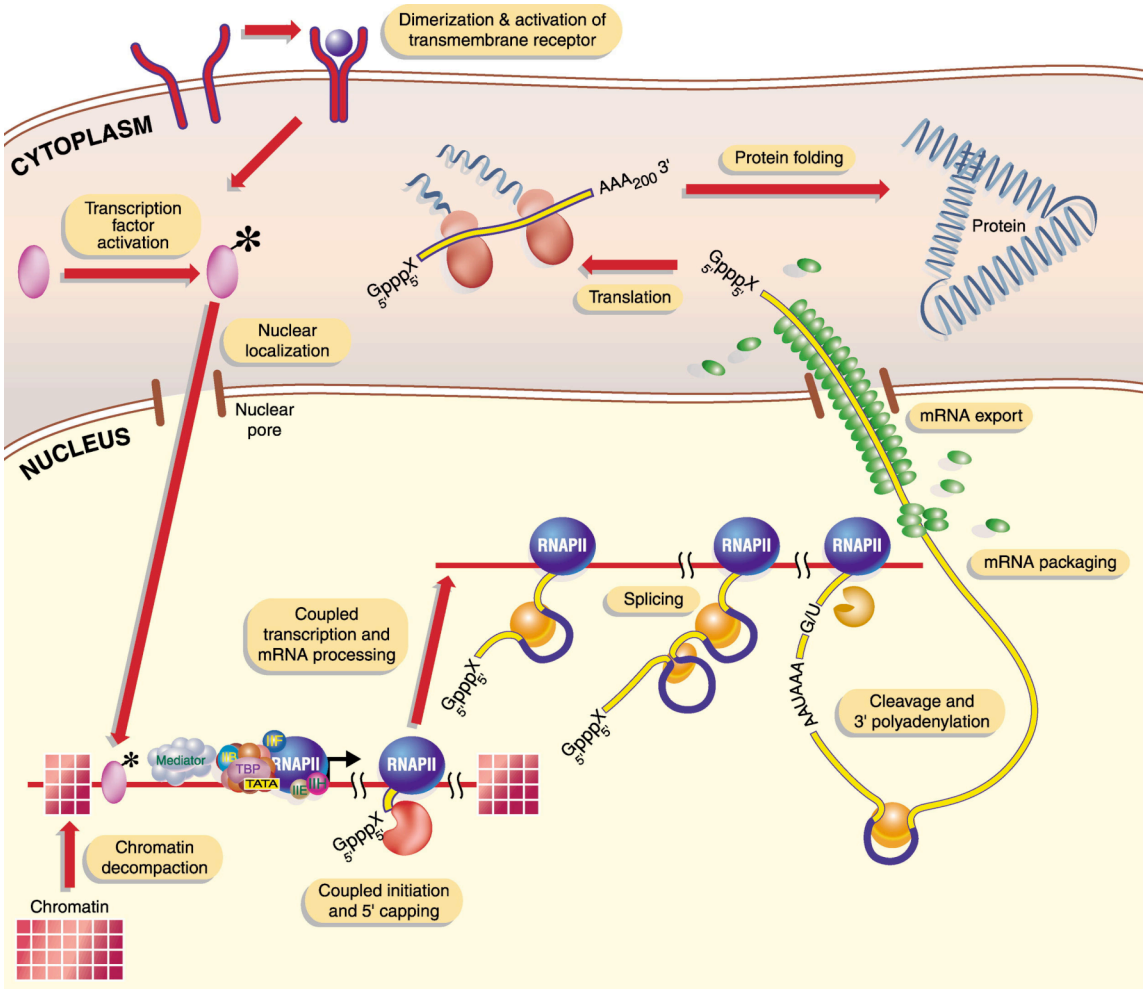


Figure 1.9 Messenger RNA processing and maturation reactions occur co-transcriptionally

Each phase of gene expression from transcription to translation is a subdivision of a continuous process. Each stage is physically and functionally connected to the next, ensuring that there is efficient transfer between manipulations and that no individual step is omitted (see text for details). Taken from Orphanides and Reinberg, 2002.

Capping is tightly coupled to RNAP II transcription

Capping is one of the best described examples of a pre-mRNA processing reaction coupled to transcription. The transcriptional apparatus plays an active role in recruiting the machinery that caps and processes the nascent RNA (Shatkin and Manley, 2000). In yeast, some of the first mRNA processing factors to be recruited to the CTD during the transcription cycle are capping enzymes: the RNA triphosphatase Cet1p, the guanylyltransferase Ceg1p and the 7-methyltransferase Abd1p (Bentley, 2005). After RNAP II transcribes the first 25–30 nucleotides, the cap structure is added at the 5’ end of the emerging RNA. Ceg1p, Cet1p and Abd1p directly interact with the Ser5 phosphorylated CTD *in vivo* (McCracken et al., 1997a). Removal of CTD Ser5 phosphate early in elongation is coupled to release of the capping enzymes (Komarnitsky et al., 2000). Thus, the capping enzymes influence early steps in

transcription and operate a checkpoint to ensure that the 5' end cap structure has been added before the productive elongation of the transcript begins (Bentley, 2005).

Splicing and transcription

Splicing factors are rapidly recruited to nascent pre-mRNAs and introns are removed co-transcriptionally. The CTD seems to play a central role in linking mRNA synthesis and splicing machineries by bridging the splicing factors bound to the 3′ and 5′ splice sites at the ends of each exon. The phosphorylated form of CTD promotes splicing via stimulation of the early steps of spliceosome assembly, whereas the hypophosphorylated form inhibits the formation of these complexes (Hirose et al., 1999). The significance of the CTD role in splicing is underscored by the fact that RNAs generated by truncated versions of RNAP II lacking the CTD cannot be spliced *in vivo* (McCracken et al., 1997b). Furthermore, like capping enzymes, spliceosomal U snRNPs promote and enhance transcription elongation (Fong and Zhou, 2001). The transcriptional recruitment of splicing factors differs between organisms, with more protein–protein contacts observed in metazoans than in yeast. Typical yeast introns are short and only few genes contain more than one of them, whereas the vast majority of metazoan genes contain multiple introns. Consistently, metazoan splicing is enhanced by the phosphorylated CTD, whereas the CTD appears to be far less important for splicing in yeast (Bird et al., 2004; Licatalosi et al., 2002; Millhouse and Manley, 2005). Interestingly, promoter structure can influence splicing in metazoans, suggesting that splicing factors bind at promoters (Kornblihtt et al., 2004). Moreover, pre-mRNA splicing is required for efficient export of the mRNA into the cytoplasm (Reed and Hurt, 2002), thus coupling these two events. Pre-mRNA splicing and packaging of mRNAs for export occurs even as the nascent transcript is spooling off the transcribing RNAP II (Kornblihtt et al., 2004).

Transcription and 3' end processing are interdependent events

Pre-mRNA 3' formation is tightly coupled to transcription and other processing events (Minvielle-Sebastia and Keller, 1999). 3' end processing factors are well established to interact with the CTD of RNAP II and at least in some cases may be carried along the gene with the elongation complex all the way from the initiation to the termination stage of transcription. This suggests that the CTD serves as a landing platform for the pre-mRNA 3' end processing factors and thereby acts as a direct physical link between RNAP II transcription and RNA processing.

The transcriptional recruitment of cleavage and polyadenylation factors is progressive and occurs not only at the 3′ end, but also with a small but significant quantity at the promoter which then increases toward the 3′ end of the gene (Kim et al., 2004a). Such 5′-3′ progressive recruitment results in part from specific recognition of the Ser2-phosphorylated CTD heptads, which are enriched at the 3′ end (Komarnitsky et al., 2000), and of specific binding of mRNA processing factors to the poly(A) site (Kim et al., 2004a). The surprising recruitment of 3′ end processing factors to promoter regions is in some cases associated with their second function

in early transcriptional events. Yeast TFIIB and its coactivator Sub1p interact with Rna15p and Ssu72p (Calvo and Manley, 2001; Krishnamurthy et al., 2004), suggesting an involvement of the 3' end processing factors in the initiation stage of transcription.

Ser2 CTD phosphorylation appears to have a special role in pre-mRNA 3′ end maturation. Two yeast pre-mRNA 3′ end processing factors, Pcf11p and Rtt103p, preferentially bind Ser2-phosphorylated CTD heptads via their CTD-interacting domains (Kim et al., 2004b; Licatalosi et al., 2002). Additionally, other 3′ end mRNA processing factors, such as CPF components Ydh1p and Yhh1p, CF IA subunits Rna14p and Pcf11p, and mammalian CstF50, were shown to bind to the phospho-CTD (Barilla et al., 2001; Dichtl et al., 2002b; Kyburz et al., 2003; Proudfoot, 2004; Sadowski, 2003). Pcf11p carries a specific CTD-binding domain (CID), mutations in which block termination but not pre-mRNA 3′ end cleavage (Sadowski, 2003). Pcf11p was also shown to preferentially bind Ser2 P-CTD, consistent with the role of CF IA in transcription termination (Proudfoot, 2004). Interestingly, Nrd1p, a known termination factor for snoRNA and snRNA genes in yeast, also contains a conserved CID domain, through which it binds to the CTD (Meinhart and Cramer, 2004). This suggests that snoRNA 3' end formation and termination are also coupled to transcription. Moreover, CPF subunit Ssu72p is directly required for pre-mRNA 3′ end cleavage and is involved in RNAP II termination especially on snRNA and snoRNA genes (Dichtl et al., 2002a; Steinmetz and Brow, 2003). Ssu72p has Ser5-specific CTD phosphatase activity, which is required for its function in transcription initiation but not for 3′ end formation (Krishnamurthy et al., 2004). This suggests a role for Ssu72p in coupling transcription initiation, elongation and termination to 3' end processing. Other 3' end processing factors such as Pta1p and Nab4p, also contact the phospho-CTD, at least indirectly (Komarnitsky et al., 2000; Rodriguez et al., 2000). Recruitment of 3′ end pre-mRNA processing factors is inhibited by inactivation of the Ser2/5 kinase Ctk1p (Jones et al., 2004).

The mammalian CTD domain, even if alone and unphosphorylated, directly stimulates 3′ cleavage reaction *in vitro* (Hirose and Manley, 1998), suggesting that this domain is an essential component of the 3′ end processing apparatus. Mammalian CPSF, the factor responsible for recognition of the conserved AAUAAA poly(A) signal sequence, interacts with the CTD. Notably, three subunits of CPSF (CPSF160, CPSF100, and CPSF73) are also associated with the general transcription factor TFIID, and are carried along by the elongating RNAP II (Dantonel et al., 1997), thus pointing to the general requirement of 3' end processing factors in early stages of transcription.

In summary, similarly to other mRNA processing activities, pre-mRNA 3' end processing is tightly coupled to transcription steps. The fact that the CTD of RNAP II has a direct effect on pre-mRNA 3′ end formation is highly consistent with the observations that poly(A) signal recognition is required to terminate transcription (Birse et al., 1998).

Transcription termination and 3' end processing

Transcription termination is defined as release of the transcription complex and nascent RNA from the DNA template. In contrast to RNAP I and III, there are no consensus sequences for the actual site of RNAP II termination as it apparently occurs stochastically within a window downstream of the polyadenylation site (Kim and Martinson, 2003). Correct transcription termination downstream of the poly(A) site is crucial, as it avoids transcriptional interference at downstream promoters (Greger and Proudfoot, 1998) and at chromosomal elements such as centromers and origins of replication. Furthermore, termination facilitates RNAP II recycling and by maintaining a pool of available polymerases enhances gene expression (Dye and Proudfoot, 1999). In metazoan cells, termination occurs in two steps, CTD-independent pausing followed by CTD-dependent polymerase release (Park et al., 2004). The pausing possibly echoes a rearrangement within the elongation complex, while the second step, CTD-dependent release, is related to recruitment of the polyadenylation factors by CTD phosphorylation (Buratowski, 2005).

The connection between 3' end processing and transcription termination was first suggested by the observation that mutations of the *cis*-acting poly(A) signals not only inhibit cleavage and polyadenylation, but also impair the correct RNAP II termination process (reviewed in: Buratowski, 2005; Proudfoot, 2000; Proudfoot, 2004; Proudfoot et al., 2002). Moreover, CPF subunits Yhh1p/Cft1p and Ssu72p, and CF IA subunits Rna15p, Rna14p, and Pcf11p that are co-transcriptionally recruited to the CTD of RNAP II, were also shown to be directly required for transcription termination on mRNA genes (Barilla et al., 2001; Birse et al., 1998; Dichtl et al., 2002a; Dichtl et al., 2002b).

Two general models for coupling of the 3' end processing and transcription termination on mRNA genes have been proposed (reviewed in: Buratowski, 2005; Luo and Bentley, 2004; Proudfoot et al., 2002; see also Figure 1.10). In the 'anti-terminator' model, the elongating RNAP II transcription complex undergoes certain rearrangements in response to the emerging polyadenylation sequences on the RNA, likely by recruiting polyadenylation and termination factors, or by displacing positive elongation factors (Kim et al., 2004a; Logan et al., 1987). In consequence, this renders the RNAP II complex 'termination competent', leading to the loss of processivity and gradual termination. It is unclear at which point in the transcription process the 3' end cleavage actually occurs, and therefore whether cleavage is needed for termination. Several lines of evidence support the anti-terminator model of termination. Run-on analyses of the human ε- and β-globin genes transcripts (Dye and Proudfoot, 1999), RT-PCR analysis of the uncleaved nascent transcripts of the Balbiani ring 1 gene of *Chironomus tentans* (Bauren et al., 1998), and electron microscopic analysis of mini-chromosomes injected into *Xenopus leavis* oocytes (Osheim et al., 1999) showed that transcription termination occurs in the 3' end flanking region of the gene, approximately 1 kB downstream of the polyadenylation signal.

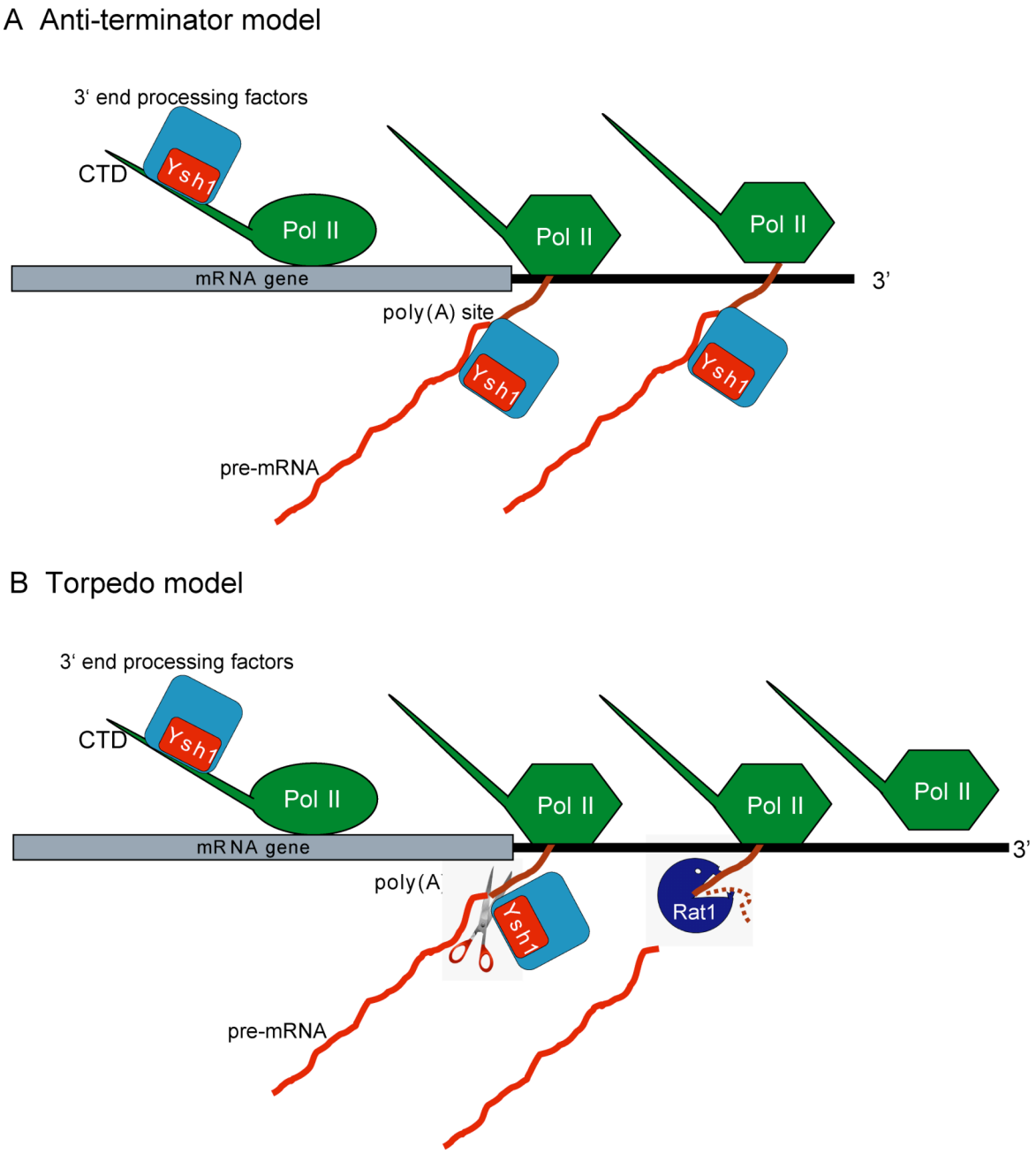


Figure 1.10 Anti-terminator and torpedo models for coupling 3' end processing and RNAP II transcription termination

(**A**) The anti-terminator model: the emergence of the polyadenylation signal on the RNA transcript and its recognition by CTD-associated 3' end processing factors (blue, with the 3' endonuclease marked in red) causes the RNAP II (green) to become less processive and triggers termination independently of the endonucleolytic cleavage.

(**B**) The torpedo model: cleavage of the transcript by the polyadenylation machinery (blue, with the 3' endonuclease marked in red) generates a new, uncapped 5′ end that is a substrate for degradation by the Rat1p nuclease. This in turn interferes with the paused RNAP II and triggers transcription termination.

Importantly, in each case the large majority of nascent RNAs detected right up to the apparent termination site were uncleaved. Moreover, a termination deficient mutant of the Rna15p homologue in *S. pombe* showed no defect in pre-mRNA 3' end processing (Aranda and Proudfoot, 2001). This suggests that transcription termination occurs before cleavage at the poly(A) site, even though the requirement of a functional poly(A) site for termination was clearly shown (Proudfoot et al., 2002).

The second model, named the 'torpedo' model, proposes that RNAP II transcription termination is triggered by poly(A) site cleavage and subsequent degradation of the downstream RNA by Rat1p 5'–3' exonuclease. Rat1p catches up with the elongation complex and causes it to terminate (Connelly and Manley, 1988; Kim et al., 2004b; Proudfoot, 1989; West et al., 2004). The yeast Rat1p 5′–3′ RNA exonuclease is progressively recruited to mRNA genes in a pattern similar to 3' end processing factors. Rat1p associates with the RNAP II complex via interaction with Rtt103 protein, which contains a CID related to that of Pcf11p, and binds to Ser2 phosphorylated CTD. Consistently, Rat1p and Rtt103p crosslink at the polyadenylation site, suggesting that they play a role in 3′ end processing. Mutant *rat1* and *rtt103* strains are competent for 3' end formation, but termination is strikingly defective (Kim et al., 2004b). Xrn2 is the homolog of yeast Rat1p in higher eukaryotes. RNAi knockdown of Xrn2 also leads to termination defects in mammalian cells, suggesting conservation of an exonuclease-mediated termination mechanism (West et al., 2004). Apparently, following the poly(A) site cleavage, Rat1p/Xrn2 dissociates from the polymerase and attacks the 5′ end of the nascent RNA generated by cleavage, and digests it exonucleolytically, which eventually leads to the RNAP II release from the DNA template (Buratowski, 2005). It is however unclear whether Rat1p can directly trigger transcription termination, or if other factors are required. Interestingly, Sen1p helicase, implicated in snoRNA 3′ end formation and in termination on short mRNA genes (Steinmetz et al., 2001; Steinmetz et al., 2006b), is a strong candidate for the enzyme that may actually trigger transcript release.

The most important difference between the two models is that cleavage of the nascent RNA at the poly(A) site is a necessary step in the torpedo model, while it is not required for transcription termination in the anti-terminator model. Moreover, whereas an allosteric change in polymerase could be effected by the cleavage factors themselves, the torpedo model requires that Rat1p exonuclease function as a dedicated termination factor that must act after cleavage has been completed. Results of studies on RNAP II transcription termination provide support for both the anti-terminator and the torpedo models, and indicate that these two models are not mutually exclusive. It is possible that termination can occur by more than one mechanism. An integrated model of transcription termination has been proposed recently, in which, as a consequence of cleavage at the poly(A) site and degradation of the nascent downstream RNA, the RNAP II complex undergoes allosteric changes that promote its release from the template (Luo et al., 2006).

In respect to Ysh1p, the yeast endonuclease acting in the 3' end processing, chapter 2 of this work demonstrates that cleavage is essential for correct transcription termination on mRNA genes, and possibly also on a subset of snoRNA genes. These observations strongly support the torpedo model of RNAP II transcription termination.

Coupling of snRNA and snoRNA 3' end formation to transcription termination

Although most of the snRNAs and snoRNAs are transcribed by RNAP II, they are neither spliced nor polyadenylated. Many snoRNAs, especially in mammals, are cleaved out from the introns of other RNAP II genes and therefore do not possess termination signals. SnoRNA genes have a similar termination process as mRNA genes, requiring RNA processing and termination elements (Steinmetz et al., 2001; Steinmetz et al., 2006a). The most important factor acting in coupling of snRNA/snoRNA 3' end formation and termination is Nrd1p (Steinmetz et al., 2001). The CID of Nrd1p interacts with the CTD of RNAP II. It also contacts an RNA binding protein Nab3p, which in turn is thought to recruit the CTD kinase, Ctk1p. *nrd1* mutant strains are defective in both 3' end formation and termination of snoRNA genes (Steinmetz et al., 2001), suggesting that the mechanism of transcriptional termination for many yeast snRNA/snoRNA genes is similar to that of mRNA genes. The mechanism of snRNA/snoRNA 3' end formation, and factors and sequences involved in both 3' end processing and transcription termination on yeast snRNA and snoRNA genes are described in part 1.2 of this chapter.

Histone 3' end processing and termination

In metazoans replication-dependent histone mRNAs are neither spliced nor polyadenylated, but instead are cleaved at a highly conserved stem-loop structure (Marzluff, 1992). Mature histone 3′ ends are formed from longer pre-mRNAs by endonucleolytic cleavage downstream of the stem-loop by the 3' end processing endonuclease CPSF73 (Dominski et al., 2005). Similarly to other mRNA genes, RNAP II transcription terminates in a certain distance downstream of the histone 3′ processing site. However, termination of a proportion of histone pre-mRNAs is dependent on the presence of functional histone processing signals rather than the cleavage reaction itself (Dominski et al., 1999). Notably, deletion of either the 3' end stem-loop, or the histone downstream element (HDE), required for 3' end processing of replication-dependent histone pre-mRNAs, abolishes transcriptional termination. This suggests that 3' end processing of replication-dependent histone pre-mRNAs is also mechanistically coupled to transcription termination (Chodchoy et al., 1987). Interestingly, hZFP100, a factor associated with the U7 snRNP and stimulating histone 3' end processing, possibly plays a role in transcription termination on the histone genes, thus coupling this event to the nascent pre-mRNA 3' end processing (Dominski and Marzluff, 1999).

Nevertheless, histone mRNAs in yeast, protists, trypanosomes, fungi and plants are all polyadenylated (Dominski et al., 1999). The yeast histone pre-mRNAs are processed by

the canonical 3' end cleavage and polyadenylation apparatus (Butler et al., 1990), and similarly, transcription termination on the histone genes is coupled to 3' end processing. Interestingly, similarly to the metazoan histone mRNAs, plant histone mRNAs lack introns, although introns are present in the histone genes of *Neurospora* (Dominski et al., 1999). This suggests that the unique 3′ end found in metazoan histone mRNAs arose late in evolution after the divergence of animals and plants.

Coupling mRNA export with transcription, splicing and pre-mRNA processing

Fully processed mRNAs must be transported to the site of protein translation in the cytoplasm. mRNAs are transported as mRNP particles through the same nuclear pore channels used to transport proteins to and from the nucleus (Reed and Hurt, 2002). Inefficient mRNP assembly impairs transcription elongation, and transport factors are deposited onto the RNA during splicing and 3' end formation, indicating tight interdependence of these processes (Vinciguerra and Stutz, 2004). Moreover, pre-mRNA splicing was shown to promote nuclear export, and members of the SR family of splicing factors are involved in mRNA transport and in regulating mRNA stability, thus suggesting that the two processes may be coupled (Luo and Reed, 1999; Huang and Steitz, 2001; Lemaire et al., 2002). It is possible that this coupling is facilitated by direct interactions between mRNA splicing and export factors. Both the export and splicing machineries are tightly connected in metazoans, where most genes contain introns. In contrast, in *S. cerevisiae*, where only a few genes contain introns, it appears that mRNA export is more extensively coupled to 3' end processing and other steps in gene expression (Reed, 2003).

Genetic analysis in yeast pointed to the coordination between polyadenylation and mRNA transport. mRNA transport is enhanced as a result of more efficient polyadenylation (Huang et al., 1999). Certain alleles of *RNA14*, *RNA15* and *PAP1* cause defects in both mRNA export and *in vitro* processing. Deletion of the *cis*-acting sequences required to couple 3' processing and termination also produces RNAs that fail to be exported from the nucleus (Hammell et al., 2002). This suggests that transcription termination, 3′ processing and mRNA export are all coupled.

CHAPTER 2

The role of the Ysh1p endonuclease in 3' end processing and in transcription termination of RNA polymerase II transcripts

2.1 Introduction

2.2 Results

The β-lactamase and β-CASP domains of Ysh1p are evolutionary conserved

The conserved β-lactamase HXHXDH signature is essential for cell viability

YSH1 mutant strains display temperature sensitive phenotypes

Ysh1p is required for both the cleavage and polyadenylation steps of the pre-mRNA 3' end formation

mRNA levels are reduced in the *ysh1-32* strain at non-permissive temperature

The average length of poly(A) tails is shortened in *ysh1-32* mutant cells

Ysh1p is required for correct poly(A) site recognition of *ACT1* pre-mRNA

Ysh1p is required for the RNAP II transcription termination on mRNA genes *in vivo*

Read-through mRNA transcripts accumulate in *ysh1-12* mutant cells

3' extended snoRNAs are produced in the *ysh1-12* mutant strain

Transcription termination of the SNR3 snoRNA gene is impaired in the *ysh1-12* strain

The levels of snoRNAs remain stable in *ysh1-12* and *ysh1-32* mutant strains *in vivo*

Conditional depletion of Ysh1p does not affect snoRNA 3' end processing

The snoRNA terminator sequences are cleaved by CPF *in vitro*

Involvement of Ysh1p in the autoregulation of *NRD1* mRNA levels

YSH1 genetically interacts with the nuclear exosome subunit *RRP6*

Splicing is impaired in the *ysh1-12* mutant strain

ysh1 mutants are deficient in early pre-rRNA processing steps

2.3 Discussion

2.4 Experimental procedures

2.5 Additional results

Cross-linking CPF/Ysh1p to the pre-mRNA 3' end cleavage site

Ysh1p interacts with subunits of CPF and CF IA complexes.

Tethering Ysh1p to RNA substrates

A subset of the *ysh1* temperature sensitive mutants is deficient in 3' end processing *in vitro*

2.6 Additional experimental procedures

2.1 Introduction

RNA polymerase II (RNAP II) transcribes protein encoding genes and a subset of non-coding small nuclear RNAs (snRNA) and small nucleolar RNAs (snoRNA). In order to be biologically functional all primary transcripts of RNAP II need to undergo extensive processing. For mRNAs this includes capping at the 5' end, removal of intronic sequences and 3' poly(A) addition.

Pre-mRNA 3' end processing is initiated by endonucleolytic cleavage at the poly(A) site, followed by polyadenylation of the upstream cleavage product. In yeast, 3' end processing is performed by a large complex of proteins, including cleavage and polyadenylation factor (CPF), cleavage factor IA (CF IA) and cleavage factor IB (CF IB) (Zhao et al., 1999a). Ysh1p/Brr5p is associated with CPF and was first identified based on sequence homologue to mammalian CPSF73 (Chanfreau et al., 1996; Jenny et al., 1996). It was recognized early on that the protein carries a highly conserved β-lactamase fold commonly found in metal-dependent hydrolytic enzymes (Aravind, 1999; Daiyasu et al., 2001; Callebaut et al., 2002). Although this suggested that catalytic activity was associated with Ysh1p, convincing structural and biochemical evidence for endonucleolytic activity is only recently emerging (Ryan et al., 2004; Wickens & Gonzalez, 2004; Mandel et al., 2006).

Cleavage and polyadenylation factors are required for termination of RNAP II transcription (Connelly & Manley, 1988; Proudfoot, 1989; Proudfoot, 2004; Buratowski, 2005). Two general models have emerged to mechanistically explain termination (Proudfoot et al., 2002; Luo & Bentley, 2004; Bentley, 2005; Buratowski, 2005). In the 'anti-terminator' model, RNAP II complex undergoes conformational changes in response to the emerging terminator sequences on RNA, either by recruiting termination factors, and/or by displacing positive elongation factors (Logan et al., 1987; Zhang et al., 2005; Zhang & Gilmour, 2006). In consequence, this renders the RNAP II complex termination competent, leading to the loss of processivity and gradual termination. The 'torpedo' model proposes that RNAP II transcription termination is triggered by poly(A) site cleavage and subsequent degradation of the 3' downstream RNA by Rat1p 5'-3' exonuclease (Connelly & Manley, 1988; Kim et al., 2004; West et al., 2004). Experimental support for both models has been provided in diverse eukaryotic experimental systems (Gilmour & Fan, 2008) and an unified model of termination has been proposed that combines mechanistic predictions of both models (Luo et al., 2006).

All snRNAs and most snoRNAs in yeast are synthesized independently by RNAP II. 3' ends of snoRNA transcripts are produced by 3'-5' exonucleolytic trimming that follows either endonucleolytic cleavage or RNAP II termination (Chanfreau et al., 1998; Allmang et al., 1999; van Hoof et al., 2000; Butler, 2002). Numerous factors were reported to be essential for transcription termination on snoRNA genes, including the RNAP II subunits Rpb3p and Rpb11p, the sequence-specific RNA binding proteins Nab3p and Nrd1p, the Sen1p helicase, the CTD kinase Ctk1p, and the RNAP II associated Paf1 complex (Ursic et al., 1997; Conrad et al., 2000; Steinmetz et al., 2001; Sheldon et al., 2005; Steinmetz et al., 2006). Furthermore, components of the 3' end processing complexes CF IA (Pcf11p, Rna15p and Rna14p) and several CPF subunits (Pta1p, Pti1p, Ref2p, the CTD phosphatase Ssu72p, Swd2p and Glc7p) were implicated in

snoRNA termination (Fatica et al., 2000; Morlando et al., 2002; Dheur et al., 2003; Ganem et al., 2003; Nedea et al., 2003; Steinmetz & Brow, 2003; Cheng et al., 2004; Dichtl et al., 2004; Kim et al., 2006; Nedea et al., 2008).

In this work we isolated a number of conditional *ysh1* mutant alleles and characterized them for defects in mRNA and snoRNA synthesis. We found that all analyzed *ysh1* mutants were deficient in pre-mRNA 3' end formation and RNAP II transcription termination on mRNA genes. Moreover, a cold-sensitive *ysh1* mutant strain displayed distinct defects in snoRNA 3' end formation, termination on snoRNA genes and RNA splicing. Most interestingly, we also provide evidence that imply endonucleolytic cleavage and functional Ysh1p in the process of regulated premature termination at the *NRD1* locus. Altogether, this work underscores the central role that is played by the 3' endonuclease in cellular RNA metabolism.

2.2 Results

The β-lactamase and β-CASP domains of Ysh1p are evolutionary conserved

The β-lactamase and the β-CASP domains of Ysh1p are highly conserved throughout evolution. Comparative sequence analysis revealed orthologues of Ysh1p in essentially every eukaryotic species, and in several Archaea (summarized in Table 2.1). One of the closest homologues of Ysh1p is the *AGR279C* gene of *Ashbya gossypii*, being 70% identical and 86% similar over the entire amino acid sequence. The Ysh1p homologue in *Schizosaccharomyces pombe* shares 48% identity and 66% similarity with its *S. cerevisiae* counterpart. Mammalian, invertebrate and plant homologues share a general level of homology with Ysh1p of around 40% identity and 60% similarity (see alignment in Figure 2.1). The C-terminal domain of Ysh1p is present exclusively in other *Saccharomyces* species and in only one other fungus (*Candida glabrata*) but it is not found in metazoans. Interestingly, the yeast cleavage and polyadenylation machinery contains a component, Syc1p, that shares a high level of homology with the C-terminus of Ysh1p.

Table 2.1 Eukaryotic and archaean homologues of Ysh1p

Species	Gene name	Swiss-Prot accession no.	% identity to Ysh1p	% similarity to Ysh1p
Ashbya gossypii	AGR279C	Q74ZC0	70	86
Schizosaccharomyces pombe	SPAC17G6.16c	O13794	48	66
Homo sapiens	CPSF73	Q9UKF6	40	60
Bos taurus	CPSF73	P79101	43	63
Mus musculus	CPSF73	Q9QXK7	40	59
Caenorabditis elegans	Y67H2A.1	NM_070152	42	60
Drosophila melanogaster	CG7698-PA	Q9VE51	41	62
Oryza sativa	Os03g0852900	Q84JJ2	44	63
Arabidopsis thaliana	At1g61010	AY074280	41	59
Methanococcus jannaschii	MJ1236	Q60355	32	50

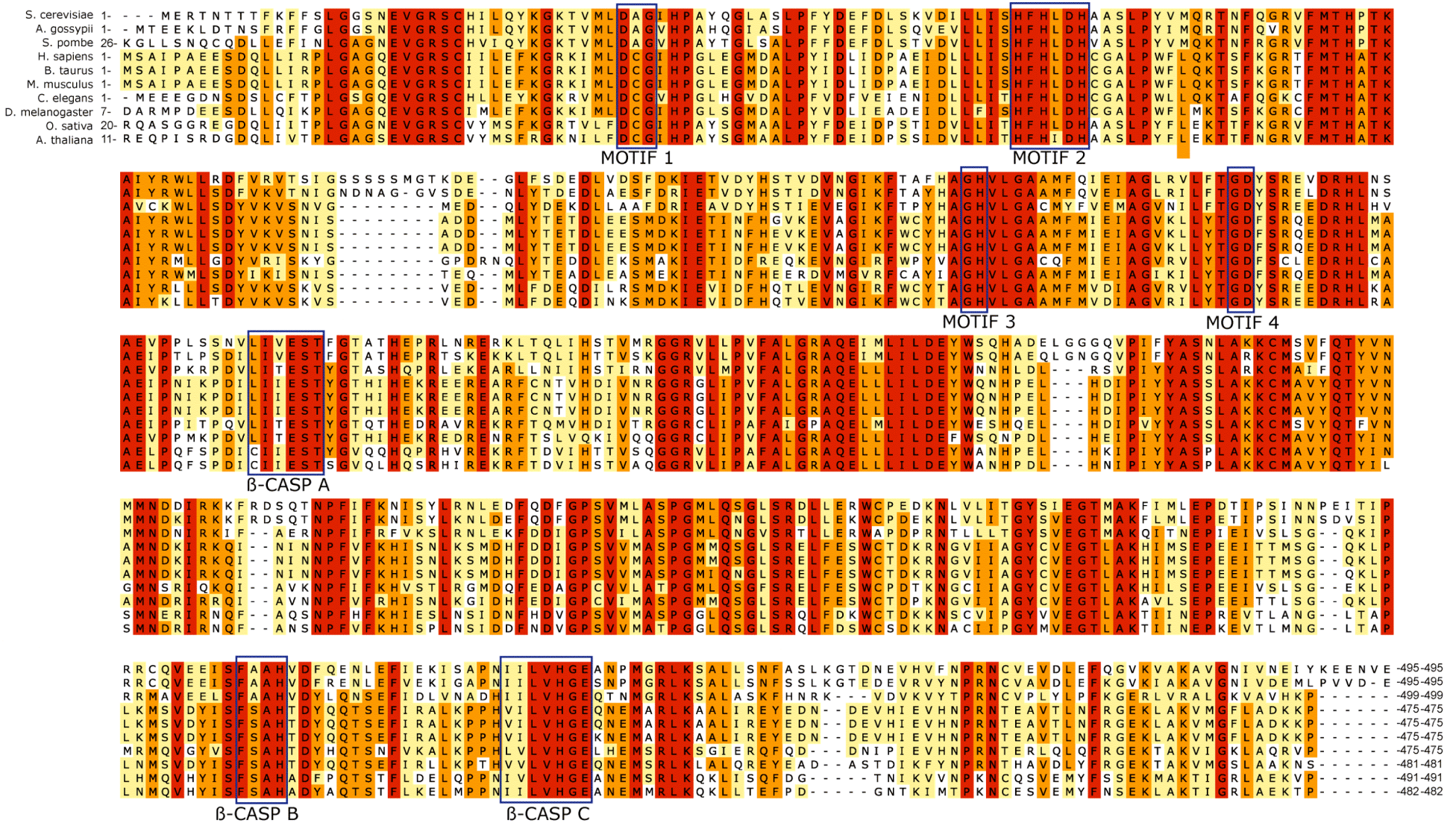


Figure 2.1 The β-lactamase and β-CASP domains of Ysh1p are evolutionary conserved

Sequence of β-lactamase and β-CASP domains of Ysh1p (YLR277C, Swiss-Prot accession number: Q06224) is aligned with its homologues in *Ashbya gossypii* (AGR279Cp, acc. no. Q74ZC0), *Schizosaccharomyces pombe* (SPAC17G6.16c, acc. no. O13794), *Homo sapiens* (CPSF73, acc. no. Q9UKF6), *Bos taurus* (CPSF73, acc. no. P79101), *Mus musculus* (CPSF73, acc. no. Q9QXK7), *Caenorabditis elegans* (Y67H2A.1, acc. no. NM_070152), *Drosophila melanogaster* (CG7698-PA, acc.no. Q9VE51), *Oryza sativa* (Os03g0852900, acc. no. Q84JJ2), *Arabidopsis thaliana* (At1g61010, acc. no. AY074280). Invariant residues are shown in red boxes (100% conservation); orange boxes indicate conserved and similar amino acids, which are present in 70-100% of the aligned sequences, and yellow boxes mark conserved and similar residues common in 30% to 70% of the aligned sequences. Conserved motifs of β-lactamase (motifs 1-4) and β-CASP (A-C) domains are marked in frames.

The conserved β-lactamase signature $H_{68}X_{69}H_{70}X_{71}D_{72}H_{73}$ is essential for cell viability

To address the cellular role of the putative 3' endonuclease Ysh1p we initially produced point mutations within the β-lactamase consensus signature motif $H_{68}X_{69}H_{70}X_{71}D_{72}H_{73}$, which is located at the amino-terminal end of the protein and which contributes to formation of the catalytic core of the protein (Figure 1C; Aravind, 1999; Mandel et al., 2006). Single alanine substitutions of highly conserved histidines H_{68} and H_{70} or of aspartate D_{72} were introduced into the *YSH1* gene. Such mutations would presumably destroy the zinc ligands in motif 2 of the β-lactamase fold, and thus abort the catalytic activity of the enzyme. Indeed, these substitutions resulted in cell lethality as none of them were able to complement for the chromosomal *YSH1* gene deletion when the plasmid harbouring wild-type *YSH1* gene and *URA3* marker was counterselected on medium containing 5-fluoro-orotic acid (5-FOA) (Figure 2.2). This result supported the significance of the β-lactamase signature motif $H_{68}X_{69}H_{70}X_{71}D_{72}H_{73}$ and thus of the whole β-lactamase domain for the function of Ysh1p *in vivo*.

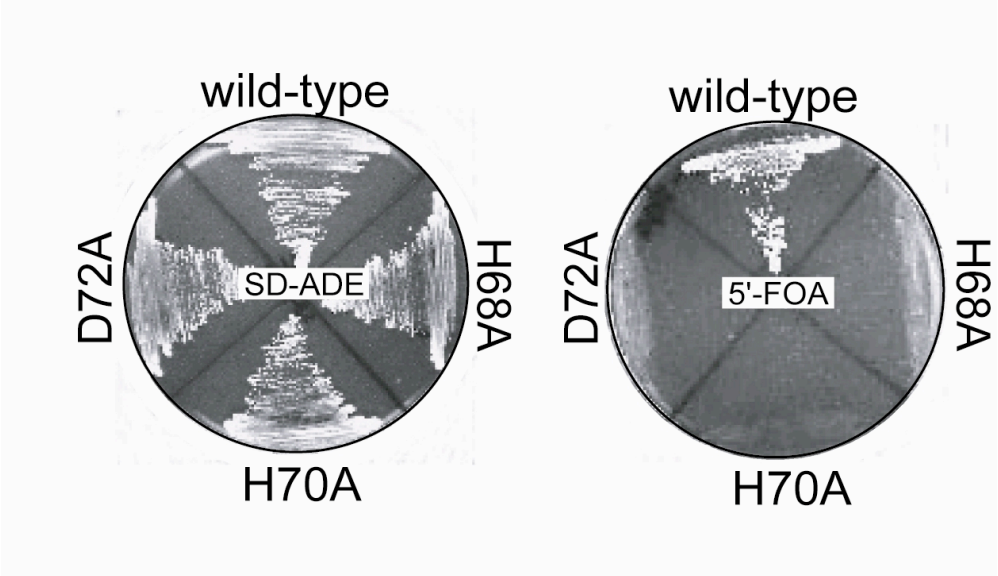


Figure 2.2 Conserved β-lactamase signature $H_{68}X_{69}H_{70}X_{71}D_{72}H_{73}$ is essential for cell viability

Conserved β-lactamase signature $H_{68}X_{69}H_{70}X_{71}D_{72}H_{73}$ is essential for cell viability. Plasmid shuffling was used to test the requirement of H_{68}, H_{70} and D_{72} residues of the β-lactamase consensus motif for cell viability. LM109 strain with disrupted chromosomal *YSH1* gene and carrying *YSH1-URA3* plasmid was transformed with *ADE2*-plasmid bearing either wild-type *YSH1* gene or its mutant version $H_{68}A$, $H_{70}A$ and $D_{72}A$, followed by counterselection on 5' FOA plates. The $H_{68}X_{69}H_{70}X_{71}D_{72}H_{73}$ mutant strains were constructed by B. Dichtl.

YSH1 mutant strains display temperature sensitive phenotypes

The observation on the β-lactamase $H_{68}X_{69}H_{70}X_{71}D_{72}H_{73}$ signature being essential for yeast viability underscored the functional significance of this motif but hampered further functional analyses. Therefore, we isolated temperature-sensitive (ts) and cold-sensitive (cs) alleles of the *YSH1/BRR5* gene using random mutagenesis (Figure 2.3 B; see Materials and Methods). Of the mutant strains chosen for further analysis three were ts, with *ysh1-32* being lethal at 33ºC, and *ysh1-13* and *ysh1-15* being lethal at 37ºC; in contrast, the cs *ysh1-12* strain ceased growth at 15ºC (Figure 2.3 A). Amino acid changes within mutant Ysh1p proteins were revealed by DNA sequencing and found to localize predominantly to the non-conserved C-terminus (Figure 2.3 B); this part of the protein is involved in both catalytic steps of 3' end processing and mediates the interaction with Pta1p (Zhelkovsky et al., 2006). Most likely, in each of the alleles, all the mutations together contribute to the observed ts or cs phenotype, as reduction of the number of mutations in each mutant allele resulted in reverting to a temperature-resistant phenotype (data not shown).

Next, we analyzed growth behaviour of the *ysh1* ts and cs alleles upon shift to restrictive temperatures. Growth of the *ysh1-15* mutant cells slowed down after 2 to 3 hours incubation at 37ºC when compared to the wild-type cells, but the cells did not cease to divide. The other *ysh1* temperature sensitive alleles, *ysh1-13* and *ysh1-32* ceased growth after approximately 3 and 4 hours shift to 37ºC, respectively. *rna15-1* strain was previously reported to be severely deficient in 3' end formation (Minvielle-Sebastia et al., 1994). Here, it was used as a reference control. In contrast to the *ysh1* mutant cells, *rna15-1* strain discontinued growth within 1 hour of incubation at 37ºC (Figure 2.3 C). The cold-sensitive *ysh1-12* mutant cells ceased growth only after 5 hours of incubation at 15ºC, when compared to the wild-type strain (Figure 2.3 D).

Ysh1p is required for both the cleavage and polyadenylation steps of pre-mRNA 3' end formation

Initially, we tested the *ysh1* mutant strains for cleavage and polyadenylation *in vitro* (Figure 2.4). Total cell extracts were produced from wild-type and mutant strains *(ysh1-12, ysh1-13, ysh1-15, ysh1-32*) and assayed on a synthetic *CYC1* substrate for cleavage, and on pre-cleaved *CYC1-pre* RNA for polyadenylation, respectively (Minvielle-Sebastia et al., 1994). Comparing the effect of the mutations to wild-type, we observed that both cleavage and polyadenylation were reduced in all the *ysh1* mutant strains, with the defects being more pronounced at respective restrictive temperatures (Figure 2.4 A, B).

At permissive temperature (30ºC), cleavage occurred efficiently only in the wild-type extract (Figure 2.4 A, upper panel, lane 2). *ysh1-12* showed some residual, partially unspecific cleavage activity (Figure 2.4 A, upper panel, lane 3). Such residual activity was even less pronounced in *ysh1-13, ysh1-15* and *ysh1-32* strains (Figure 2.4 A, upper panel, lanes 4-6, respectively).

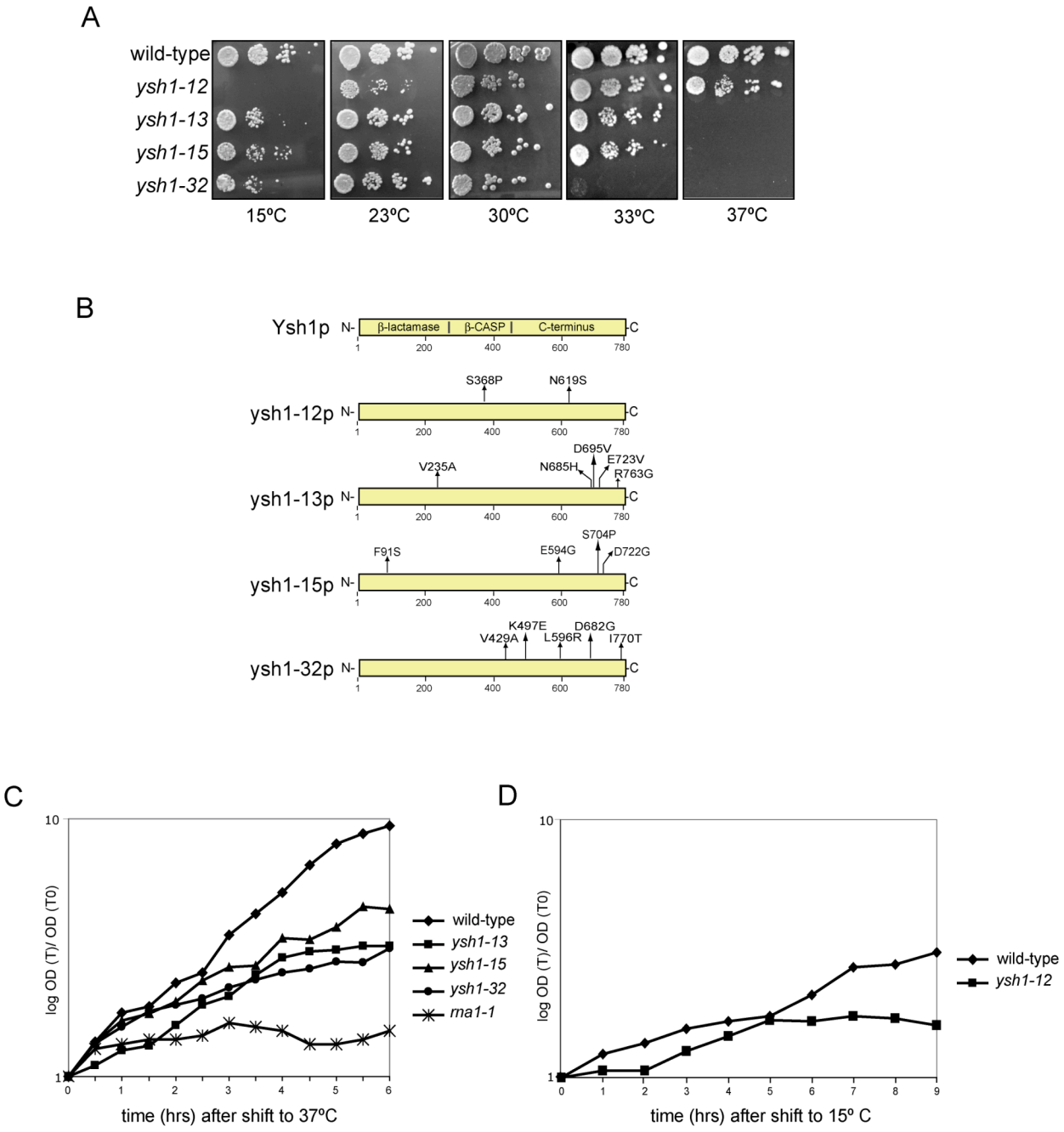


Figure 2.3 *YSH1* mutant strains display temperature sensitive phenotypes

(**A**) Tenfold serial dilutions of cultures of wild-type and *ysh1* mutant strains were spotted onto YPD medium and incubated for 3 days at 23-37ºC or 5 days at 15ºC. Battery of the *ysh1* conditional alleles was constructed by B. Dichtl. (**B**) Schematic representation of the *ysh1* mutant sequences underlying respective temperature-sensitive/cold-sensitive phenotypes. (**C**) Growth curves of wild-type and temperature-sensitive mutant *ysh1* strains at 37ºC. *rna15-1* strain was used in parallel. (**D**) Growth curves of wild-type and cold-sensitive mutant *ysh1-12* strains at 15ºC.

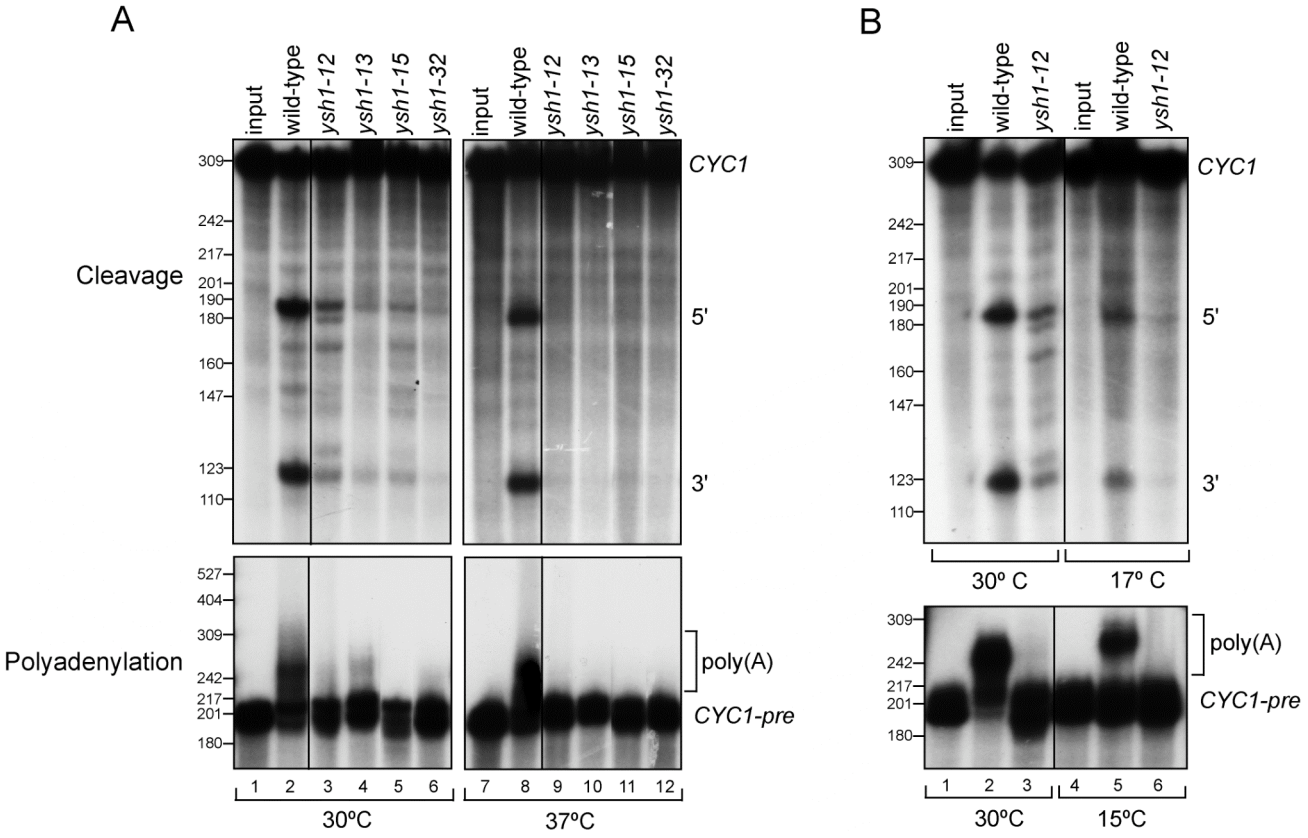


Figure 2.4 Ysh1p is required for pre-mRNA 3' end cleavage and polyadenylation *in vitro*

(**A**) *In vitro* cleavage (upper panel) and polyadenylation (lower panel) assays with extracts prepared from wild-type and *ysh1* temperature sensitive strains as indicated. Input indicates a control where no protein was added. Positions of substrate RNAs, 5' and 3' end cleavage products, and polyadenylation products bands are showed. *Hpa*II-digested pBR322 fragments were 5' end labelled and served as markers. Internally [^{32}P]-labelled substrate RNAs, CYC1 for cleavage assay and CYC1-pre for specific polyadenylation, were used. Reactions were performed either at 30ºC (lines 1-6) or at 37ºC (lines 7-12). Extracts were pre-incubated at restrictive temperatures for 10 min. prior to assaying. (**B**) as in (A), except reactions were performed either at permissive temperature 30ºC (lines 1-3) or at non-permissive temperatures 17ºC (lines 4-6) for cleavage, and at 15ºC for polyadenylation.

Cleavage activity was completely abolished in all four mutant strains at 37ºC (Figure 2.4 A, upper panel, lanes 9-12). In the polyadenylation assay, weak activity was detectable at 30ºC for *ysh1-12*, *ysh1-13* and *ysh1-32*, but not *ysh1-15* allele (Figure 2.4 A, lower panel, lanes 3-6). This residual polyadenylation activity was lost when the above extracts were assayed at 37ºC (Figure 2.4 A, lower panel, lanes 9-12). To correlate the observed lethality of the *ysh1-12* cells *in vivo* at 15ºC with an *in vitro* phenotype, we tested the *ysh1-12* cell extract for cleavage and polyadenylation activity at its restrictive temperature (Figure 2.4 B). Total loss of cleavage activity was observed for the *ysh1-12* allele, compared to wild-type, when incubated at 17ºC (Figure 2.4 B, upper panel, compare lanes 5 to 6). The cleavage assay presented here was performed at

17ºC, and not at 15˚C, as extracts of several wild-type strains were inactive in cleavage (although being active in the polyadenylation step) when assayed at 15ºC. Polyadenylation activity was abolished in the *ysh1-12* strain at 15ºC (Figure 2.4 B, lower panel, lane 6).

To confirm that the observed defects in pre-mRNA 3' end formation were due to inactive CPF, the factor which contains Ysh1p, we aimed to reconstitute activities by addition of purified wild-type CPF to the *ysh1-32* extract (Figure 2.5). We found that CPF partially rescued the cleavage defect of the mutant extract at both permissive and restrictive temperatures (Figure 2.5 A, lanes 4, 9). Likewise, specific polyadenylation activity of the *ysh1-32* extract was restored by addition of CPF (Figure 2.5 B, lanes 4, 9). CPF alone was not able to cleave the *CYC1* RNA (Figure 2.5 A, lanes 5 and 10) and gave unspecific hyperadenylation of the *pre-CYC1* substrate as previously observed (Figure 2.5, lanes 5 and 10; Preker et al., 1997). Taken together, the above results underscore an important role for Ysh1p in both catalytic steps of pre-mRNA 3' end formation.

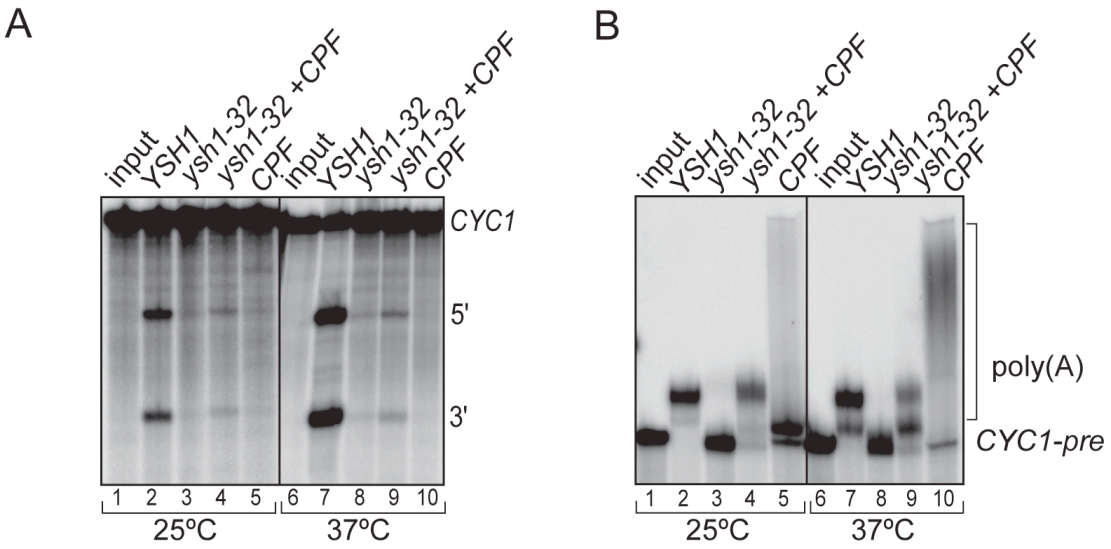


Figure 2.5 Reconstitution of specific cleavage and polyadenylation activities of the *ysh1-32* extract *in vitro*
Reconstitution of specific cleavage and polyadenylation activities of the *ysh1-32* extract *in vitro*. Cleavage (A) and polyadenylation (B) assays were performed essentially as in (A), with extracts prepared from wild-type and *ysh1-32* strain as indicated. The *ysh1-32* extract was combined with purified CPF to restore the specific 3' end processing activity. 2-times more CPF was used for the reconstitution of polyadenylation than for the reconstitution of cleavage. Reactions were performed as in Figure 2.4.

Distinct effects of mutations within Ysh1p on mRNA abundance

Yeast strains defective in pre-mRNA 3' end processing produce mRNAs without functional poly(A) tails that are unstable and rapidly directed for degradation. Therefore, we performed Northern blotting to analyze the steady-state levels of mRNAs in *ysh1* mutant strains before and after shift to restrictive temperatures. We found that several tested mRNAs (*ASC1, ASN1, PGK1, RPS16A*) were significantly reduced in *ysh1-32* strains following growth at 37ºC (Figure 2.6, lanes 6-8); similar phenotypes were observed for *ysh1-13* and *ysh1-15* mutant strains (data not shown). The well-characterized *rna15-1* strain was used as a control and showed the expected strong reduction of mRNAs after 1 h shift to 37ºC (lane 4; Minvielle-Sebastia et al., 1994). Unexpectedly, a reverse effect was observed in *ysh1-12* cells, where levels of all tested mRNAs were relatively stable or even increased upon shift to non-permissive temperature (lanes 12-14); this included *NAB2* mRNA, which increased up to 13-fold. Interestingly, Nab2p is believed to autoregulate levels of its own mRNA in a process that requires the nuclear exosome component Rrp6p (Roth et al., 2005) and our results suggest a role for Ysh1p in this process (see Discussion). Levels of RNAP I transcript *18S* rRNA and RNAP III transcript S*CR1* served as loading controls.

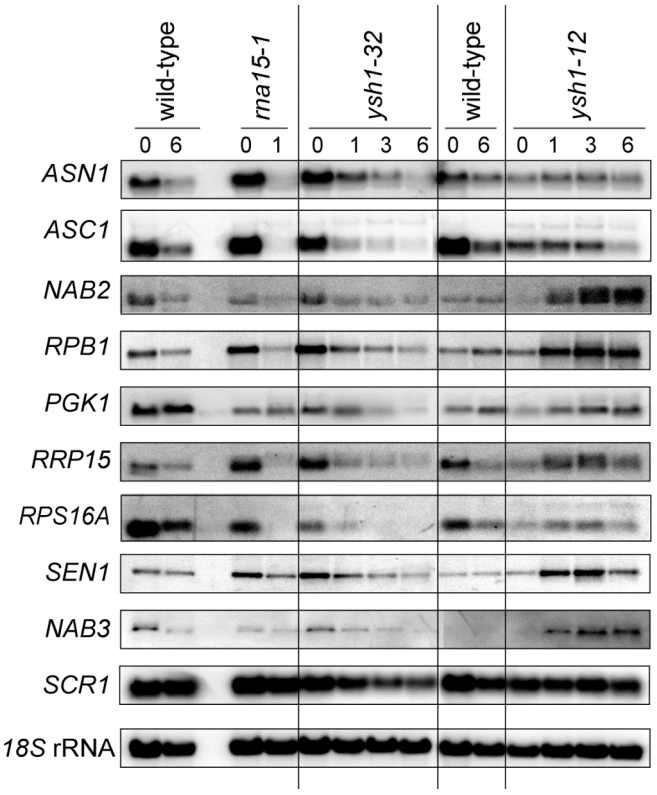


Figure 2.6 mRNA stability in the *ysh1* mutant alleles

Northern analysis of total RNA extracted from wild-type and mutant *ysh1* cells. Wild-type, *ysh1-32* and *rna15-1* strains were grown at 25ºC and shifted to restrictive temperature (37ºC) for 1, 3 and 6 hours. Cold-sensitive mutant strain *ysh1-12* was grown at 30ºC and shifted to 15ºC for 1, 3 and 6 hours. RNAs were separated on formaldehyde/1.2% agarose gels. Filters were hybridized to either random-primed labelled DNA probes or end-labelled oligonucleotides directed against RNA species indicated on the left. Levels of RNAP I transcript *18S* rRNA and RNAP III transcript *SCR1* served as loading controls.

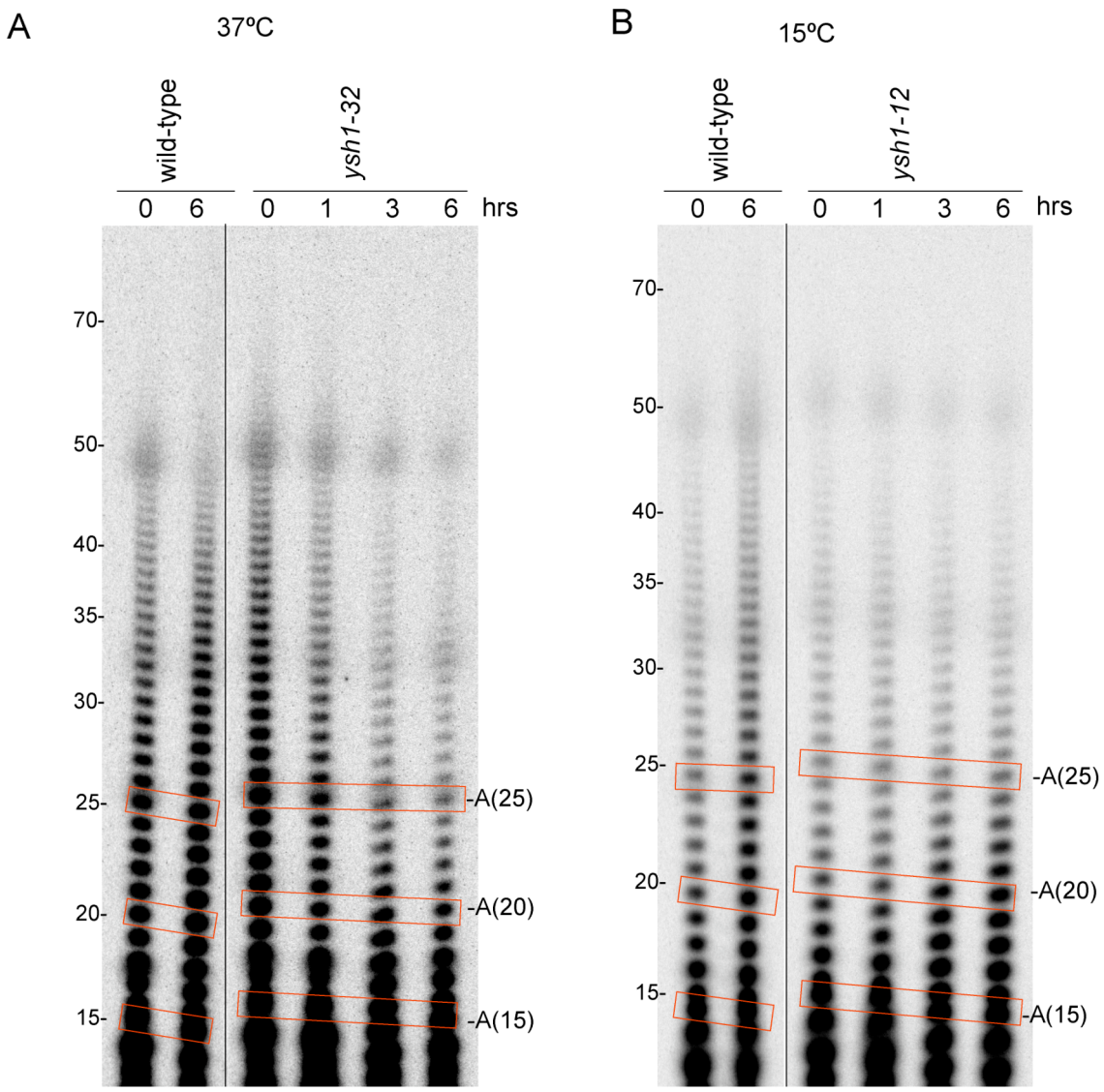
The average length of poly(A) tails is shortened in the *ysh1-32* mutant cells

In several 3' end processing mutants, general reduction of mRNA levels correlated with the reduced length of poly(A) tails (Minvielle-Sebastia et al., 1990; Minvielle-Sebastia et al., 1994). To evaluate the global phenotypes of our mutants we analyzed the length distribution of cellular poly(A). The ts *ysh1-32, ysh1-13* and *ysh1-15* strains showed a marked reduction of poly(A) length after growth at 37ºC (Figure 2.7 A; data not shown). In contrast, poly(A) length distribution remained unchanged following shift to the restrictive temperature for *ysh1-12* (Figure 2.7 B). Quantitative comparison of signal intensity of bands $A_{(15)}$, $A_{(20)}$ and $A_{(25)}$ showed that in the *ysh1-32* mutant cells poly(A) tails shortened significantly (5-10 fold) within 3 to 6 h growth at 37ºC, in contrast to the *ysh1-12* mutant strain, where no relevant change was observed. Thus, global poly(A) length reflected trends of mRNA levels observed in the Northern blot analyses for both the *ysh1* ts strains (reduced mRNA) and the *ysh1-12* cs strain (unchanged or increased mRNA).

To test whether the observed phenotypes were due to under-accumulation of Ysh1p or other subunits of the 3' end processing complex we performed Western blot analyses (Figure 2.8). We observed reduced levels of Ysh1p and Ssu72p in the *ysh1-32* strain after 1 h growth at 37ºC (lanes 4-6), whereas other CPF components (Pfs2p, Pap1p, Fip1p) and Rna15p remained largely unchanged. Interestingly, levels of all analyzed proteins remained relatively stable in the *ysh1-12* strain during a 6 h shift to 15˚C (lanes 9-12). Taken together, our data support the view that ts and cs *ysh1* mutations gave rise to distinct functional deficiencies (see Discussion).

Ysh1p is required for correct poly(A) site recognition of *ACT1* pre-mRNA

Several subunits of the CPF complex, including Yhh1p and Ydh1p, were previously shown to be required for poly(A) site recognition (Dichtl et al., 2002b; Kyburz et al., 2003). The well characterized 3' UTR of the *ACT1* gene includes five alternative polyadenylation sites and is therefore useful to examine poly(A) site selection in 3' end processing mutants (see scheme in Figure 2.9 A; Mandart & Parker, 1995). Low resolution Northern blot analysis revealed a change in the *ACT1* hybridization pattern in *ysh1* mutants compared to wild-type (Figure 2.9 A). To corroborate this phenotype we further analyzed RNA obtained from two mutant strains (*ysh1-13* and *ysh1-15)* on high resolution Northern blots following RNase H treatment (Mandart & Parker, 1995). In these experiments total RNA was targeted by RNase H and oligonucleotides complementary to the 3' end of *ACT1* mRNA and to the poly(A) tail allowing resolution of the five poly(A) sites (Figure 2.9 B). Quantification revealed that the most proximal poly(A) site (site I) was preferentially used in wild-type cells, as expected; upstream sites (sites II-V) were utilized four to seven times less frequently. In contrast, *ysh1* mutant strains displayed a change in poly(A) site usage towards the minor site V, both at permissive and restrictive temperatures. Taken together, the above results point out the involvement of Ysh1p in correct poly(A) site recognition.



	A(15)	A(20)	A(25)
wild-type 0h	100%	100%	100%
6h	140%	170%	150%
ysh1-32 0h	100%	100%	100%
1h	70%	30%	40%
3h	120%	70%	15%
6h	86%	65%	10%

	A(15)	A(20)	A(25)
wild-type 0h	100%	100%	100%
6h	100%	210%	250%
ysh1-12 0h	100%	100%	100%
1h	110%	260%	100%
3h	130%	480%	100%
6h	130%	350%	140%

Figure 2.7 Poly(A) tails shorten in the *ysh1-32* strain after shift to restrictive temperature

3' end labelling of poly(A) tails with total RNA extracted from wild-type and *ysh1-32* strains **(A)** or wild-type and *ysh1-12* strains **(B)** after growth at permissive temperatures (23ºC or 30ºC, respectively), and after shift to restrictive temperatures (37ºC or 15ºC, respectively) for 1, 3 and 6 hours. Poly(A) tail length is indicated on the left. Intensity of single poly(A) bands A(15), A(20) and A(25) were quantified by PhosphorImager analysis (Molecular Dynamics) and recalculated relative to time point 0 h. The band corresponding to the size of A15 (15nt) was used as a reference to control sample loading.

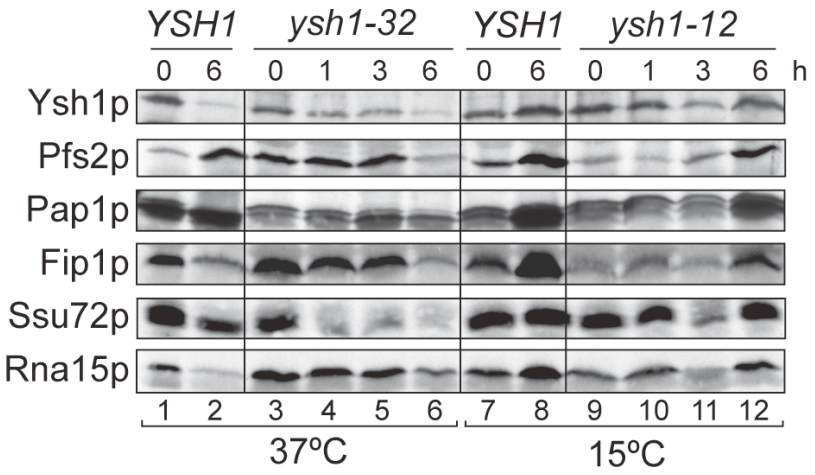
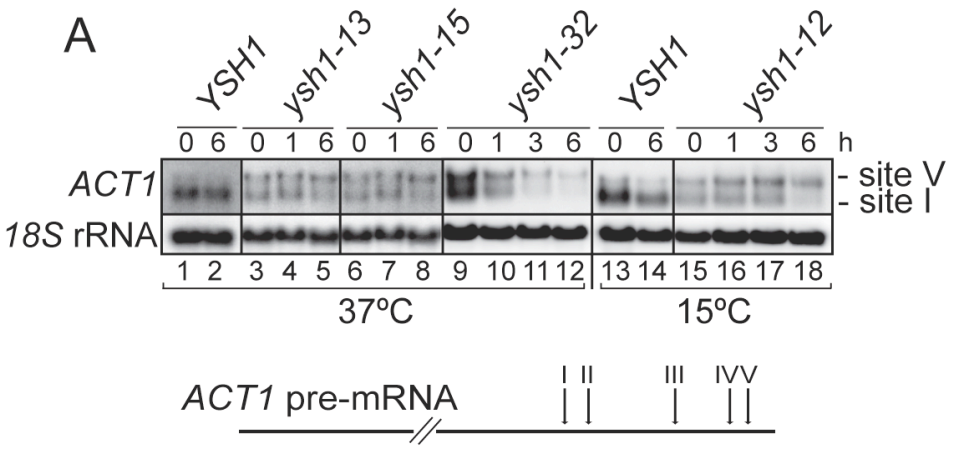


Figure 2.8 Protein stability in the *ysh1* mutant strains

Western blot analysis of wild-type, *ysh1-32*, and *ysh1-12* mutant extracts prepared from cells grown analogically as in Figure 2.6. Equal amounts of total protein were loaded in each lane. Blots were probed with antibodies directed against the proteins indicated on the left.



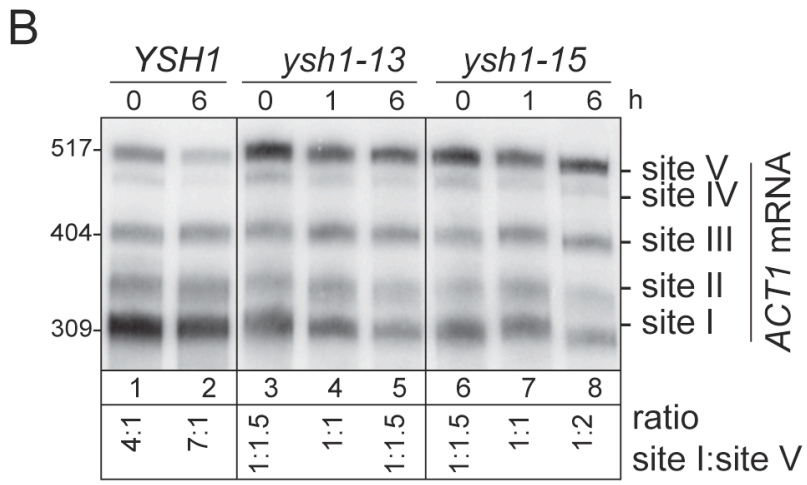


Figure 2.9 *ysh1* mutant strains are deficient in poly(A) site selection of *ACT1* pre-mRNA *in vivo*

(A) Northern blot analysis of total RNAs extracted from wild-type and mutant *ysh1* cells. *ysh1-13*, *ysh1-15*, *ysh1-32* and *rna15-1* strains were grown in YPD at 23ºC and shifted to 37ºC for 1, 3, and 6 h. *ysh1-12* was grown at 30ºC and shifted to 15ºC for 1, 3 and 6 h. The analysis was performed essentially as in Figure 2.6. Levels of 18S rRNA served as a loading control. *ACT1* poly(A) sites I – V positions are marked. The scheme below the panel represents relative positions of *ACT1* different poly(A) sites. (**B**) Analysis of *ACT1* poly(A) site usage in wild-type, *ysh1-13* and *ysh1-15* cells. Total RNAs extracted from strains grown as described in (A) were treated with RNase H and oligonucleotides *ACT1*-RNaseH and oligo(dT), and analyzed by polyacrylamide Northern blotting with a probe specific for the 3' end of *ACT1* mRNA. Positions of the resolved poly(A) sites I to V are indicated on the right. RNA levels were quantified by PhosphorImager scanning (Molecular Dynamics). The ratios of poly(A) site I versus site V usage for each lane are indicated at the bottom.

Ysh1p is required for RNAP II transcription termination on mRNA genes *in vivo*

A functional poly(A) signal is required for a correct transcription termination *in vivo* (Connelly and Manley, 1988). Subunits of yeast 3' end processing machinery such as Rna15p, Pcf11p, Yhh1p and Ssu72p were implied in RNAP II transcription termination on protein coding genes (Birse et al., 1998; Dichtl et al., 2002a; Dichtl et al., 2002b; Sadowski, 2003). To address the role of Ysh1p in RNAP II transcription termination, we carried out transcriptional run-on analysis (TRO) on a plasmid-borne *CYC1* gene, which is placed under the control of *GAL1/10* promoter (Birse et al., 1998). Correct termination produces run-on signals over probes P1 to P3, whereas defective termination results in increased signals also over probes P4-P6 (see scheme in Figure 2.10 A). TRO was done with wild-type and *ysh1-12* cells following growth in galactose-containing medium at permissive and restrictive conditions. Results of slot hybridizations showed that in wild-type cells the density of RNAP II ceased significantly downstream of the *CYC1* terminator (Figure 2.10 B). On the contrary, the *ysh1-12* mutant cells displayed an increased run-on signal over probes P4-P6, both at permissive and restrictive temperatures (Figure 2.10 B). Quantitatively, in the *ysh1-12* mutant cells, read-through signal was increased at least 2-3 fold (Figure 2.10 C). Similar results were obtained from TRO performed on the *ysh1-32* mutant cells, where statistical increase of the read-through signal over probes P4-P6 was 3-5 fold higher than in the wild-type situation (Figure 2.10 D). The above reflects a higher relative density of actively transcribing RNAP II molecules past the poly(A) site in the *ysh1-12* and *ysh1-32* mutant strains *in vivo*.

In summary, mutations in the *ysh1-12* and *ysh1-32* mutant strains impair the correct RNAP II transcription termination *in vivo*. The observed termination defect, together with the impairment in pre-mRNA 3' end processing in the *ysh1* mutants, points to the involvement of Ysh1p in coupling of transcription termination and 3' end formation events.

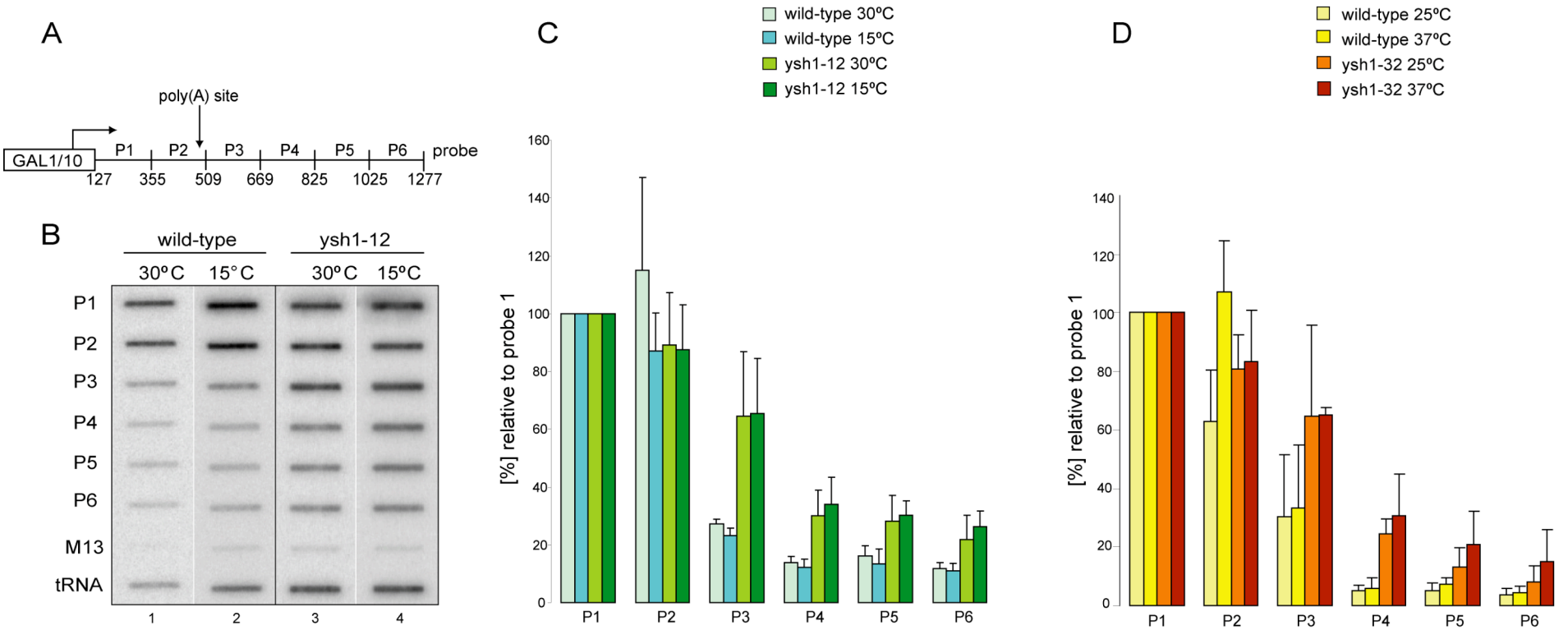


Figure 2.10 Mutations in *YSH1* impair correct RNAP II transcription termination on mRNA genes *in vivo*

(**A**) Schematic representation of the p*GAL-CYC1* gene construct that was used for the transcriptional run-on analysis (TRO). Order of M13 probes spanning the *CYC1* gene (P1-P6) and poly(A) site (position 506.) are depicted. (**B**) Slot hybridizations of run-on transcripts obtained after TRO (Birse et al., 1998a). Wild-type and *ysh1-12* cells transformed with p*UGCYC1* plasmid were grown in synthetic medium lacking uracil and containing 2% galactose under permissive growth conditions (30ºC) and after shift to restrictive temperature (15ºC) for 60 min. P1-P6 represent probes complementary to the *CYC1* transcripts as indicated in (A). Empty M13 was used as a background hybridization control. Hybridization of RNAP III transcripts to the tRNA probe is shown. (**C**) Quantitative analysis of transcriptional run-on profiles for *ysh1-12*. Values obtained by PhosphorImager scanning (Molecular Dynamics) were corrected by subtraction of the M13 background signal and normalized to the value of P1, which was fixed at 100%. Results shown in the diagram represent the average value of three independent experiments. (**D**) Quantitative analysis of transcriptional run-on profiles for *ysh1-32* strain. Values were obtained as described in (C).

Read-through mRNA transcripts accumulate in the *ysh1-12* mutant

It was previously shown that defective termination can result in production of polycistronic transcripts (Connelly & Manley, 1988). Supporting this notion, in our Northern blot analysis we observed occurrence of read-through RNAs in *ysh1-12* cells. Using probes complementary to the coding regions of mRNAs, we detected extended species of *NRD1, YSH1* and *ADH1* mRNAs appearing in this mutant strain at restrictive temperature (Figure 2.11). To confirm that these elongated transcripts resulted of impaired transcription, we probed for the products of the respective downstream genes (*MRPL17, DBP9* and *YOL087C).* Our results indicated that in the *ysh1-12* cells, a portion of *NRD1* transcript failed to terminate properly; instead, RNAP II transcription continued to a downstream adjacent gene (*MRPL17*), until it encountered its cleavage and polyadenylation site and terminated eventually (Figure 2.11 A). Similarly, the *YSH1* transcript extended into the downstream gene *DBP9* (Figure 2.11 B) and *ADH1* mRNA extended into the downstream gene *YOL087C* (Figure 2.11 C). The read-through transcripts were not detectable in the ts *ysh1* mutant strains (*ysh1-13, ysh1-15, ysh1-32*; results not shown), which is presumably resulting from a reduced stability of read-through RNAs in these strains. Taken together, the above results support a role for Ysh1p in termination of RNAP II transcription on mRNA genes.

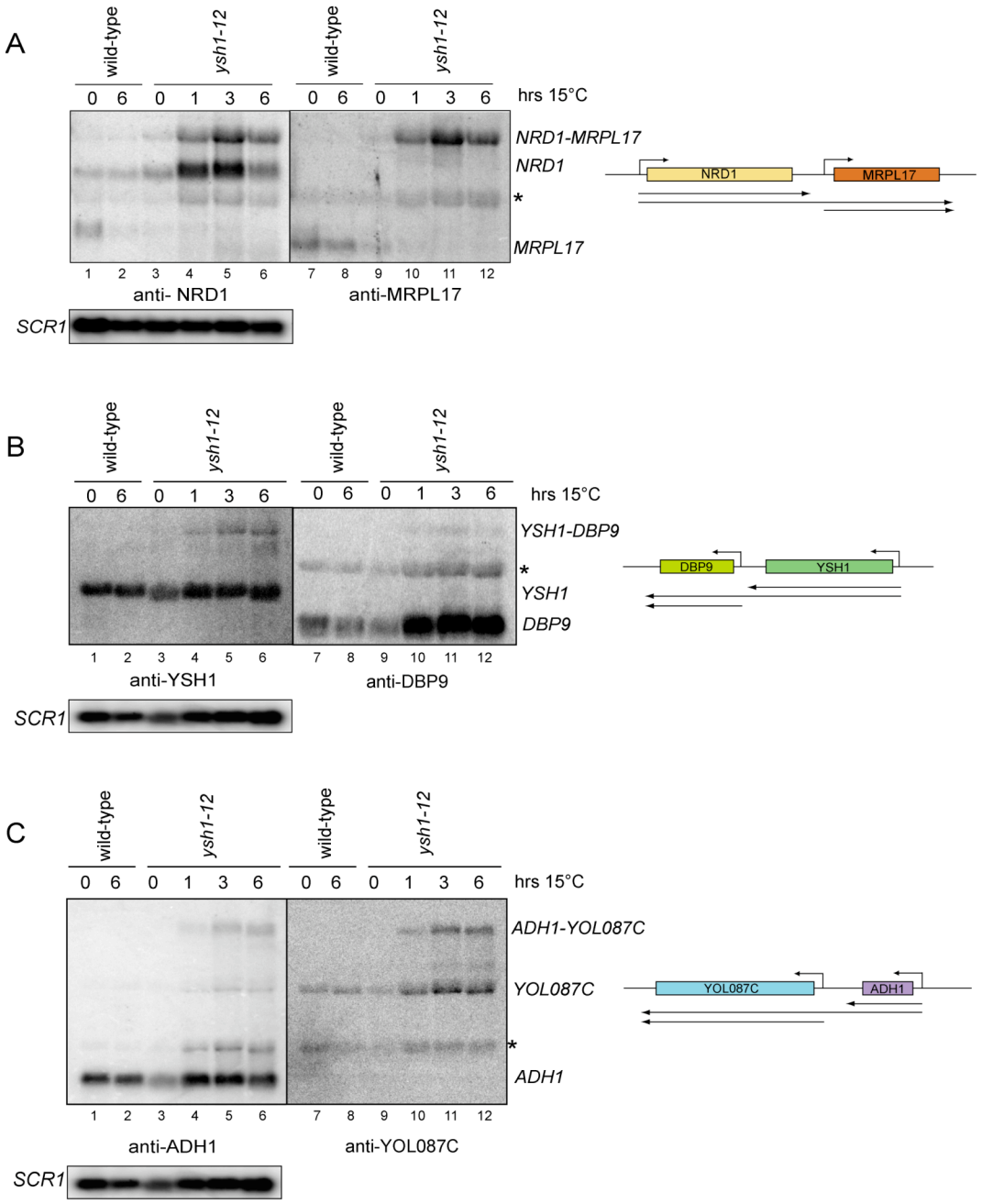


Figure 2.11 Read-through mRNA transcripts accumulate in the *ysh1-12* strain *in vivo*

(**A**) Northern blot analysis of total RNA extracted from the wild-type and *ysh1-12* cells grown in YPD medium at permissive temperature (30ºC; lanes 1-3, and 7-9) and followed by the shift to 15ºC for 1, 3 or 6 h (lanes 4-6 and 10-12). RNAs were resolved as described in Figure 2.6. Filter was developed with random-primed labelled probe directed against *NRD1*, then subsequently washed and probed for the RNA of the downstream gene *MRPL17*, as indicated. The asterisk indicates a non-specific cross-reaction with *18S* rRNA. Probing to RNAP III transcript *SCR1* served as a loading control; it revealed a lower amount of total RNA present in the *ysh1-12* strain at the time-point 0 h (**B**) Northern blot of *YSH1* and *DBP9* transcripts performed as described in (A). (**C**) Northern blot of *ADH1* and *YOL087C* transcripts performed analogically to (A).

3' extended snoRNAs are produced in the *ysh1-12* mutant strain

Subunits of the yeast APT subcomplex associated with CPF (Ssu72p, Swd2p, Pti1p, Ref2p) and CF1A (Rna15p, Rna14p) were previously implied in RNAP II transcription termination at snoRNA genes (Dheur et al., 2003; Dichtl et al., 2004; Ganem et al., 2003; Morlando et al., 2002; Nedea et al., 2003). Moreover, yeast snoRNA and snRNA 3' end formation was reported to be mediated by a cleavage that was uncoupled from polyadenylation (Fatica et al., 2000; Morlando et al., 2002).

The first evidence pointing towards the involvement of Ysh1p in the pre-snoRNA 3' end formation came from our Northern analyses where a probe directed against *TRS31* mRNA detected additional larger species of RNA that appeared after shifting the *ysh1-12* strain to restrictive temperature (data not shown). These extended transcripts could possibly be the read-through products of a faulty termination from the upstream *SNR13* snoRNA gene (the dicistronic *SNR13-TRS31* transcript), thus suggesting the involvement of Ysh1p in snoRNA synthesis.

To address the question whether Ysh1p is indeed required for the 3' end processing of snoRNA, we performed primer extension analysis on total RNA; in these experiments the appearance of an extension signal is an indication of a defect in 3' end formation of these non-coding RNAs. In the *ysh1-12* strain, we detected accumulation of 3' extended transcripts of snR13, snR33, snR39B, snR46, snR47 and snR128 snoRNAs (Figure 2.12). Interestingly, we did not detect any defects in the 3' end formation of several other snoRNAs such as snR3, snR50, snR71 and snR45 (results not shown) suggesting that the *ysh1-12* mutation displayed gene specific defects. Moreover, no extended RNAs were observed for the *ysh1-32* strain, when snR13, snR33 and snR47 were analyzed (results not shown). The *nrd1-102* and *sen1-1* mutants, previously shown to be deficient in snoRNA 3' end formation (Steinmetz et al., 2001) were used as controls. Notably, several of the snoRNA transcripts affected by mutations in *ysh1-12* (e.g. *snR33, snR39b, snR46)*, were only partially or not at all aberrant in the *nrd1-102* and *sen1-1* strains (Figure 2.12). In summary, the above data showed that mutations in the *ysh1-12* mutant strain impair the correct snoRNA 3' end formation.

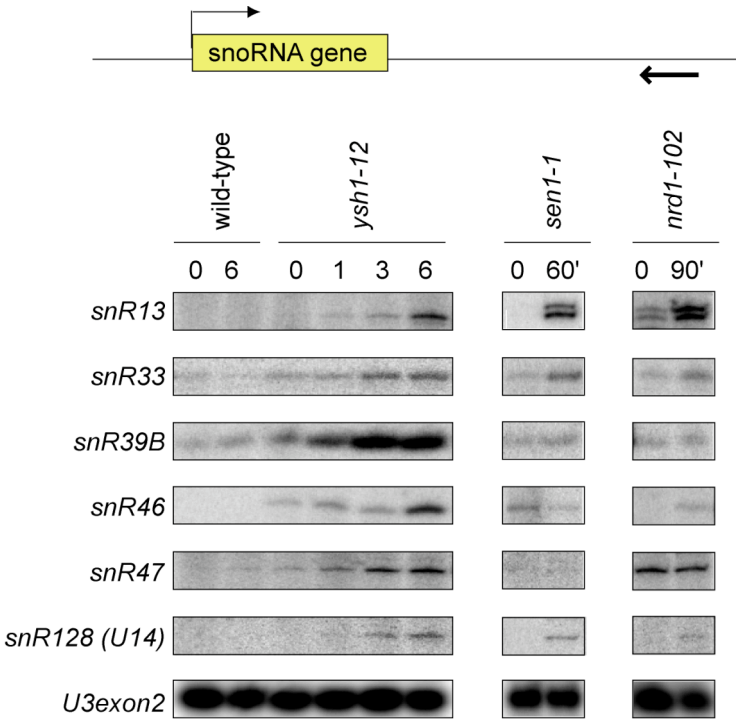


Figure 2.12 3' end extended snoRNA transcripts are produced in the *ysh1-12* mutant strain

Primer extension analysis of extended transcripts of several snoRNAs in the *ysh1-12*, *sen1-1* and *nrd1-102* mutant strains was carried out using radioactively labelled oligonucleotides complementary to sequences located downstream of the mature 3' ends of the indicated snoRNAs (scheme). An oligonucleotide complementary to the mature U3 snoRNA was used as control.

Transcription termination of the *SNR3* snoRNA gene is impaired in the *ysh1-12* strain

The accumulation of 3' end extended snoRNA transcripts in the *ysh1-12* mutant strain results possibly either from transcription termination defect on snoRNA genes and/or from impaired 3' end processing of snoRNA transcripts. It might also be an indirect effect of the *ysh1-12* mutation on the snoRNA 3' end processing factors. We therefore examined the RNAP II transcription termination efficiency on the endogenous *SNR3* snoRNA locus (Steinmetz et al., 2001) in wild-type and *ysh1-12* mutant cells. Two probes were used, one spanning the *SNR3* gene and its terminator sequences (probe A), and one downstream of this position (probe B; Figure 2.13 A). These single-stranded DNAs do not hybridize with mRNA derived from the downstream ORF *YJR129c*, as this gene is encoded in a convergent orientation. Interestingly, *ysh1-12* mutant cells revealed up to 6-fold higher density of transcribing RNAP II over the downstream probe compared to wild-type (Figure 2.13 B, C) indicating defective termination on this transcription unit. In contrast, termination on the *SNR3* gene was unaffected in the *ysh1-32* strain (results not shown), which is consistent with the primer extension analysis above. The above results of the *ysh1-12* transcription run-on analysis suggest a role for Ysh1p in preventing transcriptional read-through at snoRNA terminators.

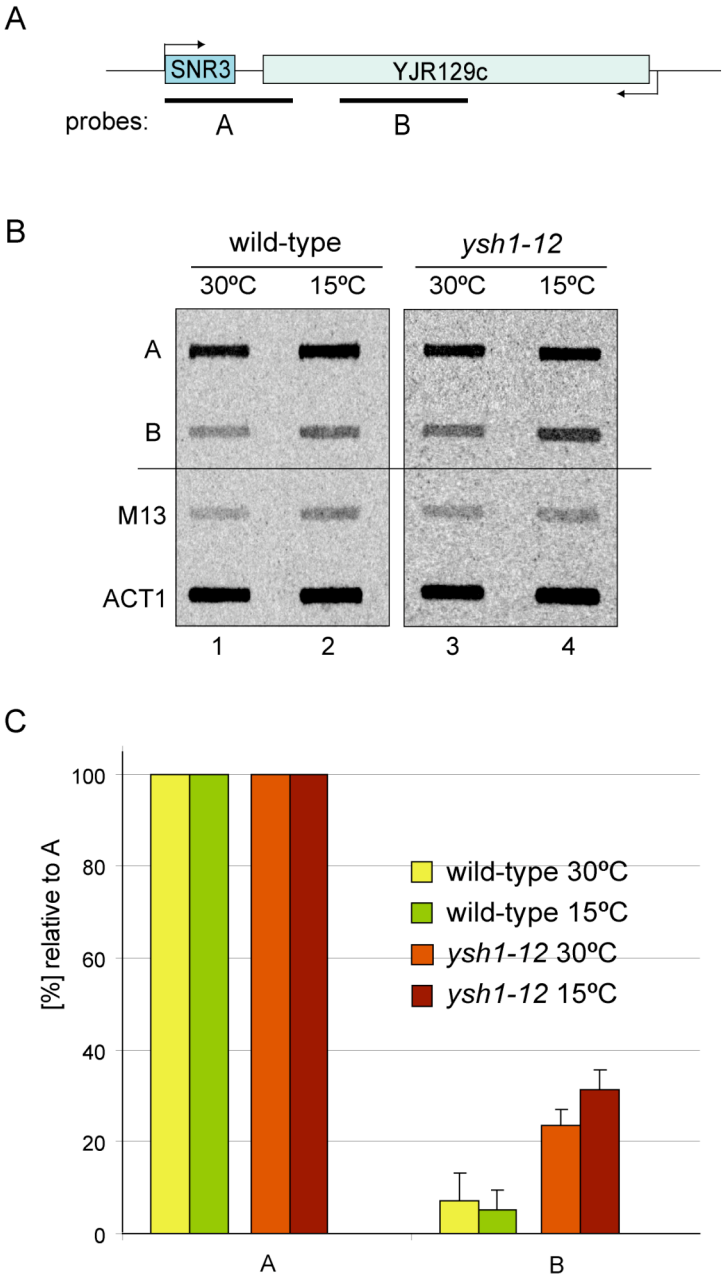


Figure 2.13 Transcription termination of the SNR3 snoRNA gene is impaired in the *ysh1-12* strain

(**A**) Schematic representation of the *SNR3-YJR129c* genomic locus. Note that *SNR3* and *YJR129c* are transcribed from opposite strands of the DNA, as indicated by the arrows. Probes A and B are shown below their locations. (**B**) Slot hybridizations of run-on transcripts obtained after TRO (Birse et al., 1998). Wild-type and *ysh1-12* cells were grown in YPD medium containing 2% glucose under permissive growth conditions (30ºC, lane 1, 2) and after the shift to restrictive temperature (15ºC, lane 3, 4) for 60 min. M13 slots are single-stranded phagemids with no insert and were used as a background hybridization control. ACT1 probe served as a positive control. (**C**) Quantitative analysis of transcriptional run-on profiles for the *ysh1-12* mutant. Values obtained by PhosphorImager scanning (Molecular Dynamics) were corrected by subtraction of the M13 background signal and normalized to the value of probe A, which was set at 100%. Results shown in the diagram represent the average value of three independent experiments.

Levels of snoRNAs remain stable in the *ysh1-12* and *ysh1-32* mutant strains *in vivo*

Deficient snoRNA 3' end formation could possibly have an effect on the stability of snoRNA transcripts in the *ysh1* mutant strains. To address this, we performed Northern blot analysis of the steady-state levels of several snoRNAs in the *ysh1-12* and *ysh1-32* strains following growth at respective restrictive temperatures. The *rna15-1* strain was previously shown to be defective in snoRNA 3' end formation (Fatica et al., 2000; Morlando et al., 2002) and was analyzed as control. We observed that overall levels of RNAP II transcribed *SNR3, SNR13, SNR33, SNR45* and *U3* snoRNAs were stable in all mutants tested, even after the 6 h incubation at restrictive temperatures (Figure 2.14). Levels of RNAP III transcripts *U6* and *5S* rRNA served as loading controls.

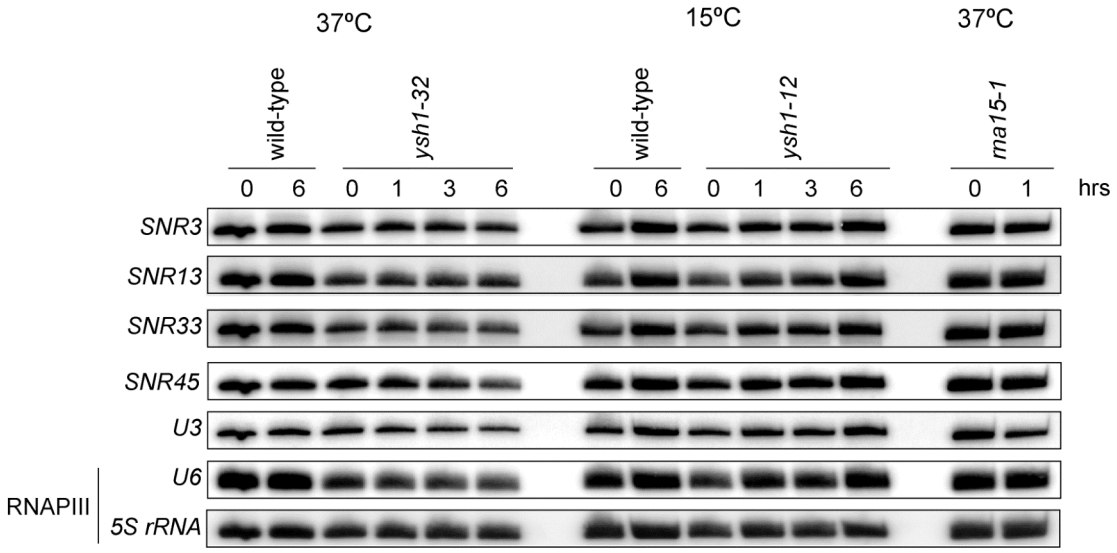


Figure 2.14 Levels of snoRNAs remain stable in the *ysh1-12* and *ysh1-32* mutant strains *in vivo*

Northern analysis of total RNAs extracted from wild-type and mutant *ysh1* cells. Wild-type, *ysh1-32* and *rna15-1* strains were grown at 25ºC and shifted to 37ºC for 1, 3 and 6 h. Cold-sensitive mutant strain *ysh1-12* was grown at 30ºC and shifted to 15ºC for 1, 3 and 6 h. The *rna15-1* strain was analyzed in parallel. RNAs were separated on 8.3 M Urea/8% polyacrylamide gels. Filters were developed with [^{32}P] end-labelled oligonucleotide directed against RNA species indicated on the left. Levels of RNAP III transcripts *U6* and *5S* rRNA served as loading controls.

Conditional depletion of Ysh1p does not affect snoRNA 3' end processing

Conditional depletion of Ysh1p in the *brr5-td* degron strain was reported not to affect termination at snoRNA genes, as determined by RNAP II chromatin immunoprecipitation (ChIP) experiments (Kim et al., 2006). As these data conflicted with our observations, we included the *brr5-td* strain in our analysis. The degron-Ysh1p levels were significantly reduced after 1 h growth at 37˚C, and almost completely diminished after 3 h (Figure 2.15). In contrast, the levels of Fip1p, Ssu72p and Rna15p remained unchanged following depletion of Ysh1p. We compared snoRNA 3' end formation in *brr5-td*, its isogenic wild-type (FY23) and *ysh1-12* strains by primer extension analysis as described above. 3' extended forms of the tested snoRNAs (snR13, snR33, snR39B and snR47) were readily detected in the *ysh1-12* strain (Figure 2.16) but not following depletion of Ysh1p. These observations indicated that defects in snoRNA 3' end formation were apparent only with distinct *ysh1* mutant strains on a subset of snoRNA transcription units (see Discussion).

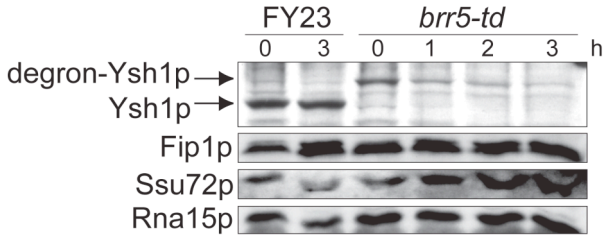


Figure 2.15 Degron-depletion of Ysh1p

Western blot analysis of isogenic wild-type (FY23) and *brr5-td* degron strain extracts prepared from cells grown at 25ºC and shifted for 1, 2, or 3 h to 37ºC. Equal amounts of total protein were loaded in each lane. The blot was probed with antibody directed against Ysh1p and subsequently re-probed with antibodies directed against other CPF subunits, as indicated.

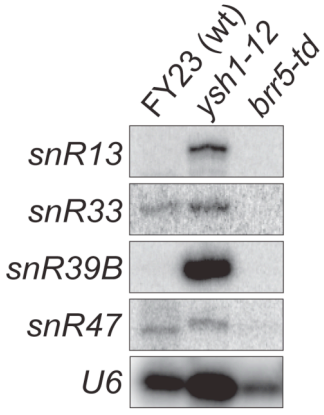


Figure 2.16 snoRNA processing remains unaffected in the *brr5-td* degron strain

Primer extension analysis of extended transcripts of several snoRNAs in the wild-type FY23 and *brr5-td* (both shifted to 37˚C for 2 h), and *ysh1-12* (shifted to 15˚C for 6 h) strains was carried out with radioactively labelled oligonucleotides complementary to sequences located downstream of the mature 3' ends of the indicated snoRNAs. Levels of the RNAP III transcript U6 served as a loading control and showed that lower amounts of total RNA extracted from the FY23 and *brr5-td* strains were used in this experiment (lanes 1 and 3).

The snoRNA terminator sequences are cleaved by CPF *in vitro*

A previous report showed that cleavage uncoupled from polyadenylation is required for 3' snoRNA formation (Fatica et al., 2000). To address whether the endonucleolytic activity of Ysh1p could be required for the 3' end formation on snoRNA genes, we performed *in vitro* cleavage assays on RNA precursors derived from snoRNA gene loci.

Our initial *in silico* analysis of the terminator sequences of *SNR39B*, *SNR47* and *SNR65* genes predicted one or more putative 3' end cleavage sites in those regions (algorithm by Graber et al., 2002). Three [^{32}P]-labelled snoRNA substrates were obtained via an *in vitro* run-off transcription and subsequently incubated with the yeast extracts isolated from the wild type and *ysh1* mutant strains. The SNR39B+ RNA substrate contains the *SNR39B* coding region and 396 nucleotides of its 3' downstream sequence (Figure 2.17 A, scheme on the left). The SNR47+ RNA substrate contains the *SNR47* coding region and 205 nucleotides of its 3' downstream sequence (Figure 2.17 A, scheme in the middle). The SNR65+ RNA substrate contains the *SNR39B* coding region and 300 nucleotides of its 3' downstream sequence (Figure 2.17 A, scheme on the right). Cleavage reactions were performed using conditions established for pre-mRNA substrates (Minvielle-Sebastia et al., 1994). Interestingly, all the three RNA substrates (SNR39B+, SNR47+ and SNR65+) were efficiently cleaved by the wild-type extract as well as by the *ysh1-1* mutant extract (*ysh1-1,* mutant allele that is <u>not</u> deficient in pre-mRNA 3' end processing), resulting in one or more cleavage products (Figure 2.17 B, lanes 2-3, 6-7, 10-11). In contrast, an extract derived from the *ysh1-32* mutant strain that is deficient in pre-mRNA 3' end processing was unable to cleave any of the SNR+ RNA transcripts (Figure 2.17 B, lanes 4, 8, 12). To further confirm that the observed cleavage was catalyzed by the canonical 3' end processing machinery, cleavage reactions were performed with purified 3' end processing factors CPF and CF 1A. All snoRNA substrates were efficiently cleaved; however, the cleavage pattern differed substantially from products of the cleavage reactions with yeast cell extracts (Figure 2.17 C, lanes 2-3, 5-6, 8-9). The cleavage specificity of the snoRNA substrates remains to be further corroborated. These initial results are in agreement with the report of Morlando et al., 2002, which showed that *SNR13* 3' downstream sequences can be cleaved by the 3' end processing machinery *in vitro*. Moreover, different cleavage patterns observed for yeast extracts and purified factors suggested that certain factors present in non-fractionated extracts but absent from purified CPF/CF IA might be involved in cleavage at the 3' ends of pre-snoRNAs.

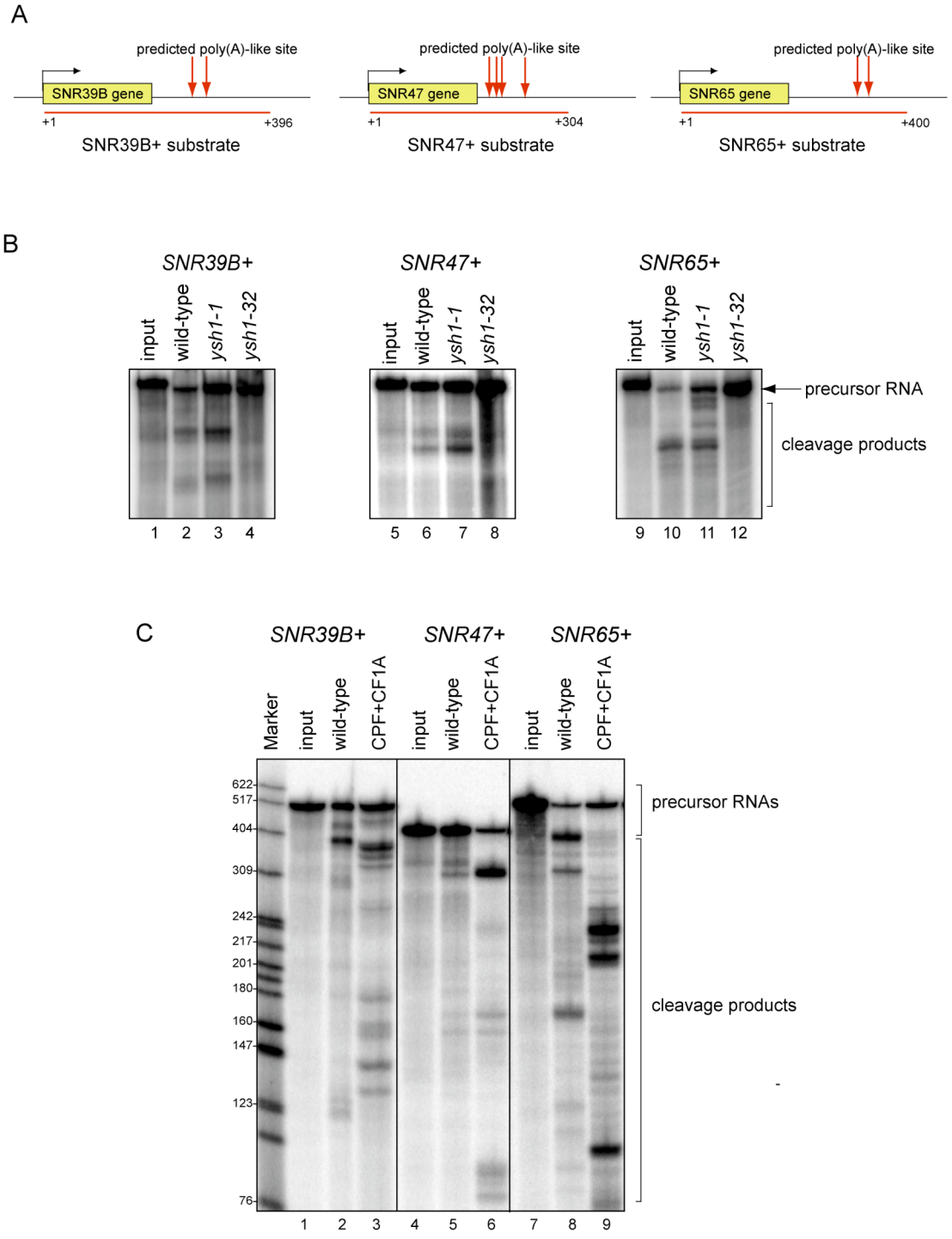


Figure 2.17 snoRNA terminator sequences are cleaved by the 3' end processing machinery *in vitro*

(**A**) Schematic representation of the *in vitro* SNR39B+, SNR47+, and SNR65+ snoRNA transcripts used in the experiment. (**B**) *In vitro* cleavage assay: internally [^{32}P]-labelled RNA substrates were incubated with extracts prepared from wild-type, *ysh1-1* and *ysh1-32* strains at 30ºC for 90 min and resolved on an 8% polyacrylamide/8.3 M urea gel. Positions of the RNA precursors and processing products are indicated on the right.

(**C**) As in (**B**), except wild-type extract and purified CPF and CF IA factors were analyzed. Sizes of the marker are given on the left.

Involvement of Ysh1p in the autoregulation of *NRD1* mRNA levels

Nrd1p controls its own expression through regulated premature termination and levels of *NRD1* mRNA are increased in *nrd1, nab3* and *sen1* mutant strains (Steinmetz et al., 2001; Arigo et al., 2006). Interestingly, we observed increased levels of *NRD1* mRNA in *ysh1-12* cells at the non-permissive temperature (Figure 2.18), which resulted in increased amounts of Nrd1p protein (see Figure 2.20). Such a phenotype was not observed with the three *ysh1* ts mutant strains (results not shown). Since a similar increase in Nrd1p protein was observed in the *nrd1-102* strain this indicated that Ysh1p is involved in regulating Nrd1p levels. Moreover, some *NRD1* transcripts extended into the downstream gene *MRPL17*, resulting in a dicistronic *NRD1-MRPL17* transcript (Figure 2.18, lanes 4-6). We also observed a ~1 kb RNA that corresponded to 3' truncated *NRD1* RNA in the wild-type strain as previously reported (Arigo et al., 2006). We detected this RNA with a probe directed against the 5' end of *NRD1* mRNA but not with a probe directed against the 3' end (results not shown). Interestingly, we found that this RNA was absent in *ysh1-12* (Figure 2.18).

The above observations on the *ysh1-12* phenotype suggested that Ysh1p might be involved in the regulation of *NRD1* expression. We speculated that Ysh1p could facilitate premature termination through endonucleolytic cleavage of the *NRD1* transcript. Direct evidence for occurrence of cleavage within *NRD1* would be the identification of downstream 3' cut-off products. Since such intermediates are predicted to be rapidly degraded by 5'-3' exonucleases, we analyzed total RNA extracted from the *rat1-1, xrn1Δ* double mutant strain (Henry et al., 1994). Primer extension analysis was carried out with oligonucleotide D that is complementary to sequences located downstream of the predicted premature termination region (see scheme in Figure 2.19). Most interestingly, a primer extension stop was obtained that corresponded in size to a predicted downstream 3' cut-off product (Figure 2.19, lane 4). An equivalent primer extension stop was not detected with wild type RNA, where a higher molecular weight band was detectable instead (lane 3); this stop most likely represented the 5' end of the *NRD1* mRNA. As control we used oligonucleotide U, which is complementary to the 5' end of the *NRD1* transcript and which gave primer extension stops that corresponded to the 5' end of *NRD1* mRNA, as expected (lanes 1-2). These data are consistent with the proposal that premature termination within the *NRD1* gene involves endonucleolytic cleavage of the primary transcript and that the putative endonuclease Ysh1p may provide this activity.

We showed by Northern blot analyses that the abundance of *NRD1* mRNA was increased in the *ysh1-12* mutant. To test whether the Nrd1 protein levels are co-ordinately increased in this strain, we performed Western analysis of total protein extracts from single and double mutant strains as shown in Figure 2.20. We found that levels of Nrd1p were increased to the same extent in the *ysh1-12* mutant strain as in the *nrd1-102* strain (Figure 2.20, lanes 2-3, 10-11). Similar effect was observed in the *ysh1-12 nrd1-102* and *ysh1-12 rrp6Δ* double mutant strains (lanes 4-7). The exosome subunit *RRP6* deletion strain (*rrp6Δ)* exhibited the same phenotype in Nrd1p accumulation. This is in an accordance with the role of the nuclear exosome in *NRD1* mRNA autoregulation (Arigo et al., 2006). In contrast, the levels of Nrd1 protein were

reduced in the *ysh1-32* strain grown at restrictive temperature, consistent with the diminishing amounts of *NRD1* mRNA in that strain at restrictive temperature (lane 12). Additionally, the amounts of Ysh1p and Act1p proteins in the above mutant strains were analyzed, in order to serve as relative loading controls.

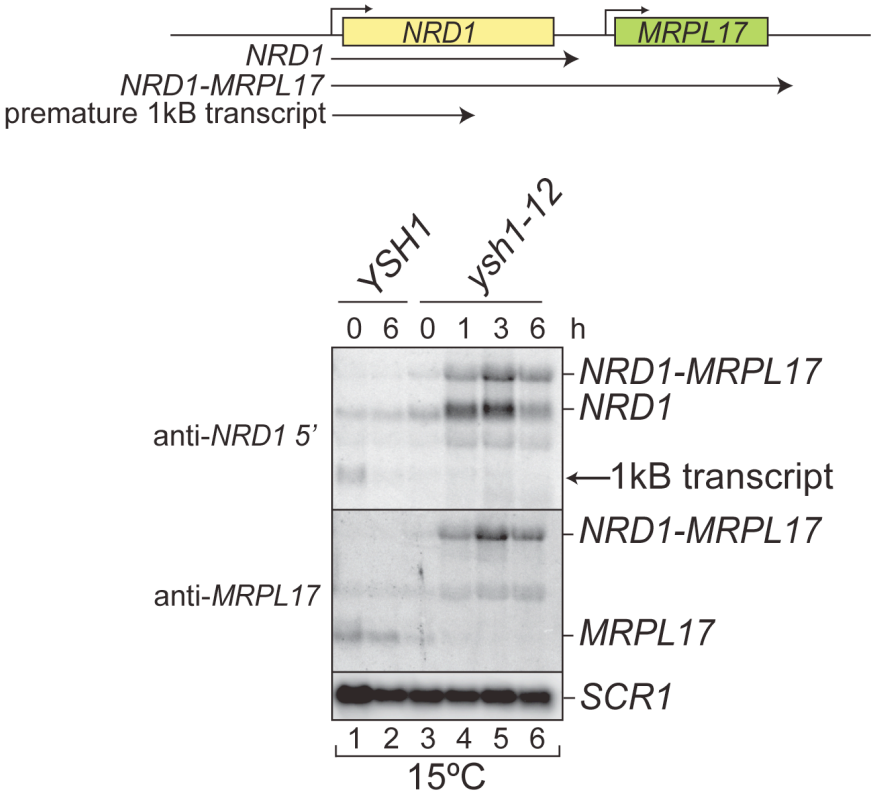


Figure 2.18 Increased levels of *NRD1* mRNA accumulate in the *ysh1-12* mutant strain

Northern blot analysis of *NRD1* mRNA in *ysh1* strains was performed using probes directed against 5' or 3' ends of the message as indicated on the left. Wild-type, *ysh1-32* and *rna15-1* strains were grown at 25ºC and shifted to restrictive temperature 37ºC for 1, 3 and 6 h. Cold-sensitive mutant strain *ysh1-12* was grown at 30ºC and shifted to 15ºC for 1, 3 and 6 h. RNAs were separated on formaldehyde/1.2% agarose gels. Filters were developed with probes directed against the 5' end of *NRD1* mRNA, or against *MRPL17*, as indicated on the left. Hybridisation to the RNAP III transcript *SCR1* served as a loading control.

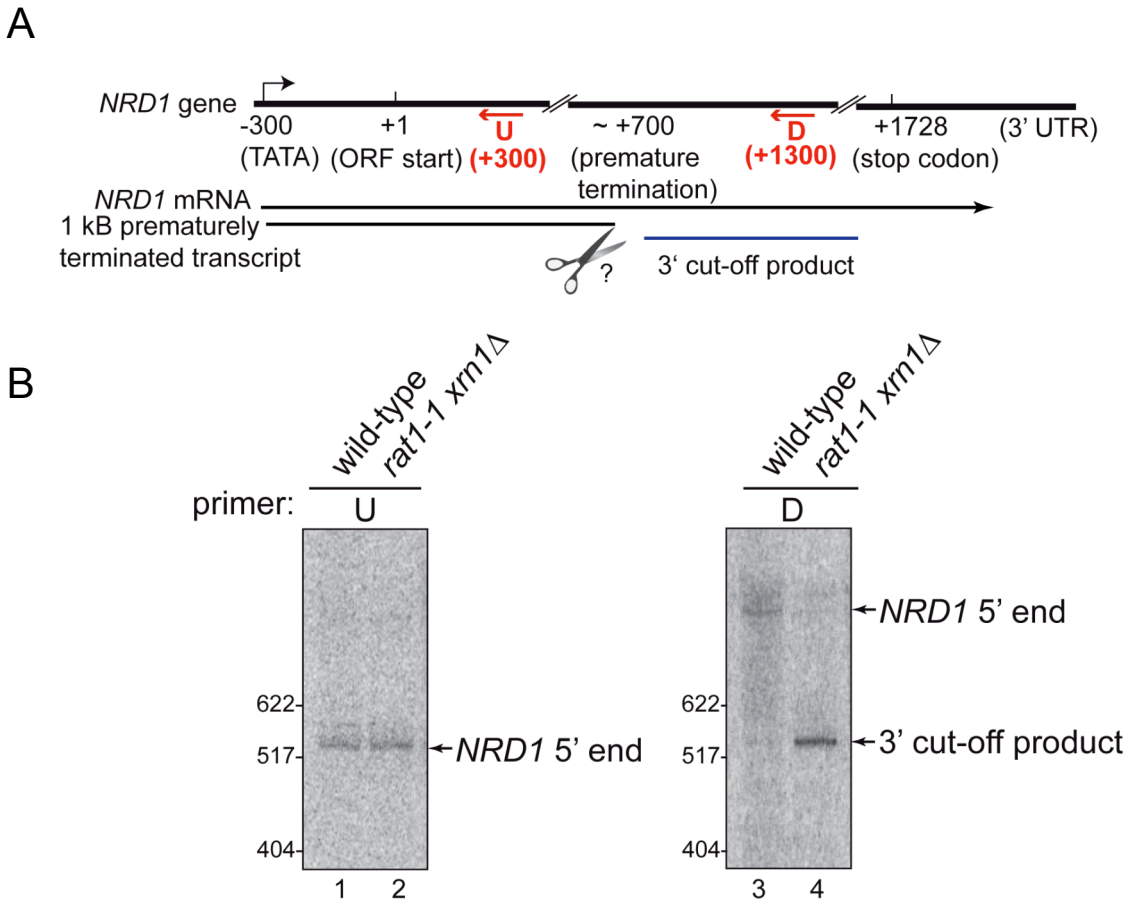


Figure 2.19 *NRD1* premature termination occurs via an endonucleolytic cleavage of the *NRD1* transcript

(A) Schematic representation of the *NRD1* gene locus. Positions of the oligonucleotides U (+300 nt) and D (+1300 nt; relative to the ORF start) used in the primer extension analyses, the full-length and prematurely terminated *NRD1* mRNA transcripts, and of the putative 3' end cut-off product are shown.

(B) Primer extension analysis of total RNA extracted from wild-type (W303) and *rat1-1, xrn1Δ* strains (grown at 37˚C for 2 h), carried out with radioactively labelled oligonucleotide D complementary to sequences located downstream of the predicted premature termination region (right panel). Oligonucleotide U complementary to the 5' end of *NRD1* transcript was used as an internal control (left panel).

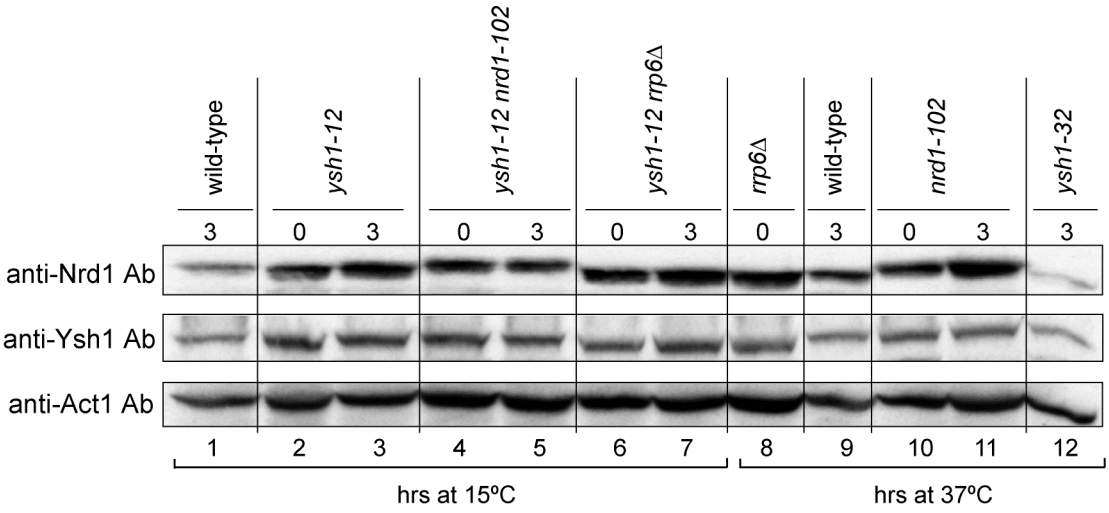


Figure 2.19 Amount of Nrd1 protein increases in the *ysh1-12* and *rrp6Δ* mutant strains

Western analysis of wild-type and mutant extracts prepared from cells grown at 25ºC and shifted for 3 h to restrictive temperatures (15ºC or 37ºC, respectively). Equal amounts of total protein were loaded in each lane. Blots were probed with antibodies directed against the proteins indicated on the left. Act1p levels served as control for equal loading.

YSH1 genetically interacts with the nuclear exosome subunit *RRP6*

Rrp6p is a non-essential component of the nuclear exosome responsible for 3' end processing of small stable transcripts such as small nuclear RNAs (snRNAs), snoRNAs and rRNAs (Allmang et al., 1999). To further analyze the possible function of Ysh1p in snoRNA 3' end formation, we generated *ysh1-12 rrp6Δ* and *ysh1-32 rrp6Δ* double mutants. As shown in Figure 2.21, deletion of the *RRP6* gene suppressed the lethality of the *ysh1-12* mutations at 17ºC. In contrast, *RRP6* deletion caused a slow growth phenotype in the *ysh1-32* background at 17-30ºC. The above results indicated that Ysh1p interacts genetically with Rrp6p.

Mutations in *RRP6* lead to the accumulation of polyadenylated 3' end extended forms of snRNAs, snoRNAs, and rRNAs (van Hoof et al., 2000). To further analyze phenotypes of the *ysh1-12 rrp6Δ* and *ysh1-32 rrp6* alleles, we performed Northern blot experiments with total RNAs extracted from both the single and the double mutant strains grown at permissive and non-permissive temperatures. Consistently, we found that the *ysh1-32 rrp6Δ* double mutant accumulated the polyadenylated species of the *SNR33* transcript at permissive temperature similarly as in the *rrp6Δ* strain, but not at 37ºC, the temperature restrictive for the *ysh1-32* allele (data not shown). Such effect was not observed in the wild-type and *ysh1-12 rrp6Δ* double mutant strain, neither at permissive (30ºC), nor restrictive (15ºC) temperatures. This preliminary result implied that Ysh1p (possibly together with the 3' end processing machinery) is involved in the synthesis of the poly(A)$^+$ *SNR33* transcript, and mutations in *ysh1-32* abolish that process.

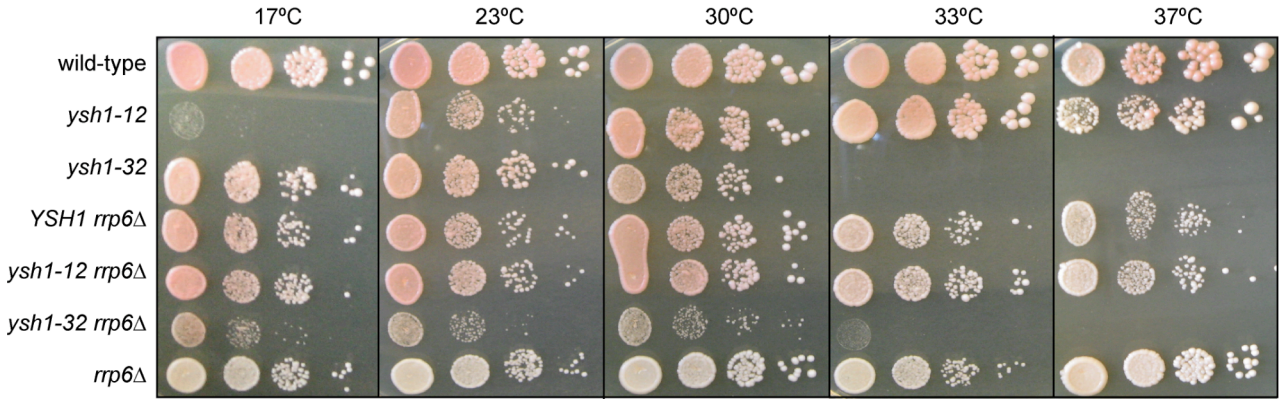


Figure 2.21 *YSH1* genetically interacts with a nuclear exosome subunit, *RRP6*

Tenfold serial dilutions of cultures of wild-type and *ysh1* mutant strains were spotted onto YPD medium and incubated for 3 days at 23ºC, 30ºC, 33ºC, and 37ºC, or for 5 days at 17ºC.

Splicing is impaired in the *ysh1-12* mutant strain

The only cold-sensitive *ysh1* mutant allele described so far, *brr5-1*, originated from a screen for splicing factors and was found to be defective in splicing (Noble and Guthrie, 1996). To determine whether *ysh1* mutants could exhibit general defects in splicing, we examined the *SNR17A* and *SNR17B* transcripts. Levels of *SNR17A* and *SNR17B* pre-snoRNAs are highly sensitive indicators of defects in the first chemical step of splicing (Noble and Guthrie, 1996). *SNR17A* encodes for the nucleolar *U3* snoRNA, as does *SNR17B*. The two genes differ in the size and sequence of the introns that interrupt the coding sequence. Using a primer extension assay, we analyzed the *ysh1* mutant alleles for their ability to splice *SNR17A* and *SNR17B*. We found that the *ysh1-12* strain exhibited a gradual accumulation of unspliced precursors derived from both *U3* genes following shift to non-permissive temperature (Figure 2.22 A). Splicing defect was present already at permissive temperature (Figure 2.22 B). The *rse1-1* mutant strain was analyzed as control and found to exhibit a similar splicing defect, as expected (Figure 2.22 C; Chen et al., 1998).

Moreover, Northern analysis of several intron-containing gene transcripts in the *ysh1-12* strain detected extended species that were likely to be the unspliced forms of *ACT1, CYH2* and *RPS16A* pre-mRNAs (Figure 2.22 D; data not shown). Sizes of the extended RNA transcripts corresponded to the sizes of the respective unspliced pre-mRNAs. To confirm that these transcripts resulted of the impaired splicing, and not of impaired transcription termination, primer extension analysis was performed using an oligonucleotide complementary to the *ACT1* exon 2. Unspliced pre-mRNA was detected in *ysh1-12* strain grown at restrictive temperature, thus confirming splicing defect present in this mutant (Figure 2. 22 E). We conclude that mutations in the *ysh1-12* allele impair the mRNA/snoRNA splicing in yeast.

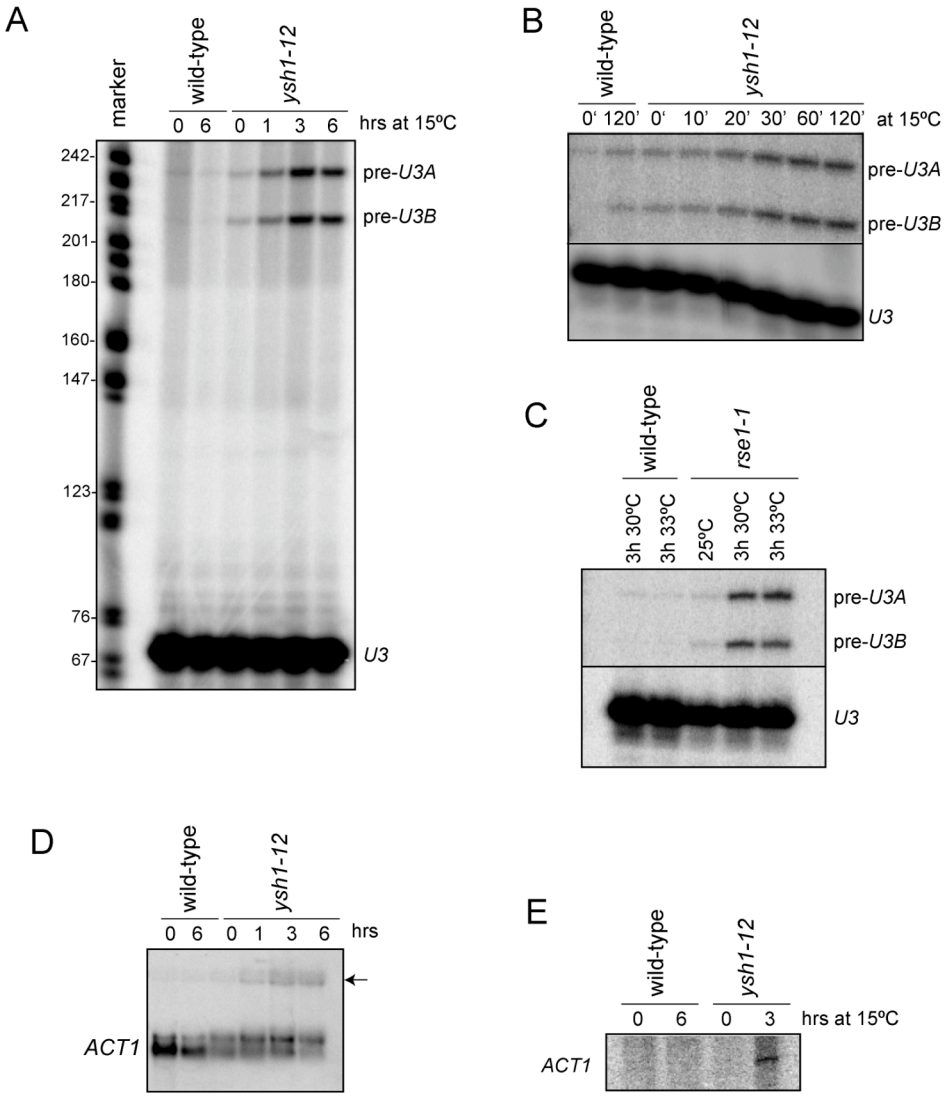


Figure 2.22 The *ysh1-12* mutant allele is impaired in splicing

(**A**) Primer extension analysis of spliced and unspliced *U3* transcripts in wild-type and *ysh1-12* mutant strain grown at 15ºC for up to 6 h was carried out using radioactively labelled oligonucleotide complementary to sequence located in *U3* exon 2. Bands corresponding to pre-*U3A,* pre-*U3B*, and mature *U3* snoRNA are indicated on the right.

(**B**) As in (**A**), except total RNA from *ysh1-12* strain shifted to 15ºC for shorter time points (0', 10', 20', 30', 60', 120') was analyzed.

(**C**) Primer extension analysis was used to visualize spliced and unspliced *U3* transcripts in isogenic wild-type and *rse1-1* mutant strain grown at semi-permissive (30ºC) and restrictive (33ºC) temperatures for 3 h. Radioactively labelled oligonucleotide complementary to the sequence located in *U3* exon 2 was used. Bands corresponding to pre-*U3A*, pre-*U3B* and mature *U3* snoRNA are indicated.

(**D**) Northern blot analysis of total RNA extracted from the wild-type and mutant *ysh1-12* cells grown at 30ºC and shifted to 15ºC for 1, 3 and 6 hours. RNAs were separated on formaldehyde/1.2% agarose gels. Filters were developed with random-primed labelled probe directed against the *ACT1* mRNA transcript.

(**E**) Primer extension analysis of *ACT1* pre-mRNA transcripts in wild-type and *ysh1-12* mutant strain was carried out using radioactively labelled oligonucleotide complementary to sequence located in the *ACT1* exon 2.

ysh1 mutants are deficient in early pre-rRNA processing steps

We showed that the *ysh1-12* mutant strain is defective in 3' end processing of snoRNA transcripts. As snoRNAs are generally required for rRNA synthesis, we analyzed processing of the 35S pre-rRNA in the *ysh1* mutant. Ribosomal RNAs are transcribed by RNA polymerase I as a polycistronic precursor, the 35S pre-rRNA. The synthesis of mature 5.8S, 18S and 25S rRNAs requires a complex series of post-transcriptional processing steps of the precursor rRNA. A large number of *trans*-acting factors are required for the maturation of rRNA transcripts, such as small nucleolar ribonucleoprotein (snoRNP) particles, RNA helicases, endonucleases (RNase MRP, Rnt1p), 5'→3' exonucleases (Rat1p; Xrn1p) and 3'→5' exonucleases (the exosome complex) (Venema and Tollervey, 1999). Schematic structure of the 35S pre-rRNA and its processing sites is shown in Figure 2.23 A. The primary transcript 35S is processed by a series of sequential cleavages. Initial cleavage in the 3'-ETS by Rnt1p yields the 35S pre-rRNA. The snoRNP-dependent cleavage at site A_0 in the 5'-ETS then generates 33S pre-rRNA, which is rapidly cleaved at site A_1, producing the 32S pre-rRNA. Cleavage at site A_2 in ITS1 then splits the 32S pre-rRNA into the 20S and $27SA_2$ pre-rRNAs. The 5' part of the molecule, 20S pre-rRNA, is exported to the cytoplasm and endonucleolytically cleaved at site D to generate mature 18S rRNA. Processing at sites C_1 and C_2 releases the mature 25S rRNA and the 7S pre-rRNA. The 7S pre-rRNA is 3' processed by the exosome complex, generating the 6S pre-rRNA, which is then trimmed to the mature 5.8S. The exosome also degrades the excised spacer region from the 5' end of the primary transcript to site A_0 (Allmang et al., 2000).

In our Northern analyses, several pre-rRNA processing defects were seen upon shifting the *ysh1-12, ysh1-32* and *rna15-1* strains to restrictive temperatures. The 35S pre-rRNA was accumulated while the 32S pre-rRNA, the product of A_1 cleavage, was depleted in all the three mutants (Figure 2.23 B, C, D, E). Aberrant 23S and 21S RNAs were also accumulated in the mutants, which is due cleavage at site A_3 in ITS1 in the absence of prior processing at sites A_0–A_2 (Henry et al., 1994). Cleavage at site A_0 is also inhibited, as the 23S RNA appears in the mutants (Figure 2.23 C, D). We conclude that the early pre-rRNA processing at sites A_0, A_1 and A_2 is inhibited in *ysh1-12, ysh1-32* as well as in *rna15-1* mutant strains.

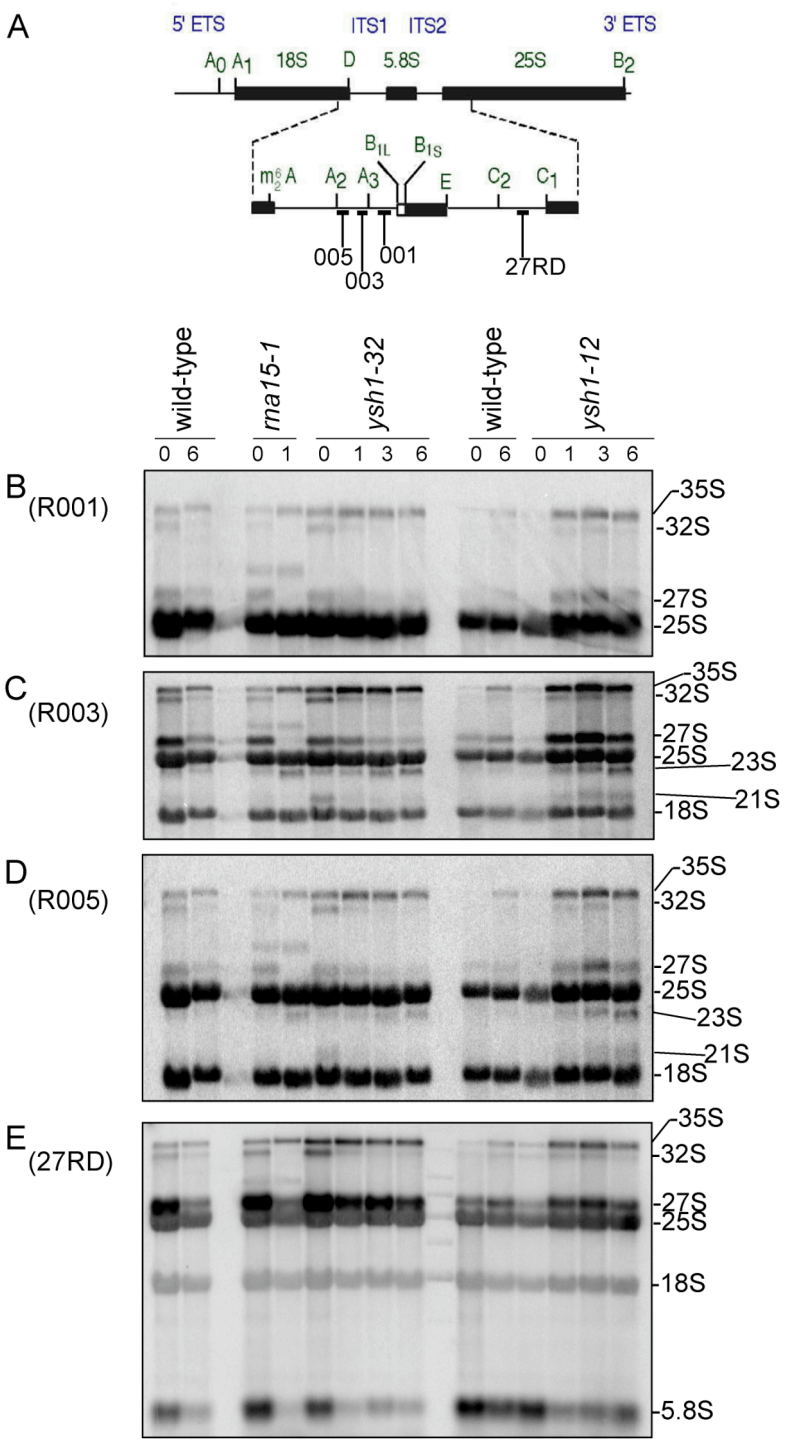


Figure 2.23 Northern analysis of pre-rRNA processing in the *ysh1* mutants

(**A**) Structure of the 35S pre-rRNA with the location of oligonucleotide probes used for Northern hybridization.

(**B**) Hybridization with probe 001, complementary to ITS1 downstream of site A_3.

(**C**) Hybridization with probe 003, complementary to ITS1 upstream of site A_3.

(**D**) Hybridization with probe 005, complementary to ITS1 downstream of site A_2.

(**E**) Hybridization with probe 27RD, complementary to ITS2.

Northern analysis was performed essentially as in Figure 2.6.

2.4 Discussion

In this study we investigated the role of Ysh1p in mRNA and snoRNA metabolism. Our results are consistent with the view that Ysh1p plays an essential role in 3' end formation, probably through catalysis of endonucleolytic cleavage of the pre-mRNA (schematically summarized in Figure 2.24). In addition to this, we provide evidence that Ysh1p is involved in termination of RNAP II both at mRNA and snoRNA genes, that it may contribute to the accumulation of distinct mRNAs independent of its role in 3' end formation and that it may act in pre-mRNA splicing.

A striking outcome of our study is that distinct sets of phenotypes were associated with specific groups of *ysh1* mutant strains. Table 2.2 summarizes a broad range of phenotypic observations, which were either resulting from our work or were previously described by others. Based on biochemical phenotypes two groups, class I and class II, can be formed by the ts strains (*ysh1-13, ysh1-15, ysh1-32* and *brr5-td*) and the cs strains (*ysh1-12* and *brr5-1*), respectively. Extracts prepared from strains of both classes displayed defects in 3' end processing *in vitro*. The only exception is the *brr5-1* extract, which was reported not to have a clear cleavage defect (Chanfreau et al., 1996). Moreover, all analyzed ts and cs alleles of *YSH1* showed an aberrant pattern of *ACT1* poly(A) site usage. These observations are consistent with the proposal that Ysh1p indeed represents the 3' endonuclease. However, defects in 3' end cleavage and polyadenylation are generally associated with reduced mRNA stability. Consistently, class I mutants show under-accumulation of several mRNAs and a reduction of the global poly(A) tail length distribution. In stark contrast, class II mutations did not cause a general loss of mRNA levels. Moreover, phenotypes in snoRNA 3' end formation and termination, in the control of *NRD1* mRNA accumulation and in RNA splicing were observed in the *ysh1-12* cs mutant strain. Notably, these phenotypes were not seen with class I mutants. Such striking differences of mutant phenotypes cannot easily be explained by a simple loss of function model and thus complicates the mechanistic interpretation of the role of Ysh1p in different pathways of cellular RNA synthesis. Consistent with this idea it was previously suggested that cold sensitive phenotypes can unravel physiologic features that may be resistant to mutation to temperature sensitivity (Moir et al., 1982; Noble and Guthrie, 1996).

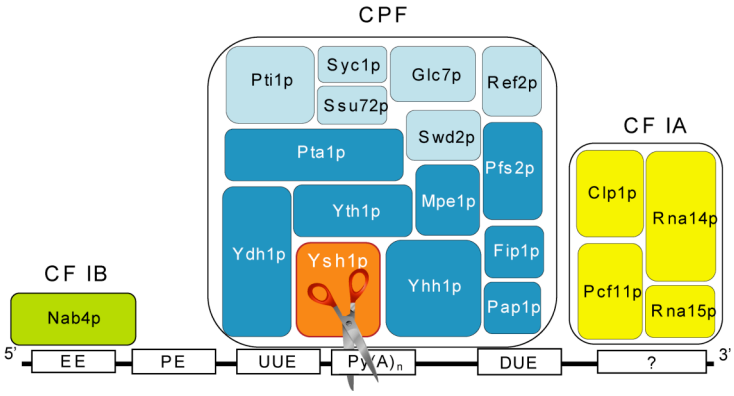


Figure 2.24 Ysh1p is the yeast 3' end processing endonuclease

A simplified model of the yeast core polyadenylation machinery (CPF, CF IA, CF IB) assembling in a pre-cleavage complex on an mRNA precursor, with Ysh1p acting at the site of cleavage (symbolically marked with scissors).

Table 2.2 *ysh1/brr5* phenotypes

	class I				***class II***	
allele	***ysh1-13***	***ysh1-15***	***ysh1-32***	***brr5-td***	***ysh1-12***	***brr5-1***
growth phenotype	ts[a]	ts[a]	ts[a]	ts-degron[d]	cs[a]	cs[b]
cleavage defect *in vitro*	yes[a]	yes[a]	yes[a]	yes[d]	yes[a]	no[c]
polyadenylation defect *in vitro*	yes[a]	yes[a]	yes[a]	yes[d]	yes[a]	yes[c]
mRNA stability	unstable[a]	unstable[a]	unstable[a]		stable[a]	
ACT1 poly(A) site selection defect	yes[a]	yes[a]	yes[a]		yes[a]	
termination defect on mRNA genes	ND	yes[a]	yes[a]	yes[e]	yes[a]	no[d]
read-through mRNA transcripts	no[a]	no[a]	no[a]		yes[a]	yes[c]
NRD1 mRNA upregulation	no[a]	no[a]	no[a]		yes[a]	
termination defect on snoRNA genes	ND	ND	no[a]	no[e]	yes[a]	
3' end extended snoRNA transcripts	ND	ND	no[a]		yes[a]	
splicing defect	ND	ND	no[a]		yes[a]	yes[b]

References:
a) This work
b) Noble and Guthrie, 1996
c) Chanfreau et al., 1996
d) Zhelkovsky et al., 2006
e) Kim et al., 2006
ND – not determined

In the 'torpedo' model of RNAP II termination, cleavage at the poly(A) site is an obligatory step to provide an entry point for the Rat1/Xrn2 exonuclease, which contributes to termination through exonucleolytic degradation of the RNA fragment that remains associated with the polymerase following 3' end cleavage (Kim et al., 2004; West et al., 2004; West et al., 2008). Recently, this model has been challenged by the observation that although Rat1p associates with 3' end factors and may contribute to the assembly of the processing machinery, degradation of the downstream RNA appears not to be essential for termination (Luo et al., 2006). Interestingly, it was also shown that degradation of the 3' cleavage fragment is stimulated when coupled to 3' end cleavage in mammalian cells (Kaneko et al., 2007). Because it was demonstrated previously that exonuclease activity is required for Rat1p termination activity (Kim et al., 2004), the role of Rat1p in this process remains paradoxical (Rosonina et al., 2006). We found that class I and class II *ysh1* mutants were defective in transcription termination when assayed on the *CYC1* gene via transcriptional run-on. These results suggest that Ysh1p, the endonuclease acting in the 3' end processing of pre-mRNA transcripts, is also essential for RNAP II termination. Unlike

other 3' end processing factors previously shown to be involved in termination, Ysh1p interacts neither with the CTD of RNAP II, nor with RNA (Kyburz et al., 2003; Sadowski et al., 2003; our unpublished results). We therefore predict that it is the endonucleolytic cleavage that contributes mechanistically to transcription termination, possibly by providing the entry site for exonucleases (Figure 2.25). Moreover, transcriptional read-through products were detected in the class II *ysh1-12* and *brr5-1* strains but not in the class I strains. As both classes of mutants were defective in 3' end cleavage and in RNAP II termination it seems that aberrant transcripts were somehow stabilized in class II mutants, pointing to a possible role of Ysh1p in RNA quality control.

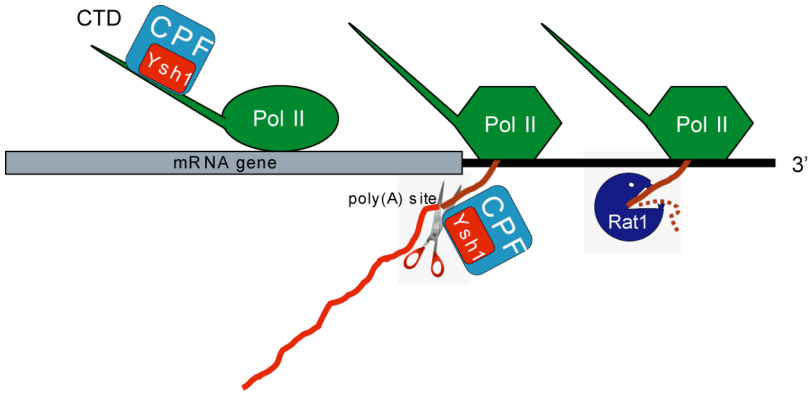


Figure 2.25 Ysh1p is involved in RNAP II transcription termination on mRNA genes
A simplified scheme of the 'torpedo' model of transcription termination. RNAP II transcribes past the poly(A) site. 3' end processing machinery, including CPF, recognizes the poly(A) site and Ysh1p endonucleolyticaly cleaves the RNA transcript (symbolically marked with scissors). This contributes to the allosteric changes of the RNAP II complex that make it termination-prone. Additionally, the cleavage step provides an entry site for the Rat1p 5'-3' exonuclease that digests the 3' cut-off product and catches up with the transcribing polymerase. This, together with the dissociation of CPF, leads to RNAP II transcription termination.

The mechanism of 3' end processing and transcription termination on independently transcribed snoRNA genes remains unclear. Several lines of evidence support the notion that the canonical 3' end processing complex is involved in pre-snoRNA 3' end formation and that an endonucleolytic cleavage step may be associated with this process. Specific cleavage sites were identified on *SNR13* and *SNR47* snoRNAs as primer extension stops in *rat1-1 Dxrn1* double mutant cells (Fatica et al., 2000). Several subunits of CFIA and CPF complexes are important for snoRNA termination (see Introduction) and ChIP experiments clearly show that poly(A) factors and Rat1p exonuclease are recruited to snoRNA genes (Kim et al., 2006). ChIP analysis of elongating RNAP II suggested that degron-mediated depletion of Ysh1p did not affect snoRNA transcription termination (Kim et al., 2006); however, these experiments did not reveal whether cleavage occurs at the snoRNA 3' ends or not. In our analyses the *brr5-td* strain did not accumulate 3' extended snoRNA transcripts following depletion of Ysh1p suggesting that the protein may not play an essential role in this process. Nevertheless, our analysis of the class II *ysh1-12* allele demonstrated that Ysh1p has a function in snoRNA 3' end formation and

transcription termination. Canonical 3' end processing factors and Rat1p are recruited to genes transcribed by RNAP II regardless of whether they encode mRNA or snoRNA (Kim et al., 2006). It is possible that an early stage of termination is similar for both types of transcripts and that in a later phase of termination a combination of factors used by RNAP II for 3' end formation may vary depending on the specific snoRNA gene (Kim et al., 2006). A speculative scenario therefore is that the phenotype associated with the cs *ysh1-12* strain might unravel an early stage of the snoRNA transcription termination that requires a certain function of Ysh1p/CPF.

Whereas a mechanistic understanding of RNAP II termination at the end of transcription units is emerging there is little information available on regulated premature termination within a transcription unit. Indeed, the case of *NRD1* autoregulation is the best documented example (Arigo et al., 2006). It has been suggested that the RNA binding proteins Nrd1p and Nab3p recognize sequence elements found within the *NRD1* mRNA to trigger RNAP II termination prematurely, i.e. before the 'regular' poly(A)-dependent terminator in the 3' UTR is transcribed (Arigo et al., 2006). It was shown that this autoregulation involves the exosome and the TRAMP complex (Arigo et al., 2006) but it remains unclear whether endonucleolytic cleavage and possibly a torpedo-like termination mechanism may play a role as well. We observed elevated levels of the full length *NRD1* transcript in the *ysh1-12* strain suggesting a negative role for Ysh1p in regulation of *NRD1* expression. Importantly, we were also able to demonstrate the existence of a *NRD1* 3' cut-off product in *rat1-1, Dxrn1* cells, which is consistent with a mechanism involving endonucleolytic cleavage. These results imply a role for Ysh1p in *NRD1* premature termination and thus for regulation of *NRD1* mRNA levels, possibly via its cleavage activity. Therefore, Ysh1p could be the first example of a 3' end processing factor involved in regulated premature RNAP II termination.

Interestingly, Northern blot analyses showed that the levels of several other mRNAs were elevated in the *ysh1-12* strain as well. Among the messages accumulating in the *ysh1-12* mutant we identified *NAB2*, which is believed to autoregulate the levels of its own mRNA in a process that requires the nuclear exosome component Rrp6p (Roth et al., 2005). This post-transcriptional mechanism involves a sequence of 26 adenosines (A_{26}) in the *NAB2* 3' UTR, which represses *NAB2* 3' end formation and subsequently directs the transcript for degradation by Rrp6p (Roth et al., 2005). The increased concentration of *NAB2* mRNA in the *ysh1-12* strain suggests a contribution of Ysh1p to the autogenic control of *NAB2* levels. Interestingly, we found that *YSH1* genetically interacts with the nuclear exosome subunit Rrp6p, as deletions of *RRP6* suppressed the cs growth phenotype of *ysh1-12* cells (results not shown). However, further investigations of the involvement of Ysh1p in *NAB2* autoregulation are required. Our results indicate that Ysh1p directly or indirectly influences the nuclear turnover of mRNAs. Class II mutations in *ysh1* alleles could inhibit correct mRNPs assembly that in turn may lead to impaired mRNA export and retention of the mRNA in the nucleus. An alternative and more intriguing possibility would be that the endonucleolytic cleavage activity of Ysh1p directly contributes to RNA surveillance. In this case, we would predict that also class I mutations result in some upregulation of *NAB2* and/or *NRD1* mRNAs; however, such phenotypes were clearly not observed in these mutant strains.

Taken together, our results support the idea that Ysh1p is involved in the negative regulation of certain mRNAs and that it may contribute to RNA surveillance.

Our analyses of the *ysh1-12* allele revealed a splicing defect that correlates well with the phenotype observed previously for the cs *brr5-1* allele (Chanfreau et al., 1996; Noble and Guthrie, 1996). However, the molecular basis of these defects remains unclear. snRNA synthesis appeared normal in *ysh1* mutant strains (result not shown) and this excludes the possibility that snRNA depletion underlies the defect. We cannot rule out the possibility that secondary effects resulting from under-accumulation of critical splicing components contribute to the phenotype. Nevertheless the possibility remains that Ysh1p acts in the coupling of splicing and 3' end processing. Evidence for such coupling is ample in mammalian cells but so far was not reported in yeast. Mechanisms of coupling include physical interactions between components of the 3' end formation and splicing machineries, e.g. an interaction between CPSF and the U2 snRNP (Kyburz et al., 2006), and between U2AF 65 and CF I_m (Millevoi et al., 2006). Interestingly, Rse1p, the yeast homologue of SF3b130 and component of U2 snRNP, genetically interacts with different subunits of CPF (A. Kyburz and W. Keller, unpublished observations). Future experiments will have to address the question whether there is indeed a functional interdependence of splicing and poly(A) addition in yeast.

The complicated process of pre-rRNA processing and pre-ribosome assembly has been extensively studied in yeast and resulted in detection of more than 150 non-ribosomal proteins and myriad of small nucleolar RNAs (snoRNAs) participating in these processes (Kressler et al., 1999). However, the function of the vast majority of these proteins remains unknown. Surprisingly, we found that mutations in the *ysh1* alleles inhibit early pre-rRNA processing steps. Moreover, a similar effect was seen in *rna15-1,* the mutant allele of a CF IA subunit Rna15p. This is highly unexpected and suggests rather an indirect involvement of the canonical 3' end processing factors in this process. rRNA genes are transcribed by RNAP I complex, therefore factors primarily associated with transcribing RNAP II such as 3' end processing machinery are most certainly not recruited to those genes. Notably, depletion of the U3 snoRNA strongly inhibits pre-rRNA processing at sites A_0, A_1 and A_2 (Hughes and Ares, 1991), sites that are also inhibited in the *ysh1-12, ysh1-32* and *rna15-1* strains. Of all the *ysh1* alleles analyzed, only *ysh1-12* is impaired in 3' end formation of many snoRNAs involved in pre-rRNA formation, therefore this defect does not account for the pre-rRNA processing inhibition which was observed also in *ysh1-32* and *rna15-1* mutant strains. Intriguingly, processing at each of the sites A_0, A_1 and A_2 is by endonucleolytic cleavage (Venema et al., 1995). However, any direct role of Ysh1p or CPF in these cleavage steps seems most improbable. Similar phenotype was seen for the nuclear exosome mutants (Allmang et al., 2000); however, it is likely an indirect effect as none of these processing steps involves 3'-5' exonuclease activities. The most probable explanation for the phenotype in rRNA processing is that mutations in Ysh1p cause alterations in the mRNA levels, and thus in the protein levels, of either factors directly involved in those rRNA processing steps or of factors involved in a quality control system that in turn inhibits pre-rRNA processing.

2.4 Experimental procedures

Yeast strains

Manipulations and growth of *S. cerevisiae* were carried out by established procedures. The *S. cerevisiae* strains used in this study and their genotypes are listed in Table 2.3. For random mutagenesis of *YSH1*, mutagenic PCR was carried out with a low concentration of ATP (Ohnacker et al., 2000). The fragment containing the pRS314 plasmid sequences and sequences flanking the *YSH1* gene were co-transformed together with the mutagenized PCR product into the yeast strain LM109. Transformants were selected on minimal medium lacking histidine and replica-plated onto 5'-FOA plates in order to shuffle out the *YSH1-URA3* rescue plasmid. The colonies were then replica-plated onto YPD plates and incubated at 15, 23, 30 or 37°C, respectively. Plasmids of candidate colonies that showed a ts or cs phenotype were isolated by standard procedures. The conditional growth phenotype was then verified by retransformation of the isolated plasmids (pBD33, pMG56, pMG57, pMG58, and pMG59) into the LM109 strain, 5'-FOA treatment and growth tests at low or elevated temperatures. To generate the double mutants of *YSH1* and *RRP6*, the LM109 strain was crossed with the YMG97 (*rrp6Δ)* strain. The spores were dissected and the resulting strains were tested for the auxotrophic phenotype to leucin. The resulting strain YMG156 was transformed with plasmids carrying either the wild-type *YSH1*, or the mutant *ysh1-12* (pMG56) or *ysh1-32* (pMG59) genes. Transformants were selected on minimal medium lacking histidine and replica-plated onto 5'-FOA plates in order to shuffle out the *YSH1-URA3* rescue plasmid. The genotypes of the resulting *rrp6Δ ysh1* mutant strains (YMG152, YMG154, and YMG155, respectively) were verified by PCR.

Droplet test

The strains were grown over night and diluted to an OD_{600} of 0.1, 0.01, 0.001, 0.0001. 5 µl of each dilution were spotted on five different YPD plates and incubated at 15, 23, 30, 33 and 37ºC.

Growth curve

The yeast strains were grown to an OD_{600} 0.4-0.6 and transferred to a water bath at 37ºC or 15ºC (time point 0). The OD_{600} was measured every 30 min. and the cells were diluted with fresh medium (pre-warmed to 37ºC or pre-cooled to 15ºC, respectively).

Extract preparation and *in vitro* cleavage and polyadenylation assays

3' end processing extracts were made following the procedure described previously (Ohnacker *et al*, 2001). Cleavage and polyadenylation assays were carried out as previously described (Minvielle-Sebastia et al., 1994). When cleavage only was assayed, EDTA replaced magnesium acetate (MgAc) and ATP was omitted. Internally ^{32}P- labelled *CYC1* was produced by *in vitro* run-off transcription. For reactions at restrictive temperatures, both the reaction mix and protein extracts were first pre-incubated separately for 15 min at respective temperatures, then combined and assayed for 60 min.

Table 2.3 Yeast stains used in this study

Strain	Genotype	Reference
W303-1A	*MAT* **a** *ade2 leu2 ura3 trp1-1 his3*	Thomas and Rothstein, 1989
LM109	*MAT* **a** *ura3-1 trp1Δ ade2-1 leu2-3,112 his3-11,15 TRP1::ysh1 [YSH1–URA1-CEN]*	Jenny et al., 1996
YMG44 (*YSH1*)	*MAT* **a** *ura3-1 trp1Δ ade2-1 leu2-3,112 his3-11,15 TRP1::ysh1 [YSH-HIS3-CEN]*	This work
YMG17 (*ysh1-12*)	*MAT* **a** *ura3-1 trp1Δ ade2-1 leu2-3,112 his3-11,15 TRP1::ysh1 [ysh1-12-HIS3-CEN]*	This work
YMG29 (*ysh1-32*)	*MAT* **a** *ura3-1 trp1Δ ade2-1 leu2-3,112 his3-11,15 TRP1::ysh1 [ysh1-32-HIS3-CEN]*	This work
YMG16 (*ysh1-13*)	*MAT* **a** *ura3-1 trp1Δ ade2-1 leu2-3,112 his3-11,15 TRP1::ysh1 [ysh1-13-HIS3-CEN]*	This work
YMG27 (*ysh1-15*)	*MAT* **a** *ura3-1 trp1Δ ade2-1 leu2-3,112 his3-11,15 TRP1::ysh1 [ysh1-15-HIS3-CEN]*	This work
YJC1166 (*nrd1-102*)	*MAT* **α** *ura3Δ0 his3Δ1 leu2Δ0 met15Δ0 nrd1-102*	Steinmetz et al., 2001
FY23	*MAT* **a**, *ura3-52, leu2Δ1, trp1Δ63*	Winston et al., 1995
YAZ212 (*brr5-td*)	*MAT* **a** *ura3-52 leu2Δ1 trp1Δ63 brr5::DHFR^{ts}-BRR5*	Zhelkovsky et al., 2006
rna15-1	*MAT* **a** *ura3-1 trp1-1 ade2-1 leu2-3,112 his3-11,15 rna15–1*	Minvielle-Sebastia et al., 1994
YMG156	*rrp6Δ ysh1Δ [YSH1-URA3]*	This work
YMG152 *(YSH1 ∆rrp6)*	ND	This work
YMG154 *(ysh1-32, ∆rrp6)*	ND	This work
YMG155 *(ysh1-12, ∆rrp6)*	ND	This work
CYK568 *(rse1-1)*	*MAT* **a**, *ura3-52, leu2Δ1, rse1-1*	Chen et al., 1998
D174 (*rat1-1, xrn1Δ*)	*MAT* **a** *ade2-1 xrn1::URA3 rat1-1*	Henry et al., 1994

ND – not determined

RNA analyses

Yeast total RNA was extracted using the hot phenol method. Northern analyses and RNase H experiments were carried out as described previously (Dichtl et al., 2004; Dichtl et al., 2002b). Oligonucleotides: anti-SNR3, anti-SNR13, anti-SNR33, anti-SNR45, anti-18S rRNA, anti-U3, anti-5S rRNA, and anti-U6, used as probes in Northern analyses, were routinely labelled with 30 µCi [γ-^{32}P] ATP and T4 polynucleotide kinase (New England Biolabs). Total poly(A) tail labelling was performed as described previously (Minvielle-Sebastia et al., 1998). The poly(A) tails were separated on a 10% denaturing polyacrylamide/8.3 M Urea/TBE gel. The reaction products were

visualized by autoradiography and an image was generated by a PhosphorImager (STORM). Transcriptional run-on analysis on *CYC1* gene was performed according to (Birse et al., 1998). The multi-copy plasmid pUG*CYC*, in which the *CYC1* gene is controlled by the strong GAL1/10 promoter, was introduced into cells of the wild-type, *ysh1-12* and *ysh1-32* strains to achieve high level of transcription. Transcriptional run-on analysis on endogenous *SNR3* gene was performed according to (Steinmetz et al., 2001), except only probes A and B were used. Strains were shifted to respective restrictive temperatures for 60' prior to run-on analysis. Primer extension analysis was performed accordingly to (Beltrame and Tollervey, 1992), with 6 µg of total RNA used. Sequences of the oligonucleotides used for primer extension analyses were as published (Steinmetz et al., 2001). Sequences of additional oligonucleotides used in those analyses [ACT1, U3exon2, U6, U (NRD1 +300), D (NRD1 +1300)] are listed in Table 2.4. All radioactive experiments were performed using [^{32}P] isotope; except for the transcriptional run-on analysis on SNR3 gene, where [^{33}P] isotope was used.

Computing analyses

Sequences have been found by several BLASTP searches at the NCBI-Genbank Website (http://www.ncbi.nlm.nih.gov/BLAST/). Protein sequences have been aligned with ClustalX (Jeanmougin et al., 1998) and T-Coffee (Notredame et al., 2000), and refined manually.

Plasmids and oligonucleotides

All the plasmids and oligonucleotides used in this study are listed in Tables 2.4 and 2.5, respectively. Plasmids pMG60, pMG61, and pMG62 encoding for the terminator 'SNR+' RNA templates were created as follows: products of PCR reactions carried out on yeast total genomic DNA with the respective primers (SNR39B+ For and SNR39B+ Rev; SNR47+ For and SNR47+ Rev; SNR65+ For and SNR65+ Rev) were cloned into the pDrive vector (Qiagen). Correct clones were confirmed by sequencing and used as templates for an *in vitro* transcription. The pBD33, pMG56, pMG57, pMG58, and pMG59 plasmids were derived in the procedure of the random mutagenesis of *YSH1* gene (see: Yeast strains).

Table 2.4 Plasmids used in this study

Plasmid	Description
pBD33 / pMG55	pRS314-*YSH1*
pMG56	pRS314-*ysh1-12*
pMG57	pRS314-*ysh1-13*
pMG58	pRS314-*ysh1-15*
pMG59	pRS314-*ysh1-32*
pMG60	pDrive-SNR39B+
pMG61	pDrive-SNR47+
pMG62	pDrive-SNR65+

Table 2.5 Oligonucleotides used in this study

Oligonucleotide	Sequence	Reference
ASN1-5'	5'-ATGTGTGGTATTTTCGCCGC	This work
ASN1-3'	5'-GAATCGTATACGTGGCCCGG	This work
DBP9-5'	5'-AGCTTGGAAGGCGAGCCTG	This work
DBP9-3'	5'-TCATTTGAAGTTCTTCAACG	This work
Nab3-5'	5'-GGTAATTTGCCGCTAAAGAACGT	This work
Nab3-3'	5'-CTGTTGCTGTTGTTGTTGTGGAG	This work
Nrd1-3'end For	5'-TGTTTCTTCCAAGGCAATATCTC	This work
Nrd1-3'end Rev	5'-TGTACTATGAGCAAATAAAGGG	This work
U-NRD1 +300	5'-CGTGTTTATGGCATGGGCACAAG	This work
D-NRD1 +1300	5'-CAGCGTCCGTGAGCCTGTGCATA	This work
R001	5'-CCAGTTACGAAAATTCTTG	Allmang et al., 2000
R003	5'-TGTTACCTCTGGGCCC	Allmang et al., 2000
R005	5'-ATGAAAACTCCACAGTG	Allmang et al., 2000
27RD	5′-GGCCAGCAATTTCAAGTTA	Kuai et al., 2004
PE snR46	5'-ATCGACCAGCTCTTTAGCATCC	This work
PE snR3	5'-GATGCAAACTGTGTCTCCTA	Steinmetz et al., 2001
PE snR13	5'-GACTCTGAGTATATTCTAGAGG	Steinmetz et al., 2001
PE snR47	5'-ACTCCAACTACTTTCGAGCA	Steinmetz et al., 2001
PE snR65	5'-GTTATGCATGTATTGTACTTG	This work
PE snR71	5'-GTTCAGGATGCTATAGAAGAGG	Steinmetz et al., 2001
PE snR128	5'-CCATTCCTCCAAGTAGTGAT	Steinmetz et al., 2001
PE snR33	5'-CAGATAAACAAGCTCAGTAGTAA	Steinmetz et al., 2001
PE snR39b	5'-AACGTAATACGTAGAAGAATGGA	Steinmetz et al., 2001
PE snR45	5'-ACAGATGAGATGACTACTCCCA	Steinmetz et al., 2001
PE snR50	5'- AAGAGTACACTACGGGCTAGAA	Steinmetz et al., 2001
PE U3exon2	5'-GAGCCACTGAATCCAACTTGGTTGA	This work
PE ACT1	5'-GATACCTCTCTTGGATTGAGCTTC	This work
SNR39B+ For	5'-TATAATACGACTCACTATAGATTTA GTGATGATACTGCCG	This work
SNR39B+ Rev	5'-TTATTTCAATATTTCAGAAAGG	This work
SNR47+ For	5'-TAATACGACTCACTATAGTATCTTT TTCCTTATTATACATTC	This work
SNR47+ Rev	5'-AGCCATTAGTAAGTACGCTTTG	This work
SNR65+ For	5'-TATAATACGACTCACTATAGTAAAA TGATGATTTTTTTAAAC	This work
SNR65+ Rev	5'-GAATTGGCCGTATCGAAGATG	This work
anti-SNR3	5'-CGCTTATCACGAATAAGACCGAG TGT	This work
anti-SNR13	5'-GGCAAAAGCCAAACAGCAACTCG CCAAATGCACTCATATTCATCATAT	Kyburz et al., 2003
anti-18S rRNA	5'-CAGACAAATCACTCCA	Dichtl et al., 2002a
anti-U3	5'-GAGCCACTGAATCCAACTTGGTT GAT	
anti-5S rRNA	5'-CTACTCGGTCAGGCTC	Dichtl et al., 2002a
anti-U6	5'-GGGGAACTGCTGATCATCTCTGTA TTG	This work
RPB1-5'	5'-CAGAGGCTAAAAAGAAAGTTTTG	This work
RPB1-3'	5'-GATCTATGTGGAAAGTTTGTTG	This work
SCR1 5'	5'-AGGCTGTAATGGCTTTCTGGTGG	This work
SCR1 3'	5'-TATGGTTCAGGACACACTCCATCC	This work
SEN1-5'	5'-CTCAATACGTCGCAGGCTGAGGC	This work
SEN1-3'	5'-GAAAGAACAGTTGGTGGTAGTTG	This work
YOL087C-5'	5'-GTCCTGATAATGTACACGATGG	This work
YOL087C-3'	5'-GTTCATGGTGATGATGATGATGG	This work
YSH1-5'	5'-GTCAATGGTATCAAATTTACGGC	This work
YSH1-3'	5'-GGTTATTTCCGGATTATTTATGG	This work

2.5 Additional results

Cross-linking CPF/Ysh1p to the pre-mRNA 3' end cleavage site

In order to unambiguously show that Ysh1p is the enzyme cleaving the 3' ends of pre-mRNAs, we undertook site-specific cross-linking experiments. Such approach enables determination of proteins interacting with a single internal nucleotide of interest by purposely incorporating a photo-cross-linkable group and associated radioactive tag at this position. Upon UV irradiation (365 nm), the 4-thiouridine (4-thioU) is activated and can form a tertiary interaction via a covalent bond (cross-link) to a stacked, neighbouring nucleotide (~3 Å). A method for site-specific incorporation of modified oligonucleotides via splinted ligation was developed in the early 1990s (Moore and Sharp, 1992). In general, two RNA oligonucleotides, one of which carries one or more labels, are ligated together with the use of a guiding DNA oligonucleotide that is complementary to the site of ligation. A mayor advantage of splinted RNA ligation is the incorporation of single radiolabelled nucleotides or photoactivatable groups for cross-linking analysis. This has been particularly important in probing protein-RNA and RNA-RNA interactions at or near sites of chemistry. Applications of this technique to studies of protein translation and pre-mRNA splicing have provided detailed information as to how a short mRNA makes close contacts with the ribosomal constituents (Rinke-Appel et al., 1991; Stade et al., 1989), and how the conserved elements in pre-mRNA (the 5' and 3' splice sites and branch site) are recognized (Sontheimer and Steitz, 1993; Wyatt et al., 1992).

Cleavage of the *CYC1* 3' UTR is believed to occur within the CAAA sequence, downstream of the efficiency and positioning elements (Dichtl and Keller, 2001). Using 3' UTR sequence of the *CYC1* gene as a template (sCYC1), we designed three internally 4-thioU labelled 66 nt RNA substrates (called SL 1-3, for splinted ligation product; Figure 2.26 A, B) and obtained them using the splinted ligation method. While all the substrates have the same backbone based on the sCYC1 sequence, their cleavage site sequence differs. In SL1 RNA, 4-thioU is incorporated in position +1 of the cleavage site (C to 4-thioU substitution); in SL2 – in position +5 (U to 4-thioU substitution), and in SL3 – in position +2 (A to 4-thioU substitution). They are summarized schematically in Figure 2.26 B. A typical autoradiogram of the radioactively labelled SL substrates obtained via splinted ligation method is shown in Figure 2.26 C. Note that only a part of the [^{32}P]-4-thioU RNA oligonucleotide 2 is efficiently ligated to the upstream RNA oligonucleotide 1.

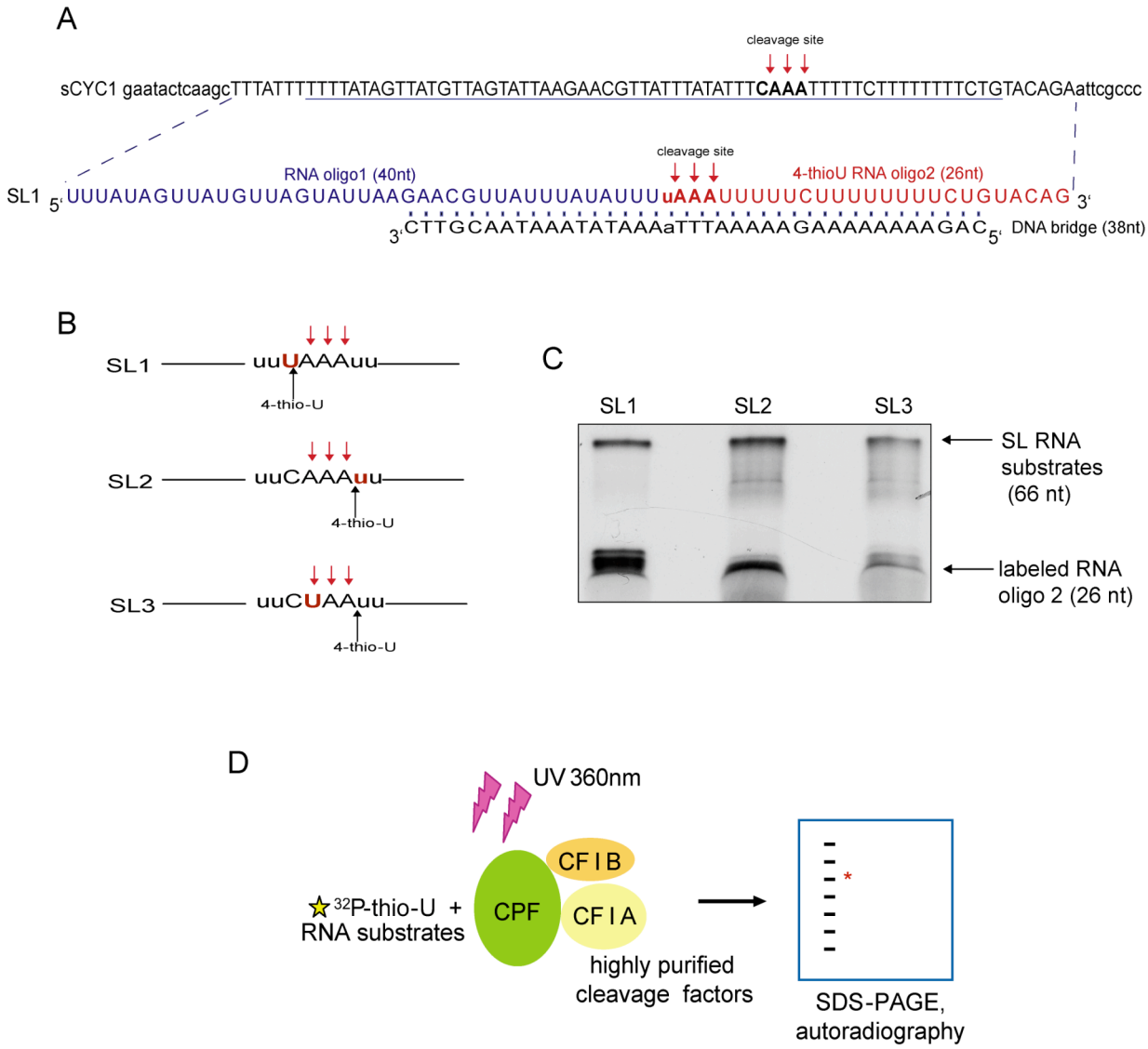


Figure 2.26 SL RNA substrates used for the specific cross-linking experiments

(**A**) Sequences of the sCYC1 RNA used as a template to create the SL substrates, the upstream RNA oligonucleotide, the downstream 4-thioU RNA oligonucleotide and the DNA bridge oligonucleotide; see text for explanations. (**B**) Schematic representation of the different SL substrates and sequences of their cleavage sites harbouring 4-thioU. The 4-thioU nucleotide is marked in red. Arrows indicate where cleavage occurs under natural conditions. (**C**) A typical autoradiogram of the radioactively labelled SL substrates obtained via splinted ligation method. Positions of SL substrates and of the labelled downstream RNA oligonucleotide are indicated. (**D**) Scheme representing the principle of the specific cross-linking experiment; see text for explanations.

To ensure that the introduction of 4-thioU into the RNA molecule did not change the substrate's cleavage competency, the SL1 substrate was tested in a cleavage assay *in vitro* (Figure 2.27). *In vitro* transcribed 5' end labelled sCYC1 RNA was used as a control. Both RNA substrates, SL1 and sCYC1, were almost equally well cleaved by wild-type and *syc1Δ* cell extract (Figure 2.27 A) as well as by purified 3' end processing factors CPF, CF IA and IB (Nab4p) (Figure 2.27 B). CPF and CF IA devoid of Nab4p cleaved the substrates more efficiently, but non-specifically. These results showed that the SL1 RNA is competent for cleavage reaction *in vitro,* and that the SL substrates carrying a 4-thioU have the potential to probe the active catalytic centre of the 3' end processing endonuclease.

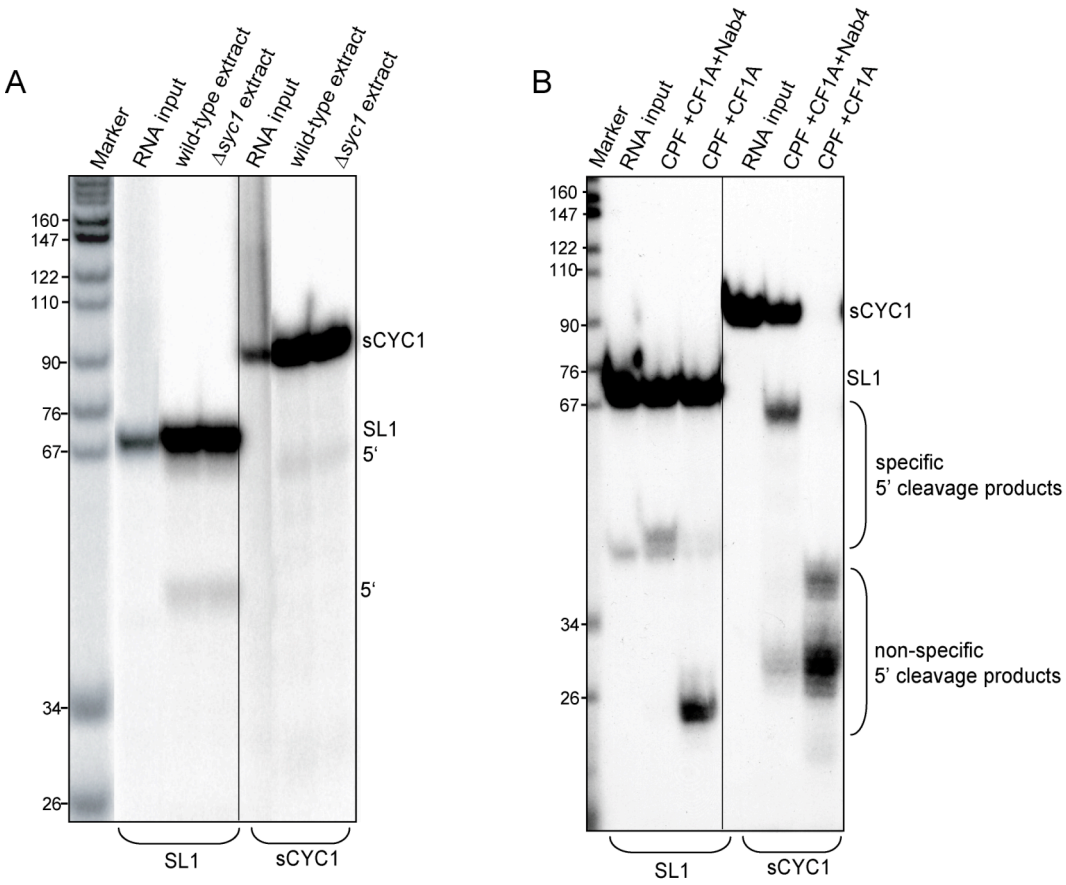


Figure 2.27 SL RNA substrate is efficiently cleaved by the 3' end processing machinery *in vitro*

(**A**) *In vitro* cleavage assay with wild-type cell extract and SL1 (left panel) or *in vitro* translated 5' end labelled sCYC1 (right panel) RNA substrates, respectively. Positions of the substrates and 5' end cleavage products are indicated on the right. Marker size is given on the left. Reactions were performed at 30ºC for 90 min. (**B**) As in (**A**), except purified 3' end processing factors (CPF, CF IA, Nab4p) were used. Positions of the specific and unspecific 5' end cleavage products are indicated on the right.

Using SL RNA substrates we performed RNA–protein cross-linking with the 3' end processing factors. Essentially, a cross-linking procedure that reflected the conditions of the cleavage reaction catalyzed by the 3' end processing machinery *in vitro* was adopted. Once the 3' end processing factors or total cell extracts and RNA substrate were mixed together in the reaction mix, they were preincubated at 30ºC to activate the protein factors and initiate the cleavage reaction. To increase the probability of cross-linking the 3' end endonucleolytic event, all manipulation after this preliminary incubation were performed on ice to reduce the cleavage rate of the substrate by the 3' end processing factors and to prevent the evaporation of the reaction mix exposed to UV irradiation. Shortly after the irradiation with UV (365 nm), the reactions were treated with RNases A and T1. As the 4-thioU nucleotide was site-specifically [^{32}P]-labelled, RNases treatment allows label transfer from RNA to its protein partner(s), which can subsequently be revealed on SDS-PAGE gels and autoradiograms. The cross-linking procedure was optimized with respect to several aspects (e.g., time of irradiation, reaction buffer composition, addition of RNA binding competitors).

We compared cross-linking profiles of the various SL RNA substrates, using cell extracts from wild-type and protein A-tagged Ysh1p, Ydh1p and Yhh1p strains (Figure 2.28). One of the repeatable results of the cross-linking experiments was a presence of an unidentified and highly unspecific ~55 kDa band. This band/protein did not have any specific affinity to the 4-thioU SL RNA substrates, as it appeared also under 'hard UV' light 280nm cross-linking conditions with internally labelled *in vitro* transcribed sCYC1 RNA substrate (data not shown). Heparin added to the wild-type extract reactions was unable to specifically compete with RNA for binding of the ~55 kDa protein (Figure 2.28). Instead, the cross-linking experiments in the presence of heparin led to decrease in the high molecular weight bands. Cross-linking bands approximately of 100 kDa (the size of Ysh1p) were analyzed thoroughly. They appeared for all of the SL RNA used in cross-linking. The cross-linking pattern of SL1 was similar to the one of SL3. Interestingly, cross-linking of SL2 RNA to the ProteinA-tagged Ysh1p resulted in a distinct double band that was absent in the other strains tested (Figure 2.28 B, marked with an asterisk).

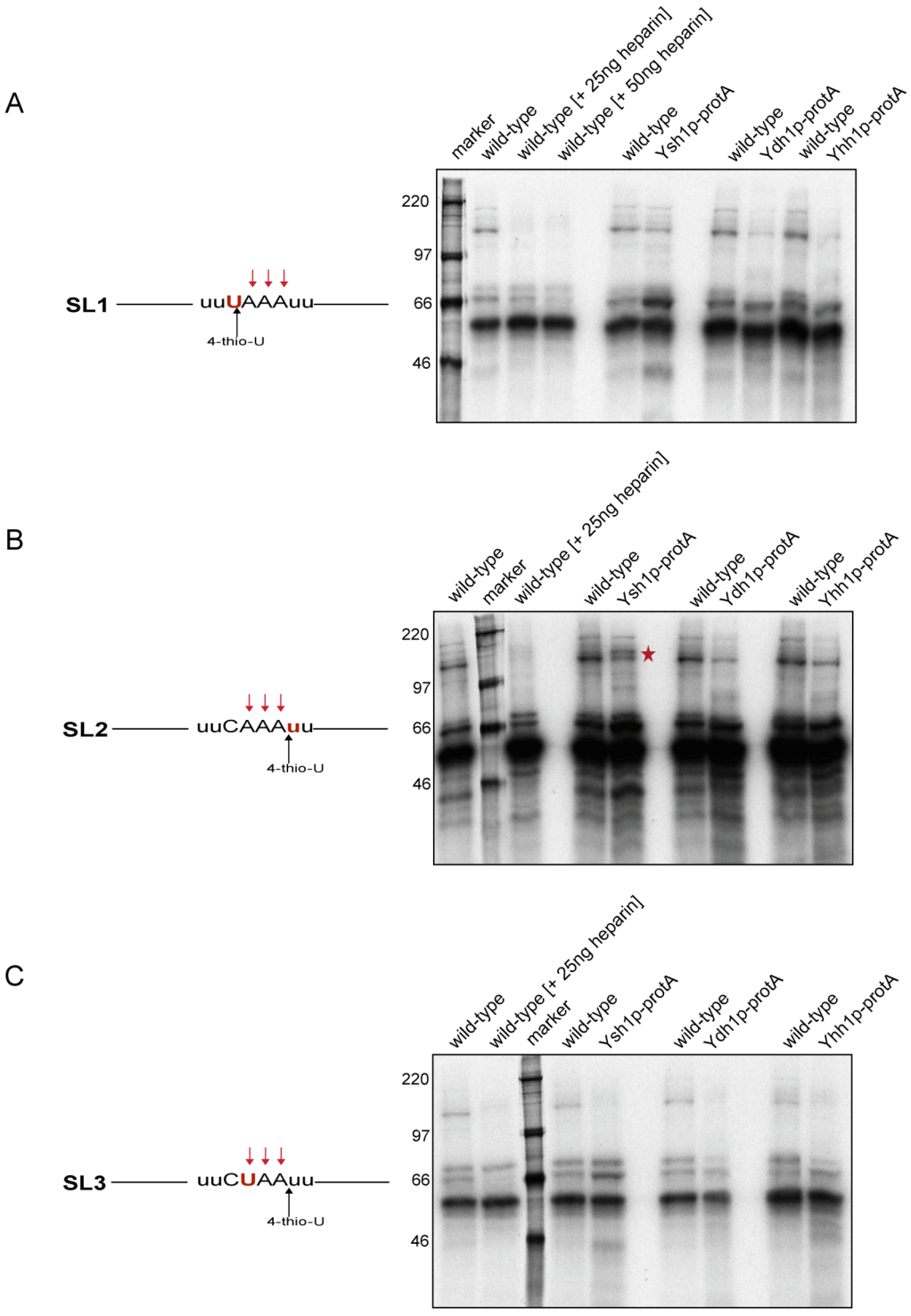


Figure 2.28 Comparison of the cross-linking experiments using various SL substrates

(**A**) Cross-linking experiments with SL1 RNA and extracts prepared from wild-type, Ysh1p-Noppata, Ydh1p-Noppata or Yhh1p-Noppata cells. For wild-type extract, heparin was used to compete for the unspecific binding to SL RNA. Schematic representations of the SL substrate and its cleavage site are given on the left. C14 protein marker sizes are indicated. (**B**) and (**C**), as in (**A**), except SL2 and SL3 RNA substrates were used, respectively.

Disappointingly, we were unable to obtain any conclusive cross-linking results using purified 3' end processing factors CPF and CF IA, even when the factors were provided in large excess to the RNA molecules. An over-exposition of the autoradiogram allowed the detection of the ~55 kDa non-specific band, however none around 100 kDa as expected for Ysh1p (data not shown).

To verify whether or not the cross-linked high-molecular bands observed in Figure 2.28 B correspond to Ysh1p, we combined the specific cross-linking experiment with Western blot detection. Cross-linking SL2 RNA with extracts prepared from wild-type, protein A-tagged- Ysh1p, Ydh1p and Yhh1p strains were separated by SDS-PAGE, followed by protein transfer to a nitrocellulose membrane and autoradiographic signal detection. Finally, a Western blot was performed using the anti-ProteinA or the anti-Ysh1p antibodies. Results from both steps of the experiment were marked and juxtaposed together to determine the exact positions of Ysh1p. An over-exposition of the autoradiogram was required to detect the >100 kDa cross-linked complex in ProteinA-tagged Ysh1p extract. The corresponding autoradiograms and chemiluminescent films are presented in Figure 2.29. The above experiment unambiguously showed that the high molecular range cross-linking bands *do not* correspond to Ysh1p.

Ysh1p interacts with subunits of CPF and CF IA

Ysh1p is an integral part of the CPF complex. We were interested with which of the other CPF subunits Ysh1p interacts and whether those interactions can be assigned to distinct regions of this protein. Therefore, N-terminally GST-tagged full-length and truncated versions of Ysh1p were expressed in *E.coli* (summarized in Figure 2.30 A). GST pull-down experiments were carried out with radioactively labelled *in vitro* translated proteins. Full-length GST-Ysh1p interacted with the CPF subunits Yhh1p, Ydh1p, Pfs2p and Pta1p. Moreover, it interacted also with the CF IA subunits Pcf11p and Clp1p (Figure 2.30 B), consistent with previous report (Kyburz et al., 2003). Intrestingly, Ysh1p interacted also with itself. The full length GST-Ysh1p did not pull down the CPF subunits Mpe1p, Pap1p, Pti1p, Ssu72p, nor the CF IA subunits Rna15p and Rna14p, Nab4p (CF IB) or Pab1p. From the above results we conclude that Ysh1p contributes to the stability and structural order of CPF, and it also bridges the two factors, CPF and CF IA, together.

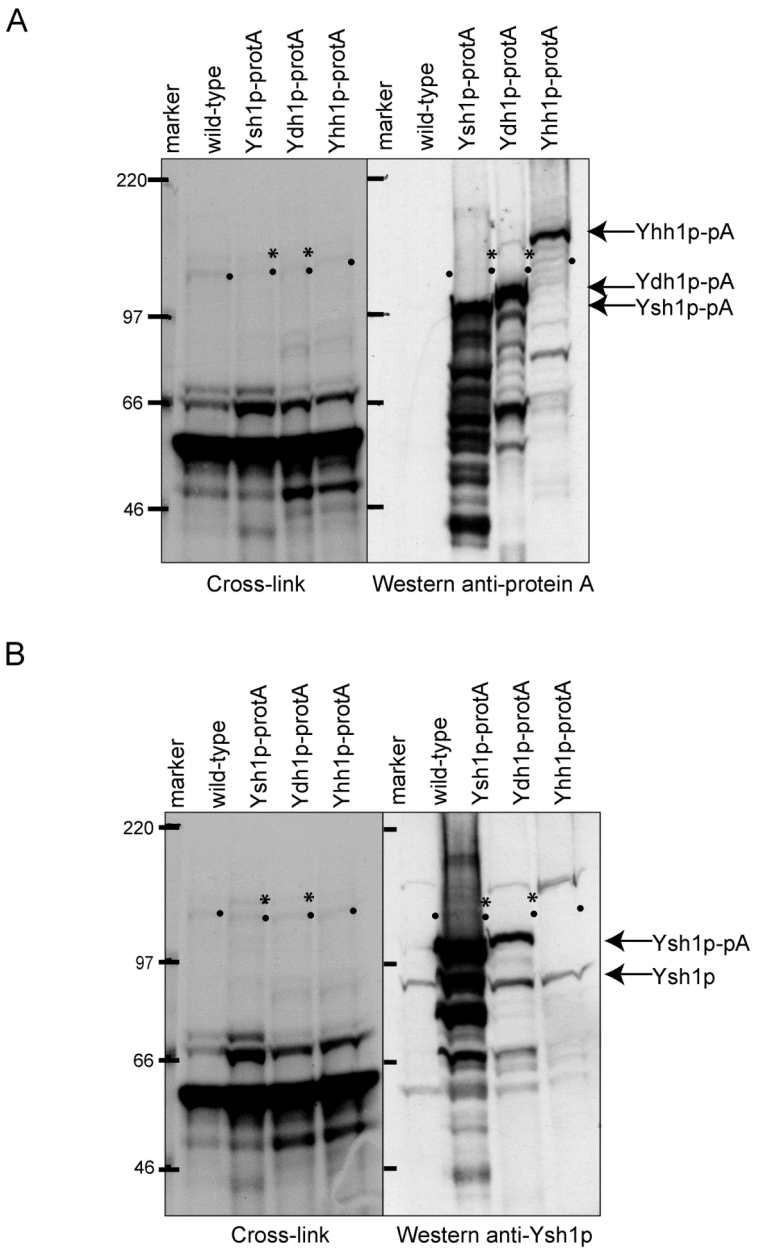


Figure 2.29 The ~ 100 kDa cross-linking band does not correspond to Ysh1p

Cross-linking experiments with SL2 RNA and extracts prepared from wild-type, Ysh1p-ProteinA, Ydh1p-ProteinA and Yhh1p-ProteinA cells. After the cross-linking proteins were separated on the SDS-PAGE and transferred to the Protran nitrocellulose membrane. The membrane was exposed O/N, followed by Western blot detection of the proteins of interest by the use of the anti-ProteinA (**A**) or the anti-Ysh1p (**B**) antibodies. Results (the autoradiogram and the chemiluminescent film) were juxtaposed together and positions of the cross-link bands were marked with an asterisk (*) (upper band) or with a dot (•) (lower band). [C^{14}] protein marker sizes are shown.

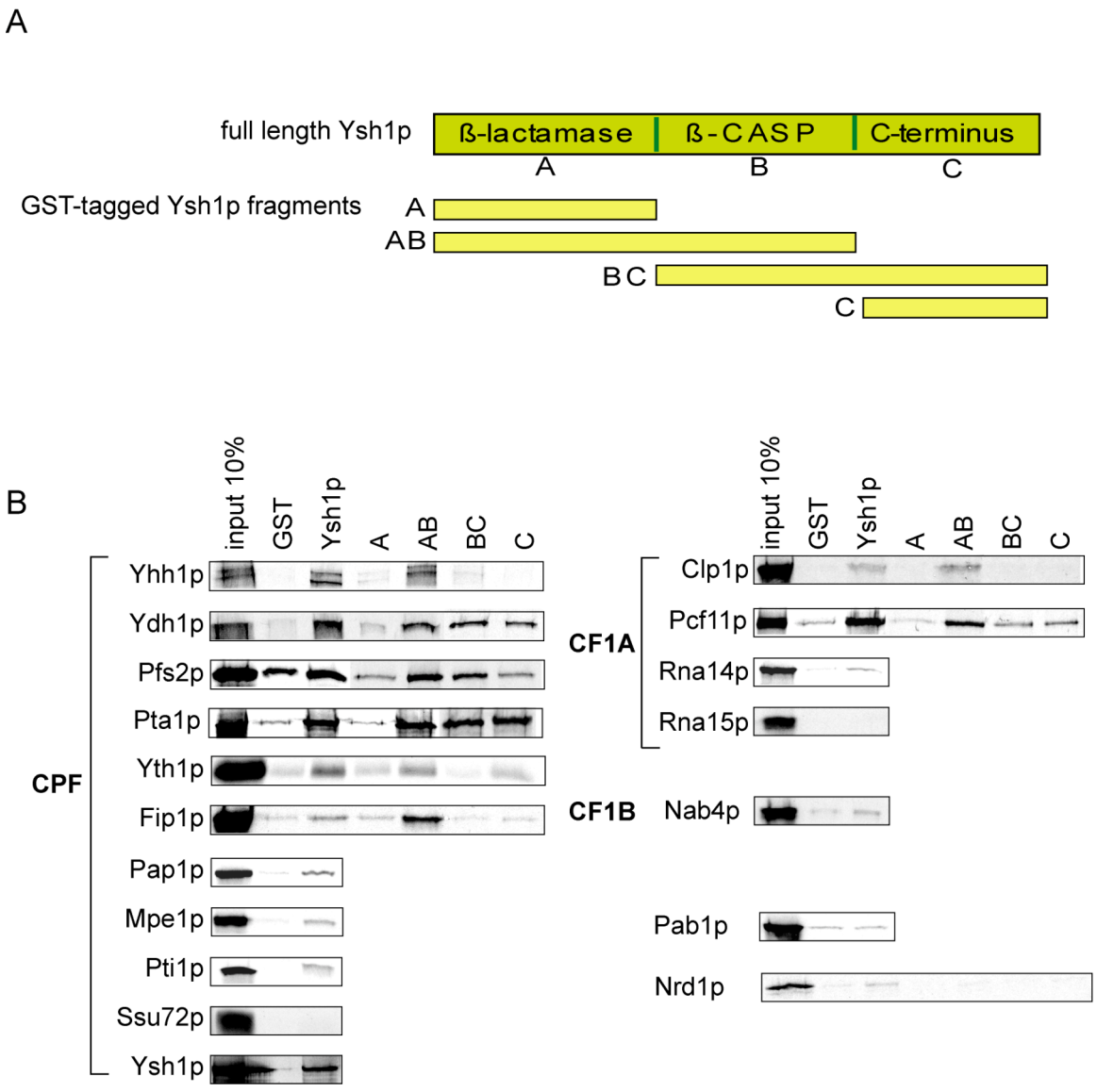


Figure 2.30 Ysh1p interacts with subunits of CPF and CF IA

(**A**) Schematic representation of the full-length and truncated fragments of Ysh1p. (**B**) GST pull-down experiments with 300 ng GST, 150 ng GST-Ysh1 tagged Ysh1p protein or 150 ng of the different truncated GST- tagged Ysh1p fragments (A, B, C) and *in vitro* translated [^{35}S]methionine-labelled proteins (indicated on the left). The first lane shows 10% of the *in vitro* translated protein used in the binding reactions.

Tethering Ysh1p to RNA substrates

Protein sequence analysis of Ysh1p detected a conserved β-lactamase domain within its N-terminus part; however, none of the known RNA binding motif was found. Recombinant N-terminally GST-tagged Ysh1p did not bind to RNA in the *in vitro* assays; neither was any endonucleolytic activity of this protein detected (B. Dichtl, personal communication). Potentially, Ysh1p might be an endonuclease that is not able to interact with the substrate in the absence of RNA binding factors (many of which are constituted by other CPF subunits) and therefore no endonucleolytic activity would be observed.

To address this possibility, we fused Ysh1p to various RNA binding domains. Two of the domains that we used bind RNA specifically (MS2, lambdaN) and one of them unspecifically (KH domain). The use of a non-specific RNA binding domain might increase the probability of success in case when the desired activity can be observed only in the context of specific RNA structures or sequence elements. The MS2 coat protein of the RNA bacteriophage MS2 binds a specific and well-characterized RNA hairpin sequence (Witherell et al., 1991). It is a rather large 22 kDa protein that forms a dimer in order to bind RNA; it may therefore interfere with the Ysh1p activity which may be undesired for our analyses. The alternative tethering system we tested was the lambdaN peptide, the 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N. The specific 19 nt binding site (boxB) of the lambdaN peptide is inserted into the target RNA, thus recruiting the properties of the fusion protein to RNA. The major advantage of this system derives from the small size of the peptide and its target sequence, which diminishes possible interferences with the fused protein (Baron-Benhamou et al., 2004). The KH-RNA binding domain used here was derived from the *MJ1236* gene of *Methanococcus jannaschii*. The properties of the KH-Ysh1p fusion are described to a larger extend in Chapter 4.2.

The hybrid constructs of Ysh1p were recombinantly expressed in *E. coli* and purified. *In vitro* assays were then performed to test whether binding of Ysh1p to the RNA substrate would uncover nuclease activity. Recombinant Ysh1-MS2 protein was unable to bind RNA, perhaps due to structural hindrance (data not shown). The recombinant LaN-Ysh1p and KH-Ysh1p proteins were found to bind to RNA substrate *in vitro*; however, no endonucleolytic activity was observed either (data not shown). In summary, the recombinant Ysh1p tethered to RNA via various RNA binding domains remained endonucleolytically inactive. We conclude that Ysh1p most likely requires other protein cofactors in order to be catalytically active.

A subset of the *ysh1* temperature sensitive mutants is deficient in 3' end processing *in vitro*

To analyze functions of Ysh1p in pre-mRNA 3' end formation, we created a battery of temperature-sensitive (ts) and cold-sensitive (cs) alleles of *YSH1*. To verify phenotypes of those *ysh1* temperature sensitive mutants in 3' end formation, total cell extracts derived from them were tested for cleavage and polyadenylation activities *in vitro*, analogically as described in Figure 2.4. We found several of them to be deficient in 3' end formation. Extracts derived from the ts strains

ysh1-7, *ysh1-13, ysh1-15, ysh1-32* and from and *ysh1-48* were impaired in the 3' end cleavage at the non-permissive temperature (Figure 2.31). Polyadenylation was inhibited in the extracts from the ts mutant strains *ysh1-3, ysh1-7, ysh1-9, ysh1-13, ysh1-15, ysh1-17, ysh1-32* and *ysh1-41,* and from the cs strain *ysh1-12* at the non-permissive temperature (Figure 2.32). Interestingly, the *ysh1-3*, *ysh1-9* and *ysh1-41* mutants were defective only in polyadenylation, but not in cleavage *in vitro*. The above results confirm the requirement for Ysh1p in 3' end cleavage and polyadenylation reactions *in vitro*. Different *ysh1* mutant alleles correspond to various phenotypes being observed (see also: 2.3 Discussion).

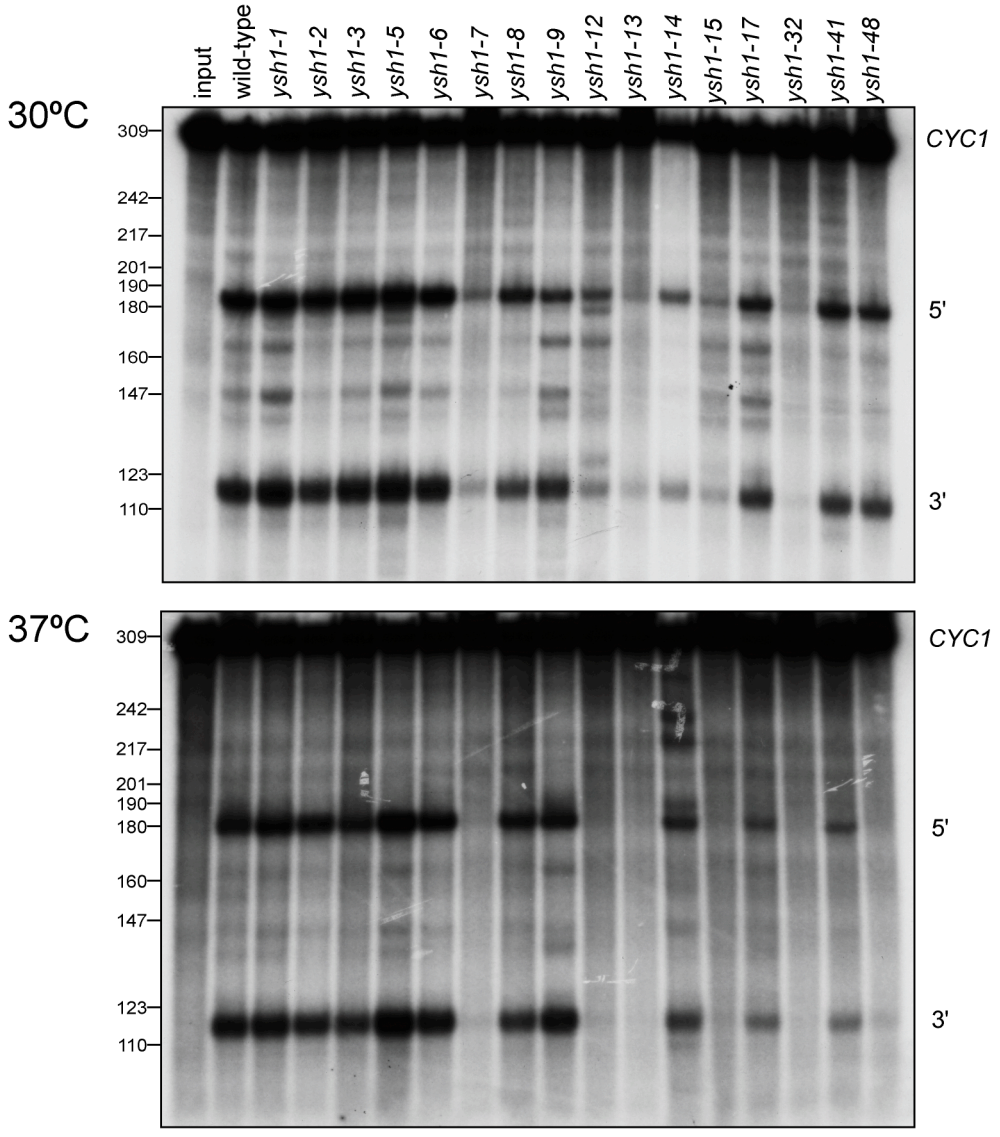


Figure 2.31 *ysh1* mutant strains are defective in pre-mRNA 3' end cleavage *in vitro*

In vitro cleavage assay with extracts prepared from *ysh1* temperature sensitive strains as indicated. Input indicates a mock-treated reaction. Positions of substrate RNAs, 5' and 3' end cleavage products, and polyadenylation products bands are indicated. *Hpa*II-digested pBR322 fragments were 5' end labelled and served as markers. Internally [^{32}P]-labelled RNA substrate CYC1 was used for cleavage. Reactions were performed either at 30ºC (**A**) or at 37ºC (**B**). Extracts were pre-incubated at the restrictive temperature for 10 min prior to assaying.

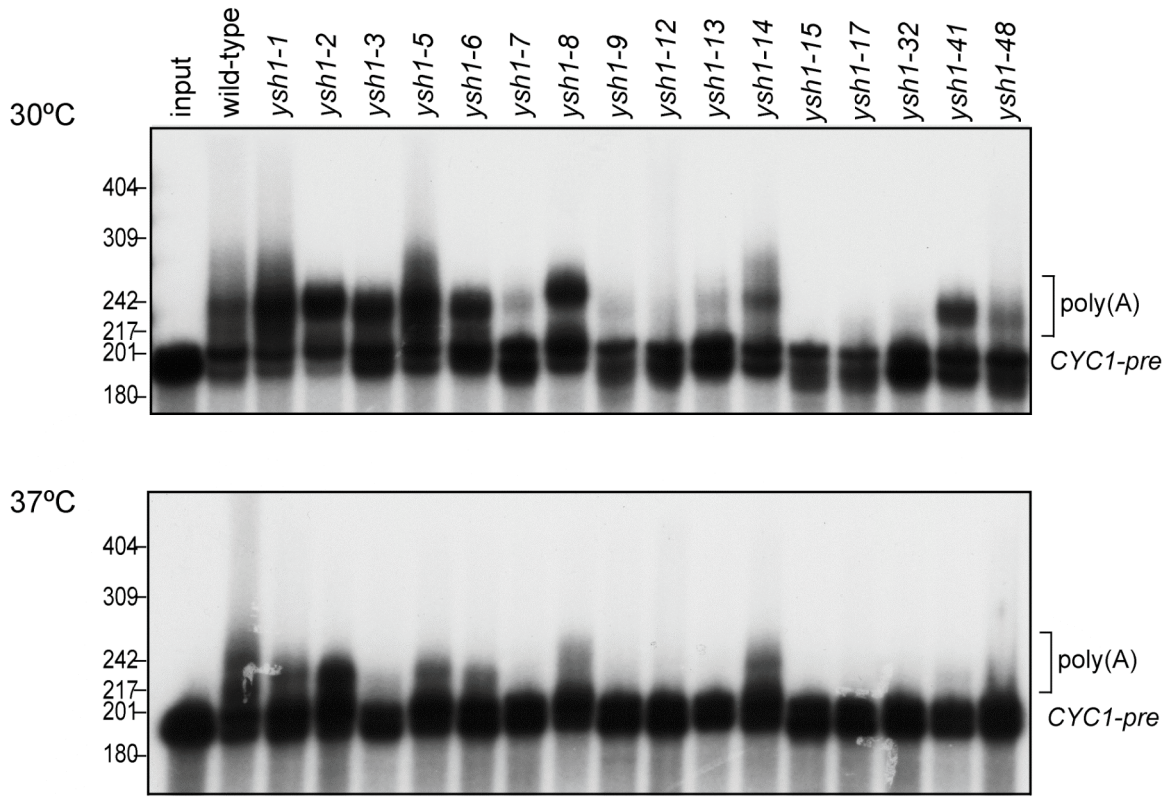


Figure 2.32 *ysh1* mutant strains are defective in pre-mRNA 3' end polyadenylation *in vitro*

In vitro polyadenylation assay with extracts prepared from *ysh1* temperature sensitive strains as indicated. Input indicates a mock-treated reaction. Positions of substrate RNAs, 5' and 3' end cleavage products, and polyadenylation products bands are indicated. *Hpa*II-digested pBR322 fragments were 5' end labelled and served as markers. Internally [^{32}P]-labelled CYC1-pre substrate RNA was used. Reactions were performed either at 30ºC (**A**) or at 37ºC (**B**). Extracts were pre-incubated at the restrictive temperature for 10 min prior to assaying.

2.6 Additional experimental procedures

Additional plasmids and oligonucleotides used in this study

Constructs conferring the following fragments of the *YSH1* gene: A, base pairs 1.-900. (amino acids 1.-300.); AB, base pairs 1.-1360. (amino acids 1.-450.); BC, base pairs 901.-2340. (amino acids 300.-780.); and C, base pairs 1350.-2340. (amino acids 450.-780.), were PCR-amplified using the respective primers (see below) and ligated to the pDrive vector (Qiagen), followed by religation to the GST-expression vector pGDV1. This resulted in the following constructs: pMG51 (GST-Ysh1-A), pMG52 (GST-Ysh1-AB), pMG53 (GST-Ysh1-BC) and pMG54 (GST-Ysh1-BC), which were subsequently used for the recombinant expression. The pNOPPATA-Ysh1 plasmid (ProteinA-tagged Ysh1p) was constructed as following: the NdeI-YSH1-BamHI fragment of the pBD73 construct was subcloned into the NdeI-BamHI sites of pMO111 (Ohnacker et al., 2000), resulting in pMG50. All the constructs were confirmed by DNA sequencing prior to further use.

RNA oligonucleotides, including the 4-thioUridine-modified ones, that were used as substrates for the splinted ligation, were synthesized commercially (Dharmacon). Their sequences were as following:

SL1 RNA oligo1, UUUUAUAGUUAUGUUAGUAUUAAGAACGUUAUUUAUAUUU;

SL2 RNA oligo1, UUUAUAGUUAUGUUAGUAUUAAGAACGUUAUUUAUAUUUCAA;

SL3 RNA oligo1, UUUUAUAGUUAUGUUAGUAUUAAGAACGUUAUUUAUAUUUC;

SL1 RNA 4-thioU oligo2, 4thio-UAAAUUUUUCUUUUUUUUCUGUACAG;

SL2 RNA 4-thioU oligo2, 4thio-UUUUUUCUUUUUUUUCUGUACAG;

SL3 RNA 4-thioU oligo2, 4thio-UAAUUUUUCUUUUUUUUCUGUACAG.

Sequences of the DNA oligonucleotides that served as 'bridges' were as following:

SL1 bridge, CAGAAAAAAAAGAAAAATTTAAAATATAAATAACGTTC;

SL2 bridge, CTGTACAGAAAAAAAAGAAAAAATTGAAATATAAATAACGTTCTTAATAC;

SL3 bridge, CTGTACAGAAAAAAAAGAAAAATTAGAAATATAAATAACGTTCTTAATAC.

DNA oligonucleotides used for the cloning of Ysh1p fragments were as following:

Ysh1-A 5'NdeI, GGAATTCCATATGGAGCGAACAAATACAACAAC;

Ysh1-A 3'BamHI, CCATCGCGGATCCTTAGGATTAGTCTGAGAGTCTCT;

Ysh1-B 5'NdeI, GGAATTCCATATGAGAGACTCTCAGACTAATCC;

Ysh1-B 3'BamHI, CCATCGCGGATCCTTAGGATTAAAAACATGGACTTC;

Ysh1-C 5'NdeI, GGAATTCCATATGGAAGTCCATGTTTTTAATCC;

Ysh1-C 3'BamHI, CCATCGCGGATCCTTAATAGCGGTGTGACCAAATTAC.

Yeast strains used in this study

The YMG88 strain (Ysh1p-NOPPATAA1, ProteinA-tagged Ysh1p) was constructed by transformation of the LM109 strain with pMG50 construct, followed by 5' FOA counterselection of the *YSH1-URA3* plasmid. The correct clones of YMG88 were confirmed by Western blotting.

Splinted ligation

General procedure of splinted ligation was adapted from Moore and Query, 2000; Yu, 1999. 120 pM of RNA 4-thioU oligonucleotide 2 was 5' end labelled with [^{32}P] using 50 µCi γ-[^{32}P] ATP and T4 PNK (New England Biolabs) in the presence of 2U RNA Guard (Amersham) for 1 h at 37ºC. 120 pmol RNA oligonucleotide 1, 120 pmol DNA bridge oligonucleotide and 1x T4 DNA ligase (Roche) buffer were added to the labeling reaction, followed by denaturation of the samples at 92ºC for 2 min. Samples were gradually cooled down for 10 min at 65ºC, 35 min at 37ºC and 20 min at 25ºC to allow hybridization. Hybridized oligonucleotides were then ligated together using 5-10 U T4 DNA ligase (Roche), 1 µM ATP and 1 mM DTT at 37ºC for up to 90 min. Ligated radiolabelled RNA was then precipitated and purified from a 8% denaturing polyacrylamide/8.3M urea gel and eluted for 1 h at 37ºC.

Specific 4-thioU cross linking

UV cross-linking reactions were routinely performed in microtiter plates and contained (in 20 µl) 50-200 fmol of [^{32}P]-4-thioU labelled transcript (SL1, SL2 or SL3) in 2% PEG-800, 75 mM potassium glutamate, 0,02% NP40, 20 mM Creatine P, 2 mM DTT, 2 mM EDTA, 250 ug/ml creatine kinase, 2U RNA Guard and 15 µl of the tested yeast cell extracts. Samples were incubated for 10 min at 32ºC in order to activate the cleavage reaction. The microtiter plate was then put on ice 5-8 cm from the UV bulb and samples were irradiated for 10 min at 365 nm wavelength using the Stratalinker UV lamp. After the irradiation, samples were treated with RNases A (2.5 µg/µl) and T1 (0.25 U/µl) for 30-60 min at 37ºC and electrophoresed using 8-10% SDS-PAGE gels. [C^{14}]-labelled proteins were used as a marker (Amersham). Gels were fixed and vaccum-dried. Cross-linked reaction products were visualized by autoradiography and an image was generated by a PhosphorImager (STORM).

Expression and purification of recombinant proteins

E. coli BL21(DE3) strain (Studier, 1991) was transformed with the respective plasmids pGDV1, pBD73, pMG51, pMG52, pMG53, or pMG54, carrying either GST, GST-Ysh1p or its truncated fragments GST-Ysh1-A, GST-Ysh1-AB, GST-Ysh1-BC and GST-Ysh1-C, and grown in 2xYT containing 100 µg/ml ampicilin and 50 µM $ZnCl_2$ at 37ºC to an OD_{600} of 0.6-0.8. The recombinant protein expression was induced by the addition of isopropyl-D-thiogalctoside (IPTG) to the final concentration of 0.5 mM. After 3 h of expression at 37ºC cells were harvested, lysed, and proteins were purified by affinity chromatography on glutathione Sepharose 4B (Pharmacia) according to the suppliers manual.

GST pull down experiments

Plasmid DNAs encoding the different 3' end processing subunits were *in vitro* transcribed and translated with the TNT coupled transcription-translation system (Amersham) in a total volume of 50 µl according to the manufacturer's instructions. Approximately 100 ng of the recombinant GST, GST-Ysh1p and its truncated fragments GST-Ysh1-A, GST-Ysh1-AB, GST-Ysh1-BC and

GST-Ysh1-C were incubated with 3 µl of the *in vitro* translated [^{35}S]-labelled proteins for 1 h at 23ºC and subsequently immobilized onto 100 µl glutathione Sepharose 4B (Pharmacia) in buffer P (phosphate-buffered saline (PBS) containing 0.01% NP40, 1 mM DTT, 0.4 µg/ml leupeptin, 0.7 µg/ml pepstatin, 0.5 µg/ml PMSF) by shaking for 1 h at 4ºC. The beads were pelleted and washed extensively 5 times with buffer P for 10 min at 4ºC. Proteins were eluted in 20 µl SDS-PAGE sample buffer and resolved by 8 % SDS-PAGE. Gels were fixed, treated with 1 M Na-salicylate, and vacuum-dried. The [^{35}S] proteins were visualized by autoradiography.

CHAPTER 3

Functional analysis of Syc1p, the homologue of the C-terminal domain of Ysh1p

3.1 Introduction

3.2 Results

Syc1p physically interacts with Pta1p and other components of the APT subcomplex *in vitro*

Syc1p is dispensable for pre-mRNA 3' end cleavage and polyadenylation *in vitro*

Syc1p is not required for 3' end processing of snoRNAs

Deletion of *SYC1* gene affects the viability of mutants of the APT complex subunits

Genetic interaction partners of *SYC1*
- -Yeast 2-hybrid screen with *SYC1* as a bait
- -Synthetic lethal screen with the *SYC1* gene

DNA microarray analysis of the *syc1Δ* profile

3.3 Discussion

3.4 Experimental procedures

3.1 Introduction

RNA polymerase II transcripts need to be extensively processed before they can be exported to the cytoplasm and function in translation. For the 3' end of pre-mRNAs, the processing involves endonucleolytic cleavage at the poly(A) site, followed by addition of the polyadenosine track. In *S. cerevisiae*, this process is mediated by the 3' end cleavage and polyadenylation apparatus, composed of a surprisingly high number of subunits. The exact function of several of them remains unclear. This machinery can be separated into three complexes: cleavage and polyadenylation factor (CPF), cleavage factor IA (CF IA), and cleavage factor IB (CF IB). Systematic tagging and purification of CPF subunits revealed its modular architecture (Nedea et al., 2003). Six polypeptides associated with the CPF form a distinct complex, APT, in which Pti1p, Swd2p, a type I protein phosphatase Glc7p, CTD-phosphatase Ssu72p, Ref2p and Syc1p are associated with the Pta1p subunit of CPF (Figure 3.1). APT subunits were shown to localise primarily near the polyadenylation signals of the constitutively expressed genes and on snoRNA genes. Various studies revealed that most of the APT subunits are important for preventing transcriptional read-through at terminators of independently transcribed snoRNA genes (Dheur et al., 2003; Dichtl et al., 2004; Ganem et al., 2003; Morlando et al., 2002; Nedea et al., 2003; Nedea et al., 2008; S. Roeck, personal communication).

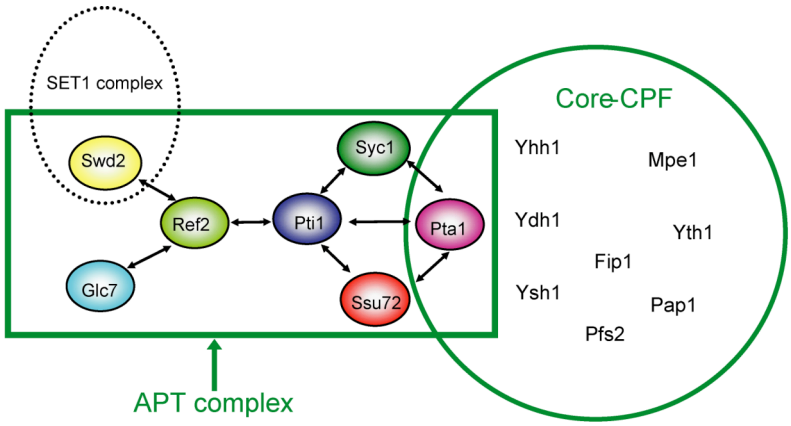


Figure 3.1 Architecture of the APT subcomplex within CPF

Double arrows indicate direct protein interactions. The green rectangle represents APT complex obtained from the Syc1-TAP purification. Pta1p is required for bridging core-CPF to APT and is a component of both complexes. Pti1p is required for joining the two submodules of APT. Swd2p is shown as a common component of the distinct APT/CPF and Set1p complexes. Ref2p is responsible for recruiting the majority of Swd2p to CPF (after Nedea et al., 2003; modified).

Initially, protein sequence analysis of the *S. cerevisiae* genome revealed an open reading frame (*YOR179c*) that was highly homologous to the C-terminus of Ysh1p, and therefore named *SYC1*, for similar to Ysh1p C-terminus (Nedea et al., 2003). It might have evolved as a gene duplication of *YSH1*. Syc1p has no apparent homologues in metazoans; neither does the C-terminus of Ysh1p. Searches of the protein sequence databases resulted in detection of several

Syc1p homologues only in other *Saccharomyces* species such as *S. bayanus*, *S. milatae*, *S. paradoxus*, and *S. kudriavzevii*, and additionally in the fungus *Candida glabrata*.

The level of conservation between the C-terminus of Ysh1p (amino acids 582 -780) and Syc1 is strikingly high, exhibiting 38% identity and 55 % similarity (Figure 3.2). Syc1p is a 21 kDa protein of 188 amino acids, non-essential for yeast viability. Syc1p carries a conserved P-loop motif commonly found in proteins binding ATP/GTP. Affinity purification of TAP-tagged Syc1p resulted a pull down of the whole APT sub-complex consisting of Pta1p, Pti1p, Ref2p, Glc7p, Ssu72p and Swd2p (Gavin et al., 2002; Nedea et al., 2003).

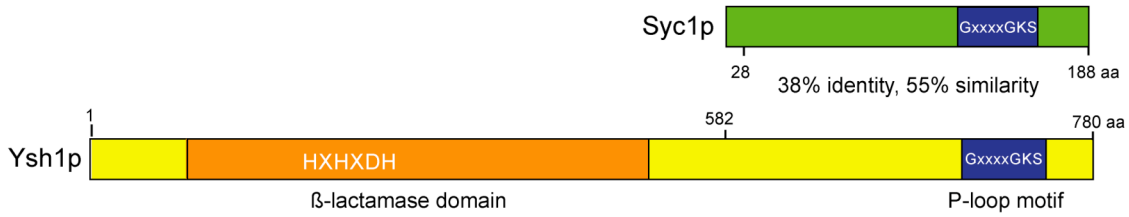


Figure 3.2 Syc1p is homologous to the C-terminal part of Ysh1p
Schematic representation of Syc1p protein structure in its relation to Ysh1p. Regions of similarity and conserved motifs (the β-lactamase and P-loop) are depicted.

Because of its sequence homology to Ysh1p, we analyzed possible functions of Syc1p within the CPF complex, and in a larger cellular context. In this work, we show that although Syc1p is not essential for pre-mRNA 3' end processing, it is an integral part of the cleavage and polyadenylation complex and its absence exacerbates growth defects of other 3' end processing mutants. Moreover, Syc1p forms a network of genetic interactions that possibly anchor the 3' end processing event into a wider cellular context.

3.2 Results

Syc1p physically interacts with Pta1p and other APT components

To unravel the function of Syc1p in the APT subcomplex, we searched for its interacting partners. We constructed the Syc1p N-terminal GST-fusion, expressed and purified it from *E.coli* (Figure 3.3 A). To check for specific protein-protein interactions, GST pull-down experiments were carried out with radioactively labelled *in vitro* translated APT subunits. We found that Syc1p interacted strongly with Pta1p and moderately with Pti1p, Glc7p, and Ssu72p (Figure 3.3 B). No interactions were observed between Syc1p and Ref2p, or Swd2p. The above results confirm stable association of Syc1p within the APT subcomplex.

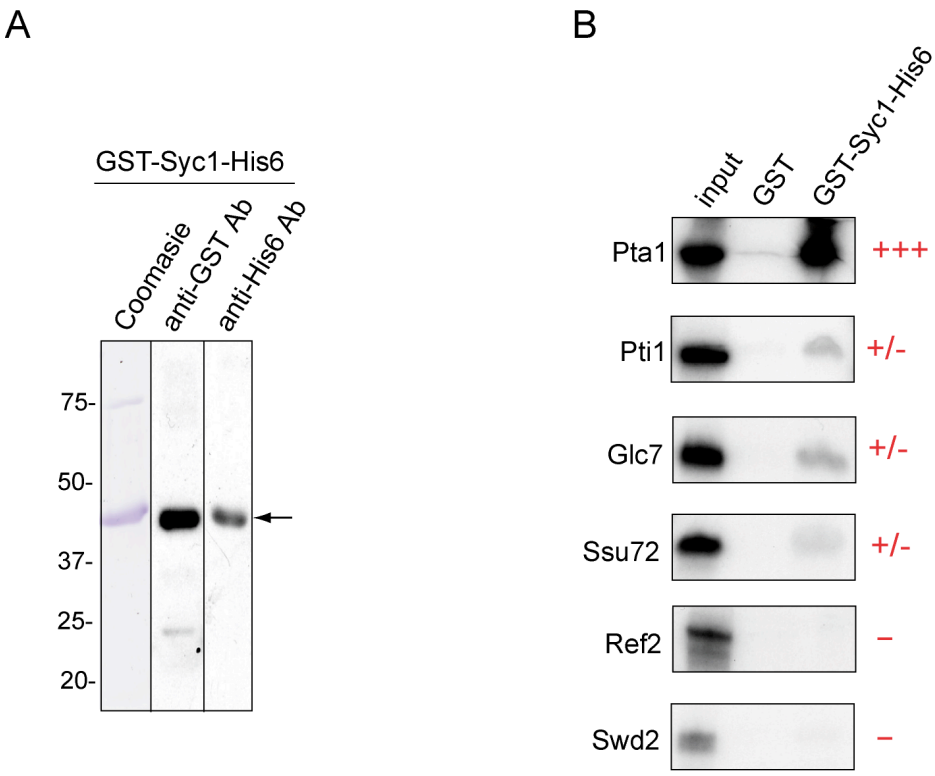


Figure 3.3 Syc1p interacts with components of the APT subcomplex *in vitro*

(**A**) Recombinant GST-Syc1-His6 protein (indicated with an arrow) was purified over the Glutathione Sepharose column and its identity was confirmed by Western blotting using anti-GST and anti-His6 antibodies. (**B**) GST pull-down experiments with either 0.5 µg GST or 100 ng GST-Syc1-His6 recombinant proteins and *in vitro* translated [^{35}S]methionine-labelled proteins (indicated on the left). The first line shows 10% of the *in vitro* translated protein used in the binding reactions. GST alone was used as a background control. Relative efficiency of binding is indicated in red on the right side of the panel (+++, strong interaction; +/-, moderate interaction; -, no interaction).

Syc1p is dispensable for pre-mRNA 3' end cleavage and polyadenylation *in vitro*

Next, we tested whether Syc1p is essential for 3' end cleavage and polyadenylation *in vitro*. Total yeast cell extracts were obtained from wild-type and *syc1* deletion strains, and assayed on synthetic *CYC1* RNA for cleavage, and *CYC1-pre* RNA for polyadenylation. *Syc1Δ* mutant strain was efficiently cleaving and polyadenylating RNA substrates compared to wild-type (Figure 3.4). The *in vitro* polyadenylation pattern observed for the extracts derived from both *syc1Δ* and its isogenic wild-type strain (BY4741) was repeatedly unspecific, when compared to W303-derived strains (Figure 3.4, right panel; and data not shown). We conclude that Syc1p is not directly required for the pre-mRNA 3' end cleavage and polyadenylation reactions.

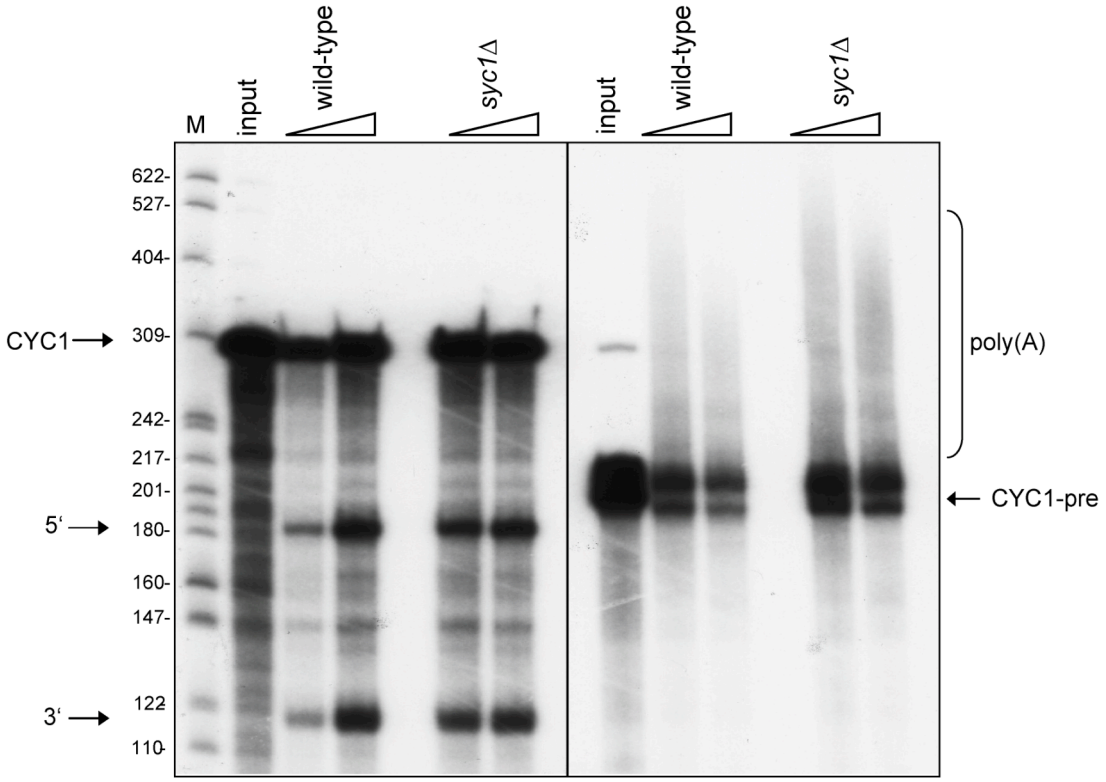


Figure 3.4 Syc1p is dispensable for pre-mRNA 3' end cleavage and polyadenylation *in vitro*

In vitro cleavage (left panel) and polyadenylation (right panel) assays with extracts prepared from the isogenic wild-type (BY4741) and *syc1Δ* strains as indicated. Input indicates a mock-treated reaction. Positions of substrate RNAs, 5' and 3' end cleavage products, polyadenylation products, and of marker bands are indicated. Internally ^{32}P-labelled substrate RNAs, CYC1 for cleavage assay and CYC1-pre for polyadenylation, were used. Reactions were performed at 30ºC.

Syc1p is not required for 3' end processing of snoRNAs

Several APT subunits (Ssu72p, Swd2p, Pti1p, Ref2p, Glc7p) were shown to be required for the correct processing of snoRNA transcripts at their 3' ends (Dheur et al., 2003; Dichtl et al., 2004; Ganem et al., 2003; Morlando et al., 2002; Nedea et al., 2003; Nedea et al., 2008; S. Roeck, personal communication). We hypothesized that Syc1p, by association with APT subcomplex, could have a similar function in that process.

To corroborate this hypothesis, we performed Northern blot analyses with total RNAs extracted from wild-type and *syc1Δ* strains. Several snoRNA transcripts were analyzed, such as *SNR13, SNR33, SNR45, SNR71,* and *SNR128*, with specific radioactively labelled oligonucleotide probes. All of the tested snoRNA transcripts were terminated and processed correctly in the *syc1Δ* strain compared to wild-type (data not shown). To further confirm these observations, the analysis was repeated for the respective mRNAs encoded downstream of *SNR13* and *SNR33* snoRNA loci (Figure 3.5).

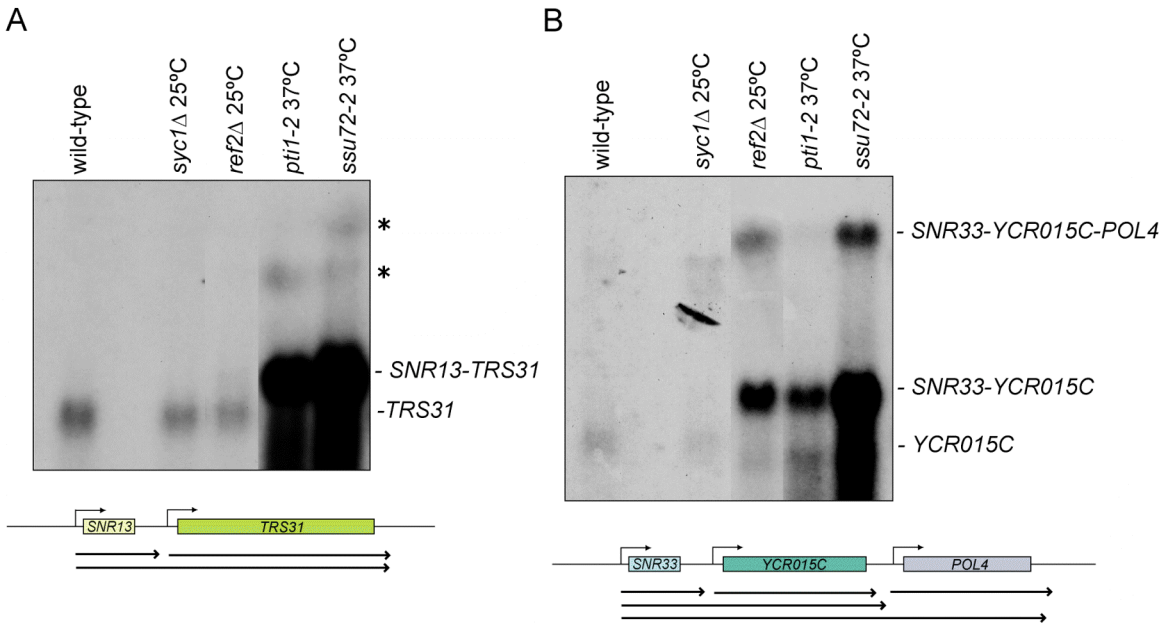


Figure 3.5 Syc1p is not required for snoRNA 3' end formation

Northern blot analysis of total RNA extracted from wild-type and mutant *syc1Δ, ref2Δ, ssu72-2* and *pti1-2* cells. *ssu72-2* and *pti1-2* strains were grown at 23ºC and shifted to restrictive temperature (37ºC) for 2 h. *syc1Δ* strain does not exhibit sensitivity to elevated temperatures, therefore *syc1Δ* total RNA was extracted from cells grown at 25ºC only. RNAs were separated on formaldehyde/1.2% agarose gels. Filters were developed with random-primed labelled probe directed against *TRS31* (**A**) or *YCR015c* (**B**) mRNA, respectively. Migration of the correct-size and read-through-size RNAs is indicated. Additional read-through products detected in *ssu72-2* and *pti1-2* strains are indicated with asterisks (*). Schemes below the panels represent genomic arrangements of the analysed genes and their direction of transcription.

No read-through transcripts originating from these snoRNAs were detected in the *syc1*Δ strain, either. Mutants in APT subunits known to be required for snoRNA 3' end formation (*ref2Δ, pti1-2* and *ssu72-2*) were included as positive controls; they revealed defects in formation of several snoRNAs as previously described (Dheur et al., 2003; Dichtl et al., 2002a). From the above results we conclude that Syc1p is not required for 3' end formation/transcription termination of snoRNAs.

Deletion of *SYC1* gene affects the viability of other APT mutant strains

We showed that Syc1p is not directly required for pre-mRNA and snoRNA 3' end formation. However, Syc1p sequence displays a significant homology to the C-terminus of Ysh1p, suggesting that Syc1p could have a regulatory function in 3' end formation. Therefore, we generated double mutant strains of *syc1Δ* and other APT subunits and analyzed them for alterations in growth behavior. Additionally, double mutants were treated with cordycepin (3' deoxyadenosine), which interferes with poly(A) tail production but does not significantly affect the general RNA synthesis within the concentrations used (20-40 µg/ml). Deletion of *SYC1* gene alone did not cause any growth defect at 25°C, 30°C, or 37°C (Figure 3.6). Interestingly, deletion of *SYC1* suppressed the ts lethality of *glc7-12* mutant at 33º and 37°C. Similarly, a strong suppression of growth defect was observed in *ref2Δ syc1Δ* double mutant strain. The suppression caused by *SYC1* deletion in both *glc7-12* and *ref2Δ* alleles was further exacerbated on cordycepin-containing medium. Notably, *syc1Δ* aggravated the growth defect of the *pti1-2* mutant at 33ºC. This effect was further enhanced when *pti1-2 syc1Δ* was grown on cordycepin-containing medium. In contrast, deletion of *SYC1* gene had no effect on *pta1-1* and *ssu72-2* mutant strains. Taken together, these results indicate that Syc1p genetically interacts with Pti1p, Ref2p and Glc7p components of APT, and that it could be indirectly involved in poly(A) tail synthesis.

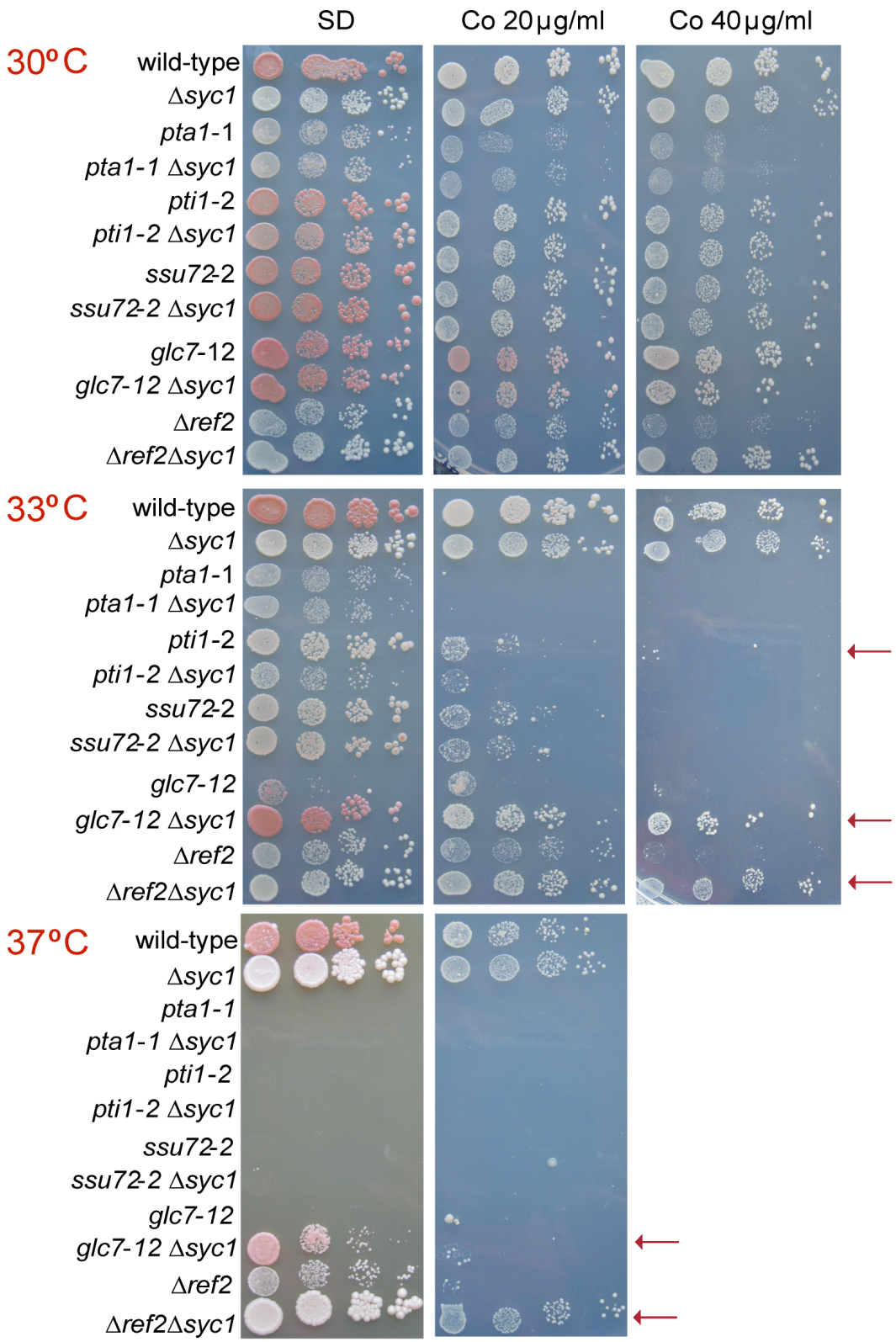


Figure 3.6 Deletion of *SYC1* gene affects the viability of other APT mutant strains

Ten fold serial dilutions of wild-type and mutant strains were spotted onto synthetic complete medium alone or with addition of 20 µg/ml or 40 µg/ml cordycepin (Co), and incubated at the respective temperatures for 3 days. Noteworthy phenotypes are marked with red arrows on the right; see text for explanations.

Genetic interaction partners of *SYC1*

Genetic screens take the advantage of protein properties and interactions to identify new interacting factors or factors that have redundant functions. To discover potential interaction partners of Syc1p within the cell, two distinct genetic screens were performed. The first screen, yeast 2-hybrid analysis, was performed in collaboration with Hybrigenics, Paris, France. The second, a novel type of the synthetic lethal screen, was performed in collaboration with A. Jacquier and L. Decourty, Institut Pasteur, Paris, France. The initial objective of both of the screens was to gain insights into the function of Syc1p in pre-mRNA 3' end processing, identification of putative new components of this pathway, determination of interactions between these components, and uncovering new links to other RNA processing pathways.

Yeast 2-hybrid screen with *SYC1* as the bait

2-hybrid screening technique is used to discover protein-protein interactions by testing for physical interactions (such as binding) between two proteins (reviewed in: Young, 1998). In our screen, with *SYC1* used as the bait, 123 clones were recovered and 7 genes identified (summarized in Table 3.1). 95 clones (ca. 77%) represented regions of the *PTA1* gene. Another commonly fished prey in the yeast 2-hybrid analysis encoded the gene *PTI1*, another component of APT, involved in pre-mRNA and snoRNA 3' end formation. This was consistent with our *in vitro* protein-protein interaction analysis, which showed a strong interaction between Syc1p and Pta1p as well as Pti1p, respectively (GST pull-down experiments, Figure 3.3). As both factors belong to APT, it is possible that at least one of them recruits Syc1p to the complex.

Furtermore, we identified five genes encoding for new interacting partners of Syc1p. The *YNL078W* gene was originally identified by the yeast 2-hybrid screening method with several septin genes as baits (Iwase and Toh-e, 2001), and therefore named *NIS1* (Neck protein Interacting with Septins). Septins are a conserved family of proteins forming a ring structure of the mother bud-neck; they are required for cytokinesis in numerous species, from fungi to human. Neither disruption nor overexpression of *NIS1* caused a prominent phenotypic change. Nis1p is phosphorylated throughout the cell cycle and is less abundant in G2/M phase (Iwase and Toh-e, 2001). GFP-Nis1p localizes in the nucleus throughout the cell cycle and to the bud neck at G2/M phase in a septin-dependent manner. Additionally, a number of proteins interacting with Nis1p are members of the mitotic signaling network, suggesting that Nis1p plays a non-essential role in this network (Iwase and Toh-e, 2001).

The second gene identified by our screen, *TDP1* (*YBR223C*), encodes for a tyrosine-DNA phosphodiesterase that hydrolyzes 3' phosphotyrosyl bonds to generate 3' phosphate DNA and tyrosine *in vitro*. Tdp1p is involved in double strand break repair of DNA lesions created by topoisomerase I, both in yeast and mammalian cells. Topoisomerase I regulates the degree of DNA supercoiling by causing single-strand breaks and re-ligation. The activity of this enzyme is especially crucial during DNA transcription and replication, when the DNA helix must be unwound to allow proper functioning of the transcriptosom (Wang, 1991). Mutations in human Tdp1 have been linked to patients with spinocerebellar ataxia, and over-expression of Tdp1 results in resistance to known anti-cancer compounds (Raymond and Burgin, 2006; Raymond et al., 2005).

Table 3.1 Results of yeast 2-hybrid screen with *SYC1* as the bait gene

Bait gene	Prey gene	Clones	PBS	Gene function
SYC1 (YOR179c)	*PTA1 (YAL043C)*	95	A	pre-mRNA 3' end cleavage and polyadenylation
	NIS1 (YNL078W)	11	A	protein localized in the bud neck at G2/M phase; possibly involved in mitotic signalling network
	PTI1 (YGR156W)	8	A	pre-mRNA 3' end cleavage and polyadenylation, snoRNA 3' end formation
	TDP1 (YBR223C)	4	A	tyrosyl-DNA phosphodiesterase I, involved in the repair of DNA lesions created by topoisomerase I
	PRP16 (YKR086W)	1	D	RNA helicase of the DEAH-box family, involved in the second catalytic step of splicing
	WSS1 (YHR134W)	1	D	sumoylated protein of unknown function, interacts genetically with *SMT3*; possible involvement in DNA damage response
	FIR1 (YER032W)	2	E	pre-mRNA 3' end processing - a positive regulator of poly(A) synthesis; interacts with Ref2p

Legend:

Clones - number of clones detected;

PBS - Predicted Biological Score. A statistical confidence score that is assigned to each protein interaction; it ranks interacting proteins according to technical parameters and identifies sticky proteins and other false positives (Rain et al., 2001). The PBS score reflects the probability that an interaction is found by chance. It is grouped in five categories (A, B, C, D, and E):

A, B, C - technically very reliable; they correspond to interactions found in two reciprocal and independent screens or interactions found in a single screen with many overlapping prey fragments;

D - defined by a single bait-SID (Selected Interaction Domain) interaction, the SID being defined by a singleton fragment instead of a family of several overlapping fragments. It can include false-positive interactions (background noise), as well as interactions hardly detectable by 2-hybrid systems because of conformation, toxicity in yeast, very low representation of the mRNA, etc.;

E - these interactions involve SID that have been found non-specifically as prey in many independent screens; they are likely to be false-positives of the 2-hybrid system.

Three other interaction candidates found in the 2-hybrid screen were *PRP16, WSS1* and *FIR1.* They were found with a rather low number of clones and categorized as possibly non-specific, false-positive interactions (see: Table 3.1).

PRP16 (YKR086W) gene encodes for an essential RNA helicase of the DEAH-box family involved in the second catalytic step of splicing. Prp16p exhibits an ATP-dependent RNA unwinding activity that disrupts a duplexed RNA structure on the spliceosome (Wang et al., 1998).

WSS1 (YHR134W) gene was found by its genetic interaction with *SMT3*, gene encoding for the yeast SUMO-tag protein. *WSS1* (Weak Suppressor of *SMT3*) encodes for a sumoylated protein of unknown function. UV-sensitive mutant phenotype and genetic interactions suggest a role in the DNA damage response, processing stalled or collapsed replication forks (O'Neill et al., 2004).

FIR1 (YER032W) gene was originally identified in a yeast 2-hybrid screen using *REF2* as bait (Russnak et al., 1996). *FIR1* (Factor Interacting with *REF2*) interacts with two independent regions of *REF2*. Fir1p and Ref2p are required for the formation of a normal-length poly(A) tail on pre-cleaved CYC1 pre-mRNA *in vitro* (Mangus et al., 2004). Kinetic analyses of polyadenylation reactions indicated that Fir1p and Ref2p act as a positive-, and a negative regulator of poly(A) synthesis, respectively (Mangus et al., 2004; Russnak et al., 1996). In addition, Fir1p was found to interact genetically with poly(A) polymerase (Pap1p) and poly(A) binding protein (Pab1p) in yeast 2-hybrid screen (del Olmo et al., 1997). Nonetheless, in many published genetic screens *FIR1* was found to interact with several other factors, including *UTP15*, *HAL7, SPO12, CLB5* and *SPC24,* suggesting that the 2-hybrid interactions of Fir1p may be partially unspecific.

Synthetic lethal screen with *SYC1* gene

Yeast synthetic lethal screens are generally used to search for combinations of gene deletions or mutations that lead to cell death, but that individually have little or no effect on the viability of the organism. In this case, the term synthetic lethal is used to denote that the lethal event derives from the synthesis (*i.e.*, combining together) of two different deletions, rather than from a single deletion. Such synthetic lethal screens are useful to identify genetic interactions and to map networks and pathways regulating cellular processes.

Here, we have used a new methodology that allowed us to perform the synthetic lethal screen with higher efficiency and higher sensitivity than the previously reported methods. This strategy is designed for screening of the non-lethal deletions present within the systematic yeast deletion library. Combining the colony-colour assay with a synthetic lethal screen offers means to visually detect a mutant that depends on a plasmid for survival. Principles of this screen are depicted in Figure 3.7. Briefly, a reporter plasmid containing the colour marker *MET15* gene, a tet-regulated centromer and the gene of interest was transformed into the *syc1Δ*::NatMX Matα deletion strain. Next, the above strain was crossed with a pool of deletion library, diploids were selected on hygromycin- and kanamycin- containing medium, and subsequently sporulated. Haploid strains obtained in this way were then grown on led-containing medium. Any cell with a mutation that causes dependence on the gene of interest for survival must maintain the plasmid.

Cells that lost the reporter plasmid acquire black/brown colour, while cells that retain the plasmid produce milky white colonies. Mutants maintaining the reporter plasmid are dependent on the gene (*SYC1*) carried on the plasmid and they are therefore synthetic lethal with it. The synthetically lethal gene can then be identified by isolating and sequencing the unique DNA tag attached to the deletion cassette.

The screen we performed with *SYC1* gene identified only two new candidate genes that showed a slow synthetic growth when their deletion was combined with the *SYC1* deletion: *ARP8* and *SSN2* (Figure 3.8). No synthetic lethal combinations were observed. *ARP8* (*YOR141C*) gene encodes for Arp8p (Actin-Related Protein 8), a component of the yeast INO80 ATP-dependent chromatin-remodelling complex. GST-Arp8 binds preferentially to histones H3 and H4 *in vitro* suggesting a histone chaperone function (Shen et al., 2003).

A recent report by Kawashima et al., 2007, has shown that Arp8p is involved in damage-induced sister chromatid recombination and interchromosomal recombination between heteroalleles.

The second tag identified in the screen was initially *YDR442W,* a dubious open reading frame, which is unlikely to encode a functional protein, as based on available experimental and comparative sequence data. Nevertheless, analyzing its gene locus, we found that the *YDR443C* gene was positioned in a converse orientation. We predicted that systematic genomic deletion of *YDR442W* caused also the deletion of 360 base pair fragment of *YDR443C*. Tests for genetic interactions between *YDR443C* and *SYC1* revealed the synthetic slow growth defect. *YDR443C (SSN2)* encodes for Ssn2 protein (also known as Srb9p/Med13p) that belongs to the conserved Srb module of the RNAP II Mediator complex. Ssn2p is required for stable association of Srb10p-Srb11p kinase with RNAP II holoenzyme. Ssn2p is essential for transcriptional regulation, and *ssn2* mutations can suppress CTD truncations or phosphorylation mutants. The Srb complex, like many components of the RNAP II machinery, is responsible for regulating the expression of a relatively large number of genes. Genetic and biochemical data indicate that Ssn2p is a substrate for PKA and that its phosphorylation modulates the activity of the Srb complex (Chang et al., 2004).

Growth alterations resulting of double gene deletions in the *syc1Δ ssn2Δ* and *syc1Δ arp8Δ* strains are presented in Figure 3.8. Differences in growth rate of *syc1Δ YDR442WΔ* and *syc1Δ arp8Δ* haploid deletion strains were monitored following transformation with either empty vector or plasmid bearing the *SYC1* gene. Growth rate of both of those strains was slightly impaired when compared to the growth of those strains complemented with plasmid-borne *SYC1*. Similar situation was observed when synthetic slow growth rate of *syc1Δ YDR442WΔ, syc1Δ ssn2Δ,* and *syc1Δ arp8Δ* double deletion strains was compared to the growth of *syc1Δ, ssn2Δ* and *arp8Δ* mutants combined with the test deletion of *EDC3* gene (used here as a negative control; Badis et al., 2004). While the growth rate of *syc1Δ ssn2Δ* strain was visibly impaired, the growth of *syc1Δ arp8Δ* was comparable to the *syc1Δ edc3Δ* control strain.

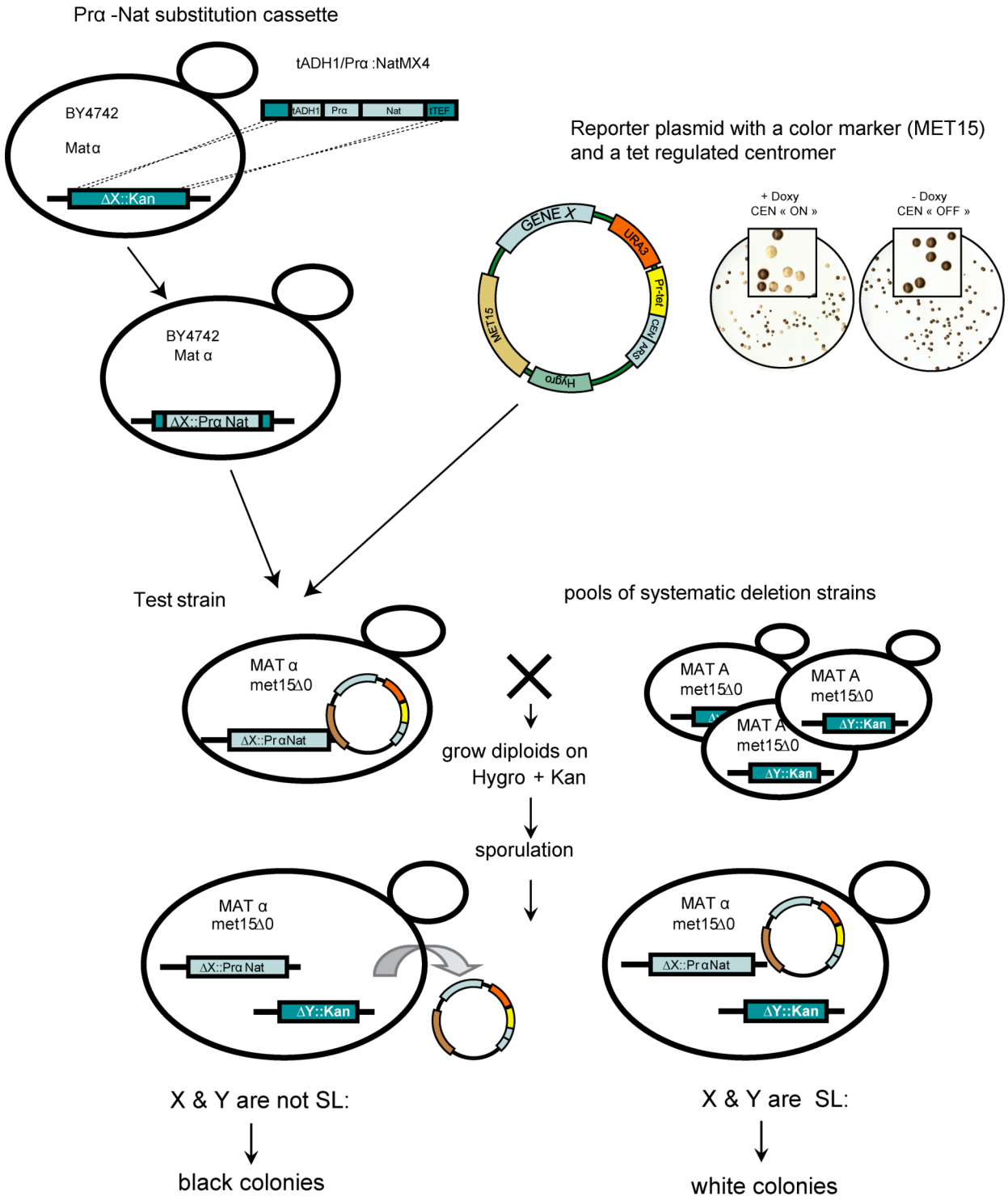


Figure 3.7 Principles of the synthetic lethal screen used in this study

See text for explanations.

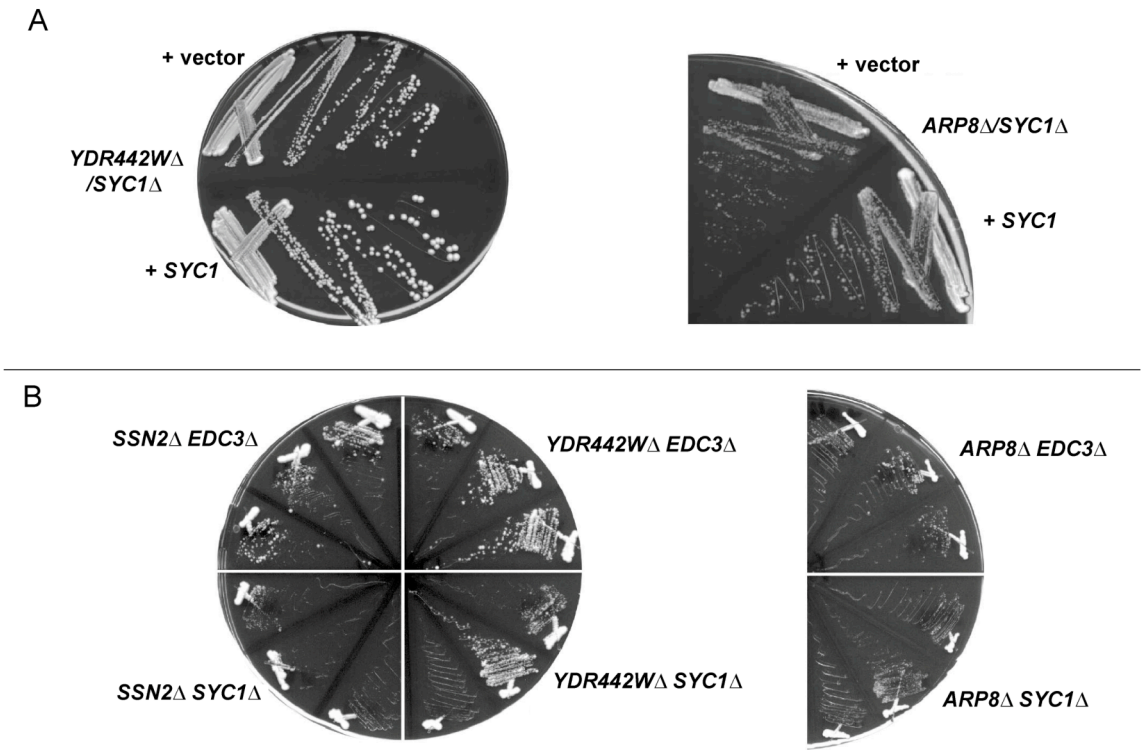


Figure 3.8 *SYC1* genetically interacts with *SSN2* and *ARP8*

(**A**) Synthetic slow growth phenotypes of *syc1Δ YDR442WΔ* and *syc1Δ arp8Δ* deletion strains. Haploid strains containing the *syc1Δ* and *YDR442WΔ* (left) or the *syc1Δ* and *arp8Δ* (right) alleles were transformed with either vector or plasmid bearing the *SYC1* gene, and incubated at room temperature for 3 days.

(**B**) Synthetic slow growth rate of *syc1Δ YDR442WΔ, syc1Δ ssn2Δ* and *syc1Δ arp8Δ* double deletion strains was compared to the growth rate of *syc1Δ, ssn2Δ* and *arp8Δ* mutants combined with *EDC3* gene deletion, for control. Three independent clones were tested for each of the deletion combinations.

Results of both genetic screens with *SYC1* gene are summarized graphically in Figure 3.9. Genetic interaction network of *SYC1* includes links to several cellular pathways. The main link points towards the 3' end cleavage and polyadenylation apparatus, as *SYC1* was found to interact both physically and genetically with *PTA1*, *PTI1* and *REF2* components of APT subcomplex within CPF. *SYC1* may also interact with *FIR1*. Additional links point towards chromatin remodeling (*ARP8*), RNAP II transcription initiation (*SSN2*), mitotic signaling network (*NIS1*), DNA repair and damage response (*TDP1, WSS1*), and splicing (*PRP16*).

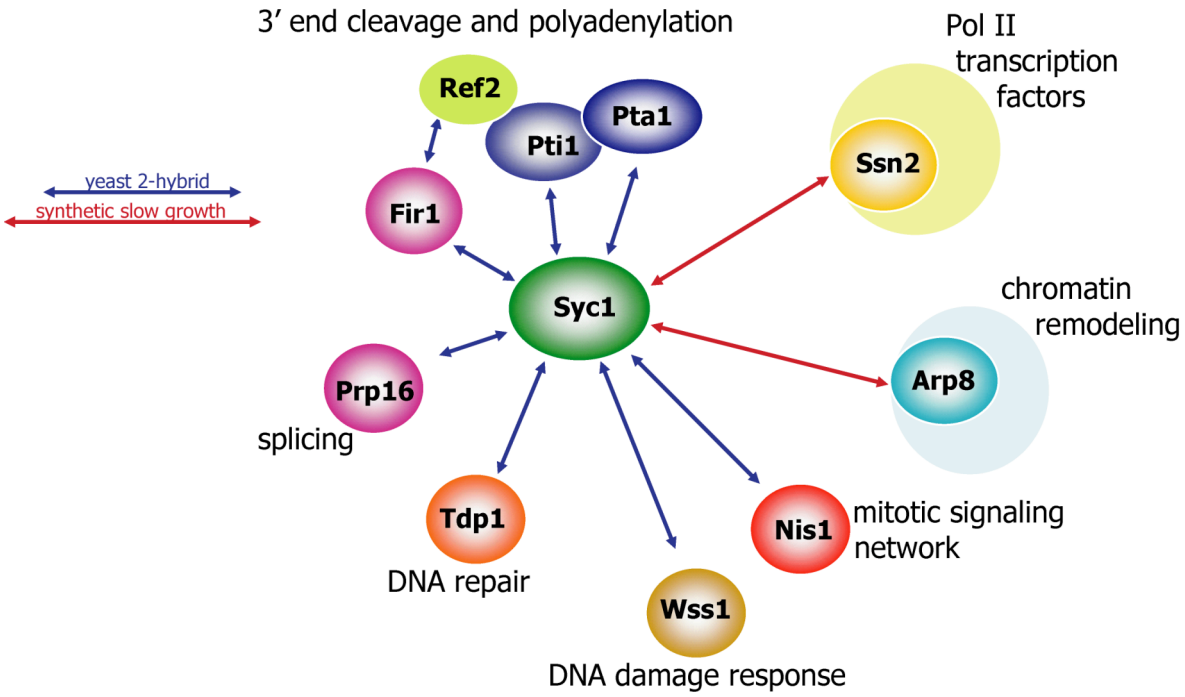


Figure 3.9 *SYC1* genetic interaction network

Results of genetic screens with *SYC1* gene (yeast 2-hybrid and synthetic lethal screen) are schematically summarized. Blue arrows - yeast 2-hybrid interacting partners; red arrows - synthetic slow growth.

DNA microarray analysis of the *syc1Δ* profile

In order to analyze the effect of *SYC1* gene deletion on RNA metabolism in the cell, we compared RNA profiles of *syc1Δ* and wild type cells using microarray analyses. Total RNA was extracted from wild type (BY4741) and *syc1Δ* cells grown at permissive temperatures. Aliquots were tested with Agilent 2100 bioanalyzer system to assure high quality of the RNA samples. Two initial co-hybridization experiments were carried out, with Cy3-labelled wild-type RNA versus Cy5-labelled *syc1Δ* RNA, and reversed. The MWG Yeast Array used here contained 6,250 gene specific oligonucleotide probes (40-oligomers) representing the complete *Saccharomyces cerevisiae* genome. The synthesis of single stranded cDNA, fluorescent dye coupling, hybridization, microarray scanning and data analysis were performed by the MWG Biotech Array Services (Mannheim, Germany). Arrays were scanned using ImaGene software and the resulting data was pre-analyzed and normalized with MAVI Pro 2.6.0 tool. The microarray analysis resulted in detection of several putative or pseudo genes located in the telomere proximities (*YIL058W, YIR043C, YIR044C, YMR172C*) that were upregulated in the *syc1Δ* strain. Apart from that, no significant differences were found in the RNA profile of *syc1Δ* deletion strain compared to wild-type.

3.3 Discussion

In this work, we examined the potential involvement of Syc1p in 3' end formation of mRNAs and snoRNAs. Additionally, we analyzed the net of genetic interaction partners of *SYC1*. Syc1 protein exhibits a high level of homology with the C-terminus of Ysh1p, the endonuclease cleaving pre-mRNA 3' ends. Protein sequence of Syc1p contains a conserved putative P-loop motif (GX_{4-8}GKS/T) found in proteins that bind ATP or GTP. This would suggest a function for Syc1p in binding and perhaps hydrolyzing ATP/GTP. Both cleavage and polyadenylation steps of pre-mRNA 3' end processing were previously reported to require ATP (Zhao et al., 1999). The 3' end endonucleolytic cleavage depends on hydrolysis of the β–γ bond of ATP (Butler and Platt, 1988). Ysh1p also contains a P-loop in its C-terminus, however mutations in this motif have no effect on cell viability, ruling out the possibility that Ysh1p binds or hydrolyzes ATP (Zhelkovsky et al., 2006). Therefore, a function of Syc1p that involves ATP hydrolysis is rather improbable.

We showed that Syc1p is an integral subunit of APT that physically and genetically interacts with the CPF/APT factors Pta1p and Pti1p. Physical interactions between Syc1p and other subunits of APT, Ref1p and Glc7p, were also detected. Most of the APT subunits were implied in 3' end formation of snoRNAs; however, we found that deletion of *SYC1* had no influence on this process. Deletion of *SYC1* gene does not affect yeast cell viability; we observed however that deletion of *SYC1* caused a strong suppression of growth defects in several 3' end processing mutants, such as *glc7-12, pti1-2,* and *ref2Δ*. No effects of *SYC1* deletion were seen in *pta1-1* and *ssu72-2* mutant alleles. Nevertheless, temperature lethality of a different allele of *PTA1*, namely *pta1-2* was reported to be suppressed by *SYC1* deletion (Zhelkovsky et al., 2006), implying the existence of allele-specific genetic interactions between *syc1Δ* and *pta1* mutants. Importantly, the suppression caused by *SYC1* deletion in both *glc7-12* and *ref2Δ* alleles was further exacerbated on cordycepin-containing medium. Both Glc7p and Ref2p factors were previously implied in the polyadenylation step of 3' end formation (He and Moore, 2005; Mangus et al., 2004). The reverse effect of cordycepin was observed in the double mutant strain *pti1-2 syc1Δ,* where the *SYC1* deletion aggravated the growth defect of the *pti1-2* mutant strain at 33ºC. Moreover, the presence of Syc1p exacerbates the growth defects of *rna14-3* and *rna15-2* mutants at semi-permissive temperatures. *SYC1* gene deletion enhances growth defects of *pta1-3, pfs2-1* and *fip1-206*, but has no effect on the growth of *pap1-1* mutant alleles (Zhelkovsky et al., 2006). The alleviation of growth defects of the 3' end processing mutants when combined with *SYC1* gene deletion, and also with the presence of the poly(A) chain terminators (cordycepin), suggests that Syc1p has a possible regulatory function in the polyadenylation step, or perhaps also in the cleavage step of pre-mRNA 3' end processing. As mentioned above, protein structure of Syc1p is highly similar to the C-terminal domain of Ysh1p. The Ysh1p C-terminus is critical for both cleavage and poly(A) addition steps. Importantly, *SYC1* can substitute for the *YSH1* 3' end truncation *in vivo* (Zhelkovsky et al., 2006). This suggests that Syc1p may compete with Ysh1p C-terminus for interactions with other CPF subunits and thus work as a negative regulator of 3' end

processing. In this way, Syc1p could provide a fine-tuning mechanism of mRNA 3' end processing steps, for example to improve efficiency of the cellular responses to environmental changes.

Furthermore, we showed that Syc1p forms a net of genetic interactions that may serve the purpose of linking the 3' end processing machinery with other cellular events. We performed two genetic screens that resulted in discovery of new interaction partners of *SYC1*. The yeast 2-hybrid screen revealed interactions with 3 proteins involved in pre-mRNA 3' end processing: Pta1p, Pti1p, and Fir1p. Several other genes found in the screen point to a wider context of cellular interactions and influences. The correct analysis/prediction of the relevance of such interactions is rather difficult. Worth remembering is the fact that a yeast 2-hybrid is a genetic method with no biochemical value. Expression of the fusion construct is dependent on the nature of such fusion and on yeast genetic background; it can be poorly controlled by conditional promoters. Moreover, the method is not exhaustive; a lot of existing protein-protein interactions cannot be monitored. No common rules in regards to the amount of positive clones can be drawn because of that uncertainty. Therefore all results should be considered potentially significant.

NIS1 was one of the prey genes found to interact with *SYC1* with high statistic relevance. Nis1p plays a role in the mitotic signalling pathway; therefore we hypothesize that the link between *SYC1* (and thus the 3' end processing apparatus) and *NIS1* could include the fine-tuning of 3' end processing. Syc1p might act as a checkpoint to slow processing down when the cell has to adjust to the changing environmental conditions and delay the mitotic process.

An interesting link between the 3' end processing and transcription-related events was suggested by the interaction of *SYC1* with *TDP1*. Tdp1p is a tyrosine-DNA phosphodiesterase that is involved in double strand break repair of DNA lesions created by topoisomerase I (Wang, 1991). One could envision a possibility that the Tdp1 phosphodiesterase, apart from its function in double strand break repair, transmits a signal (directly or indirectly, perhaps via Syc1p) to the RNAP II transcriptional machinery about the current state of euchromatin coiling. Therefore, the signal of transcriptional competency of the euchromatin could enhance the transcriptional events, thus including 3' end processing.

Intriguingly, the yeast 2-hybrid approach identified a splicing factor, *PRP16*, as a genetic interaction partner of *SYC1*. As discussed earlier in this work (Chapter 2), we propose that pre-mRNA 3' end formation and splicing are coupled processes in yeast. Up to date, there is little data available concerning coupling of 3' end processing and splicing in yeast, and the exact molecular basis of this coupling interaction remains unknown. Genetic interaction of Syc1p with Prp16p could provide an additional hint towards such a coupling mechanism. However, Prp16p did not interact physically with Syc1p nor with Ysh1p in the *in vitro* GST pull down assay (data not shown), thus suggesting an indirect mechanism of interaction in the functional coordination of splicing and 3' end processing. Further investigations are required to address potential involvement of Syc1p and Prp16p in this mechanism.

Another genetic interaction of Syc1p, Fir1p, was reported to play an active positive role in regulation of the poly(A) tail length in cooperation with APT subunit Ref2p (Mangus et al., 2004). Moreover, *FIR1* gene was shown to interact genetically with *PAP1* (del Olmo et al., 1997), consistent with the function in poly(A) tail regulation. As both Fir1p and Syc1p might possibly act

in regulating the 3' end processing steps, it is plausible that their influence in this process is of opposite nature, for example Fir1p being a positive regulator of the poly(A) tail length, and Syc1p - putatively - being a negative one. Interestingly, Fir1p was found to interact genetically with a splicing factor, Prp45p in a yeast 2-hybrid screen with a high confidence score (Albers et al., 2003). This could point to another link between the splicing and 3' end processing machineries. One might hypothesize of a model in which Syc1p, Ref2p, Fir1p and other factors are part of a regulatory complex that exploits several, direct or indirect, interactions with the splicing factors (possibly Prp16p and Prp45p), to link cleavage and polyadenylation to splicing. However, the existence of such hypothetical interaction network remains to be experimentally verified.

To shed more light on functional genetic interactions of *SYC1*, we performed an extensive synthetic lethal screen that resulted in identification of two genes, *ARP8* and *SSN2*. Deletion of each of those genes resulted in a slow growth phenotype when combined with a genomic deletion of *SYC1*. Genetic interaction of *ARP8* with *SYC1* suggests a link between transcription, 3' end formation and the chromatin remodelling state. Ssn2p belongs to the Srb module of the RNAP II Mediator complex. Perhaps recruitment of 3' end processing factors to the RNAP II complex during transcription initiation is also facilitated by the Mediator activity.

In summary, we propose that Syc1p has a regulatory function in the pre-mRNA 3' end processing steps and possibly links 3' end formation to other transcriptional events in the nucleus. It is rather difficult to fully explain the genetic interactions of *SYC1* without a better understanding of the architecture of the 3' end processing complex and its functions in relation to other nuclear processes not exclusively involved in RNA metabolism. Nevertheless, the data presented here add to this repertoire by highlighting the importance of Syc1p for optimal functioning of the 3' end processing machinery as well as for its co-regulation in a wider nuclear context.

3.4 Experimental procedures

Expression and purification of the recombinant Syc1 protein

E. coli BL21(DE3) strain (Studier, 1991) was transformed with the respective plasmids pBD26 or pMG37 carrying either GST or GST-Syc1p-His6 and grown in 2xYT medium containing 100 µg/ml ampicilin at 37ºC to an OD_{600} of 0.8-1.0 and expression was induced by the addition of isopropyl-D-thiogalctoside (IPTG) to final concentration of 0.5 mM. After 3 h of expression at 37ºC, cells were harvested, lysed and proteins were purified by affinity chromatography on Ni-NTA agarose (nickel-nitrilotriacetic acid agarose; Qiagen), followed by second affinity purification on glutathione Sepharose 4B (Pharmacia), according to the suppliers manual.

GST pull down experiments

Plasmid DNA encoding the different 3' end processing subunits were *in vitro* transcribed and translated with the TNT coupled transcription-translation system (Amersham) in a total volume of 50 µl, according to the manufacturer's instructions. Approximately 100 ng of the recombinant GST and GST-Syc1-His6 proteins were incubated with 3 µl of the *in vitro* translated ^{35}S-labelled proteins for 1 h at 23ºC and subsequently immobilized onto 100 µl glutathione Sepharose 4B (Pharmacia) in buffer P (phosphate-buffered saline (PBS) containing 0.01% NP40, 1 mM DTT, 0.4 µg/ml leupeptin, 0.7 µg/ml pepstatin, 0.5 µg/ml PMSF) by shaking for 1 h at 4ºC. The beads were pelleted and washed extensively 5 times with buffer P for 10 min at 4ºC. Proteins were eluted in 20 µl SDS-PAGE sample buffer and resolved by 8 % SDS-PAGE. Gels were fixed, treated with 1 M Na-salicylate, and vacuum-dried. The ^{35}S proteins were visualized by autoradiography.

Extract preparation and *in vitro* cleavage and polyadenylation assays

3' end processing extracts were made following the procedure described previously (Ohnacker et al., 2001). Cleavage and polyadenylation assays were carried out as previously described (Minvielle-Sebastia et al., 1994). When cleavage only was assayed, EDTA replaced magnesium acetate (MgAc) and ATP was omitted. Internally ^{32}P-labelled CYC1 was produced by *in vitro* run-off transcription.

Droplet test

Strains were grown overnight in YPD medium and diluted to OD_{600} of 0.1, 0.01, 0.001 and 0.0001. 5 µl of each dilution were spotted on SD plates, without or containing 20 µg/ml or 40 µg/ml cordycepin and incubated at 30, 33 and 37ºC for 3 days.

Yeast strains

Manipulations and growth of *S. cerevisiae* were carried out by established procedures. *S. cerevisiae* strains used in this study and their genotypes are listed in Table 3.2. Double mutant strains were obtained by standard crossing and sporulation procedures. Positive spores were tested for the presence of the respective auxotrophic markers and temperature sensitivity.

Table 3.2 Yeast stains used in this study

Strain	Background genotype	Reference
W303-1A	*MATa ade2; leu2; ura3; trp1-1; his3*	Thomas and Rothstein, 1989
BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	ResGen Inc.
syc1Δ (syc1Δ::KanMX)	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 syc1Δ::KanMX*	ResGen Inc.
syc1Δ (syc1Δ::NatMX)	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 syc1Δ::NatMX*	ResGen Inc.
ref2Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ref2Δ::KanMX*	ResGen Inc.
arp8Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 arp8Δ::KanMX*	ResGen Inc.
ssn2Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ssn2Δ::KanMX*	ResGen Inc.
edc3Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 edc3Δ::KanMX*	ResGen Inc.
pti1-2	n. d.	Dheur et al., 2003
glc7-12	*MATa W303 glc7::LEU2 trp1:: glc7-12*	M. Stark
YWK186 (ssu72-2)	*MATa ura3-1 trp1-1 ade2-1, leu2-3,112 his3-11,15 ssu72-2*	Dichtl et al., 2002a
pta1-1	*MATa ade2-1 leu2-Δ1 lys2-801 trp1-Δ101 ura3-5 pta1-1*	O'Connor and Peebles, 1992
YMG33 (pta1-1/syc1Δ)	n. d.	This work
YMG34 (pti1-2/syc1Δ)	n. d.	This work
YMG35 (ssu72-2/syc1Δ)	n. d.	This work
YMG50 (syc1Δ/ref2Δ)	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ref2Δ::KanMX Δsyc1::NatMX*	This work

n.d. – not determined

RNA analyses

Yeast total RNA was extracted using hot-phenol method and Northern blot analyses were carried out as described previously (Dichtl et al., 2002b). Oligonucleotide probes were routinely labelled with 30 µCi [γ-^{32}P]ATP and T4 polynucleotide kinase. Random-primed DNA probes were generated with [α-^{32}P]dATP using the Roche kit.

Yeast 2-hybrid screen

The yeast 2-hybrid screen with *SYC1* gene as bait was performed in collaboration with Hybrigenics, Paris, France, as part of the RNOMICS network. For details please refer to the webpage: http://www.hybrigenics.com/services/Y2HService.html.

Synthetic lethal screen

The synthetic lethal screen with *SYC1* gene was performed in collaboration with A. Jacquier and L. Decourty (Institut Pasteur, Paris, France), as part of the RNOMICS network. The details of the method are partially described elsewhere (Decourty et al., 2008).

Plasmids and primers

Plasmids used in this study are listed in Table 3.3. The sequence of the *SYC1* gene (*YOR179C*) was PCR amplified from genomic DNA with the primers: YOR179c-5'NdeI (5'-GGGAATTCCATATGGATTTACCCAAGG) and YOR179c-3'BamHI (5'-CCGCGGATCCTTAATG ATGATGATGATGATGGCACAATGGAGTTGTTAATT). PCR product was ligated into pGDV1 and pET21a vectors using NdeI and BamHI restriction sites, resulting in pMG37 and pMG38 constructs, respectively. pMG31, the bait entry plasmid for yeast 2-hybrid screen, was constructed by amplification of the *SYC1* genomic locus (*SYC1* ORF plus 300 nucleotides upstream and downstream of the coding region) by PCR using primers YOR179c-5'SacII (5'-TCCCCGCGGGGATGGATTTACCCAAGGACAAAAG) and YOR179c-3'SpeI (5'-GGACTAGTCA GCACAATGGAGTTGTTAATTTC). The resulting PCR fragment was cloned into the pB27 vector (Hybrigenics) using SacII and SpeI restriction sites. pMG40 and pMG41, constructs used in the synthetic lethal screen, are derivatives of pFL36cII and pY2-MET15 vectors, respectively. They contain *SYC1* genomic locus (analogically to pMG31) and were obtained using the Gateway technology (Invitrogen), following supplier instructions.

Table 3.3 Plasmids used in this study

Plasmid	Description	Reference
pMG37	pGDV1- SYC1-His6, NdeI BamHI	This work
pBD26	pGDV1	B. Dichtl
pMG38	pET21a-SYC1-His6, NdeI BamHI	This work
pB27	Bait entry, yeast 2-hybrid	Hybrigenics
pMG31	pB27- SYC1, SacII SpeI	This work
pMG40	pFL36cII-SYC1, attB	This work
pMG41	pY2-MET15-SYC1, attB	This work

DNA microarray analysis

The microarray analysis of the *syc1Δ* RNA profile was performed using a commercial service by MWG Biotech, Germany. For details please refer to the webpage: http://www.ocimumbio.com/web/.

Computing analyses

Sequences have been found by several BLASTP searches at the NCBI-Genbank Website (http://www.ncbi.nlm.nih.gov/BLAST/).

CHAPTER 4

The archean homolog of Ysh1p, *M. jannaschii* MJ1236, has RNase activity *in vitro*

4.1 Introduction

4.2 Results

The genome of *Methanococcus jannaschii* encodes for three homologues of the yeast *YSH1* gene

Recombinant MJ1236 possesses an endonucleolytic activity *in vitro*

Mutations in the conserved β-lactamase motif of rMJ1236 reduce its endonucleolytic activity

The nucleolytic activity of rMJ1236 is sequence-independent and temperature-stable

The hybrid KH-Ysh1 protein binds to RNA but remains endonucleolytically inactive *in vitro*

4.3 Discussion

4.4 Experimental procedures

4.1 Introduction

A consensus view exists in the scientific community that all living beings are descendants of a single ancestor. Homologues of most genes can be identified within all three domains of life: Eukarya, Archaea and Bacteria, and they are therefore traceable to the universal ancestral progenote (last universal common ancestor, LUCA; Woese and Fox, 1977). The existence of Archea as the third domain of life was first proposed in 1977 by Carl Woese (Woese and Fox, 1977). Despite a superficial resemblance to bacteria, archaeons have a distinct evolutionary heritage and therefore belong in a separate category (Pace, 2006). Gene duplication analyses support the view that Bacteria branched before the Archaea and Eucarya; and consistently Archaea are closer related to Eucarya than to Bacteria (Embley and Martin, 2006; Pace, 2006; Figure 4.1).

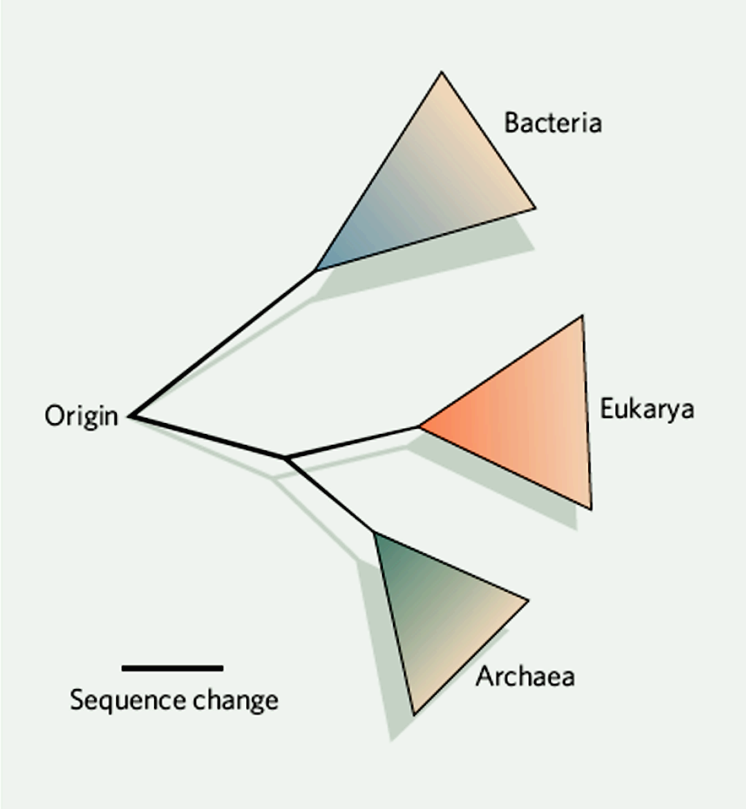


Figure 4.1 Scheme representing the three major kingdoms within the universal tree of life
The relative distance in the figure represents an evolutionary distance as determined by 16S-like ribosomal RNA analyses. Scheme taken from Pace, 2006.

Most of the archean species inhabit extreme environments, which makes them the rule breakers of biology, as they not only survive but also thrive under conditions previously thought to prohibit all forms of life. Analyses of the archaean genome sequences provide valuable evolutionary information, and in addition they offer insights into understanding the principles of various molecular pathways, as well as of the structural and chemical basis of protein thermostability. Many of these can be further exploited for research and for industrial and

pharmaceutical processes. Moreover, archaean homologues of many important eukaryotic proteins were found to be easier to handle experimentally and thus also easier to reveal their function.

Methanococcus jannaschii was the first member of the Archaea domain to have its genome completely sequenced and analyzed (Bult et al., 1996). *M. jannaschii* belongs to the euryarchaeotal kingdom of Archaea, and was originally isolated in 1983 from sediment of a deep sea hydrothermal vent in the Pacific Ocean (depth of 2,610 m, 261 atmospheres; Jones et al., 1983). This extremely thermophilic organism grows at temperatures approaching the boiling point of water (80°C to 95°C) and at pressures over 200 atmospheres. Importantly, *M. jannaschii* requires no organic nutrients for growth, as it has all the biochemical pathways and genes needed to reproduce itself from inorganic constituents, such as carbon dioxide and ammonia (Jones et al., 1983). As a methanogen, it derives energy by using molecular hydrogen to reduce carbon dioxide to methane. The genome of *M. jannaschii* is rather small as for a free-living organism, and consists of a major chromosome (1.66 Mbp) and two smaller extra-chromosomal elements. It has also a surprisingly low average G+C content (31 %). Computational comparative analysis predicted a total of 1738 protein-coding genes; however, only a part of these (38%) could be assigned a putative cellular role with high confidence. In general, *M. jannaschii* genes related to energy production, cell division, and metabolism are similar to those found in bacteria. Strikingly, however, most of the *M. jannaschii* genes involved in transcription, translation, and replication are similar to those found in Eukaryotes (Bult et al., 1996). For instance, its transcriptional apparatus, the RNA polymerase, promoter motifs, transcription factors and TATA-binding proteins, splicing apparatus and histone-like structures, all closely resemble those of eukaryotes. Similarly, ribosomal proteins found in Archaea are often related to the eukaryotic ones (Garrett, 1996).

The phenomenon of polyadenylation occurs in all three primary kingdoms. Little is however known about the 3' end processing of mRNAs in Archaea. A sequence homologue of the 3' end processing endonuclease (Ysh1p/CPSF73) is conserved throughout the archaeal lineage, and is also present in some bacteria. Moreover, the archaean homologues of the CCA-adding enzymes are closely related to the eukaryotic poly(A) polymerases, suggesting that they may act as PAPs in these organisms (Anantharaman et al., 2002).

The main aim of this thesis was to identify the endonucleolytic activity involved in pre-mRNA 3' end processing and assign it to a specific component of the 3' end processing machinery in yeast. However, in the course of experiments with the putative yeast endonuclease Ysh1p we encountered several obstacles in assaying its activity *in vitro*. Therefore, we turned to the archaeal system as a potentially simplified one, aiming to assign an endonucleolytic activity to the archaean homologue(s) of Ysh1p. We chose the archaeon *M. jannaschii*, as many of the open reading frames present in its genome have been assigned putative functions as a result of computer searches for sequence homologues and known sequence motifs. We performed computational protein sequence analyses of the *M. jannaschii* genome in order to identify potential homologues and sequence motifs of the (putative at the time) 3' end processing endonuclease, Ysh1p. Our analysis revealed three open reading frames: *MJ0047*, *MJ0162*, and

MJ1236, being notably similar to the β-lactamase and β-CASP domains of Ysh1p. Interestingly, the *MJ1236* gene encodes for a protein that carries, apart from the β-lactamase and the β-CASP domains, also a KH-RNA binding domain, thus suggesting a potential function in RNA metabolism. Heterogeneous expression and assaying of *MJ1236* gene revealed its endonucleolytic activity *in vitro*, thus strongly implying the same type of activity for its *S. cerevisiae* homologue, Ysh1p.

4.2 Results

The genome of *Methanococcus jannaschii* encodes for three homologues of the yeast *YSH1* gene

The β-lactamase and β-CASP domains of Ysh1p are highly conserved throughout evolution. Using *in silico* comparative sequence analysis, we identified three open reading frames within the genome of *M. jannaschii* that are highly homologous to Ysh1p endonuclease. Two of those genes, namely *MJ0047* and *MJ0162*, are 26 % and 29% identical with Ysh1p over the β-lactamase and β-CASP domains, respectively (Figures 4.2 and 4.3). The percent of conservation of the third homologous gene, *MJ1236*, is also comparable, being ~32% identical and ~50% similar to Ysh1p. Strikingly, the N-terminus of MJ1236 contains an RNA binding domain of the KH type (Figure 4.3). The K homology (KH) domains occur in a number of proteins that bind non-specifically to single-stranded RNA and play major roles in regulating cellular RNA metabolism (Ito et al., 1994; Musco et al., 1996; Siomi et al., 1993). Therefore, the presence of KH domain in MJ1236 is a good indication of the involvement of this protein in RNA metabolism.

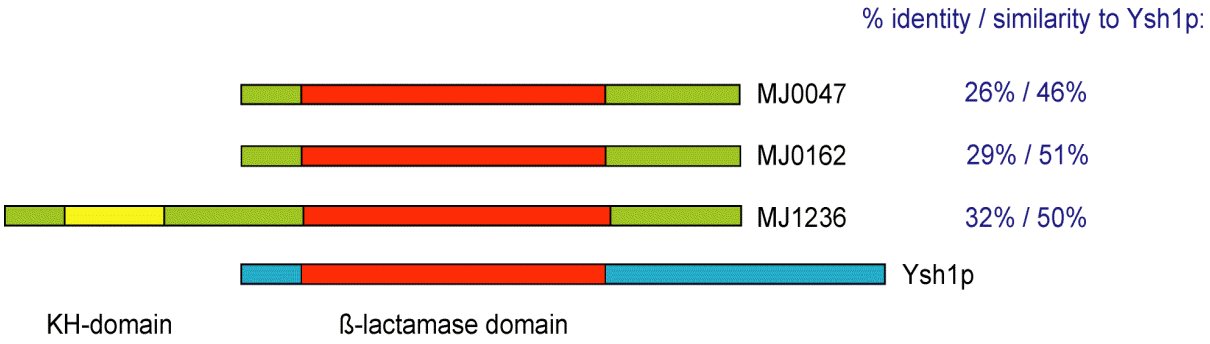


Figure 4.2 *M. jannaschii* homologues of Ysh1p share the conserved B-lactamase domain
Simplified representation of the domain organization of Ysh1p and its *M. jannaschii* homologues. All the three archaean genes (*MJ0047*, *MJ0162*, and *MJ1236)* carry the conserved β-lactamase and β-CASP motifs (marked in red), similarly to Ysh1p. Additionally, MJ1236 carries an N-terminal KH RNA binding domain (yellow). The percent of homology between Ysh1p and archaean proteins is shown on the right.

Sequences of the three *M. jannaschii* genes are highly identical with each other over the conserved β-lactamase and β-CASP motifs, but rather only similar in between them (Figure 4.3). Comparison of the protein sequences of those archaean homologues to Ysh1p revealed conserved blocks identical between them, and several similar ones; thus confirming close relationship between those orthologues, and suggesting the same type of enzymatic activity (Figure 4.4).

Recombinant MJ1236 has an endonucleolytic activity *in vitro*

In order to verify whether the archaean homologues of Ysh1p possess an endonucleolytic activity *in vitro*, we analyzed the product of the *MJ1236* gene. We chose *MJ1236* based on the presence of KH RNA binding domain within its protein sequence. *MJ1236* gene encodes for a 635 amino acid protein with predicted size of 71.7 kDa. *MJ1236* was cloned from the *M. jannaschii* genomic DNA. Heterologous expression of GST- and His6-tagged *MJ1236* in *E. coli* was followed by double affinity purification over the NiNTA agarose and gluthatione Sepharose columns (Figure 4.5 A). Next, we tested for potential nucleolytic and RNA binding activities of the purified recombinant MJ1236 protein (rMJ1236). Recombinant proteins were expressed in *E. coli*, as no endogenous pre-mRNA cleavage activity has been observed in bacteria that could account for the background activity.

Initial tests for the RNA binding activity of the recombinant MJ1236 via electro-mobility shift assays (EMSA) resulted in severe degradation of the RNA substrate (data not shown). Furthermore, *in vitro* cleavage assays with rMJ1236 using various RNA substrates revealed a strong nucleolytic activity embedded in this protein (Figures 4.5 and 4.6, and data not shown). To test whether this activity is of an exo- or an endonucleolytic type, we performed *in vitro* cleavage assays using sCYC1 RNA (Dichtl and Keller, 2001) radioactively labelled at either 5' or 3' end. Incubation of rMJ1236 with RNAs labelled both ways resulted in a similar pattern of bands, and fast accumulation of mononucleotides was not observed (Figure 4.5 B, C). This indicated that the rMJ1236 possessed an endonucleolytic, not exonucleolytic, activity.

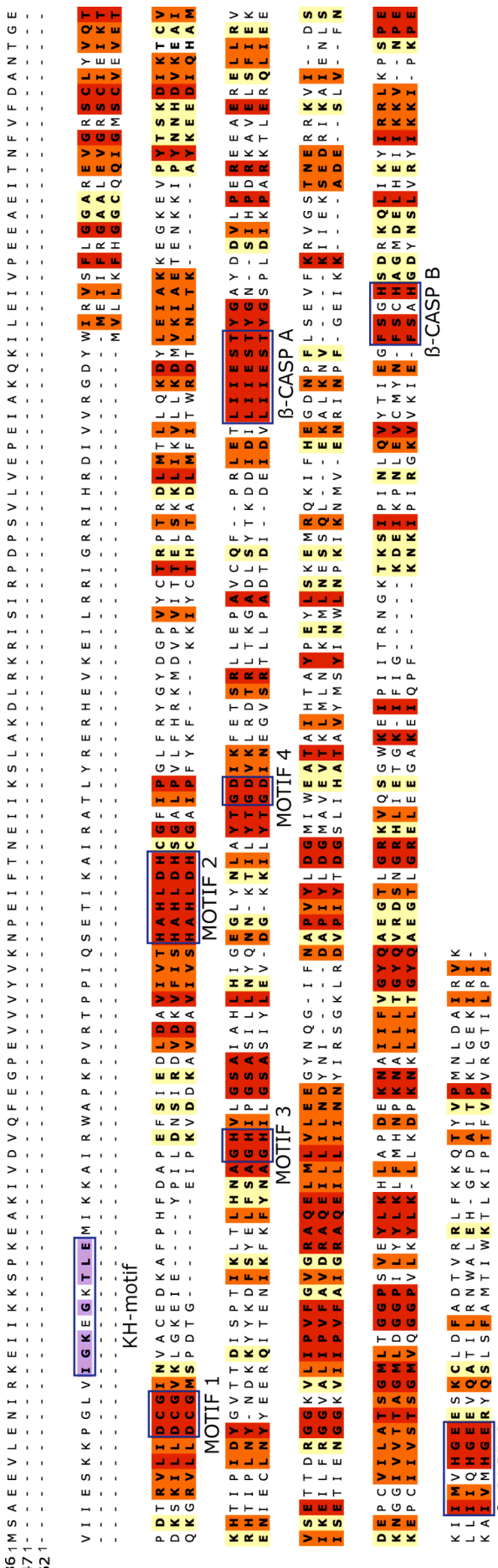


Figure 4.3 *M. jannaschii* homologues of Ysh1p carry conserved β-lactamase and β-CASP domains

Sequence alignment of *M. jannaschii* genes: *MJ0047* (Swiss-Prot accession number: Q60355), *MJ0162* (acc. no. Q57626), and *MJ1236* (acc. no. Q58633). Invariant residues are shown in red boxes (100% conservation); orange boxes indicate conserved and similar amino acids; yellow boxes mark similar residues common for all of the aligned sequences; and violet boxes indicate the conserved amino acids of the KH domain. Conserved motifs of β-lactamase (1-4), β-CASP (A-C), and KH domains are marked in frames.

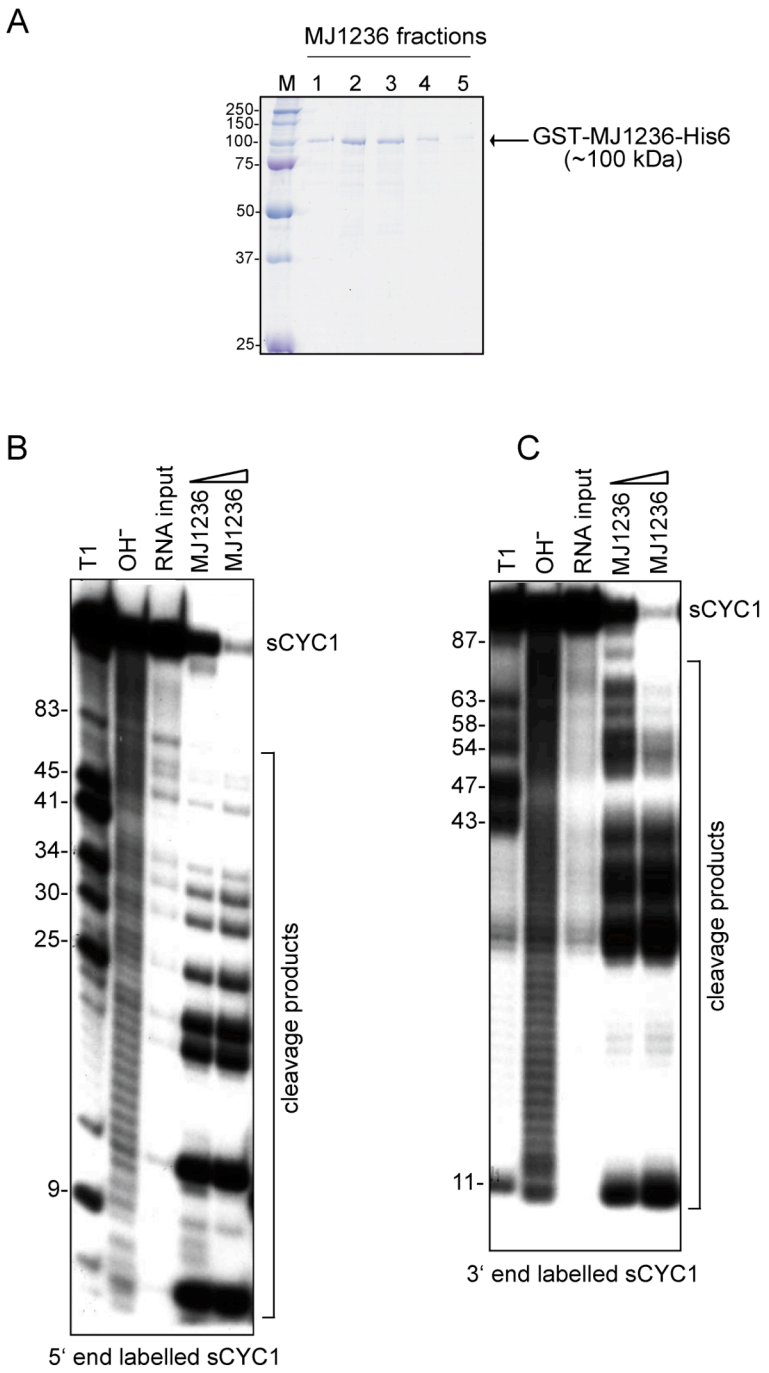


Figure 4.5 Recombinant MJ1236 possesses an endonucleolytic activity *in vitro*

(**A**) SDS-PAGE analysis of rMJ1236 obtained after double affinity purification (see: Experimental procedures). Position of rMJ1236 is indicated. Protein marker sizes are given on the left. (**B**) Nucleolytic assay with purified rMJ1236. The protein was mixed with *in vitro* transcribed 5’ end radioactively labelled sCYC1 RNA substrate, and the mixture was incubated at 37ºC for 1 h. T1, sCYC1 RNA digested with the T1 RNase, and OH-, sCYC1 RNA hydrolyzed with NaOH, were used as size markers. (**C**) Nucleolytic assay performed as in (B), except that the 3’ end labelled sCYC1 substrate was used.

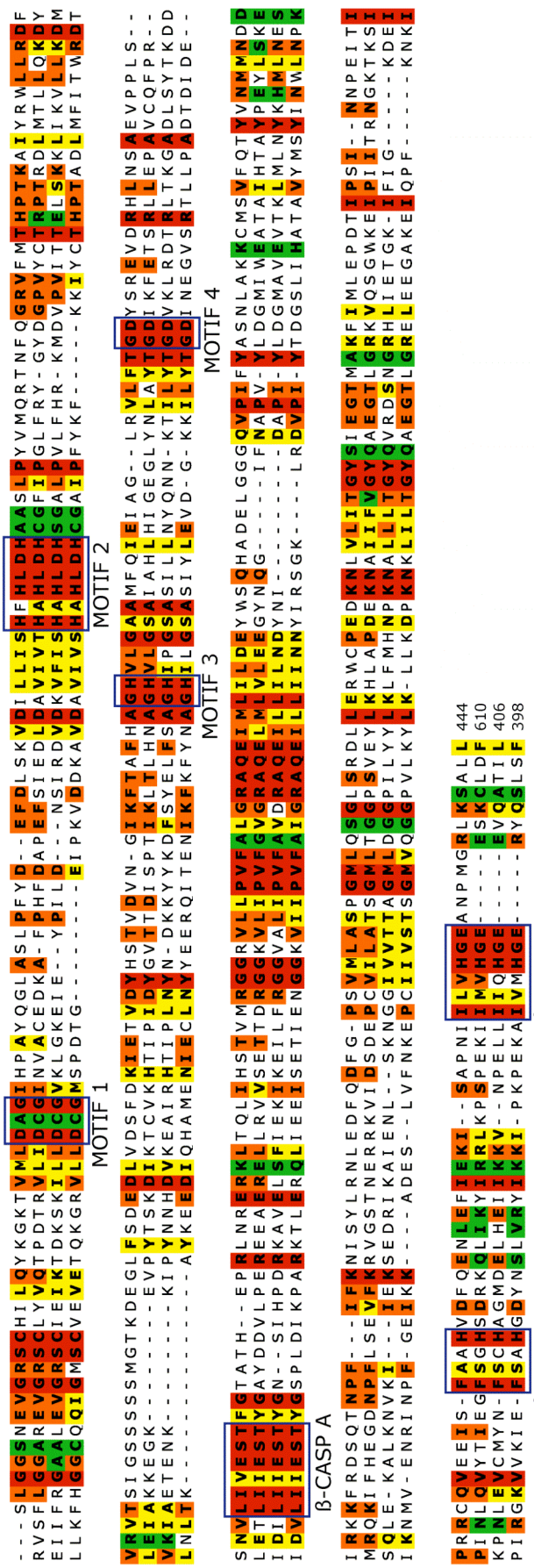


Figure 4.4 Similarities between the protein sequences of Ysh1p and its *M. jannaschii* homologues

Protein sequence of β-lactamase and β-CASP domains of Ysh1p (YLR277C, Swiss-Prot accession number: Q06224) is aligned with its homologues in *M. jannaschii: MJ0047* (acc. no. Q60355), *MJ0162* (acc. no. Q57626), and *MJ1236* (acc. no. Q58633). Invariant residues are shown in red boxes (100% conservation); orange boxes indicate amino acids conserved in Ysh1p and at least one of the *M. jannaschii* homologues; yellow boxes mark similar residues common for Ysh1p and the archaean homologues; and green boxes indicate amino acids similar between Ysh1p and at least one of the *M. jannaschii* homologues. Conserved motifs of β-lactamase (1-4), and β-CASP (A-C) domains are marked in frames.

Mutations in the conserved β-lactamase motif of rMJ1236 reduce its endonucleolytic activity

To confirm that the observed endonucleolytic activity was embedded in the rMJ1236 protein, and was not due to the impurities originating from *E. coli* cells used for its expression, we prepared and expressed a mutant version of this protein. The crucial residues H_{240}, H_{242}, and D_{244} found within the β-lactamase signature (HXHXDH) motif were substituted by alanine residues. Such mutations are predicted to abolish the enzymatic activity of the metallo-β-lactamase enzymes, as they destroy the catalytic centre (see also: Chapter 2.2, Figure 2.2). Subsequently, we performed purification of wild-type and mutant versions of rMJ1236 protein. In addition to the GST- and NiNTA affinity purification, the third purification step - ion exchange chromatography using MonoS column - was performed. This triple-purification resulted in highly pure fractions of both wild-type and mutant rMJ1236 proteins (Figure 4.7 A). When the two rMJ1236 versions were assayed for endonucleolytic activity *in vitro*, only the wild-type protein produced digestion products of CYC1 RNA; in contrast, rMJ1236 HHD mutant remained largely inactive (Figure 4.6 B; compare lanes 3-6 to 7-10). The above result confirmed that rMJ1236 protein possesses an endonucleolytic activity *in vitro* and that the β-lactamase signature motif (HXHXDH) is indispensable for this activity.

The nucleolytic activity of rMJ1236 is sequence-independent and temperature-stable

In its natural environment, *M. jannaschii* thrives at temperature range of 80-95ºC; therefore proteins encoded in its genome are believed to be thermostable. We tested rMJ1236 for an endonucleolytic activity at elevated temperatures, and found that the endonucleolytic activity was most prominent at 50ºC when compared to 37ºC (Figure 4.7 A). Interestingly, the protein was active even at 70ºC; however, as RNA substrates tend to degrade upon incubation at temperature higher than 60ºC, it was difficult to distinguish between unspecific degradation and specific endonuclease activity of rMJ1236. Nevertheless, recombinant MJ1236 exhibited properties of a thermostable protein. Additionally, *in vitro* cleavage assay was performed in a time-course fashion. Cleavage products appeared within 15 minutes of incubation of rMJ1236 protein with RNA substrate at 37ºC (Figure 4. 7 B); however, the exact kinetics analysis has not been elucidated. Next, we asked whether the endonucleolytic activity of rMJ1236 may be sequence-specific. We found that various RNA substrates (sCYC1, sADH1 and sGAL7) were equally well cleaved by rMJ1236 *in vitro* (Figure 4.7 C). This result suggested that the endonucleolytic activity of rMJ1236 is probably sequence-unspecific, consistent with the non-specific type of RNA binding provided by the KH domain embedded in this protein (see also: 4.3 Discussion).

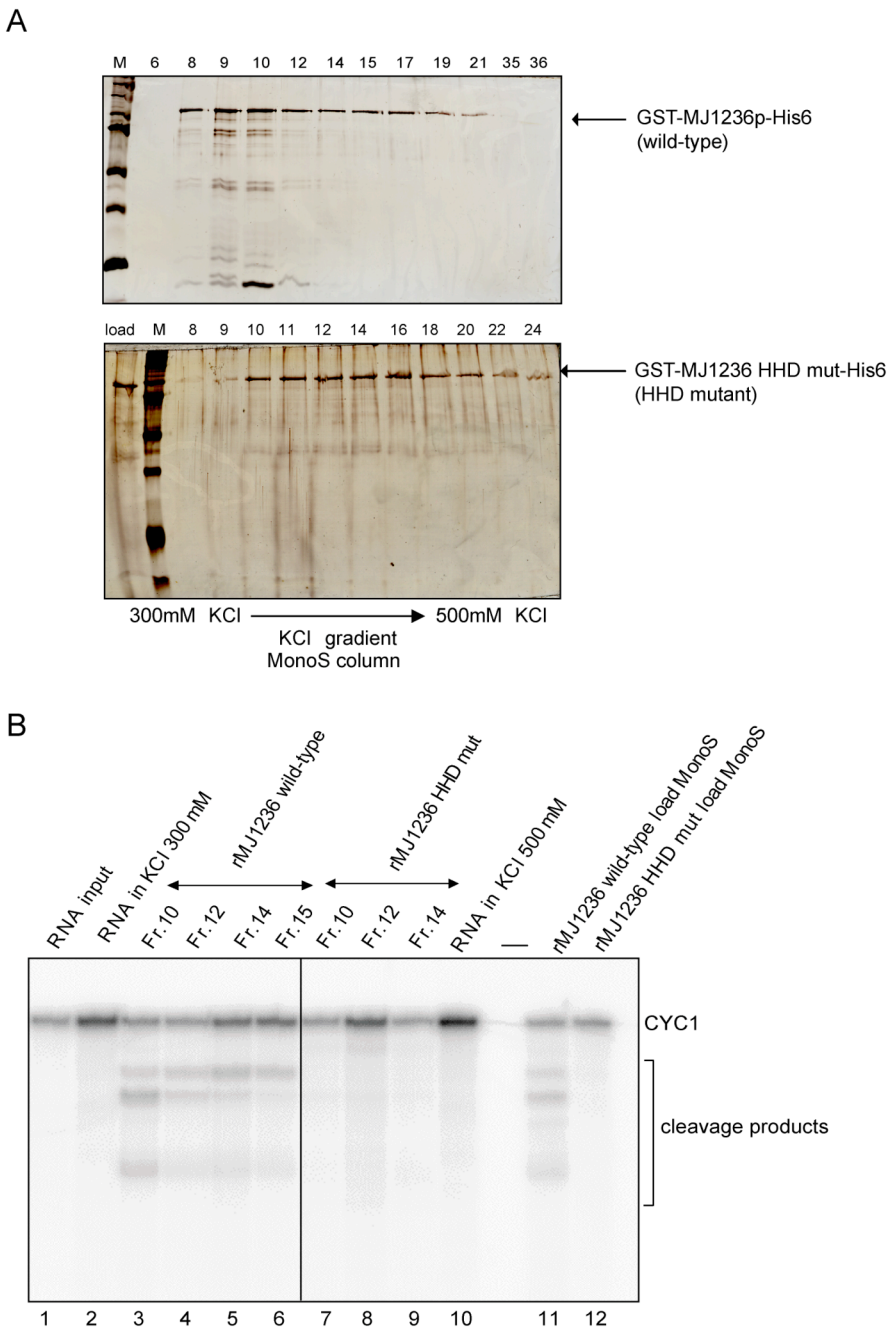


Figure 4.6 Mutations in the conserved β-lactamase motif of rMJ1236 reduce its endonucleolytic activity

(**A**) Results of the three-step purification of the recombinant wild-type and mutant MJ1236. Recombinant MJ1236, affinity-purified on glutathione Sepharose and Ni-NTA Agarose columns, was further fractionated using the ion exchange chromatography (MonoS column). Proteins bound to the column were eluted using KCl gradient ranging from 300 mM to 500 mM, in Hepes/EDTA buffer. Fractions were analyzed on silver-stained SDS-PAGE. Numbers on the top of the gels represent subsequent fractions used later for cleavage assays. (**B**) *In vitro* cleavage assay using CYC1 RNA substrate and fractions of the triple purified wild-type and mutant rMJ1236 shown in (A). 20 µl of the respective fractions were used for assaying. Positions of CYC1 RNA substrate and of cleavage products are indicated.

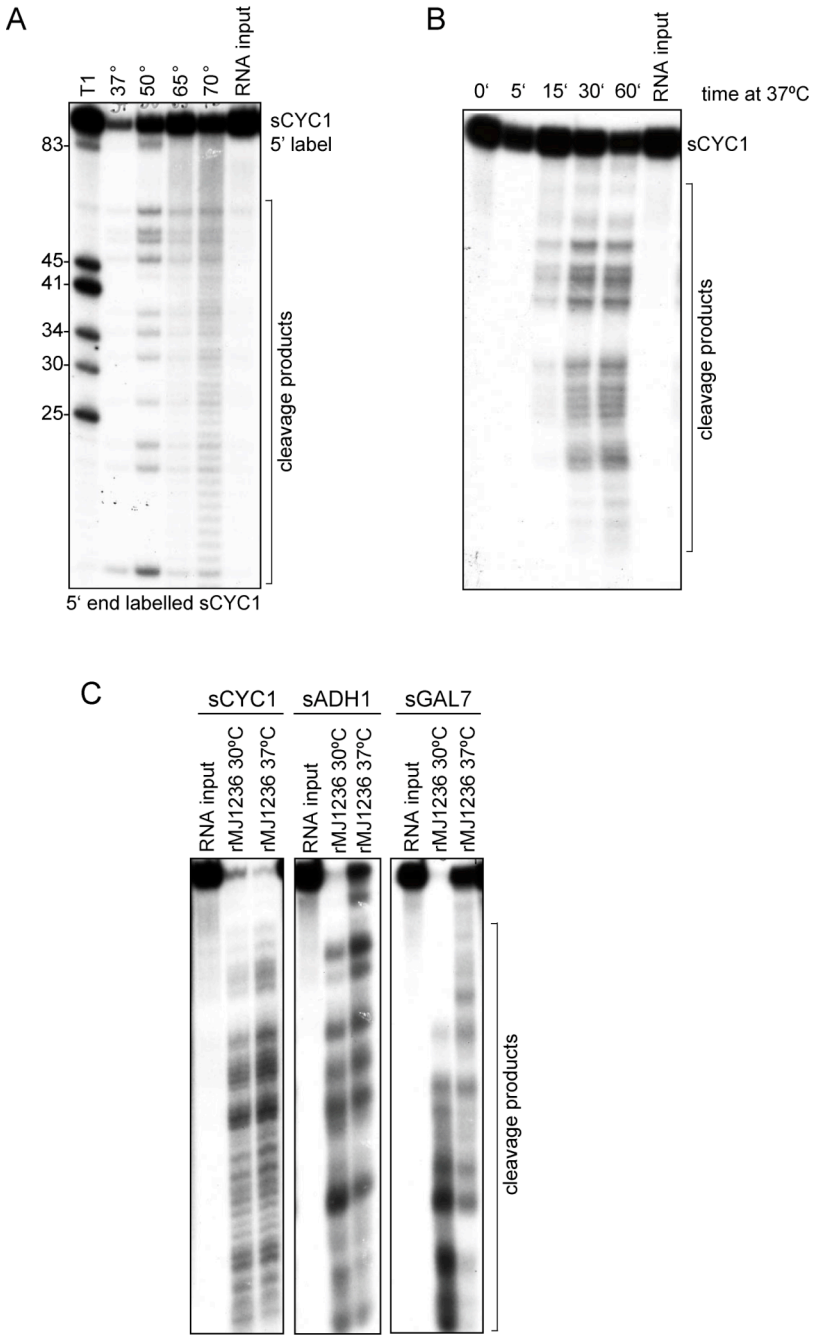


Figure 4.7 Endonucleolytic activity of rMJ1236 is temperature-stable and sequence-independent

(**A**) Recombinant MJ1236 is active at a wide range of temperatures. rMJ1236 was assayed with *in vitro* transcribed 5' end labelled sCYC1 RNA substrate for 30 min at the respective temperatures (37º, 50º, 65º, 70ºC). T1, sCYC1 RNA digested with the T1 RNase, was used as a specific size marker. (**B**) *In vitro* cleavage assay with rMJ1236 performed in a time-course fashion. The assay was performed in principle as in (A), except internally labelled sCYC1 RNA was used. Reactions were incubated for 0'-60' at 37ºC. (**C**) rMJ1236 is active on various RNA substrates. rMJ1236 fractions were assayed with *in vitro* transcribed internally labelled RNA substrates: sCYC1, sADH1, and sGAL7, respectively; using standard cleavage assay conditions. Samples were incubated for 1 h at 30º or 37ºC, as indicated.

Hybrid KH-Ysh1 protein binds to RNA but remains endonucleolytically inactive *in vitro*

The main constraint in studying the putative enzymatic activity of Ysh1p *in vitro* is the fact that Ysh1p alone, without its CPF counterparts, is unable to recognize and bind to RNA. Taking advantage of high homology of Ysh1p to its *M. jannaschii* homologues, we created a genomic construct that encoded for a hybrid protein in which the KH domain of MJ1236 was attached to the N-terminus of Ysh1p. We aimed to reconstitute the putative structural properties of such a hybrid protein in a manner most similar to the native situation in MJ1236, therefore the N-terminus of Ysh1p that is not conserved between Ysh1p and MJ1236 was removed and thoughtfully replaced with the archaean N-terminus containing the KH domain (Figure 4.8 A). Having recombinantly expressed and purified the KH-Ysh1 protein we tested it for an endonucleolytic activity *in vitro*; however, we did not detect any (data not shown). Therefore, we assayed KH-Ysh1 protein for RNA binding activity. Electro-mobility shift assay (EMSA) revealed that KH-Ysh1p bound to sCYC1 RNA in a mode similar to the recombinantly expressed KH domain alone (Figure 4.8 B, lanes 4-5); again, it did not show any endonucleolytic activity. In contrast, recombinant MJ1236, used here as a control, efficiently degraded RNA (lane 3). Recombinant Rna15p was used as a positive control for EMSA conditions, and avidly bound RNA (lane 2). The above results indicated that tethering of Ysh1p to RNA via an RNA binding domain, even in a manner that resembles evolutionarily conserved conformation, is not sufficient to 'revive' the *in vitro* endonucleolytic activity of this factor.

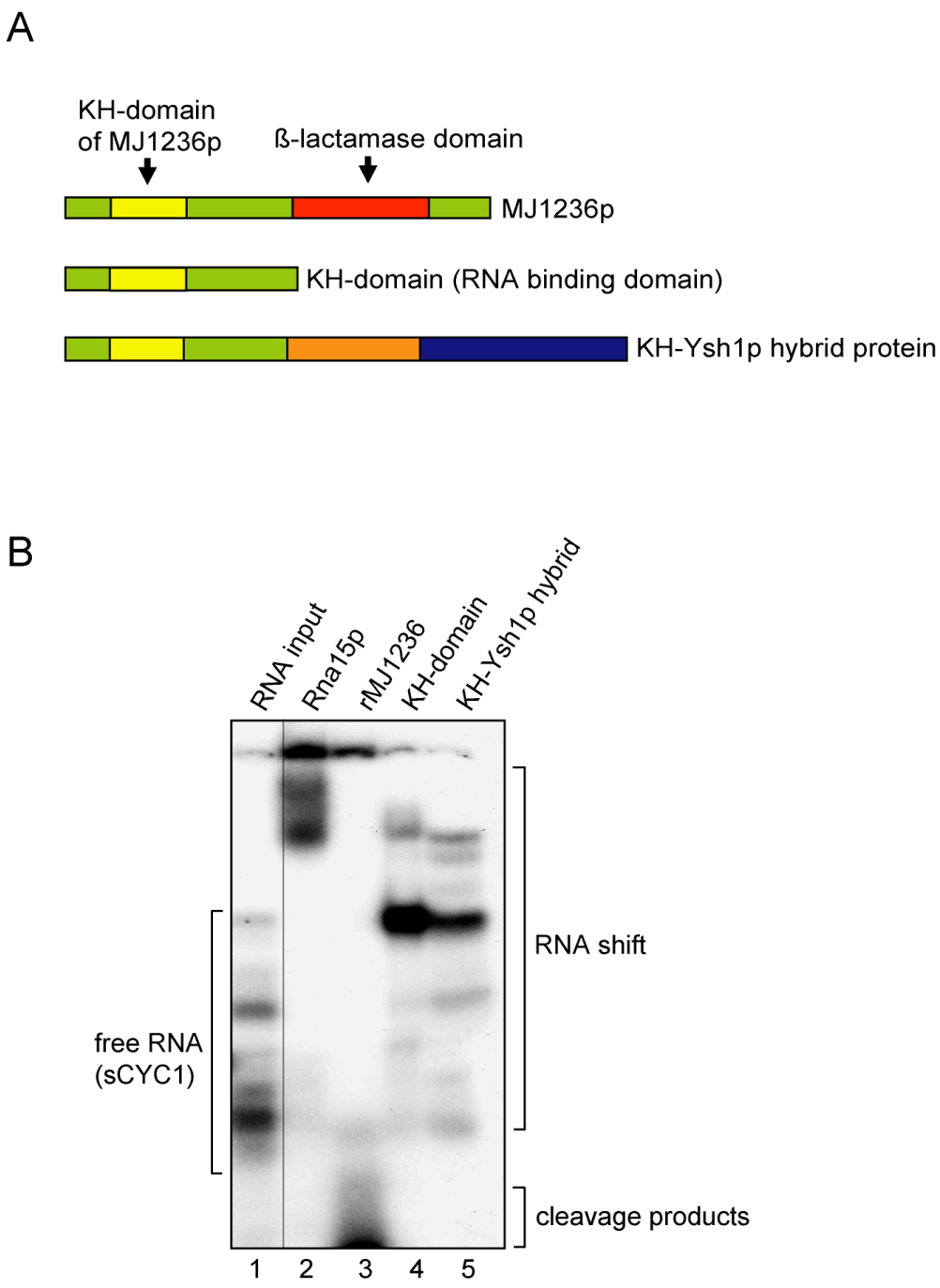


Figure 4.8 Hybrid KH-Ysh1 protein binds to RNA but remains endonucleolytically inactive *in vitro*

(**A**) Schematic representation of the hybrid constructs used in the experiment. (**B**) Electro-mobility shift assay (EMSA) using *in vitro* transcribed sCYC1 RNA substrate and purified recombinant proteins: KH-domain, MJ1236, KH-Ysh1p, and Rna15p. Positions of the unbound RNA, of the shifted (protein-bound) RNA, as well as of the rMJ1236 cleavage products are indicated.

4.3 Discussion

In this work, we searched the genome of the Archaeon *Methanococcus jannaschii* for the presence of sequences homologous to the pre-mRNA 3' end endonuclease Ysh1p. We identified three genes, *MJ0047*, *MJ0162*, and *MJ1236*, encoding for putative orthologues of Ysh1p/CPSF73. Out of these, only MJ1236 protein sequence contains a KH domain motif involved in single stranded RNA binding, thus pointing to RNA-related function of this protein. Consistently, in this work we showed that recombinantly expressed MJ1236 possessed an endonucleolytic activity *in vitro*. Notably, such domain architecture, i.e. N-terminal KH-RNA binding domain fused to C-terminal metallo-β-lactamase domain, is highly conserved in the archaean kingdom, as revealed by *in silico* analyses of the available archaean genomes (data not shown). This suggests an important function of such enzyme family in the archaean RNA metabolism.

The digestion pattern of rMJ1236 was similar to the one observed with recombinant CPSF73, published recently by Mandel et al., 2006. Although, as presented here, purified recombinant MJ1236 revealed strong endonucleolytic activity on yeast RNA substrates, the output from independent *in vitro* cleavage assays showed various results (data not shown). Fractions of rMJ1236 originating from independent expression and purification experiments differed in strength of their cleavage activity and addition of extra purification steps correlated with gradual loss of rMJ1236 activity. We speculate that perhaps native folding of MJ1236, and thus the maximum activity of this factor, can only be achieved under certain physical conditions resembling those of *M. jannaschii*'s natural inhabitat (e.g. elevated temperature and pressure, salts). It is also possible that rMJ1236 is sensitive to repeated freeze-thaw cycles; those were however unavoidable in the course of several purification steps. It has been reported before that extensive purification of enzymes may result in loss of their activity. Possibly, purification of rMJ1236 could remove unknown cofactors (e.g. Zn^{2+} or other metal ions) that are essential for its endonucleolytic activity *in vitro*. Optimal (native) requirements for rMJ1236 activity are not known. Our experiments with various *in vitro* cleavage conditions did not unravel any specific requirement for ion(s)/salt(s) that would be essential for the rMJ1236 activity (data not shown).

Mandel et al., 2006, presented an *in vitro* evidence for the endonucleolytic activity of recombinant CPSF73, the mammalian homologue of Ysh1p. This was surprising, as from our experience recombinant CPSF73 alone was unable to bind to RNA substrates and did not show any nucleolytic activity when assayed *in vitro* (S. Dettwiler, B. Dichtl and W. Keller, unpublished observations; see also: Chapter 2.3, Discussion). Consistently, recombinant Ysh1p alone did not show any endonucleolytic activity either. Furthermore, tethering of Ysh1p to RNA via the archaean KH-RNA binding domain did not restore Ysh1p endonucleolytic activity. It is possible that the structural conformation of the hybrid protein KH-Ysh1p hinders the nucleolytic activity of this factor. It is also possible that the C-terminus of Ysh1p, i.e. the part of the protein that is not evolutionarily conserved (neither with the higher eukaryotes, nor with archaea) conformationally prevents the endonucleolytic activity of the recombinant KH-Ysh1p *in vitro*. However, removal of

the Ysh1p C-terminus from the KH-Ysh1 hybrid protein did not result in any detectable nucleolytic activity (data not shown).

The three *M. jannaschii* genes, *MJ0047*, *MJ0162*, and *MJ1236*, encoding for putative endonucleases, are apparently paralogs. They could have possibly evolved from a common ancestor via gene duplications. Presence of Ysh1p/CPSF73-related genes in *Eukaryota, Archaebacteria*, and also *Eubacteria*, is consistent with a very ancient origin of that subfamily of enzymes within the metallo-β-lactamase superfamily (Aravind, 1999). Interestingly, our comparative sequence analysis of *M. jannaschii* genome revealed eight potential open reading frames having homology to known eukaryotic nucleotidyltransferases (data not shown). Currently, the function of these proteins in Archaea is not known. It would be interesting to determine experimentally, whether similar 3' end processing mechanisms operate in Archea and Eukaryotes. Based on the *in silico* data, one might speculate that archaeans have a 3' end cleavage and polyadenylation system that cleaves RNA transcripts and a nucleotidyltransferase [poly(A) polymerase] that adds the adenylates.

The *in vivo* substrate specificity of the three archaean homologues of Ysh1p is unknown. In a naïve consideration, one could speculate that in *M. jannaschii* the MJ0047 and MJ0162 proteins - as devoid of any RNA binding properties - are a part of a larger 3' end processing complex, in which function of RNA sequence binding and recognition is delegated to some other component(s) of the complex. In contrast, MJ1236 could act as a "freelancer" endonuclease that is sequence-unspecific (as it carries a sequence-independent KH-RNA binding domain) and perhaps gains specificity only by interactions with some hypothetical regulating cofactors. It could also be possible that the primary *in vivo* function of MJ1236 is of a degradation-type. These are however only hypothetical speculations, based on *in silico* genomic analyses.

Virtually, it is almost impossible to perform biochemical, genetic or physiological studies *in vivo* on *M. jannaschii* cells, due to its extreme growth conditions (temperature 85-95ºC, pressure 260 atm). Nevertheless, archaean homologues of eukaryotic proteins possess a powerful means to study their function and structural properties *in vitro*. Archaean homologues are usually 'simplified' in structure, and may be thermostable. Moreover, those proteins/complexes usually crystalize better that their eukaryotic counterparts, thus allowing insights into their structure and functions. For instance, the first crystal structure of the exosome ever obtained was from the archaeon *Sulfolobus solfataricus* (Lorentzen et al., 2005). High sequence similarity of this archaeal exosome to eukaryotic exosomes supported a common basis for RNA-degrading machineries in both domains of life.

The identity of the endonuclease acting in 3' end formation of mRNAs remained elusive for more than 20 years. The work presented here successfully identified an endonucleolytic activity within the archaean homologue of Ysh1p/CPSF73, thus strongly pointing towards the orthologues of MJ1236 in the eukaryotic kingdom as also being endoribonucleases.

4.4 Experimental procedures

Plasmids and primers

Plasmids and oligonucleotides used in this study are listed in Tables 4.1 and 4.2, respectively. *MJ1236* gene was amplified by PCR from genomic DNA of *M. jannaschii*, using primers: upstream MJ1236 5'NdeI, and downstream MJ1236 3' BamHI containing sequences encoding for the His6 tag. The PCR product was subsequently inserted into the pGDV1 vector, using the NdeI and BamHI restriction sites, to form the expression construct pMG34, from which *MJ1236* was expressed as the GST-MJ1236-His6 fusion protein. Plasmid pMG25 encodes for the HHD mutant version of MJ1236, and was constructed using the two-step mutagenesis method (Mikaelian and Sergeant, 1992), with oligonucleotides MJmut-HHD and MJmut-down, on pMG34 as backbone. The KH RNA binding domain (1.-552. nt of the MJ1236 gene) was PCR-amplified from pMG34, using primers KH 5' NdeI and KH 3' H6 XbaI. PCR product was subsequently cloned into pGDV1, using sites NdeI and XbaI, to form the pMG14 construct. Hybrid construct pKH-Ysh1p (pMG21) was created as the following: the 40.-2340. bp fragment of the *YSH1* gene sequence was amplified from the pBD73 plasmid, using primers Ysh1-5'XbaI and Ysh1-3' His6-BamH1. The resulting PCR product was inserted into pMG14 plasmid, using XbaI and BamHI restriction sites, thus creating the pMG21 construct (for expression of the GST-KH-Ysh1-His6 fusion protein). Cloning, amplifications, and ligation procedures were performed using standard methods of molecular biology. All sequences were confirmed by DNA sequencing.

Table 4.1 Plasmids used in this study

Plasmid	Description
pMG34	GST-MJ1236-His6, in pGDV1 vector, NdeI/BamHI
pMG25	GST-MJ1236 HHD mutant-His6, in pGDV1 vector, NdeI/BamHI
pMG14	GST-KH-His6 (KH-domain: 1.-552. bp fragment of the *MJ1236* gene), in pGDV1 vector, NdeI/XbaI
pMG21	GST-KH-Ysh1p-His6 (KH-domain: 1.-552. bp fragment of the MJ1236 gene; Ysh1: 40.-2340. bp of the YSH1 gene), in pGDV1 vector, NdeI /XbaI /BamHI
pBD73	GST-YSH1-His6, in pGDV1 vector, NdeI/BamHI

Table 4.2 Oligonucleotides used in this study

Oligonucleotide	Sequence
MJ1236 5'NdeI	GGAATTCCATATGTCAGCAGAGGAAGTTTTAG
MJ1236 3' BamHI	CGGGATCCCGTTACTTAACCCTTATAGCATCCAAG
MJmut-HHD	GCTGTTATAGTTACTGCTGCTGCCTTAGCTCACTGTGGTTTTATTCCC
KH-5'NdeI	GGAATTCCATATGATTAAAAAAGCAATAAG
KH 3' H6 XbaI	CGGGATCCTTAATGATGATGATGATGATGCCAATAATCTCCTCTAAC AACTATATCTC
Ysh1-5'XbaI	TGCTCTAGATTTTTTTCATTGGGAGGAAGTAACGAAG
Ysh1-3'His6 BamH1	CGGGATCCTTAATGATGATGATGATGATGACATAGCGGTGTGACCA

Computational analysis

Sequences have been found by BLAST searches on the NCBI-Genbank Website (http://www.ncbi.nlm.nih.gov/BLAST/). Protein sequences have been aligned with ClustalX (Jeanmougin et al, 1998) or T-Coffee (Notredame et al., 2000), and refined manually.

Expression and purification of recombinant proteins

E. coli BL21(DE3) strain (Studier, 1991) or Rosetta-gammi strain (Novagen) were transformed with the respective plasmids pMG34 (GST-MJ1236-His6), pMG25 (GST-MJ1236-HHDmut-His6), pMG14 (GST-KH-His6) or pMG21 (GST-KH-Ysh1-His6) and grown in 2xYT medium containing 100 µg/ml ampicilin and 50 µM $ZnCl_2$ [BL21(DE3) strain], or additionally with 10 µg/µl tetracyclin and 15 µg/µl kanamycin if the Rosetta-gammi strain was used. Strains were grown at 37ºC to an OD_{600} of 0.6 - 0.8. Protein expression was induced by addition of isopropyl-D-thiogalctoside (IPTG) to final concentration of 0.5 mM. After 3-5 h of expression at 37ºC cells were harvested and lysed. Typically, fusion proteins were first purified by affinity chromatography on Ni-NTA (nickel-nitrilotriacetic acid agarose; Qiagen), followed by affinity chromatography on glutathione Sepharose 4B (Pharmacia), according to the manufacturers' manuals. An additional step of purification was applied to GST-MJ1236-His6 and GST-MJ1236-HHDmut-His6 proteins, which was the affinity chromatography fractionation over the ion-exchange MonoS column, using the 300-500mM KCl gradient in 10 mM Hepes/0.2 mM EDTA butter. Protein fractions were then analyzed using silver-stained SDS-PAGE gels. Recombinant Rna15p was a gift from M. Sadowski.

Cleavage assays *in vitro*

Cleavage assays with rMJ1236 protein were performed similarly to the classical 3' end pre-mRNA cleavage assays (Minvielle-Sebastia et al., 1994; see also: Chapter 2.4). Typical cleavage reaction mix contained 2% PEG-800, 75 mM potassium glutamate, 0,02% NP40, 20 mM creatine

P, 2 mM DTT, 2mM EDTA, 250 µg/ml creatine kinase, 2U RNA Guard (Amersham Biosciences), and 20-100 fmol of the respective ^{32}P-labelled RNA substrate. Samples were typically incubated at 37ºC for 1 h, unless otherwise indicated. Internally ^{32}P-labelled sCYC1, CYC1, sADH1, and sGAL7 RNA substrates were produced by the *in vitro* run-off transcription using [α-^{32}P]UTP. Additionally, sCYC1 RNA produced by the cold *in vitro* run-off transcription was 5' end labelled using [γ-^{32}P]ATP and T4 PNK (New England Biolabs), or 3' end labelled using [α-^{32}P]cordycepin-TP and recombinant yeast Pap1p enzyme.

Electro-mobility shift assay (EMSA)

Tested proteins were assayed following the protocol of the cleavage reaction assays *in vitro*, incubated 10' at 4ºC, and loaded onto a pre-run 10% native polyacrylamide gel (0.5x TBE). Gel electrophoresis was performed for approx. 3 h at 4ºC and results of EMSA were visualized by autoradiography.

Conclusions and outlook

The synthesis of different RNAs by RNA polymerase II is a complex process. In order to be biologically functional, RNA transcripts require extensive processing, including capping, splicing and 3' end formation. All the processing events are interconnected and tightly coupled to transcription.

The main focus of the work presented here was to analyze the functions of yeast 3' end processing factor Ysh1p/Brr5p in 3' end formation of RNAP II transcripts and its influence on other aspects of RNA synthesis, such as transcription termination, regulation of mRNA levels and splicing. Ysh1p most likely is the endonuclease acting primarily in the 3' end processing of pre-mRNA transcripts in yeast. The protein sequence of Ysh1p contains a conserved β-lactamase motif which is involved in the catalysis and essential for cell viability. In Chapter 2 of this thesis, we presented direct evidence for Ysh1p being essential for the 3' end cleavage and polyadenylation of pre-mRNAs, as well as for the correct poly(A) site selection. Moreover, we showed that Ysh1p is involved in RNAP II transcription termination on mRNA genes. Intriguingly, we found that transcription termination and 3' end formation on snoRNA genes as well as autoregulation of NRD1 mRNA levels are impaired in the *ysh1-12* mutant allele, pointing to involvement of Ysh1p in these processes. Furthermore, the *ysh1-12* strain is defective in splicing, implying a function for Ysh1p in coupling of this event to the 3' end formation in *S. cerevisiae*.

The *S. cerevisiae* genome encodes for *SYC1* gene, the C-terminal homologue of Ysh1p; therefore we analyzed functions of the Syc1p protein in pre-mRNA 3' end formation. Similarly to Ysh1p, Syc1p is an integral component of the canonical 3' end processing machinery. Although this factor is not essential for the 3' end processing steps, its absence exacerbates growth defects of other 3' end processing mutants. Syc1p has possible regulatory functions in pre-mRNA 3' end formation and through several genetic interactions links the processing machinery to other nuclear events (Chapter 3).

The identity of the endonuclease acting in 3' end formation of mRNAs remained a puzzle for many years. The work presented in Chapter 4 of this thesis identified an endonucleolytic activity within the archaean homologue of Ysh1p/CPSF73, the *M. jannaschii* gene *MJ1236*, therefore strongly pointing to the fact that the orthologues of the MJ1236 protein in the eukaryotic kingdom are also endoribonucleases.

Our findings on the functions of the 3' end processing endonuclease Ysh1p provoke several further questions to be addressed in the near future. An important one is how exactly the 3' ends of the snoRNA and snRNA transcripts are made. Little is known about the exact mechanism(s) of this process, and of the myriad of protein factors apparently involved in it. Is the canonical 3' end processing machinery a direct prerequisite for the correct snoRNA 3' end formation and transcription termination on these genes? Moreover, can the two stages of the snoRNA synthesis: transcription termination and 3' end formation be functionally dissected from each other? Last but not least, what is the functional significance of the poly(A)-like cleavage sites in the 3' end downstream terminator regions of snoRNA genes, and is the endonucleolytic

cleavage at these sites mandatory for snoRNA 3' end formation? Future experiments will aim to address these questions.

Another important issue for future experiments is to reveal the general mechanism by which the transcribing RNAP II complex undergoes premature termination (abortive elongation) at the 3' ends of the genes. Currently, there is no direct evidence of poly(A) factors being involved in abortive elongation; however, the involvement of 3' end processing factors in this process is highly anticipated. Our results suggest that the activity of Ysh1p is required for premature termination of *NRD1* transcripts. It is possible that Ysh1p suppresses the transcriptional read-through at this gene. Further experiments are required to reveal a more detailed view of the exact mechanism by which the mRNA transcripts are terminated prematurely. It is possible that the 3' end processing-dependent regulation of gene expression by transcription attenuation may be more common than currently expected.

Moreover, our findings on Ysh1p being required for splicing imply a coupling mechanism between 3' end formation and splicing. These processing events were reported to be coupled in mammalian (Kyburz et al., 2006; Millevoi et al., 2006) but not in yeast cells. Further investigations of the underlying molecular basis in yeast are required. They might include screening of the pre-mRNA 3' end processing mutants for defects in splicing and vice versa, screening of the splicing mutants for defects in pre-mRNA 3' end processing. It will be also interesting to define whether that coupling mechanism is evolutionary conserved.

Both the Ysh1p and Syc1p components of CPF, and additionally Clp1p, the component of CF IA, carry a conserved P-loop motif commonly found in proteins that bind ATP or GTP. This suggests a putative function for those factors in binding and perhaps hydrolyzing ATP/GTP. However, mutations in the P-loop motif of Ysh1p have no effect on cell viability (Zhelkovsky et al., 2006), ruling out the possibility that Ysh1p binds or hydrolyzes ATP. Interestingly, the human homologue of Clp1p (hClp1) exhibits a kinase activity *in vitro*. hClp1 was found to associate with the human tRNA splicing endonuclease complex (Paushkin et al., 2004), and to phosphorylate the 5' end of the 3' exon during human tRNA splicing, allowing the subsequent ligation of both exon halves by an unknown tRNA ligase (Weitzer and Martinez, 2007). Further investigations will be required to unravel the putative substrate(s) of hClp1 and its yeast homologue in the 3' end processing events of various RNA transcripts. Another interesting question to be addressed in the future is how the two processes, 3' end maturation of tRNAs and mRNAs, are interconnected in mammalian cells.

References

Albers, M., Diment, A., Muraru, M., Russell, C.S. and Beggs, J.D. (2003) Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. *RNA*, **9**, 138-150.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. and Tollervey, D. (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J*, **18**, 5399-5410.

Allmang, C., Mitchell, P., Petfalski, E. and Tollervey, D. (2000) Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res*, **28**, 1684-1691.

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

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

Anantharaman, V., Koonin, E.V. and Aravind, L. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res*, **30**, 1427-1464.

Anderson, J.T., Wilson, S.M., Datar, K.V. and Swanson, M.S. (1993) NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Mol Cell Biol*, **13**, 2730-2741.

Aranda, A., Pérez-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.

Arigo, J.T., Carroll, K.L., Ames, J.M. and Corden, J.L. (2006) Regulation of yeast NRD1 expression by premature transcription termination. *Mol Cell*, **21**, 641-651.

Badis, G., Saveanu, C., Fromont-Racine, M. and Jacquier, A. (2004) Targeted mRNA degradation by deadenylation-independent decapping. *Mol Cell*, **15**, 5-15.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N. and Shiekhattar, R. (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*, **123**, 265-276.

Ballarino, M., Morlando, M., Pagano, F., Fatica, A. and Bozzoni, I. (2005) The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **25**, 5396-5403.

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

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

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

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

Baron-Benhamou, J., Gehring, N.H., Kulozik, A.E. and Hentze, M.W. (2004) Using the lambdaN peptide to tether proteins to RNAs. *Methods Mol Biol*, **257**, 135-154.

Bauren, G., Belikov, S. and Wieslander, L. (1998) Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron. *Genes Dev*, **12**, 2759-2769.

Beltrame, M. and Tollervey, D. (1992) Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J*, **11**, 1531-1542.

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

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.

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

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W. and Proudfoot, N.J. (1998) Cleavage of the primary transcript couples 3'-end RNA-processing with termination of pol II transcription. *Science*, **280**, 298-301.

Brune, C., Munchel, S.E., Fischer, N., Podtelejnikov, A.V. and Weis, K. (2005) Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. *RNA*, **11**, 517-531.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S. and Venter, J.C. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058-1073.

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

Butler, J.S. (2002) The yin and yang of the exosome. *Trends Cell Biol*, **12**, 90-96.

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

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

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

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

Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D. and Corden, J.L. (2004) Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol Cell Biol*, **24**, 6241-6252.

Chanfreau, G., Legrain, P. and Jacquier, A. (1998a) Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J Mol Biol*, **284**, 975-988.

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.

Chanfreau, G., Rotondo, G., Legrain, P. and Jacquier, A. (1998b) Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J*, **17**, 3726-3737.

Chang, Y.W., Howard, S.C. and Herman, P.K. (2004) The Ras/PKA signaling pathway directly targets the Srb9 protein, a component of the general RNA polymerase II transcription apparatus. *Mol Cell*, **15**, 107-116.

Chapon, C., Cech, T.R. and Zaug, A.J. (1997) Polyadenylation of telomerase RNA in budding yeast. *RNA*, **3**, 1337-1351.

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

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.

Chodchoy, N., Levine, B.J., Sprecher, C., Skoultchi, A.I. and Marzluff, W.F. (1987) Expression of mouse histone genes: transcription into 3' intergenic DNA and cryptic processing sites downstream from the 3' end of the H3 gene. *Mol Cell Biol*, **7**, 1039-1047.

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

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

Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S. and Corden, J.L. (2000) A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics*, **154**, 557-571.

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.

Daiyasu, H., Osaka, K., Ishino, Y. and Toh, H. (2001) Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. *FEBS Lett*, **503**, 1-6.

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

Davis, C.A. and Ares, M., Jr. (2006) Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **103**, 3262-3267.

Decourty L, Saveanu C, Zemam K, Hantraye F, Frachon E, Rousselle JC, Fromont-Racine M, Jacquier A. 2008. Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles. *Proc Natl Acad Sci U S A 105*:5821-5826.

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

del Olmo, M., Mizrahi, N., Gross, S. and Moore, C.L. (1997) The Uba2 and Ufd1 proteins of Saccharomyces cerevisiae interact with poly(A) polymerase and affect the polyadenylation activity of cell extracts. *Mol Gen Genet*, **255**, 209-218.

Dheur, S., Nykamp, K.R., Viphakone, N., Swanson, M.S. and Minvielle-Sebastia, L. (2005) Yeast mRNA poly(A) tail length control can be reconstituted in vitro in the absence of Pab1p-dependent poly(A) nuclease activity. *J Biol Chem*.

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

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

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.

Dunn, E.F., Hammell, C.M., Hodge, C.A. and Cole, C.N. (2005) Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. *Genes Dev*, **19**, 90-103.

Dye, M.J. and Proudfoot, N.J. (1999) Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol Cell*, **3**, 371-378.

Embley, T.M. and Martin, W. (2006) Eukaryotic evolution, changes and challenges. *Nature*, **440**, 623-630.

Fatica, A., Morlando, M. and Bozzoni, I. (2000) Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *EMBO J*, **19**, 6218-6229.

Filipowicz, W. and Pogacic, V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol*, **14**, 319-327.

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

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.

Ganot, P., Jady, B.E., Bortolin, M.L., Darzacq, X. and Kiss, T. (1999) Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol*, **19**, 6906-6917.

Garrett, R.A. (1996) Genomes: *Methanococcus jannaschii* and the golden fleece. *Curr Biol*, **6**, 1377-1380.

Garrity, J.D., Bennett, B. and Crowder, M.W. (2005) Direct evidence that the reaction intermediate of metallo-beta-lactamase L1 is metal bound. *Biochemistry*, **44**, 1078-1087.

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

Gilbert, W. (1986) The RNA world. *Nature*, **319**, 618.

Gilmour DS, Fan R. 2008. Derailing the locomotive: transcription termination. *J Biol Chem 283*:661-664.

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

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

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

Graber, J.H., McAllister, G.D. and Smith, T.F. (2002) Probabilistic prediction of Saccharomyces cerevisiae mRNA 3'-processing sites. *Nucleic Acids Res*, **30**, 1851-1858.

Green, D.M., Marfatia, K.A., Crafton, E.B., Zhang, X., Cheng, X. and Corbett, A.H. (2002) Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p. *J Biol Chem*, **277**, 7752-7760.

Greger, I.H. and Proudfoot, N.J. (1998) Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of *Saccharomyces cerevisiae*. *EMBO J*, **17**, 4771-4779.

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

Guo, Z. and Sherman, F. (1996) Signals sufficient for 3'-end formation of yeast mRNA. *Mol Cell Biol*, **16**, 2772-2776.

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.

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

Heidmann, S., Obermeier, B., Vogel, K. and Domdey, H. (1992) Identification of pre-mRNA polyadenylation sites in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **12**, 4215-4229.

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

Henry, Y., Wood, H., Morrissey, J.P., Petfalski, E., Kearsey, S. and Tollervey, D. (1994) The 5' end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *EMBO J*, **13**, 2452-2463.

Hernandez, N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem*, **276**, 26733-26736.

Higgs, D.R., Goodbourn, S.E., Lamb, J., Clegg, J.B., Weatherall, D.J. and Proudfoot, N.J. (1983) Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature*, **306**, 398-400.

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

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

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

Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C. and Tollervey, D. (2006) Microarray detection of novel nuclear RNA substrates for the exosome. *Yeast*, **23**, 439-454.

Huang, Y. and Steitz, J.A. (2001) Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell*, **7**, 899-905.

Huang, Y., Wimler, K.M. and Carmichael, G.G. (1999) Intronless mRNA transport elements may affect multiple steps of pre-mRNA processing. *EMBO J*, **18**, 1642-1652.

Hughes, J.M. and Ares, M., Jr. (1991) Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J*, **10**, 4231-4239.

Humphrey, T., Birse, C.E. and Proudfoot, N.J. (1994) RNA 3' end signals of the S.pombe ura4 gene comprise a site determining and efficiency element. *EMBO J*, **13**, 2441-2451.

Irniger, S., Egli, C.M. and Braus, G.H. (1991) Different classes of polyadenylation sites in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol*, **11**, 3060-3069.

Ito, K., Sato, K. and Endo, H. (1994) Cloning and characterization of a single-stranded DNA binding protein that specifically recognizes deoxycytidine stretch. *Nucleic Acids Res*, **22**, 53-58.

Iwase, M. and Toh-e, A. (2001) Nis1 encoded by YNL078W: a new neck protein of Saccharomyces cerevisiae. *Genes Genet Syst*, **76**, 335-343.

Jady, B.E. and Kiss, T. (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J*, **20**, 541-551.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. and Gibson, T.J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci*, **23**, 403-405.

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

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

Jones, J.C., Phatnani, H.P., Haystead, T.A., MacDonald, J.A., Alam, S.M. and Greenleaf, A.L. (2004) C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. *J Biol Chem*, **279**, 24957-24964.

Jones, W.J., Leigh, J.A., Mayer, F.W., R., W.C. and Wolfe, R.S. (1983) *Methanococcus jannaschii*, an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol*, **136**, 254-261.

Kaneko S, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2007. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. *Genes Dev 21*:1779-1789.

Kawashima, S., Ogiwara, H., Tada, S., Harata, M., Wintersberger, U., Enomoto, T. and Seki, M. (2007) The INO80 complex is required for damage-induced recombination. *Biochem Biophys Res Commun*, **355**; 835-41.

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

Kessler, M.M., 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. *J Biol Chem*, **271**, 27167-27175.

Kim, H.D., Choe, J. and Seo, Y.S. (1999) The sen1(+) gene of *Schizosaccharomyces pombe*, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. *Biochemistry*, **38**, 14697-14710.

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

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

Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C. and Buratowski, S. (2006) Distinct pathways for snoRNA and mRNA termination. *Mol Cell*, **24**, 723-734.

Kim, S.J. and Martinson, H.G. (2003) Poly(A)-dependent transcription termination: continued communication of the poly(A) signal with the polymerase is required long after extrusion in vivo. *J Biol Chem*, **278**, 41691-41701.

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

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

Kressler, D., Linder, P. and de La Cruz, J. (1999) Protein trans-acting factors involved in ribosome biogenesis in Saccharomyces cerevisiae. *Mol Cell Biol*, **19**, 7897-7912.

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.

Kuai, L., Fang, F., Butler, J.S. and Sherman, F. (2004) Polyadenylation of rRNA in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, **101**, 8581-8586.

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

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.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.L. and Lafyatis, R. (2002) Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev*, **16**, 594-607.

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.

Lin, C.L., Bristol, L.A., Jin, L., Dykes-Hoberg, M., Crawford, T., Clawson, L. and Rothstein, J.D. (1998) Aberrant RNA processing in a neurodegenerative disease: the cause for absent EAAT2, a glutamate transporter, in amyotrophic lateral sclerosis. *Neuron*, **20**, 589-602.

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

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

Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G. and Conti, E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat Struct Mol Biol*, **12**, 575-581.

Luo, M.J. and Reed, R. (1999) Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci U S A*, **96**, 14937-14942.

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

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

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

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

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

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

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

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

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

Marzluff, W.F. (1992) Histone 3' ends: essential and regulatory functions. *Gene Expr*, **2**, 93-97.

Maxwell, E.S. and Fournier, M.J. (1995) The small nucleolar RNAs. *Annu Rev Biochem*, **64**, 897-934.

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

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

Medlin, J.E., Uguen, P., Taylor, A., Bentley, D.L. and Murphy, S. (2003) The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO J*, **22**, 925-934.

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

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

Mikaelian, I. and Sergeant, A. (1992) A general and fast method to generate multiple site directed mutations. *Nucleic Acids Res*, **20**, 376.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M. and Vagner, S. (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *EMBO J*, **25**, 4854-4864.

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

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

Minvielle-Sebastia, L., Petitjean, A., Winsor, B., Bonneaud, N. and Lacroute, F. (1990) Mutations involved in mRNA stability and in the length of their poly(A) tails in the yeast *Saccharomyces cerevisiae*. In McCarthy, J.E.G. and Tuite, M.F. (eds.), *Post-Transcriptional Control of Gene Expression*. Springer Verlag, Berlin, Vol. 49, pp. 55-64.

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 pre-messenger 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.

Moir D, Stewart SE, Osmond BC, Botstein D. 1982. Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies. *Genetics 100*:547-563.

Moore, M.J. and Query, C.C. (2000) Joining of RNAs by splinted ligation. *Methods Enzymol*, **317**, 109-123.

Moore, M.J. and Sharp, P.A. (1992) Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science*, **256**, 992-997.

Morlando, M., Greco, P., Dichtl, B., Fatica, A., Keller, W. and Bozzoni, I. (2002) Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation. *Mol Cell Biol*, **22**, 1379-1389.

Morlando, M., Ballarino, M., Greco, P., Caffarelli, E., Dichtl, B. and Bozzoni, I. (2004) Coupling between snoRNP assembly and 3' processing controls box C/D snoRNA biosynthesis in yeast. *EMBO J*, **23**, 2392-2401.

Murthy, K.G.K. 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.

Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M.A., Nilges, M., Gibson, T.J. and Pastore, A. (1996) Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell*, **85**, 237-245.

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.

Nedea E, Nalbant D, Xia D, Theoharis NT, Suter B, Richardson CJ, Tatchell K, Kislinger T, Greenblatt JF, Nagy PL. 2008. The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol Cell 29*:577-587.

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.

Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, **302**, 205-217.

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

O'Neill, B.M., Hanway, D., Winzeler, E.A. and Romesberg, F.E. (2004) Coordinated functions of WSS1, PSY2 and TOF1 in the DNA damage response. *Nucleic Acids Res*, **32**, 6519-6530.

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.

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

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

Pace, N.R. (2006) Time for a change. *Nature*, **441**, 289.

Park, N.J., Tsao, D.C. and Martinson, H.G. (2004) The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain. *Mol Cell Biol*, **24**, 4092-4103.

Patel, D. and Butler, J.S. (1992) Conditional defect in mRNA 3' end processing caused by a mutation in the gene for poly(A) polymerase. *Mol Cell Biol*, **12**, 3297-3304.

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

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

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

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

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

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

Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., Chemama, Y., Labigne, A. and Legrain, P. (2001) The protein-protein interaction map of Helicobacter pylori. *Nature*, **409**, 211-215.

Rasmussen, T.P. and Culbertson, M.R. (1998) The putative nucleic acid helicase Sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nucleolar RNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **18**, 6885-6896.

Raymond, A.C., Staker, B.L. and Burgin, A.B., Jr. (2005) Substrate specificity of tyrosyl-DNA phosphodiesterase I (Tdp1). *J Biol Chem*, **280**, 22029-22035.

Raymond, A.C. and Burgin, A.B., Jr. (2006) Tyrosyl-DNA phosphodiesterase (Tdp1) (3'-phosphotyrosyl DNA phosphodiesterase). *Methods Enzymol*, **409**, 511-524.

Reed, R. (2003) Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol*, **15**, 326-331.

Reed, R. and Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell*, **108**, 523-531.

Rinke-Appel, J., Junke, N., Stade, K. and Brimacombe, R. (1991) The path of mRNA through the Escherichia coli ribosome; site-directed cross-linking of mRNA analogues carrying a photo-reactive label at various points 3' to the decoding site. *EMBO J*, **10**, 2195-2202.

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.

Roth, K.M., Wolf, M.K., Rossi, M. and Butler, J.S. (2005) The nuclear exosome contributes to autogenous control of NAB2 mRNA levels. *Mol Cell Biol*, **25**, 1577-1585.

Rosonina E, Kaneko S, Manley JL. 2006. Terminating the transcript: breaking up is hard to do. *Genes Dev 20*:1050-1056.

Rothnie, H.M. (1996) Plant mRNA 3'-end formation. *Plant Mol Biol*, **32**, 43-61.

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

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.

Russnak, R., Pereira, S. and Platt, T. (1996) RNA binding analysis of yeast REF2 and its two-hybrid interaction with a new gene product, FIR1. *Gene Expr*, **6**, 241-258.

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.

Sadowski, M., 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.

Schiffer, S., Rosch, S. and Marchfelder, A. (2002) Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. *EMBO J*, **21**: 2769-2777.

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.

Sheldon, K.E., Mauger, D.M. and Arndt, K.M. (2005) A requirement for the *Saccharomyces cerevisiae* Paf1 complex in snoRNA 3' end formation. *Mol Cell*, **20**, 225-236.

Shen, X., Ranallo, R., Choi, E. and Wu, C. (2003) Involvement of actin-related proteins in ATP-dependent chromatin remodeling. *Mol Cell*, **12**, 147-155.

Siomi, H., Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res*, **21**, 1193-1198.

Sontheimer, E.J. and Steitz, J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989-1996.

Stade, K., Rinke-Appel, J. and Brimacombe, R. (1989) Site-directed cross-linking of mRNA analogues to the Escherichia coli ribosome; identification of 30S ribosomal components

that can be cross-linked to the mRNA at various points 5' with respect to the decoding site. *Nucleic Acids Res*, **17**, 9889-9908.

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

Steinmetz, E.J. and Brow, D.A. (1998) Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. *Proc Natl Acad Sci U S A*, **95**, 6699-6704.

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.

Steinmetz, E.J., Ng, S.B., Cloute, J.P. and Brow, D.A. (2006a) cis- and trans-acting determinants of transcription termination by yeast RNA polymerase II. *Mol Cell Biol*, **26**, 2688-2696.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z. and Brow, D.A. (2006b) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell*, **24**, 735-746.

Studier, F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol*, **219**, 37-44.

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**, 1517-1519.

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., Manley, J.M., McDonald, 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.

Thomas, B.J. and Rothstein, R. (1989) Elevated recombination rates in transcriptionally active DNA. *Cell*, **56**, 619-630.

Ursic, D., Chinchilla, K., Finkel, J.S. and Culbertson, M.R. (2004) Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res*, **32**, 2441-2452.

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

van 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.

van Hoof, A., Lennertz, P. and Parker, R. (2000) Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol*, **20**, 441-452.

Vasiljeva, L. and Buratowski, S. (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell*, **21**, 239-248.

Venema, J., Henry, Y. and Tollervey, D. (1995) Two distinct recognition signals define the site of endonucleolytic cleavage at the 5'-end of yeast 18S rRNA. *EMBO J*, **14**, 4883-4892.

Venema, J. and Tollervey, D. (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet*, **33**, 261-311.

Vethantham V, Rao N, Manley JL. 2007. Sumoylation modulates the assembly and activity of the pre-mRNA 3' processing complex. *Mol Cell Biol 27*:8848-8858.

Vethantham V, Rao N, Manley JL. 2008. Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. *Genes Dev 22*:499-511.

Vinciguerra, P. and Stutz, F. (2004) mRNA export: an assembly line from genes to nuclear pores. *Curr Opin Cell Biol*, **16**, 285-292.

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

Wahle, E., Martin, G., Schiltz, E. and Keller, W. (1991) Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase. *EMBO J*, **10**, 4251-4257.

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

Wang, J.C. (1991) DNA topoisomerases: why so many? *J Biol Chem*, **266**, 6659-6662.

Wang, Y., Wagner, J.D. and Guthrie, C. (1998) The DEAH-box splicing factor Prp16 unwinds RNA duplexes in vitro. *Curr Biol*, **8**, 441-451.

Weiner, A.M. (2005) E Pluribus Unum: 3' end formation of polyadenylated mRNAs, histone mRNAs, and U snRNAs. *Mol Cell*, **20**, 168-170.

Weitzer S, Martinez J. 2007. The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs. *Nature 447*:222-226.

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

West S, Proudfoot NJ, Dye MJ. 2008. Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol Cell 29*:600-610.

Wickens M, Gonzalez TN. 2004. Molecular biology. Knives, accomplices, and RNA. *Science 306*:1299-1300.

Winston, F., Dollard, C. and Ricupero-Hovasse, S.L. (1995) Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast*, **11**, 53-55.

Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C. (1991) Specific interaction between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol*, **40**, 185-220.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*, **74**, 5088-5090.

Wyatt, J.R., Sontheimer, E.J. and Steitz, J.A. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev*, **6**, 2542-2553.

Young KH. 1998. Yeast two-hybrid: so many interactions, (in) so little time. *Biol Reprod 58*:302-311.

Yu, Y.T. (1999) Construction of 4-thiouridine site-specifically substituted RNAs for cross-linking studies. *Methods*, **18**, 13-21.

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

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.

Zhang Z, Gilmour DS. 2006. Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell 21*:65-74.

Zhao, J., Hyman, L. and Moore, C. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev*, **63**, 405-445.

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

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

APPENDIX

Abbreviations

APT: associated with Pta1p complex

β-CASP: metallo-β-lactamase-associated CPSF Artemis SNM1/PSO2 motif

CF IA: cleavage factor IA

CF IB: cleavage factor IB

CF I_m: cleavage factor I_m

CF II_m: cleavage factor II_m

ChIP: chromatin immunoprecipitation

CID: CTD interaction domain

CPF: cleavage and polyadenylation factor

CPSF: cleavage and polyadenylation specificity factor

CstF: cleavage stimulatory factor

CTD: C-terminal domain of the largest subunit (Rpb1p) of RNA polymerase II

DSE: downstream element

EE: efficiency element

kDa: kiloDalton

M. jannaschii: *Methanococcus jannaschii*

mRNA: messenger RNA

nt: nucleotides

ORF: open reading frame

PCR: polymerase chain reaction

PAP: poly(A) polymerase

PE: positioning element

PF I: polyadenylation factor I

pre-mRNA: pre-messenger RNA or primary transcript

protA; protein A

RBD: RNA binding domain

RNA: ribonucleic acid

rRNA: ribosomal RNA

RNAP II: RNA polymerase II

RRM: RNA recognition motif

S2: CTD serine 2

S2-P: phosphorylated CTD serine 2

S5: CTD serine 5

S5-P: phosphorylated CTD serine 5

S. cerevisiae: *Saccharomyces cerevisiae*

SDS PAGE: sodium dodecyl sulphate polyacrylamide gel electroforesis

SL:	splinted ligation
snoRNA:	small nucleolar RNA
snRNA:	small nuclear RNA
snRNPs:	small nuclear ribonucleoprotein
UTR:	untranslated region
YPD:	yeast peptone dextrose

Curriculum vitae

Name	Monika Garas
Date of birth	August 10, 1976
Citizenship	Polish
Address	Ryffstrasse 22, 4056 Basel, Switzerland

Education and professional experience

June 2007-Sept. 2008	Postdoctoral Research Fellow, Department of Cell Biology, Biozentrum, University of Basel, Basel, Switzerland
Jan. 2002-May 2007	Ph.D. Studies in Cell Biology, Biozentrum, University of Basel, Basel, Switzerland Ph.D. Thesis: "Functional Characterization of Ysh1p, the Yeast Endonuclease Involved in 3' end Processing and in Transcription Termination of RNA Polymerase II Transcripts" (performed under the supervision of Prof. Walter Keller)
Sept. 2000-Nov. 2001	Reseach Assistant, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland
Sept. 1999-June 2000	MSc Thesis performed in the Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland
Oct. 1995-June 2000	MSc Graduate Studies in Biotechnology (First Honours), University of Marie Curie-Sklodowska, Lublin, Poland

Publications

M. Garas, B. Dichtl and W. Keller, "The role of the putative 3' end endonuclease Ysh1p in mRNA and snoRNA synthesis"; The RNA Journal, 2008. Manuscript accepted.

Scientific communications

2008 EuroDyna Network Meeting. Zurich, Switzerland.
Talk: "The role of Ysh1p in 3' end processing and for transcription termination of RNA polymerase II transcripts."

Biozentrum Annual Symposium. Basel, Switzerland.
Poster: "Ysh1p/Brr5p is required for 3' end processing and in transcription termination of RNA polymerase II transcripts."

2006 Swiss RNA Meeting. Bern, Switzerland.
Talk: "The role of Ysh1p in 3' end processing and in transcription termination of RNA polymerase II transcripts."

2005 Messenger RNA 3' Ends: Interconnections with Transcription and mRNA Turnover Conference. University of Oxford, United Kingdom.
Poster: "Functions for *S.cerevisiae* Ysh1p/Brr5p and its C-terminal homologue Syc1p in 3'end processing of RNA polymerase II transcripts".

2005 European RNOMICS Network Meeting. University of Lisbon, Portugal.
Talk: "Potential roles of yeast CPF subunits Ysh1p and Syc1p in pre mRNA 3' end processing and RNA polymerase II transcription termination."

2004 Swiss RNA Meeting. Bern, Switzerland.
Poster: "Characterization of Syc1p, a subunit of the APT sub-complex within the yeast cleavage and polyadenylation factor CPF".

Biozentrum Annual Symposium. St. Chrischona, Switzerland.
Poster: "Characterization of Syc1p, a subunit of the APT sub-complex within the yeast cleavage and polyadenylation factor CPF".

2003 European RNOMICS Network Meeting. Wellcome Trust, University of Edinburgh, United Kingdom.
Talk: „Pre-mRNA 3' end processing machinery in yeast - about cleavage reaction and beyond".

2002 European RNOMICS Network Meeting. Institute Pasteur, Paris, France.
Talk: "Searching for the pre-mRNA 3' end processing endonuclease".

Teaching experience

Teaching Assistant. Courses in Cell Biology, Biozentrum, University of Basel, Basel, Switzerland, 2002-2007