Investigating the links between transcription and mRNA 3' end processing

A dissertation submitted for the degree of

MSc by Research

Department of Biochemistry

University of Oxford

Sum Po Ava Chan

Lincoln College

Trinity Term 2020

Abstract

Investigating the links between transcription and mRNA 3' end processing

Sum Po Ava Chan, Lincoln College, University of Oxford

Msc by Research in Biochemistry

Trinity Term 2020

RNA Polymerase II (Pol II) is responsible for transcribing all protein coding genes and some non-coding genes. Protein-coding mRNAs undergo extensive processing during Pol II transcription in order to become functional molecules. This includes 5'capping, splicing and 3'end cleavage and polyadenylation. mRNA processing is coordinated by the phosphorylation of the Pol II carboxy-terminal domain (CTD) and has been shown to be affected by transcription rate. However, the molecular mechanisms underpinning regulation of mRNA processing by transcription are not well understood.

Here, I aim to address this by establishing an *in vitro* system which allows to the study of processes linked to transcription such as mRNA 3' end processing. The work presented here contains detailed studies of Pol II Rpb1 mutants that show opposite effects in RNA processivity and transcription rate using an *in vitro* transcription system and state-of-the-art *in vivo* approaches such as PRO-seq (in collaboration with the postdoctoral visitor), studying how transcription rate contributes to the ability of Pol II to interact with factors and reconstitution of the 3' mRNA processing machinery that is involved in cleavage and polyadenylation.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Lidia for giving me an invaluable MSc by research experience. I enjoyed working on this project and Lidia has given me the opportunity to explore different areas of the project, which enhanced my knowledge, increased my skillset, and most importantly trained me to be an independent scientist. I believe that this MSc by research has provided me with a solid foundation to continue along the bumpy road of science.

I am grateful to all the people who have worked with me in the Vasilieva lab. I would like to give a special thanks to Krzysztof, for helping me in my early days, getting me started in the lab and his guidance and support along the way. I had the pleasure to closely work with him to reconstitute the 3' end processing machinery and understand the roles of Seb1. I also would like to thank Adrien, Dong-Hyuk and Tea for always being up for discussions about science and research. Furthermore, and Takuya were kind enough to provide me with his PRO-seq data and I feel very happy to have been working in a lab with a very friendly atmosphere between all lab members. Also, Emily, Katie and Doğukan for many nice chats during lunchtime or lab outings, I always enjoyed talking to you.

I would like to thank my thesis committee, Lars Jansen and Neil Brockdorff for their constructive comments and feedback to my project. My gratitude also goes to Lincoln College for providing me the opportunity to study here and the support at these critical times with COVID-19.

Many of the experiments would not have been possible without the help of people from outside the lab. This includes James and Menelaos from Nasmyth lab for helping me to set up the MultiBac system in lab. Furthermore, the whole third floor, predominantly the Brockdorff and Nasmyth labs, provided good company and I am sure that each of you have helped me with something over the years. The University of Oxford provided me with a very inspiring environment, with numerous of opportunities to learn about the cutting edge of science and research from joint meetings with other labs inside and outside the department as well as through many talks and seminars given by visiting speakers. I was able to gain invaluable insights into science and develop my own ideas, which will help me throughout my career as a scientist.

To my parents and brother – thank you for your unconditional love, support and sacrifices for educating and preparing me for my future. Without your help and support, I would not be here in Oxford and would not be able to pursue my dreams.

Table of Contents

List of Figures

List of Tables

Declaration

All data described in this thesis are generated by me, except as detailed below: -

- Model of Pol II mutants was generated by Dr Krzysztof Kuś, Vasilieva lab (section 2.1.1)
- Purification of *S. pombe* Spt4/5 was done by Dr Krzysztof Kuś, Vasilieva lab (section 2.1.4)
- The PRO-seq experiment for WT, fast and slow Pol II was performed, and the initial bioinformatics analysis done by Dr Takuya Kajitani, Vasilieva lab (section 2.1.6)
- Trypsin digestion and mass spectrometric analysis of the Seb1 purification was performed by the Advanced Proteomics Facility, Department of Biochemistry, University of Oxford (section 2.2.1)
- Seb1 termination assay and polyadenylation assay were performed by Dr Krzysztof Kuś, Vasilieva lab (section 2.2.4.1 & 2.2.4.4)
- Cloning and purification of recombinant PolyA module was done by Dr Krzysztof Kuś, Vasilieva lab (section 2.2.4.2)
- Cloning and purification of recombinant Phosphatase module was done by Mr Alex Au, Vasilieva lab (section 2.2.4.2)

Abbreviations

aa	amino acid
CC	catalytic centre
CPF	cleavage and polyadenylation factor
CPSF	cleavage and polyadenylation specificity factor
CID	CTD-interacting domain
CTD	C-terminal domain of Rpb1 in Pol II
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSE	downstream sequence elements
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EMMG	Edinburgh minimal medium + glutamate
EMSA	electrophoretic mobility shift assay
FAM	5'-fluorescein amidite
FBS	fetal bovine serum
GOF	gain of function
hr(s)	hour(s)
H. sapiens	*Homo sapiens* (human)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	horse radish peroxidase
IP	immunoprecipitation
kDa	kilo dalton
kb	kilobases

LB	Lysogeny broth
LOF	loss of function
min(s)	minutes
mRNA	messenger RNA
MQ water	Milli-Q water
MS	mass spectrometry
ncRNA	non-coding RNA
NiNTA	nickel-nitrilotriacetic acid
nt	nucleotide(s)
OD_{600}	optical density measured at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
Pol II	RNA polymerase II
PRO-seq	Precision Run-On sequencing
PVDF	polyvinylidene fluoride
RNA	ribonucleic acid
Rps2	40S ribosomal protein S2
RT	room temperature
RRM	RNA recognition motif
rRNA	ribosomal RNA
s	second(s)
Seq	sequencing (refers to next-generation sequencing methods)

SDS	sodium dodecyl sulfate
Sf9	a clonal isolate from *Spodoptera frugiperda* (Fall armyworm) IPLB-Sf21-AE cells
S. cerevisiae	*Saccharomyces cerevisiae* (budding yeast)
S. pombe	*Schizosaccharomyces pombe* (fission yeast)
TBE	Tris-buffered EDTA
TBST	Tris-buffered saline + Tween
TL	trigger loop
Tris	Tris(hydroxymethyl)aminomethane
UBL	ubiquitin-like domain
UTR	untranslated region
WT	wild-type
YES	yeast extract broth + supplements

1. Introduction

Transcription of DNA into RNA is an essential step to make proteins using the genetic information encoded in DNA. Transcription is performed by enzymes called RNA polymerases, which synthesizes RNA from a DNA template. In prokaryotes, transcription is performed by a single RNA polymerase, whereas in eukaryotes, three types of RNA polymerases have evolved to synthesize different RNAs. RNA polymerase I (Pol I) is responsible for transcribing ribosomal RNA (rRNA); RNA polymerase II (Pol II) transcribes messenger RNA (mRNA) and some non-coding RNA (ncRNA) such as telomerase RNA, small nuclear and small nucleolar RNAs (sn/snoRNAs); RNA polymerase III (Pol III) transcribes small RNAs including transfer RNA (tRNA),5S rRNA and U6 snRNA.

The transcription cycle consists of three stages: initiation, elongation, and termination where termination is the least understood. All transcriptional stages are highly regulated to achieve accurate gene expression. Regulation is mediated via trans-acting factors that interact with DNA, RNA or polymerase that either modulate DNA structure to make it more accessible or properties of polymerase itself. The extensive regulation of transcription is necessary for proper cell function, adaptation to environmental changes and maintaining cell identity. Deregulation of transcription leads to a broad range of diseases including cancer and developmental diseases (Izumi, 2016; Latchman, 1996; Lee and Young, 2013).

Regulation and coordination of events during Pol II transcription relies on phosphorylation of the carboxy-terminal domain (CTD) of the largest catalytic subunit Rpb1. Pol II CTD consists of a consensus sequence of heptapeptide $Y^1S^2P^3T^4S^5P^6S^7$ repeats, where Y^1, S^2, T^4, S^5 and S^7 residues can be phosphorylated. Phosphorylation of CTD residues helps to recruit proteins required for each step of transcription (Eick and Geyer, 2013; Kobor and Greenblatt, 2002;

Schwer and Shuman, 2011). For transcripts to be functional, pre-mRNA would need to undergo extensive processing including capping at 5' end, splicing, cleavage and polyadenylation (3' end processing). While a lot of research has been conducted to understand how Pol II CTD plays a role in coupling transcription and co-transcriptional events, the molecular details are largely unknown especially for 3' end processing. Therefore, my thesis focuses specifically on how 3' end processing is coupled to transcription using fission yeast *Schizosaccharomyces pombe (S. pombe)* as a model organism.

1.1 Transcription cycle of RNA Polymerase II (Pol II)

1.1.1 Pol II CTD code

Pol II is a multi-subunit complex, which consists of 12 subunits. The largest subunit of Pol II, Rpb1, contains the active site and CTD that can be differentially and reversible phosphorylated within consensus sequence of heptapeptide $Y^1S^2P^3T^4S^5P^6S^7$ throughout the transcription cycle. This helps to recruit factors and complexes that are required in each stage of transcription. Pol II CTD of budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae)* consists of 26 repeats, *S. pombe* consist of 29 repeats, whereas human Pol II CTD consist of 52 repeats, with 21 consensus repeats and 31 non-consensus repeats (Buratowski, 2009, 2003; Eick and Geyer, 2013; Jasnovidova and Stefl, 2013). Serine 2 and 5 (S^2 and S^5) are major phosphorylation sites that indicate the stage of transcription: CTD serine 5 phosphorylation (S5P) at initiation and elongation helps recruit the mRNA capping enzyme and H3K4 methyltransferase, whereas serine 2 phosphorylation (S2P) in the late elongation and termination helps recruit the 3' end processing machinery at the mRNA 3' end. CTD phosphorylation is reversibly controlled by kinases and phosphatases. Ser5 is regulated by Cdk7 kinase and Rtr1 and Ssu72 phosphatases in the beginning and at the end of transcription cycle. Cdk9, a kinase subunit of p-TEFb, phosphorylates CTD S2P, as well as Spt5 and negative elongation factors NELF. S2PCTD is

dephosphorylated by Fcp1 phosphatase at the end of transcription cycle (Buratowski, 2009; Eick and Geyer, 2013; Hsin and Manley, 2012; Jasnovidova and Stefl, 2013; Schwer et al., 2012; Schwer and Shuman, 2011). Threonine 4 phosphorylation (T4P) at the end of transcription cycle is removed by PP1/ Dis2 phosphatase (Cortazar et al., 2019; Kecman et al., 2018; Parua et al., 2018).

1.1.2 Initiation

In eukaryotes, transcription of protein coding genes starts upstream of the transcription unit in a region called the transcription start site (TSS). The core promoter located upstream of TSS is recognised by transcription initiation factors that form the pre-initiation complex (PIC), which consists of the 12-subunit Pol II and the general transcription factors TFIIB, TFIID, TFIIE, TFIIF and TFIIH. These general transcription factors help to open up the promoter DNA and initiate synthesis of RNA. Mediator co-activator complex is one of the essential co-activators that is identified in mammals and yeast (Hampsey and Reinberg, 1999). The Mediator complex bridges the core promoter to Pol II, general transcription factors and DNA-bound activators, which is important for transcription activation (Lewis and Reinberg, 2003). Pol II together with the general transcription factors melt the DNA and form the "transcription bubble", in which the template DNA strand passes nearby the active site of Pol II and RNA synthesis starts. (reviewed in Sainsbury et al., 2015).

1.1.3 Promoter proximal pausing

Transcription is not a uniform process and Pol II pauses at the promoter and 3'end. These pausing events are proposed to control Pol II transition from initiation to elongation and from elongation to termination (described in section 1.1.7). In addition, Pol II also pauses as a part of proofreading during nucleotide incorporation cycle. Pol II promoter-proximal pausing is

observed in 30-50 nucleotides (nt) downstream of TSS. This process is throught to be important for RNA processing, such as 5' end capping, as it creates a "window of opportunity" for elongation factors and RNA processing factors to bind and exert their functions. Pausing depends on transcription factors such as Spt5 that forms DRB sensitivity-inducing factor (DSIF) with Spt4. Spt5 is conserved in bacteria as well. Bacterial and archaeal Spt5 contains two domains: a bacterial elongation factor NusG-like N-terminal domain (NGN) and a C-terminal Kyrpides-Ouzounis-Woese (KOW) domain (Figure 1.1.3). The NGN domain binds to Spt4 and together forms a heterodimer. These two domains are conserved across three domains of life (Meyer et al., 2015). Eukaryotic Spt5 has additional KOW domains, and a C-terminal region (CTR). The CTR is unstructured and consist of consensus repeats that can be phosphorylated by Cdk9 in the beginning and de-phosphorylated by PP1/ Dis2 phosphatase at the end of the transcription cycle (Figure 1.1.3). Unphosphorylated Spt5 recruits the 4-subunit negative elongation factor (NELF), which forms the paused transcription elongation complex with Pol II (Vos et al., 2018b). Recent studies showed that DSIF and NELF binding requires the clearance of promoter region as their binding sites are occupied by initiation factors in the PIC, and binding of anti-pausing transcription elongation factor IIS (TFIIS) is prevented by NELF (Vos et al., 2018b, 2018a). NELF is not conserved in yeast.

Spt5 can regulate transcription both negatively and positively, which was proposed to depend on its phosphorylation status. Structural studies of Spt4/5 revealed that this complex binds to the Pol II clamp domain and the Spt5-Pol II interface is conserved (Hirtreiter et al., 2010; Klein et al., 2011; Martinez-Rucobo et al., 2011; Schwer et al., 2009). Spt4/5 can also bind nucleic acids RNA and DNA (Blythe et al., 2016). A key event to release Pol II from pausing is Cdk9 phosphorylation of NELF leading to its dissociation, phosphorylation of Spt5 and S2 on Pol II CTD (Yamada et al., 2006). Elongation factors PAF1 complex (PAF) and Spt6 are also

recruited to Pol II at this stage. PAF replaces NELF from the Pol II funnel, and Spt6 opens the RNA clamp formed by DSIF by binding to the phosphorylated C-terminal domain linker (Vos et al., 2018a).

Once Pol II enters elongation, Spt5 stimulates Pol II processivity by binding to the polymerase clamp coiled-coiled motif and helps to keep it in a closed conformation (Fitz et al., 2018; Hirtreiter et al., 2010; Klein et al., 2011).

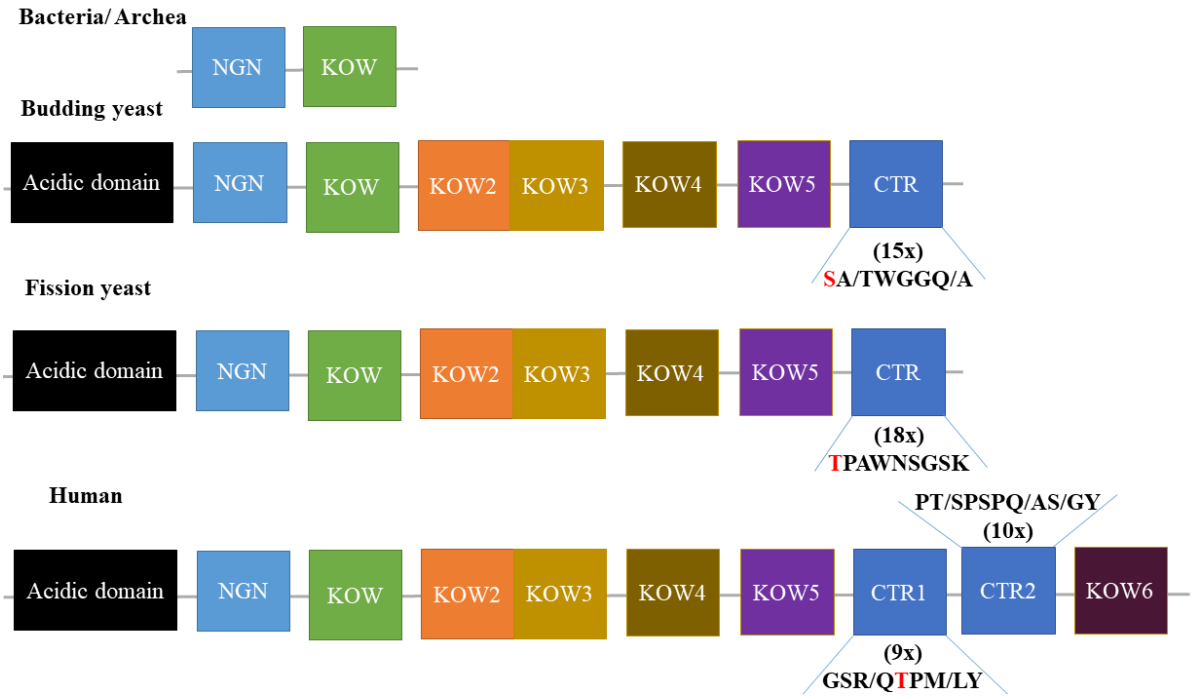


Figure 1.1.3 Organization of Spt5 protein domains across species. The organization of domains are based on (Shetty et al., 2017; Zhou et al., 2009). The CTR region of Spt5 consists of repeats, which the amino acid residues that can be phosphorylated are labelled in red.

1.1.4 Capping of the 5' end of mRNA

As mentioned in the previous section, promoter proximal pausing provides the "window of opportunity" for co-transcriptional events to happen. The phosphorylation of Pol II CTD Ser 5 helps to regulate and recruit capping enzymes to synthesize the 7-methylguanosine (m7G(5')ppp(5')N_1 cap) and add to 5'end of the nascent transcript (Fabrega et al., 2003;

McCracken et al., 1997; Shatkin, 1976). Three capping enzymes are involved: RNA triphosphatase removing gamma phosphate from the 5' end of pre-mRNA, RNA guanylyltransferase adding GMP, and RNA (guanine-7) methyltransferase, which methylates the 5' guanine (Fabrega et al., 2003; Shuman, 2001). The cap protects the nascent transcript from 5' exonucleases and facilitates splicing, export and translation of mRNA (Kachaev et al., 2020).

1.1.5 Elongation

After Pol II is released from promoter-proximal pausing, it enters "productive elongation," which is marked by the phosphorylation of Ser 2 of CTD repeats by Cdk9. Most genes require Cdk9 for elongation, inhibition of Cdk9 is known to cause global downregulation of transcription (Haberle and Stark, 2018). Transcription elongation is a tightly controlled process that involved a number of regulatory factors. As Pol II transcribes along the gene, this process is frequently interrupted by pausing. Pausing is thought to be linked to Pol II proofreading, which ensures transcription fidelity and can also be induced by certain DNA and RNA sequences. When Pol II pauses during elongation it can enter the elemental pause state from which pauses can be prolonged by entering a backtrack state (Palo et al., 2019; Saba et al., 2019). Extensive backtracking would lead to trapping of RNA and trigger loop in the pore, which in turn arrests Pol II and inhibit transcription (Cheung and Cramer, 2011). Several residues in the Pol II active site, particularly in the trigger loop, are important for elongation and pausing, mutations in these residues shown to increase or decrease Pol II pausing (described in section 1.3). Pol II pausing also occurs in sequence specific manner, in a 9-mer sequence specific motifs including GC-rich regions followed by AT-rich sequences. A genome-wide study showed that 65% of Pol II pause sites are cytosines (Watts et al., 2019). This sequence is believed to induce the backtracking and pausing in Pol II as the sequence

reduces the stability and melting temperature of the DNA-RNA hybrid inside the elongation complex (Kireeva et al., 2000; Kulish and Struhl, 2001). Backtracking also occurs when a nucleotide is mis-incorporated into the 3' end of nascent RNA. Other sequences that induce Pol II pausing includes poly 'A' sequences and secondary structure or R-loops formed by the nascent RNA (Mayer et al., 2017).

Pausing can be prevented by Spt5 that helps to keep Pol II clamp in closed conformation and thus stimulate elongation. Elongation factor TFIIS helps to release Pol II from backtracked state. TFIIS is also conserved in bacteria (GreB). TFIIS releases Pol II from backtracking by stimulating its intrinsic RNA cleavage activity. TFIIS complements Pol II active sites with its basic and acidic side chains, locks the Pol II trigger loop away from RNA, displaces and mobilizes backtracked RNA in the pore. This helps releases backtracked RNA and induces RNA cleavage, which allows reactivation of transcription (Cheung and Cramer, 2011). Pausing frequency, the efficiency of nucleotide incorporation and backtracking all contribute to Pol II processivity, which will be further discussed in section 1.3.

Apart from sequence specific pauses, Pol II pausing can be induced by nucleosomes and regulatory factors. For this type of pausing, the biophysical mechanism of how Pol II pausing occurs is poorly understood, however, it is known that nucleosomes act as barriers for transcription as increased pausing is observed in transcription through nucleosomes compared to transcription in naked DNA *in vitro* (Chang and Luse, 1997). Several factors and complexes are known to help Pol II to overcome the nucleosome barrier. PAF complex helps to recruit elongation factor complexes such as FACT and histone modifying enzymes to the paused Pol II. The 2-subunit factor FACT consists of Spt16 and SSPR1 and functions to disassemble H2A-

H2B dimer from nucleosomes. Spt6 increase Pol II intrinsic elongation speed, disassemble and reassemble H3 and H4 (reviewed in Kwak and Lis, 2013).

1.1.6 Splicing

Other co-transcriptional events such as splicing take place during elongation, as splicing is not the main focus of the thesis, it is only briefly described in this section. Splicing of the pre-mRNA is carried out by the spliceosome recruited in a manner dependent on phosphorylation on Ser2 and Ser 5 is important for the requirement of the splicing machinery, and truncations of the CTD lead to a great reduction in splicing (Misteli and Spector, 1999). Splicing is important for producing mRNA and protein isoforms that are different in length, function, localization, structure, physicochemical properties, and biological activity. Splicing occurs in close proximity to transcribing Pol II, suggesting that splicing can be affected via kinetic coupling of transcription elongation, as pausing of Pol II creates a "window of opportunity" for splicing factors to bind (Fong et al., 2014).

1.1.7 Transition from elongation to termination

The transition from elongation to termination is tightly regulated to prevent interfering transcription of neighbouring genes. The transition to termination requires dephosphorylation of S5 on Pol II and Spt5. PP1 isoform Dis2 in fission yeast and PP1 in mammals, a phosphatase component of the cleavage and polyadenylation factor (CPF), dephosphorylates Spt5 CTR (Parua et al., 2018; Kecman et al., 2018; Cortazar et al., 2019). In fission yeast, Dis2 also dephosphorylates Thr4 on CTD (Kecman et al., 2018). Co-transcriptional recruitment of the termination/ 3'end processing factor Seb1 is dependent on the dephosphorylation of Spt5, as the interaction surface of Seb1 on Pol II is largely overlapped with that of Spt5. Deletion of Dis2 led to the impaired recruitment of termination factors and defective transcription

termination (Kecman et al., 2018). It is proposed that Dis2/ PP1 functions to regulate the transition of elongation to termination.

1.1.8 RNA 3' end processing

Near the end of elongation, Pol II transcribes through the polyadenylation signal (PAS) at the 3' end of a gene. The cleavage and polyadenylation factor (CPF) in yeast (cleavage and polyadenylation specificity factor (CPSF) in humans) recognizes the PAS, cleaves and polyadenylates the pre-mRNA. The CPF is a giant multi-subunit complex, which consists of three functional modules: the nuclease module that cleaves the mRNA, the polymerase module (poly A module) that adds a polyadenylate tail and recognizes polyadenylation (pA) signal, and the phosphatase module that regulates this process by dephosphorylating proteins (Casañal et al., 2017). As 3' end processing is coupled to transcription, transcription rate was proposed to influence selection of the pA and 3' end processing (Fusby et al., 2016), although the underlying mechanism is not well understood.

1.1.8.1 Cleavage and polyadenylation sites (PAS)

Eukaryotic mRNA 3' end processing is important for mRNA maturation to produce functional mRNAs that can be transported out of the nucleus to the cytoplasm and translated into proteins. RNA 3' end processing by the CPF/ CPSF involves mainly two steps: (1) cleavage of the mRNA, followed by the (2) addition of the poly A tail, which has a median length of around 80 nt (Butler and Platt, 1988; Subtelny et al., 2014). CPF recognizes the PAS once it emerges from Pol II. Most eukaryotic genes have multiple PAS, resulting in a widespread phenomenon called alternative cleavage and polyadenylation (reviewed in Elkon et al., 2013; Gruber et al., 2016; Neve et al., 2017).

PAS is defined by surrounding motifs including efficiency element, positioning element (PE), cleavage site (CS) and downstream sequence elements (DSE). The positioning element is the most conserved and core element of PAS, and is located 10-30nt upstream of the cleavage site, it consists of the sequence A[A/U]UAAA hexamer and is recognized by the components of polyadenylation module CPSF4 (Yth1 in fission yeasts) and WDR33 (Pfs2 in fission yeast). The cleavage site (also called the poly(A) site) is usually a cytosine followed by adenosine (C/A), a study identified the nucleotide before cleavage site is often C (59% of the genes analysed) (Sheets et al., 1990) followed by A > U > C >> G in the order of preference (Chen et al., 1995). The less conserved downstream sequence element, which consists of U-rich or GU-rich region, is located downstream 0-20 nt of the cleavage site (reviewed in Zhao et al., 1999).

PAS motifs in budding yeast are loosely defined, whereas fission yeast PAS resemble those of mammals and are more clearly defined (Figure 1.1.8.1). Fission yeast PAS motifs consist of a non-essential A/UA-rich upstream element located 30-60nt upstream of the cleavage site, a position element AAU[A/G]AA and a U-rich region upstream of cleavage site, cleavage site and downstream sequence element A/ U-rich/ GUA (Graber et al., 1999; Liu et al., 2017; Schlackow et al., 2013; Tian and Graber, 2012; Zhao et al., 1999a).

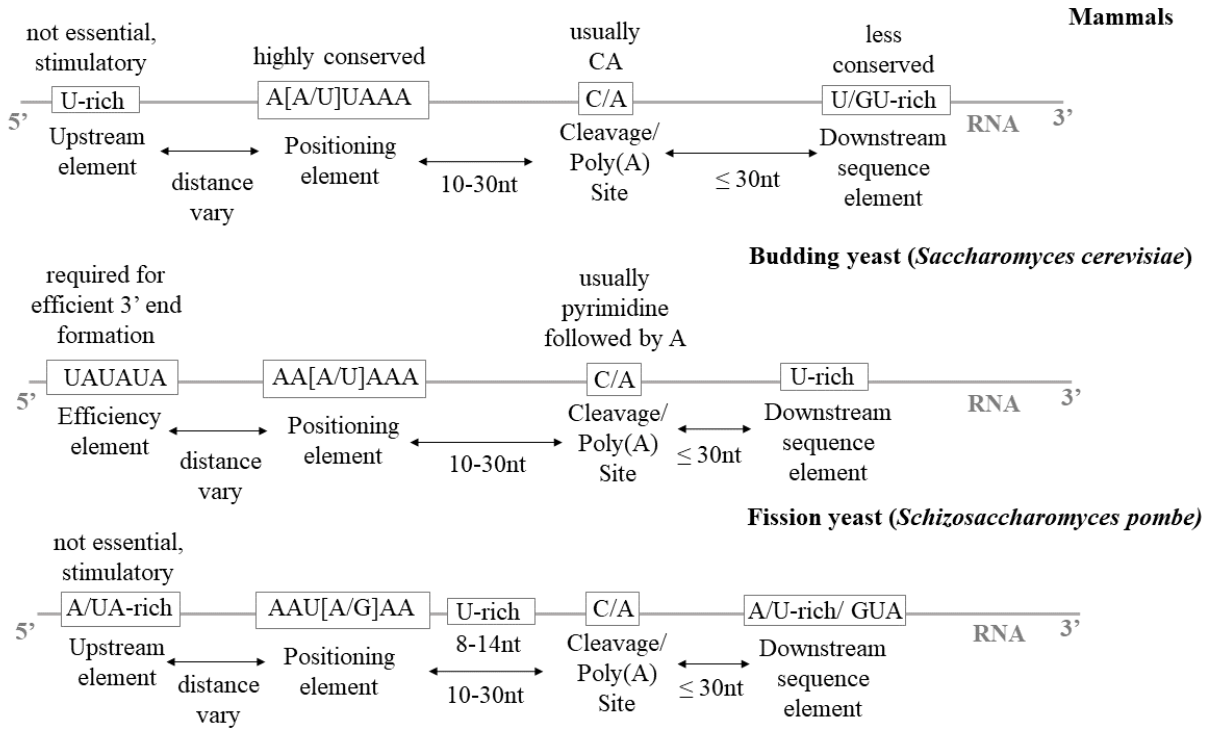


Figure 1.1.8.1 A diagram showing the PAS motifs in mammalian, budding yeast and fission yeast. This figure is based on (Chan et al., 2011; Chen et al., 1995; Graber et al., 1999; Schlackow et al., 2013; Tian and Graber, 2012; Zhao et al., 1999a).

1.1.8.2 3' end processing machinery

In eukaryotes, 3' end processing is tightly coupled to termination since endo-nucleolytic cleavage of pre-mRNA is a key event necessary to recruit transcription termination factor, 5'-3' exonuclease Rat1/ Xrn2 that degrades Pol II associated cleavage product leading to release of Pol II via so called "torpedo" mechanism (Kim et al., 2004b; West et al., 2004). Hence, compromised mRNA 3' and processing not only makes non-functional unstable RNA but also compromises transcription termination.

The 3' end RNA processing machinery has been extensively studied in budding yeast and in humans, however, not much is known about the 3' end RNA processing machinery in fission yeast. Table 1.1.8.2 detailed the complexes involved in 3' end processing in budding yeast and humans.

Table 1.1.8.2 RNA processing machinery in budding yeast and human. Table constructed based on (Clerici et al., 2017). *no exact correspondence

Budding Yeast			Human		
Complex	**Module**	**Subunit**	**Complex**	**Module**	**Subunit**
CPF	Poly(A) polymerase module	Cft1	CPSF	Poly(A) polymerase module	CPSF160
		Pfs2			WDR33
		Fip1			hFip1
		Yth1			CPSF30
		Pap1			PAPOLA*
	Nuclease module	Cft2		Nuclease module	CPSF100
		Ysh1			CPSF73
		Mpe1			RBBP6*
	Phosphatase module	Pta1		Phosphatase module	Symplekin*
		Glc7			PP1α/β
		Ssu72			Ssu72
		Swd2			WDR82
		Pti1			-
		Ref2			-
		Syc1			-
CFIA		Rna14	CstF		CstF77
		Rna15			CstF64
		-			CstF50
		Pcf11	CFIIm		hPcf11
		Clp1			hClp1
-		-	CFIm		CFIm 68/59
					CFIm 25
CFIB		Hrp1	-		-

Recent structural studies of 3’ end RNA processing machinery from budding yeast and mammalian cells (Casañal et al., 2017; Hill et al., 2019; Kumar et al., 2019; Sun et al., 2018; Xiang et al., 2014) have provided important insights into organisation of this machinery that consists of the Cleavage and Polyadenylation Factor (CPF) in yeast (Cleavage Polyadenylation Specificity Factor (CPSF) in mammals) organised into three functional modules: polymerase, nuclease and phosphatase and Cleavage Factors (CF) CFIA and CFIB.

The Poly A polymerase module of CPF contains five subunits: Cft1, Pfs2, Pap1, Fip1, and Yth1, of which Pap1 is the polymerase (Figure 1.1.8.2, denoted with a star) that is responsible for adding a poly(A) tail to the 3’ end of the upstream pre-mRNA cleavage product. The addition of the poly(A) tail is required for mRNA export from the nucleus, mRNA stability and translation regulation. Cft1 is composed of three seven-bladed β propellers (BP) and a C-terminal helical domain. Cft1, Yth1 and Pfs2 together form a scaffold for the assembly of the polymerase module (Casañal et al., 2017). From the structural data, the N-terminal region of Pfs2 is inserted into the cavity of Cft1 BP1 and BP3, and forms close contacts. Key interactions between Pfs2 and Cft1 are known to be conserved in human orthologs (WDR33 and CPSF160 respectively). Yth1 (human ortholog CPSF30) binds to the Cft1 BP3 cavity using its N-terminal segment and 2 out 5 of its zinc fingers are located in the interface between Cft1 and Pfs2. Fip1, a disordered protein, is known to bind to the C-terminal region of Yth1 and polymerase Pap1. Pfs2 contains a RNA recognition domain that consists of conserved lysines, arginines, and aromatic residues, this domain is in close contact with the zinc finger 2 of Yth1, which is known to bind RNA. Yth1 (CPSF30) and Pfs2 (WDR33) together help recognize the pA (AAUAAA) signal.

The nuclease module contains Ysh1 (human ortholog CPSF73), a highly conserved endonuclease involved in cleaving the pre-mRNA at the cleavage site. Other subunits in the nuclease module include Cft2 (human ortholog CPSF100), Ipa1 and Mpe1 (human ortholog RBBP6) For efficient cleavage, correct positioning of Ysh1 is required, and this is mediated by CFIA and CFIB, which are recruited and bound to the CPF and RNA substrates, which will be described below (Hill et al., 2019). The function of Ipa1 is unknown, however it is essential for cell viability and implicated to be involved in polyadenylation (Casañal et al., 2017; Costanzo et al., 2016). Ysh1 is only activated when incorporated into the core CPF complex, which included the eight subunits in polymerase and nuclease module (Hill et al., 2019).

Two cleavage factors CFIA and CFIB are involved in stimulating and positioning CPF on RNA. CFIB is a single RNA-binding factor, Hrp1, which binds to the efficiency element UAUAUA which located upstream of the cleavage site (Pérez-Cañadillas, 2006; Valentini et al., 1999). Rna15 of CFIA recognizes the positioning element and help CPF in positioning accurately in the RNA substrate for cleavage. Hrp1 and Rna15 interact with Rna14, another component in CFIA that stabilizes and bridges Hrp1 and Rna15 (Gross and Moore, 2001), to increase the efficiency of recognizing RNA sequences accurately (Kessler et al., 1997; Leeper et al., 2010). Pcf11 and Clp1 are two other components of the 4-subunit CFIA. Disruption of the dimer interface of Rna14 is known to affect the interaction of Rna14 or Rna15 with RNA and 3' RNA end processing but not Rna14-Rna15 interaction (Gordon et al., 2011). Apart from Hrp1 and Rna15, Rna14 also interacts with Pfs2 and Cft1 of the CPF polymerase module (Kyburz et al., 2003; Ohnacker et al., 2000). Structural studies of highly conserved Pcf11 showed that it formed a zinc finger structure at C-terminal that is important for 3' mRNA processing (Yang et al., 2017). At its N-terminus, Pcf11 contains a CTD-interacting domain (CID) that binds to Pol II CTD (section 1.4) and a domain containing 20 glutamines that is for the interaction with

the Rna14-Rna15 dimer. A short C-terminal sequence in Pcf11 interacts with Clp1, forming a heterodimer. Mutations that disrupt the interaction between Pcf11 and Clp1 resulted in growth defects and defects in 3' end processing (Ghazy et al., 2012; Noble et al., 2007). Pcf11 also interacts with Cft1, Cft2, Ysh1 and Pta1 (Kyburz et al., 2003). Clp1 is conserved and contains a P-loop domain which mediates the interaction between CFIA and CPF subunits (Holbein et al., 2011), and a Walker A motif that may be involved in nucleotide binding or catalysis (Dupin and Fribourg, 2014; Noble et al., 2007; Walker et al., 1982).

For the phosphatase module, there are two phosphatases: Ssu72 (human ortholog SSU72) and Glc7 (budding yeast), Dis2 (fission yeast) (human ortholog PP1 α and β). They are known to regulate transcription and 3' end processing. Extensive studies are conducted to study the role of Ssu72. Mutation of Ssu72 is known to have adverse effect on 3' end processing and termination (Bataille et al., 2012; Nedea et al., 2003; Zhang et al., 2012). It is a Pol II CTD-phosphatase that dephosphorylates Pol II CTD S5P and S7P (reviewed in Rosado-Lugo and Hampsey, 2014). Glc7 is proposed to de-phosphorylate Y1P in *S. cerevisiae* (Schreieck et al., 2014) and its homologue Dis2 dephosphorylates T4P of CTD in fission yeast (Kecman et al., 2018). The other factors in the phosphatase module includes Swd2, Syc1, Pti1, Ref2 and Pta1 in *S. cerevisiae*. The human orthologue of Pta1, Symplekin interacts with SSU72 (Xiang et al., 2010). Biochemical reconstitution *in vitro* showed that Pta1 is non-essential for cleavage and polyadenylation (Hill et al., 2019), however, Pta1 mutation in yeast extract showed to disrupt CPF function (Zhao et al., 1999b). Swd2, a component of the Set1 methyl transferase complex COMPASS, is known to regulate Pol II termination and histone H3 methylation at lysine 4 (Cheng et al., 2004; Mischo and Proudfoot, 2013; Soares and Buratowski, 2012). Co-purification of Pti1 or Ref2 with CPF is known to suppress Pap1-dependent polyadenylation (Dheur et al., 2003).

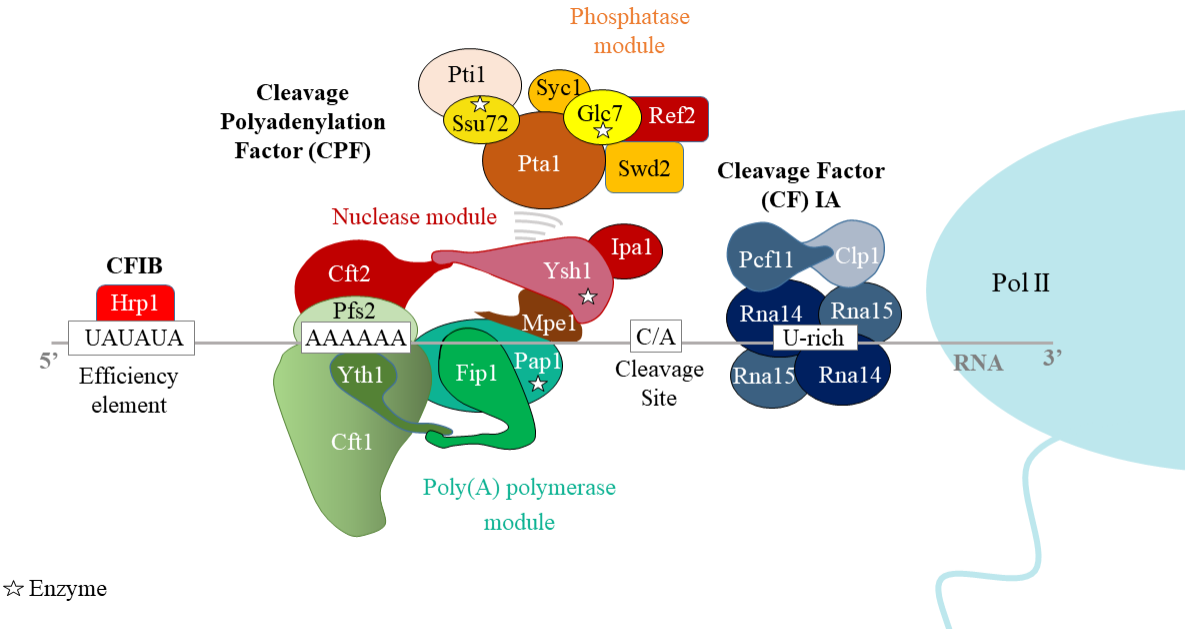


Figure 1.1.8.2 Architecture of the 3'end processing complex in budding yeast. A diagram showing the cleavage and polyadenylation factor (CPF), cleavage factor IA and IB and their recognized RNA sequences. This figure is based on (Casañal et al., 2017; Hill et al., 2019).

1.1.8.3 Model of 3' end machinery in *S. pombe*

Compared to human CPSF and budding yeast CPF, there are much less studies on the 3' end machinery of fission yeast. A hypothetical model is constructed based on the known 3' mRNA processing machinery in budding yeast (Figure 1.1.8.3). Orthologs of protein subunits in CPF and CFs are identified using the fission yeast database PomBase (https://www.pombase.org) and Uniprot (https://www.uniprot.org). However, there is no evidence that these proteins are involved in 3' end processing, one of the aims of my thesis is to understand the 3' end machinery in fission yeast.

The polymerase module of CPF is conserved and contains all subunits that are found in budding yeast: Pfs2, Cft1, Fip1, Pla1 (ortholog of Pap1) and Yth1. Subunits of nuclease module is also conserved containing Mpe1, Ysh1, Ipa1 and Cft2. For the phosphatase module, Swd22 is the ortholog of Swd2, Dis2, Pta1 and Ssu72 are conserved. Ppn1 is a non-essential protein that is

associated with the Dis2 in the phosphatase module. Pti1, Ref2, and Syc1 are not found in fission yeast.

For CFIA, Pcf11, Rna14 and Rna15 are conserved and SPAC22H10.05c (referred as Clp1 for the rest of thesis) is the fission yeast ortholog of Clp1. Ctf1 is a non-essential ortholog of Rna15 in fission yeast. Rna14 and Ctf1 are known to be involved in maintenance of genomic integrity in fission yeast (Sonkar et al., 2017, 2016) and Rna14 is important in preventing transcription readthrough (Sonkar et al., 2016). For CFIB, Msi2 is the fission yeast ortholog of Hrp1.

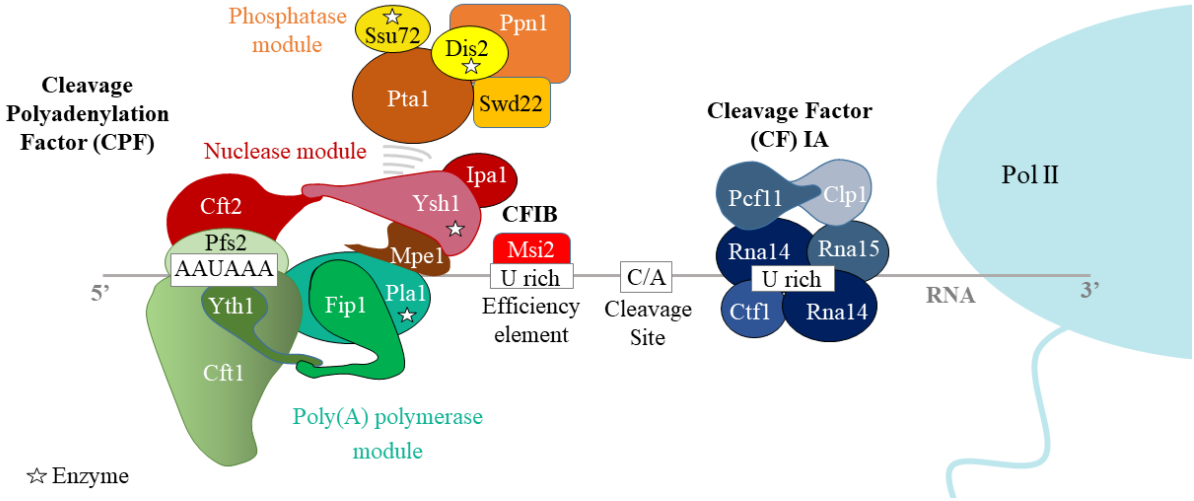


Figure 1.1.8.3 Hypothetical model of cleavage and polyadenylation factor (CPF) and cleavage factor (CF) IA and IB based on budding yeast. Three modules of CPF: Polymerase (polyA) module is labelled in green, nuclease module is labelled in red and phosphatase module is labelled in orange. The complexes are placed at their predicted binding sequences in RNA. The enzyme in the complexes are denoted as star. Predicted components of polymerase module: Pfs2, Cft1, Fip1, Pla1, Yth1; nuclease module: Cft2, Mpe1, Ysh1, Ipa1 phosphatase module: PpnI, PtaI, Swd22, Ssu72, Dis2; CFIA: Pcf11, Clp1, Rna14, Rna15, Ctf1 (non-essential); CFIB: Msi2. Seb1 and Pcf11 are known to bind preferentially to Pol II S2P CTD. This figure is created based on the previous studies in budding yeast (Casañal et al., 2017; Hill et al., 2019).

1.1.9 Termination

Termination is an important step to generate functional RNA transcripts, release mRNA and recycle Pol II. However, the mechanism of termination is poorly understood. In eukaryotes, termination, the dislodgment of Pol II from the template, is tightly coupled to mRNA 3' end processing, making it hard to dissect mechanistically *in vivo*. Endo-nucleolytic cleavage of pre-mRNA by the CPF is a key event necessary leading to recruitment of transcription termination factor, 5'-3' exonuclease Rat1/ Xrn2. Xrn2 degrades Pol II associated cleavage product leading to release of Pol II via so called "torpedo" mechanism (Connelly and Manley, 1988; Eaton and West, 2018; Kim et al., 2004b; Proudfoot, 1989; West et al., 2004). Rat1/ Xrn2 role in transcription termination has been further confirmed by recent studies (Baejen et al., 2017; Davidson et al., 2019). Also, depletion or inactivation of CPSF73 which caused PAS cleavage inhibition resulted in extensive readthrough transcription, suggesting that RNA 3' end processing is important for termination (Eaton and West, 2018). Termination occurs in a range of around 200 nt (yeast) to 1500 nt (human) downstream of PAS, mutations of CPF subunits result in termination defects (Mischo and Proudfoot, 2013).

In addition to the torpedo model, the allosteric model was proposed where factors binding to Pol II can cause conformational changes once it transitions through pA site (Logan et al., 1987), leading to its dislodgement from DNA (Connelly and Manley, 1988; Proudfoot, 1989). The two models are both supported by extensive research and they are not mutually exclusive.

The allosteric model is supported by the presence of anti-terminators in budding yeast (PC4) and humans (SCAF4 and SCAF8), and the exchange of factors/ protein modifications as evidenced (Ahn et al., 2004; Kecman et al., 2018; Kim et al., 2004a). In addition, Pol II binding

component of 3'end processing machinery such as Pcf11 was shown to release Pol II from DNA *in vitro* (Zhang and Gilmour, 2006).

In budding yeast, two distinct termination mechanisms are proposed to be used for termination of non-coding and protein-coding transcripts. In contrast to protein-coding genes that rely on CPF mediated cleavage and Xrn2, non-coding transcripts depend on Pol II and RNA binding protein Nrd1, that interacts with RNA binding protein Nab3 and RNA helicase Sen1 for transcription termination (Creamer et al., 2011; Darby et al., 2012; Steinmetz and Brow, 1996; Vasiljeva et al., 2008). The Nrd1 mechanism does not seem to require CPF and RNA cleavage. Nrd1 also regulates protein-coding genes by promoting premature transcription termination (Webb et al., 2014). *In vitro* studies showed that only Sen1 of Nrd1-Nab3-Sen1 complex is able to dislodge Pol II (Martin-Tumasz and Brow, 2015; Porrua and Libri, 2013). However, this mechanism is not found in higher eukaryotes as human homolog of Sen1 (Senataxin) requires the presence of Xrn2 to terminate Pol II (Chen et al., 2014). In fission yeast, the mechanism of termination is poorly studied, thus, one of my aims in this study is to shed light on the mechanism of termination and understand how transcription is coupled and contribute to termination.

1.2 Coupling between transcription and 3' end processing

1.2.1 Role of CID-containing proteins in 3' end processing and termination

As mentioned in the previous section (section 1.1), the Pol II CTD can help to recruit RNA processing factors during transcription (Bentley, 2014; Hsin and Manley, 2012; Phatnani and Greenleaf, 2006). CID-containing proteins can interact with the Pol II CTD via a conserved CTD-interacting domain (CID). Several CID-containing proteins have been linked to either 3' end processing or transcription termination. One of these, budding yeast Nrd1 protein, in

addition to a CID domain also possesses RNA-recognition motif (RRM) that recognize a specific sequence on RNA (Patturajan et al., 1998; Yuryev et al., 1996) (Figure 1.2.1I). Unlike Pcf11, Nrd1, however, preferentially binds to S5P (Vasiljeva et al., 2008) and is recruited to the 5' end of genes. Fission yeast CID protein Seb1 and mammalian SCAF4 and SCAF8 have domains organisation similar to Nrd1.

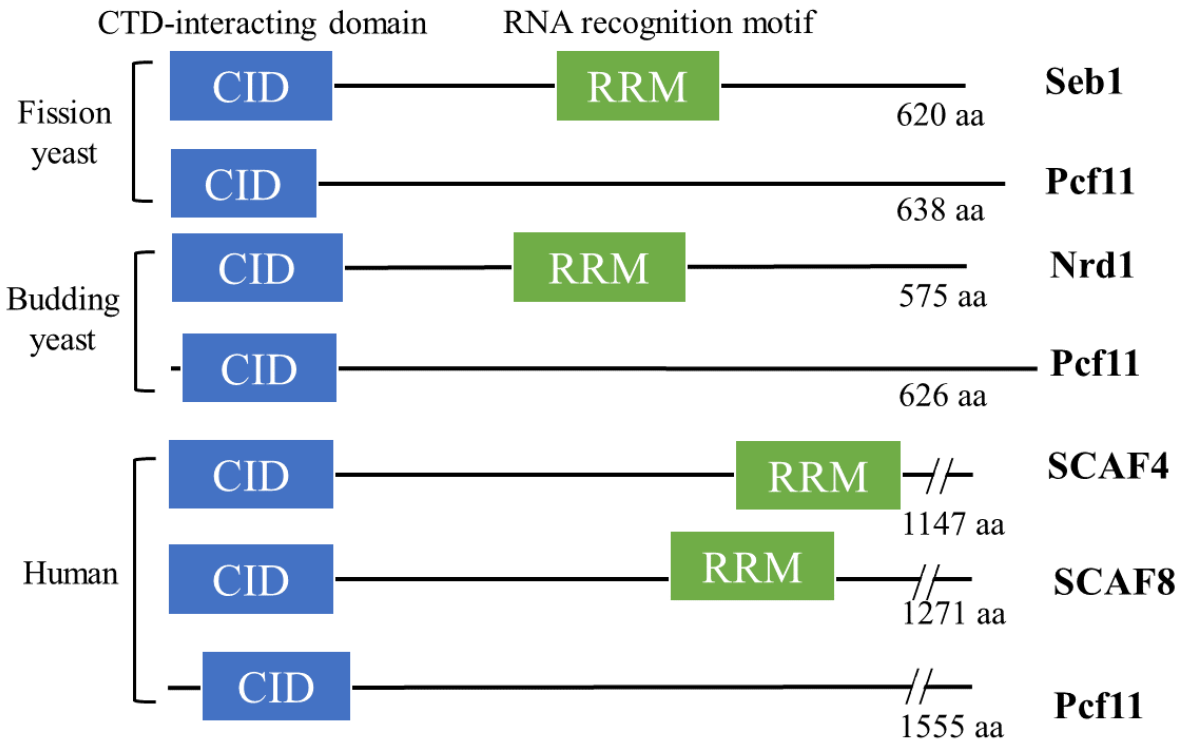


Figure 1.2.1I Conserved CID-containing proteins. CTD-interacting domain (CID) is labelled in blue and RNA recognition motif (RRM) is labelled in green. The relative positions of the domains in fission yeast, budding yeast and humans are shown in the figure. Figure created based on (Wittmann et al., 2017).

Pcf11, an essential highly conserved protein, interacts with CPF/ CPSF and S2P of Pol II CTD. In PAS-dependent (protein-coding) pathway, Pcf11 works together with CPF and CF to terminate longer RNA transcripts (Grzechnik et al., 2015). In budding yeast, Pcf11 is a component of the CFIA, and is required for Nrd1-dependent termination (Grzechnik et al., 2015). Rtt103 is a non-essential CID protein, which was proposed to assist Xrn2/ Rat1 in

releasing Pol II following transcript cleavage by CPF (Kim et al., 2004b; West et al., 2004). In mammals, Pcf11 is part of the CFIIm, which interacts with Clp1 and plays a role in transcription termination. Pcf11 orthologues in yeasts and mammals are different in sizes, but the protein domains such as N-terminal CID and protein interaction domains are conserved. In fission yeast, Pcf11 and Seb1 are recruited to the same position of 3' end of protein-coding and non-coding genes and are involved in terminating both PAS-dependent and PAS-independent genes (Wittmann et al., 2017).

In contrast to Nrd1 that binds to S5P, Seb1 preferentially binds to Pol II CTD S2P. It is a multivalent protein, in addition to interactions with Pol II CTD and Pol II core, it recognizes RNA sequence UGUAA by its RRM and co-purifies with CPF. Seb1 has a flexible linker between CID and RRM domains, deletion of either domain or regions after the domains are lethal. RNA-protein crosslinking experiments (PAR-CLIP) showed that Seb1 crosslinks at TSS and down-stream of the PAS. Mutations in the CID and RRM abolish its ability to interact with RNA resulted in severe 3'end processing as well as transcription termination defect genome-wide (Figure 1.2.1IIB, 4-6) (Wittmann et al., 2017). This suggests that both domains are important for transcription and 3' end processing. Another study suggested that Seb1 is involved in selection of pA site (Lemay et al., 2016). However, as 3' end processing is tightly coupled to termination *in vivo*, which process is being affected is unclear – it may be selection of PAS, cleavage, polyadenylation, mRNA release or termination (Pol II release). Moreover, Seb1 mutants show severe reduction in Pol II occupancy across the gene bodies (Wittmann et al., 2017), which could be indicative of transcription elongation defects. This leads to a question: what is the direct role that Seb1 plays in transcription and RNA 3'end processing?

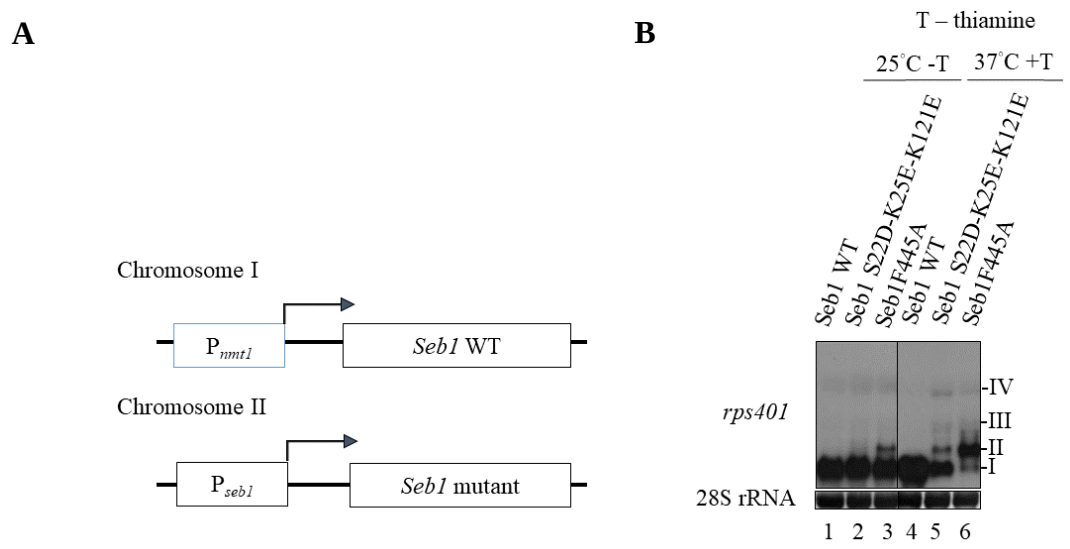


Figure 1.2.1II Seb1 mutants show mRNA 3'end processing as well as transcription termination defects

(A) **A diagram showing the yeast strains containing Seb1 mutants.** The WT copy of Seb1 is under control of the P_{nmt1} which the expression of WT Seb1 can be repressed by the addition of thiamine (T). The mutant copy of Seb1 is under the control of the Seb1 promoter P_{seb1}.

(B) **Northern blot showing the transcripts of *Rps401* with Seb1 WT and point mutants.** Yeast strains with point mutations in CID (S22D-K25E-K121E) and RRM (F445A) are constructed by Dr Sina Wittmann. Read through transcripts are indicated as 2-4. The figure is extracted from (Wittmann et al., 2017) with modifications. Lane 1:Seb1 WT at 25°C without thiamine; 2: Seb1 S22D-K25E-K121E mutant at 25°C without thiamine; 3: Seb1 F445A mutant at 25°C without thiamine; 4: Seb1 WT at 37°C with thiamine; 2: Seb1 S22D-K25E-K121E mutant 37°C with thiamine; 3: Seb1 F445A mutant at 37°C with thiamine

A recent study on SCAF4 and SCAF8 (human orthologs of Seb1) showed that they are redundant but essential factors that function as anti-terminators, which suppress early termination and use of early poly(A) sites. SCAF4 and SCAF8 bind preferentially to Pol II bi-phosphorylated CTD S5P and S2P. Double knockout of SCAF4 and SCAF8 using CRISPR is lethal and resulted in early termination and production of truncated transcripts. Both SCAF4 and SCAF8 interact with PAF complex, RPRD1-RPAP2 complex, elongation factor SUPT6H, RECQL5, CTD-associated SCAF1 and SCAF11 and other uncharacterized Pol II associated proteins. Interestingly, SCAF4 and SCAF8 do not interact with each other. SCAF4 and SCAF8 appear to play independent roles in transcription. SCAF8 was proposed to be a positive

elongation factor, which showed to stimulate Pol II transcription elongation, and even lead to read-through when SCAF4 is knocked out. The average transcription speed of WT and SCAF4 knock out are 2.2kb/ min, whereas knock out of SCAF8 is 1.9 kb/ min and double knock out is 1.8kb /min. SCAF4, in cooperation with SCAF8, functions to ensure termination occurs at distal canonical termination sites. SCAF4 and SCAF8 together regulate the selection of poly(A) sites and termination sites in human cells (Gregersen et al., 2019).

Fission yeast Rhn1 is homologues to *S. cerevisiae* Rtt103 (Kecman et al., 2018; Sugiyama et al., 2012), and RPRD1A, RPRD1B, RPRD2 in humans (Ni et al., 2014, 2011). As they are not the main focus of this thesis, they are not shown in Figure 1.2.1I.

1.3 Regulation of Pol II transcription rate

Pol II is a sophisticated multi-subunit molecular machine that is regulated in a manner specific for individual genes. Pol II transcription rate is highly dynamic in *vivo*: Pol II speed can vary from gene to gene and can vary throughout the transcription of a single gene after releasing from the promoter-proximal pause site (Jonkers and Lis, 2015). Transcription rate is important for controlling gene expression. It was proposed that slow elongation rate expands the "window of opportunity" to allow more efficient co-transcriptional recruitment of factors leading to selection of early pA site compared to WT; and fast elongation rate in contrast, compresses the "window of opportunity" resulting in selection of PAS further downstream (Figure 1.4) This will affect the transcript length as well as the function and localization of RNA transcripts. This kinetic model is also applicable to other co-transcriptional events such as splicing (Bentley, 2014; Fong et al., 2014).

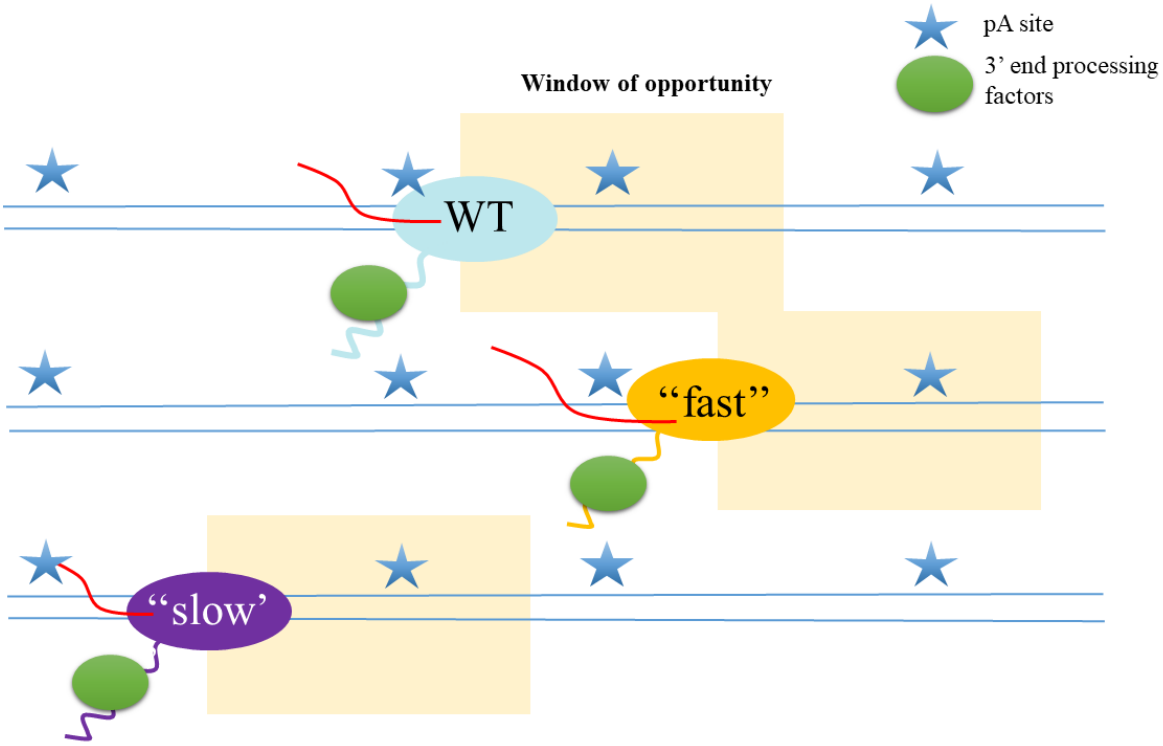


Figure 1.3 Transcription rate and window of opportunity. Transcription rate affect the selection of cleavage/ polyadenylation site (pA site – labelled as star) by moving the "window of opportunity" (labelled as yellow) for 3' end processing factors (labelled in green) to bind Pol II CTD upstream or downstream.

Transcription rate is regulated by multiple general transcription factors including Spt4/5 (Fuda et al., 2009) as mentioned in previous sections that modulate properties of Pol II and its processivity.

1.3.1 Pol II active site

X-ray crystal structures of high resolution Pol II in eukaryotes (Armache et al., 2005; Cheung and Cramer, 2011; Ehara et al., 2017; Spåhr et al., 2009) aid our understanding of how Pol II functions. The surface charge of Pol II is entirely negative, however, several regions including the Pol II active site have a uniform positive charge. This asymmetric charge distribution aided the interactions between polymerase and DNA, and restricted the path of RNA exit and template DNA entry to the active site in the transcribing complex (von Hippel et al., 1984). The active centre of Pol II consists of two regions: the active site of Rpb1 and the RNA-DNA

hybrid binding region in Rpb2. Two metal ions are found at the active site of Pol II - a magnesium ion (metal ion A), which is bound by Rpb1 aspartates invariants D481, D483 and D485; and the second metal ion (metal ion B), which is located in close proximity to metal ion A. Both metal ions are able to access the NTP substrates (Cramer et al., 2001).

Another important structure at the active site that affects the processivity of Pol II is the trigger loop, a conserved element located in Rpb1. The trigger loop is the master regulator of the three main phases of transcription elongation – substrate selection, catalysis and translocation, and it balances the rate of transcription and fidelity. The trigger loop has two states: a closed state which facilitates the interactions with the correctly match NTP at the A site during catalysis; and open state, which allows translocation to occur after the addition of NTP. These two states are regulated by the hinge regions of the trigger loop (Figure 1.5.1 labelled in dark blue). When a mismatch NTP enters the A site, catalysis does not occur as trigger loop cannot switch to the close conformation due to the destabilizing mismatched NTPs (Huang et al., 2010). In the trigger loop, there are two residues H1085 and L1081 that are known to contact the β-phosphate and base of NTP respectively (Wang et al., 2006). Both residues are located in the nucleotide interaction region (Figure 1.5.1 labelled in pink). During substrate selection, the incoming NTP is positioned by the side chain of H1085, which interacts with the NTP β-phosphate oxygen. Studies proposed that H1085 facilitates the formation of phosphodiester bond during nucleophilic attack through its protonated form. Importance of this residue is evidenced by mutation studies - tyrosine substitution of H1085 showed elongation defects and phenylalanine substitution of H1085 is lethal. L1081 contacts NTP through hydrophobic interactions, and this residue is believed to play an important role in stabilizing the position and confirmation of the Pol II active site. Mutations of this residue increase the duration of pausing by an order of magnitude (Toulokhonov et al., 2007). Moreover, a recent co-crystal structure of Pol II

elongation complex suggested that this residue may be important in translocation as it contacts the bridge helix through forming a wedged configuration (Brueckner and Cramer, 2008). Extensive studies on the functions and roles of other trigger loops residues have been conducted in budding yeast and this will be described in the following section.

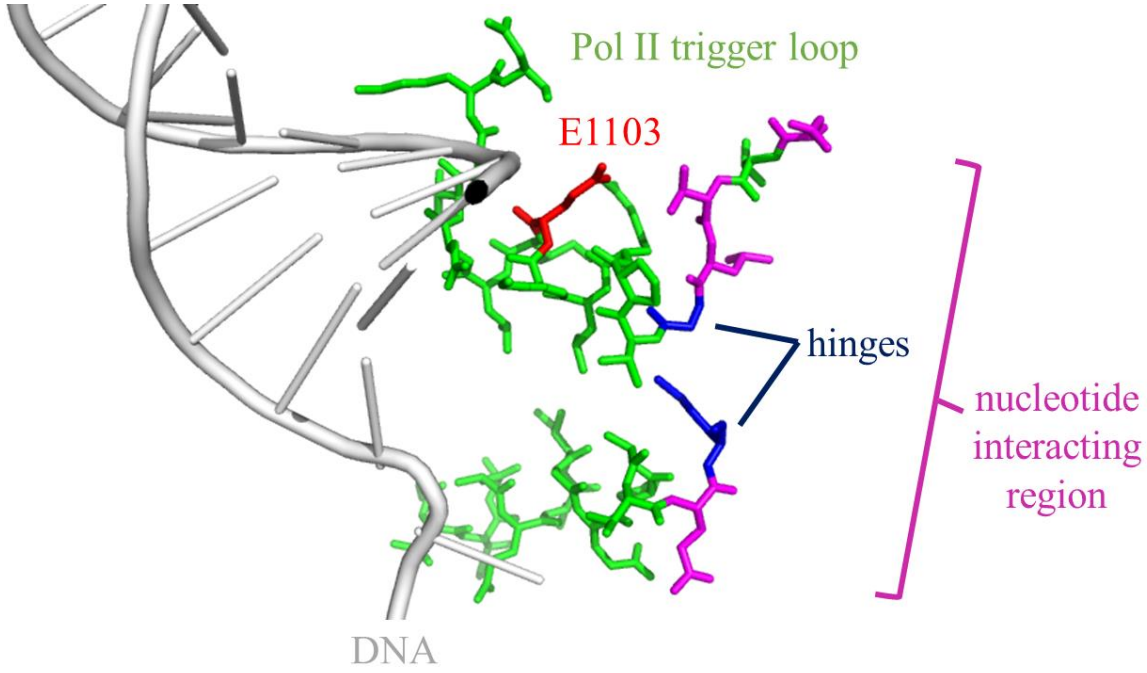


Figure 1.3.1 Budding yeast Pol II trigger loop. Pol II trigger loop was labelled in green; E1103 residue was labelled in red; nucleotide interacting region was labelled in pink; hinges of trigger loop was labelled in blue; DNA was labelled in grey. Crystal structure is obtained from PDB: 5XOG (Ehara et al., 2017) and modified using Pymol.

1.3.2 RNA Polymerase II mutants and transcription speed

Extensive studies in budding yeast have investigated the function of the trigger loop. Two mutations, *rpb1-N488D* which locates in the catalytic centre and *rpb1-E1103G*, which locates in the homology box G are the best characterized in budding yeast. These mutations showed opposite effects on the rate of *in vitro* polymerization and were first reported by (Malagon et al., 2006a). The study identified recessive mutations of *rpb1* by screening yeast strains with doxycycline repressible promoter in WT *rpb1*, and a low-copy plasmid carrying *rpb1* mutations

that can rescue the effect. Yeast clones that showed growth in the presence of doxycycline and defects in the presence of doxycycline and 6-azauracil (6AU), a drug that acts to reduce the GTP and UTP levels (Exinger and Lacroute, 1992), were analysed. Sensitivity to 6AU indicates the importance to transcription *in vivo*. N488 is located in close proximity to the NADFDGD motif, which is involved in coordinating one of two the magnesium ions (metal ion A) in the Pol II active site located downstream of the RNA–DNA hybrid. This residue is also close to the Pol II basic residues N445 and R446. Previous studies showed that impairment of transcription initiation is observed when N445 is affected, particularly the selection of transcription start site (Archambault et al., 1998; Berroteran et al., 1994). E1103 is located near the region where the template DNA strand is separated from the non-template strand (Figure 1.5.1 labelled in red), this region was shown to be involved in controlling the lateral mobility of elongation complexes (Bar-Nahum et al., 2005), and play an important role in NTP selection and incorporation (Cramer et al., 2001). *In vitro* transcription elongation assays revealed that Pol II with *rpb1-N488D* mutation showed a slower elongation rate than WT, whereas Pol II with *rpb1-E1103G* showed a faster elongation rate than WT (Malagon et al., 2006a), which was further confirmed *in vivo* (Hazelbaker et al., 2013; Jimeno-González et al., 2010). Further studies on fidelity and termination efficiency were conducted on the *rpb1-E1103G* mutant. *rpb1-E1103G* showed a 3-fold increase of transcription errors in vivo and 10-fold decrease in transcription elongation fidelity *in vitro*. *rpb1-E1103G* is also known to promote sequestration of NTP in the Pol II active centre and shifted the equilibrium towards the close confirmation of trigger loop (Kireeva et al., 2008), and has a preference for upstream TSS (Braberg et al., 2013). Moreover, *rpb1-E1103G* showed a slight increase in the number of pauses observed *in vitro* compared to WT (Larson et al., 2012). Transcription readthrough was reported in *rpb1-E1103G* for selected non-coding genes but not in *rpb1-N488D* (Hazelbaker et al., 2013). Interestingly, Pol II backtracking, which cleaves off mis-incorporated nucleotides, was shown

to be deficient in *rpb1-E1103G* (Larson et al., 2012), upregulation of chaperone and heat shock proteins were observed in mass spec data and protein aggregates were observed in yeast mutants (Vermulst et al., 2015). Introduction of *rpb1-N488D* mutation causes reduced Pol II CTD S2P levels without affecting the level of Pol II, suggesting this mutant is linked to decreased CTD phosphorylation (Jimeno-González et al., 2010).

Studies had investigated the functions of other residues in the trigger loop by introducing substitutions (Braberg et al., 2013; Kaplan et al., 2012, 2008; Qiu et al., 2016; Viktorovskaya et al., 2013). For example, all tested L1081 variants were lethal except L1081M showed a severe growth defect; and variants of H1085 A/N/D/F were lethal, K/R/W/Y showed severe growth defects, Q showed slight growth defects and L was viable and healthy (Braberg et al., 2013; Kaplan et al., 2012, 2008; Qiu et al., 2016), suggesting positively charged or polar residue could partially complement the loss of histidine. These trigger loop mutants can be divided into two main groups: loss of function (LOF) or gain of function (GOF). Both GOF and LOF mutants showed transcription defects. Table 1.5.2 showed an extracted listed of identified GOF/ LOF mutants from literature. In this study, we have created Pol II mutants E1106G and N494D in fission yeasts which are relative to E1103G and N488D in budding yeast. These mutants were chosen as they are the most established and characterized mutants in budding yeast showing opposite effects on transcription rate, and these effects are well documented in both *in vitro* and *in vivo* studies.

Table 1.3.2 Pol II active site mutants.

The table is created based on (Braberg et al., 2013; Kaplan et al., 2012, 2008; Malagon et al., 2006b; Qiu et al., 2016)

Residue/ Variant	Location Trigger Loop (TL) Catalytic Centre (CC)	Gain of function (GOF)/ Loss of function (LOF)
E1103G	TL	GOF
A1076T	TL	GOF
G1097D	TL	GOF
N488D	CC	LOF
H1085Y	TL	LOF
L1081M	TL	LOF
A1087V	TL	LOF

1.4 Objectives of this study

The primary aim of this thesis is to investigate the links between transcription and 3' end processing. My research question for this thesis is "how does transcription affect 3' end processing?" I have divided this question into two specific questions:

(a) How does transcription rate affect the function of trans-interacting factors that control 3' end processing?

Transcription rate is found to impact many processes including splicing, RNA processing and termination efficiency (Aslanzadeh et al., 2018; Hazelbaker et al., 2013). Pol II speed is highly dynamic *in vivo* (Maiuri et al., 2011), and it is known that Pol II slows down towards the gene end (Proudfoot, 2016). Thus, transcription rate plays a role in regulating 3' end processing and termination. Transcription rate will affect the "window of opportunity" for transcription factors or CTD-interacting proteins to bind Pol II and perform their functions, thus regulating the formation of 3' end and production of functional transcripts by affecting the selection of

polyadenylation site and termination window. My thesis aims to investigate how transcription rate affects 3' end processing and Pol II release by undertaking biochemical and genetic approaches.

In this thesis, I undertake *in vitro* studies to understand the links between transcription, 3' end processing and termination. Our approach uses the MultiBac system (Bieniossek et al., 2008) that allows the co-expression of multi-subunit complexes to reconstitute the whole 3' end processing machinery *in vitro*. We aimed to establish *in vitro* systems that coupled transcription to cleavage and polyadenylation to study how the speed of transcription would affect cleavage, polyadenylation and the combination of both. To study the effect of transcription rate on 3' end processing and termination, I generated "fast" and "slow" polymerases in fission yeast by making Pol II Rpb1 mutants based on previous studies on budding yeast and tested their effect in our established *in vitro* systems. This is accompanied by *in vivo* and genome-wide studies. In collaboration with Dr Krzysztof Kuś and Mr Alex Au, we have reconstituted the majority of the 3' end processing machinery in fission yeast including the polymerase module and phosphatase module in CPF, cleavage factor IB and components of cleavage factor IA.

(b) What is the role of transcription factor Seb1 in 3' end processing?

Function of the Pol II CTD is medicated by conserved CTD-interacting proteins. In mammals, it is known that SCAF4 and SCAF8 functions as mRNA anti-terminators, in which knocking out both SCAF4 and SCAF8 will result in short mRNA isoforms and a drastic change in early poly A site selection. Moreover, SCAF8, was proposed to act as positive Pol II elongation factor and it promotes transcriptional readthrough in the absence of the other factor SCAF4 (Gregersen et al., 2019). Seb1 is fission yeast ortholog of SCAF4 and SCAF8. Previous studies from Dr Sina Wittmann in Vasilieva lab (Wittmann et al., 2017) showed that Seb1 bind to CPF

and Pol II CTD, and is known to be important for cleavage and termination. Mutations in either CID or RRM domains of Seb1 resulted in 3' extended RNA species, a reduction in global gene expression as evidenced by a decrease in the total mRNA level, and reduced Pol II occupancy. I hypothesized that Seb1 would be the key factor in coupling transcription to 3' end processing. By understanding the role of Seb1 in transcription and 3' end processing, we may be able to shed light on to the mechanistic links between transcription and 3' end processing.

To help decipher the linkage between eukaryotic transcription, 3' end processing and termination, I aimed to undertake a biochemical approach to study the role of Seb1 in these processes. Using the established *in vitro* system, I studied how Seb1 affect transcription, what CPF subunits Seb1 interacts with, and how Seb1 affect polyadenylation. We aimed to further investigate the molecular linkage between transcription and 3' end processing using the established *in vitro* system that allows the coupling and uncoupling of transcription, cleavage, polyadenylation and termination.

2. Results

2.1. Setting up *in vitro* elongation assay to study transcription in fission yeast

To understand how transcription is mechanistically connected to mRNA 3' end processing, I first focused on setting up an *in vitro* transcription assay. To do that, I have generated different Pol II Rpb1 mutants based on the previous studies in *S. cerevisiae* that are predicted to affect Pol II processivity in different ways. The trigger loop and catalytic centre of Pol II is extremely conserved, not only in eukaryotic polymerases but also in bacterial polymerases. Studies in budding yeast have revealed that Rpb1 mutations E1103G and N488D show opposite effect on Pol II elongation rate *in vitro* and *vivo*. Located in homology box G in trigger loop, E1103G plays an important role in NTP selection and incorporation. N488D, located in the catalytic centre in close proximity to the NADFDGD motif, is involved in coordinating one of two the magnesium ions (metal ion A) in the Pol II active site (Kaplan et al., 2008; Kireeva et al., 2008; Malagon et al., 2006). E1103G is a gain-of-function mutation and known to produce Pol II with a faster elongation rate, whereas N488D is loss-of-function mutation and known to produce Pol II with slower elongation rate *in vitro* and *in vivo* (Malagon et al., 2006a; Qiu et al., 2016). Although there are extensive studies on these Pol II mutants on how they affect the processivity of Pol II, how these mutants contribute to the properties of Pol II and its ability to interact with and be regulated by transcription factors are not well understood. Therefore, we designed similar mutations in fission yeast using the available high-resolution x-ray structure of fission yeast Pol II (protein data bank (PDB): 3H0G) and studied how these mutations affect Pol II crosstalk with transcription factors.

To design similar Rpb1 mutations in our model organism, fission yeast, we have mapped out the relative positions of these mutations in *S. pombe* Rpb1 (Figure 2.1.1A) using the known X-ray crystal structure (PDB 3H0G) from a previous study (Spåhr et al., 2009). This analysis has

revealed that E1106G and N494D in *S. pombe* Rpb1 correspond to E1103G and N488D in *S. cerevisiae* and predicted to yield "fast" and "slow" Pol II respectively. The two mutations are highlighted in red, and the trigger loop (TL) is labelled in blue. N494D is located near to the catalytic centre, with a close proximity to the Mg^{2+} ion in the Pol II active site, whereas E1106G locates near the trigger loop and has an indirect effect on the sensitivity and incorporation of nucleotides.

2.1.1 Construction of Pol II mutants

To create "fast" and "slow" Pol II mutants, I have introduced the E1106G and N494D mutations into *S. pombe rpb1* locus in a yeast strain with Flag-tagged Rpb9 using pop-in pop-out method (Gao et al., 2014) and URA-colonies where selected and mutations were verified by DNA sequencing (Supplementary Figure 4).

To verify whether the E1106G and N494D Pol II Rpb1 mutations impact Pol II processivity, I initially employed a genetic approach. This was done by crossing the yeast strains with Rpb1 E1106G and N494D mutations with strains carrying mutations in the CTR of the essential elongation factor Spt5 T1A and T1E mutants. Spt5-T1A and Spt5-T1E mimic either unphosphorylated (T1A) or phosphorylated CTR (T1E). To investigate if there are any defects in cell growth, I spotted these strains on YES plates and incubated them at three temperatures: 25°C, 30°C (data not shown) and 37°C. This analysis has revealed that E1106G and N494D Rpb1 mutants show genetic interactions with Spt5 CTR mutants at 25°C and 37°C (selected images shown in Figure 2.1.1B). The "slow" mutation (N494D) showed an additive synthetically sick effect with Spt5 CTR mutants (T1A and T1E) when comparing to single mutants, whereas the "fast" mutation (E1106G) showed a slight rescue effect with Spt5 CTR

mutants (T1A and T1E) in contrast to single mutants. This indicated that E1106G and N494D Rpb1 mutations influence transcription and may have different processivity compared to WT.

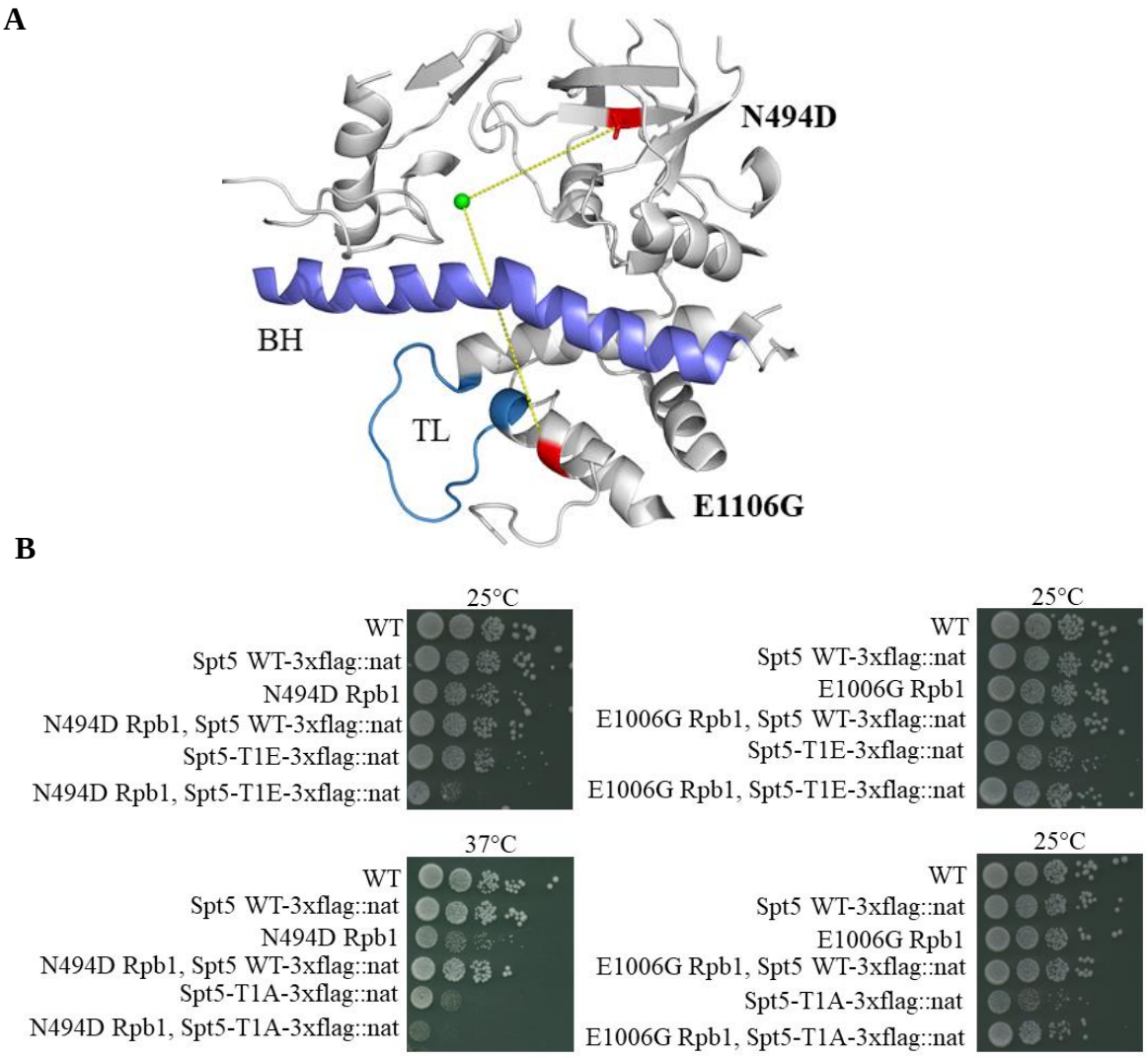


Figure 2.1.1 Construction and characterization of Pol II mutants.

(A) Model of *S. pombe* RNA Polymerase II catalytic centre with mutations. Mutations were modelled based on the X-ray structure of *S. pombe* RNA Polymerase II from Protein Data Bank (3H0G). Rpb1 showing the locations of Pol II Rpb1 mutants N494D and E1106G are highlighted in red using. Structure was rendered with Pymol software. Green sphere – one of Mg^{2+} forming catalytic centre, TL – trigger loop, BH – beta helix. Model was done by Dr Krzysztof Kuś.

(B) Spot test showing fast (E1106G) Pol II Spt5 mutants and slow (N494D) Pol II Spt5 mutants. Strains carrying fast and slow Pol II mutants were crossed with strains carrying Spt5 C-terminal region (CTR) mutants (Spt5-T1A and Spt5-T1E). Strains were backcrossed to clean the genetic background. Spot test was carried out on YES plates. Selected images showing genetic interactions between Pol II mutants and Spt5 mutants at 25°C and 37°C are shown.

2.1.2 Purification of native WT and mutant Pol II

In order to further test effect of introduced mutations on transcription, WT and mutated Pol II were purified to subsequently study their activity in *in vitro* transcription assays (Figure 2.1.2, lanes 2-4). Cells expressing WT Pol II (YP637), E1106G (YP1420) and N494D (YP1421) Pol II (Supplementary Table 3) were harvested, lysed, and purified by FLAG-affinity chromatography followed by ion exchange chromatography to remove nucleic acids. The elution fractions after final chromatography step using the HiTrap Q HP column were pulled together and their content was analysed using Coomassie stained SDS-PAGE, which nicely resolved 12 subunits of Pol II (Figure 2.1.2, lanes 2-4).

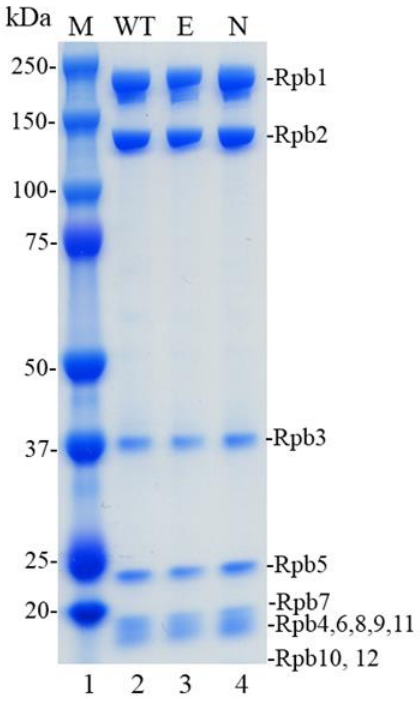


Figure 2.1.2 A Coomassie-stained SDS gel showing the 12-subunit Pol II wild type (WT) and Rpb1 mutants E1106G (E) and N494D (N). Pol II WT and mutants were subjected to 2-step purification (Flag affinity chromatography, ion exchange chromatography), and the figure showed the Pol II WT and mutants after size exclusion chromatography. Lane 1: marker, 2: wild type Pol, 3: "fast" Pol, 4: "slow" Pol

2.1.3 Studying effect of Pol II mutations on transcription *in vitro*

Next, I employed an *in vitro* approach to study effect of mutations on ability of Pol II to transcribe. Using an *in vitro* transcription elongation system previously established in the lab, I investigated how the E1106G and N494D mutations affect ability of Pol II to incorporate nucleotides, pausing frequency and ability to start transcription. The *in vitro* transcription assay uses assembled transcription complexes immobilised on beads including Pol II, template DNA, non-template DNA and RNA oligo to form the "transcription bubble". The ability of Pol II to incorporate nucleotides and initiate transcription can be assessed by the signal strength of the RNA bands corresponding to the run-off and intermediate reaction products. Two different templates are utilized to study pausing frequency, pause-induced HIV scaffold which contains HIV_T (template), HIV_NT (non-template), HIV_R (RNA) (Figure 2.1.3A); and general transcription scaffold which contains GT1_T, GT1_NT and RNA1. Pausing frequency can be studied by comparing the intensity and position of intermediate products between WT and mutated Pol IIs. The ability to initiate transcription is tested using RNA 3' labelling method (Figure 2.1.3B, Supplementary Figure 2), which visualize only transcriptionally competent Pol II by allowing Pol II to incorporate [α-^{32}P]-labelled rNTP at the 3'-end of the RNA. WT and mutant transcription complexes were assembled first on streptavidin beads in the presence of HIV RNA oligo, HIV DNA template strand and biotin labelled HIV non-template DNA strand (Figure 2.1.3A). Following assembly of the complexes, nascent RNA was labelled using [α-^{32}P]-rGTP incorporated by Pol II during 5 minutes of incubation, after which excess [α-^{32}P]-GTP was removed and reaction continued in the presence of cold rNTPs for an additional 10/ 30/ 60 s. This approach allows for detection of only transcriptionally active complexes that can incorporate nucleotides. Samples collected at time 0 correspond to the assembled complexes. Interestingly, the number of assembled complexes (as judged by time point 0), was lower for slow Pol II compared to fast and WT suggesting that slow Pol II is deficient in its ability to

initiate transcription (data not shown). In budding yeast, *rpb1-N488D* was proposed to affect transcription initiation or early elongation phases as evidenced by synthetic lethality with *soh1* deletion, which encodes for a subunit in the Mediator transcription complex (Malagon et al., 2006b).

In order to study the efficiency of elongation, we normalised the number of assembled complexes for each Pol II by using the same number of assembled complexes. Transcription was started after the addition of 0.1 mM cold rNTPs, and transcription was stopped at 0s, 10s, 30s and 60s by addition of stop solution with EDTA and formamide that chelates the magnesium ion in the Pol II catalytic centre. The analysis of the reaction products by 15% PAGE/ Urea has revealed that polymerase with E1106G mutation ("fast" Pol II) showed a slightly stronger run-off band compared to WT Pol II (Figure 2.1.3C, lanes 10-11). This indicates that this mutant incorporates nucleotides more efficiently compared to WT. In contrast, polymerases with N494D mutation ("slow" Pol II) showed slower transcription compared to WT Pol II (Figure 2.1.3C, lanes 10, 12). N494D inability to initiate transcription is hinted by studies in budding yeast N488D mutation, N494D (fission yeast) and N488D (budding yeast) residues are in close proximity to N445 and R446, and previous studies showed that impairment of transcription initiation is observed when N445 is affected, particularly the selection of transcription start site (Archambault et al., 1998; Berroteran et al., 1994). This concludes that the fission yeast Pol II Rpb1 mutants were behaving as expected (Qiu et al., 2016). Next, we assessed the pausing behaviour of mutated Pol IIs compared to WT. From Figure 2.1.3C, the pause sites (indicated by the intermediate bands) among WT and mutants Pol II are similar, with a minor difference in the strength of intermediate bands. This suggest that timing of each pause of Pol II mutants might be different from WT given their different processivity. Taken together, this data suggest that E1106G is a gain-of-function mutation,

which renders Pol II more efficient in nucleotide incorporation. Additionally, our data suggest that N494D is a loss-of-function mutation, which compromises Pol II ability to incorporate nucleotide and initiate transcription. I conclude that E1106G Pol II showed a faster elongation rate and N494D showed a slower elongation rate *in vitro* possibly due to altered ability of Pol II to incorporate nucleotides and frequency/ duration of pausing.

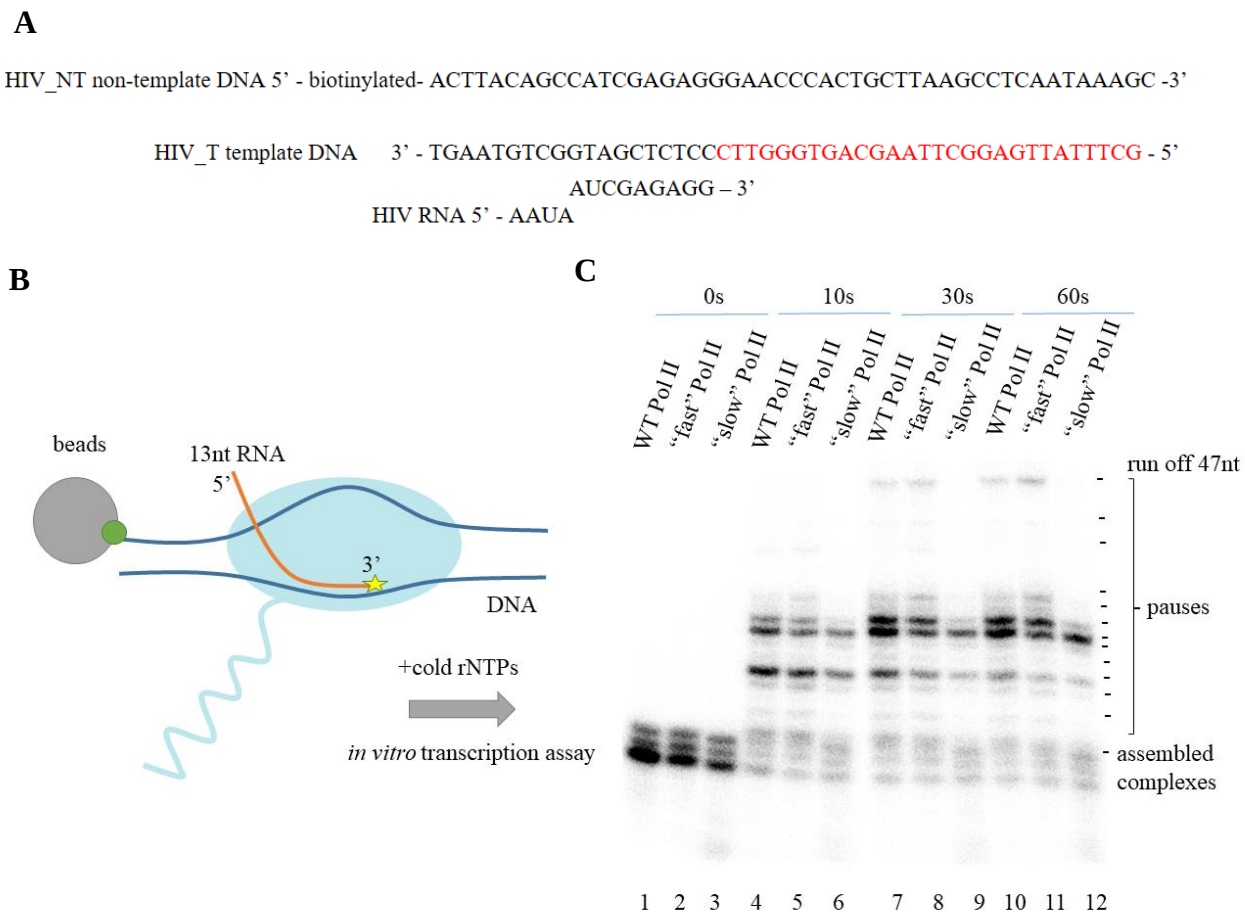


Figure 2.1.3 Characterization of Pol II mutants.

(A) **Sequences of HIV RNA, template and non-template DNA**. The location of where HIV RNA was annealed to HIV template DNA was shown.

(B) **Setup of *in vitro* transcription assay.** Complexes composed of Pol II, RNA, template and biotin labelled non-template DNA are assembled and bound to streptavidin beads by 10 minutes incubation at RT. For 3' labelling of nascent RNA in the assembled complexes, 2.5 µl of [α-^{32}P]-GTP for HIV RNA (37MBq 1 mCi in 100 µl) was added to the assembled complexes. Following 5 minutes of incubation 30°C to incorporate radioactivity at 3'end of nascent RNA, excessive [α-^{32}P]-GTP were washed away and transcription was started after addition of 0.1 mM cold rNTPs. Transcription was stopped at 0s, 10s, 30s and 60s by addition of stop solution.

(C) **A gel showing transcription assay for WT, fast and slow Pol II**. Lane1: WT Pol at 0s, 2: "fast" Pol at 0s; 3: "slow" Pol at 0s; 4: WT Pol at 10s, 5: "fast" Pol at 10s; 6: "slow" Pol at 0s; 7: WT Pol at 30s, 8: "fast" Pol at 30s; 9: "slow" Pol at 30s; 10: WT Pol at 60s, 11: "fast" Pol at 60s; 12: "slow" Pol at 60s;

2.1.4 Investigation of the ability of Pol II mutants to interact with elongation factor Spt4/5

To understand how transcription speed contributes to the ability of transcription factors to regulate transcription, I have investigated the effects of the Pol II mutants on the function of elongation factors. To do that I have performed *in vitro* transcription assay using WT, "fast" and "slow" Pol II in the presence of recombinantly expressed and purified heterodimer of elongation factors Spt4/ Spt5 (Figure 2.1.4A). In this assay, an RNA oligo was labelled at 5' with gamma ATP by PNK allowing detection of all assembled complexes (both transcriptionally active and inactive complexes) (Figure 2.1.3.1B). As mentioned in the previous section, Pol II WT and mutants have different abilities in initiating transcription, with N494D deficient in its ability to assemble and initiate transcription. Analysis of the reaction products by 15% PAGE/Urea has revealed that addition of Spt4/5 showed stimulatory effect on WT Pol II in agreement with known role of Spt4/5 in transcription elongation (Figure 2.1.4C, lanes 7-8) (Hirtreiter et al., 2010). Interestingly, both "slow" and "fast" polymerases are also stimulated by Spt4/5 (Figure 2.1.3.1C, lanes 9-10, 11-12). However, as the signal of "slow" Pol II is sub-optimal and weaker than that in WT and "fast" Pol II (Figure 2.1.3.1B, lanes 11-12) and the overall signals of the three Pol IIs are uneven (as evidenced by time point 0, assembled complexes in Figure 2.1.4C, lanes 1-6), which does not permit a conclusive comparison of the stimulatory effect of Spt4/5 between three mutants. I conclude that Pol II mutants E1106G and N494D did not lose the ability to interact with transcription factor Spt4/5, however I am unable to conclude if the mutation in Rpb1 influences stimulatory effect of Spt4/5 and to which extent.

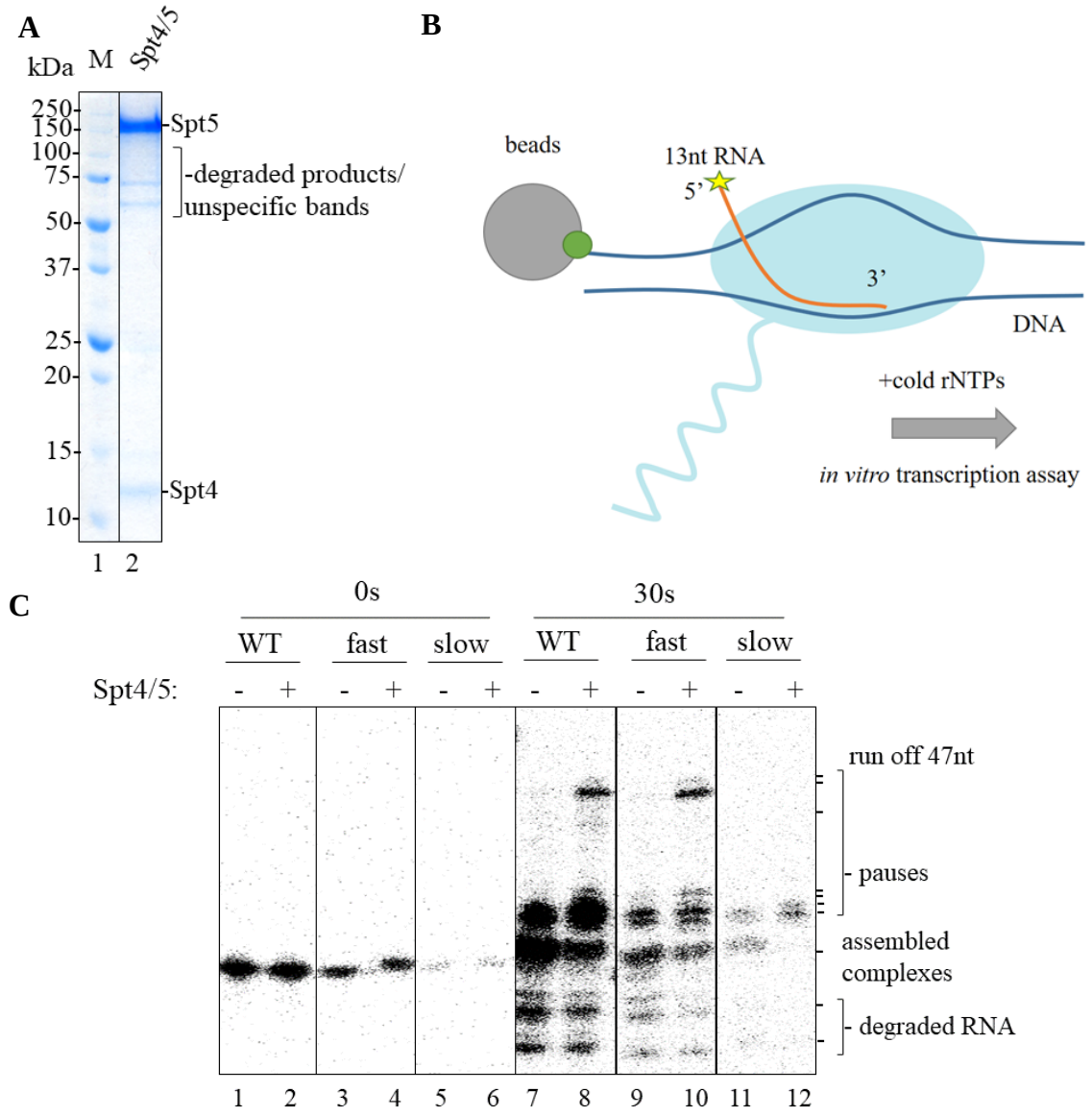


Figure 2.1.4 Spt4/5 stimulates transcription of WT, fast and slow Pol II *in vitro*.

(A) Coomassie-stained gel showing purified Spt4/5. Recombinant Spt4/5 (Spt5-6xHis) were purified from bacteria through Ni-NTA affinity chromatography followed by ion exchange chromatography and size exclusion chromatography. Molecular weight (MW) of Spt4: 12 kDa; Spt5: 110 kDa (migrated to around 150 kDa). Lane 1: marker; 2: Spt4/5

(B) Setup of *in vitro* transcription assay. Complexes involving Pol II, pre-kinated 5' labelled RNA, template and biotin labelled non-template DNA are assembled and bound to streptavidin beads by 10 minutes incubation at RT. 3 µl of buffer or Spt4/5 at 0.21 mg/ml (1.74 mM) was added to 6 µl of the assembled complexes. Following 10 minutes of incubation at 30°C, transcription was started after addition of 0.1 mM cold rNTPs. Transcription was stopped at 0s, 10s, 30s and 60s by addition of stop solution.

(C) A gel showing effect of Spt4/5 on WT, fast and slow Pol II transcription at 0s and 30s. Lane1: WT Pol II without Spt4/5 at 0s, 2: WT Pol II with Spt4/5 at 0s; 3: fast Pol II without Spt4/5 at 0s; 4: fast Pol II with Spt4/5 at 0s; 5: slow Pol II without Spt4/5 at 0s; 6: slow Pol II with Spt4/5 at 0s; 7: WT Pol II without Spt4/5 at 30s, 8: WT Pol II with Spt4/5 at 30s; 9: fast Pol II without Spt4/5 at 30s; 10: fast Pol II with Spt4/5 at 30s; 11: slow Pol II without Spt4/5 at 30s; 12: slow Pol II with Spt4/5 at 30s

2.1.5 Does Spt5 phosphorylation play a role in transcription?

As mentioned in the previous section, Spt4/5 stimulates Pol II processivity by binding to Pol II clamp coiled-coil motif, and Spt5 is phosphorylated during transcription. To test whether Spt5 phosphorylation plays a role in affect Pol II processivity (and ultimately speed of transcription elongation), we assessed the effects of Spt5 phosphorylation on transcription elongation by undertaking a biochemical approach. First, recombinantly reconstituted Spt4/5 was phosphorylated *in vitro* using recombinantly reconstituted Cdk9 associated cyclin Pch1 (Kilchert et al., 2020 accepted to Genome Research) in the presence of ATP. Phosphorylation of Spt5 was assessed by analysing the mobility of Spt5 (Spt5-P) on a 7.5% phosphotag gel. This has revealed a shift in mobility of Spt5 compared to Spt5 that was not incubated with Cdk9 (Figure 2.1.3.2A, lanes 1-2), indicating that Spt5 was fully phosphorylated. Next, I tested whether Spt5 phosphorylation affects Pol II ability in transcription.

To do that I have performed an *in vitro* assay using the same setup as Figure 2.1.4B. In this assay, the kinetics of nucleotide incorporation was assessed in the presence of phosphorylated and unphosphorylated Spt5. As can be seen from the analysis of the reaction products following 10 minutes of incubation in the presence of Spt4/Spt5-P and Spt4/Spt5 both complexes show similar ability to stimulate transcription independently of phosphorylation status of Spt5 as judged by no obvious difference between the signal strengths of the runoff product (Figure 2.1.5B, lanes 7-8). This suggest that phosphorylation of Spt5 does not change its properties to stimulate Pol II transcription at least in the defined *in vitro* system with minimal number of components. Based on our data, I speculate that effect of Spt5 phosphorylation on transcription inside the cell is likely to be mediated via trans-acting factors interacting with the CTR *in vivo*.

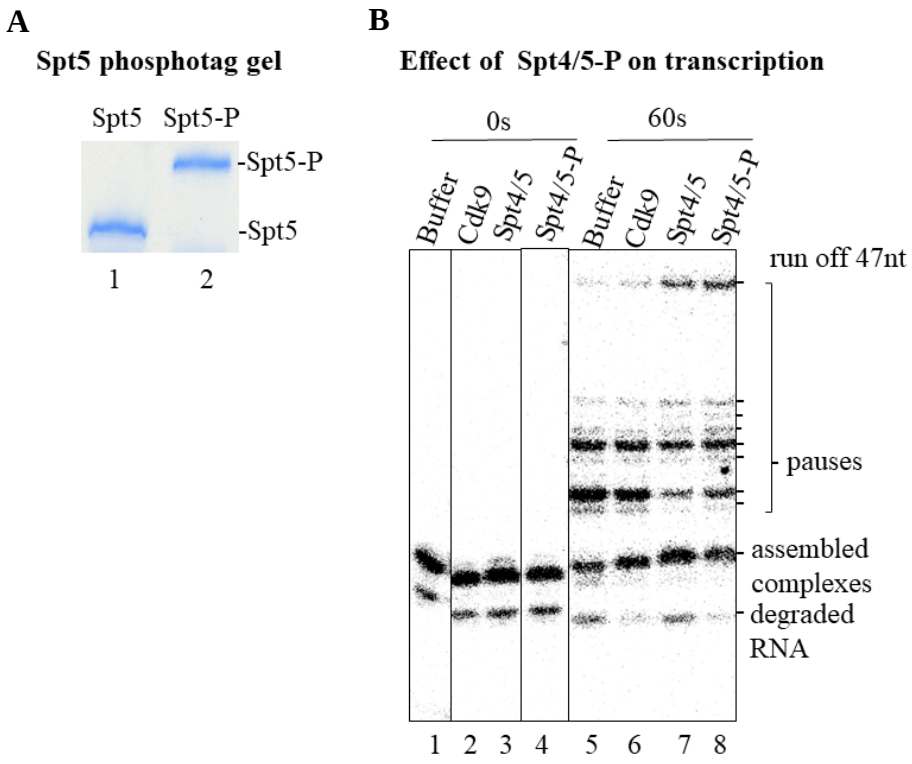


Figure 2.1.5 Spt4/5 phosphorylation has no effect on nucleotide incorporation and Pol II elongation.

(A) A phosphotag gel showing the phosphorylation of Spt5. 0.085 µg/ µl (703 µM) of Spt5 was phosphorylated by addition of 0.0075 µg/ µl (110.2 µM) Cdk9 in the presence of 0.51 mM ATP and 50 mM $MnCl_2$ and incubation at 30°C for 30 minutes. 1 µg of the phosphorylated Spt5 was boiled at 95°C for 5 minutes in 5x SDS loading buffer and loaded onto 7.5% acrylamide gel with 7.5 µM phosphor-tag and 50 µM $MnCl_2$.

(B) A gel showing the results of transcription assay with Spt4/5 at 0s and 60s. Complexes involving Pol II, pre-kinated 5' labelled RNA, template and biotin labelled non-template DNA are assembled and bound to streptavidin beads by 10 minutes incubation at RT. 3 µl of buffer, Cdk9 at 0.0075 µg/ µl (110.2 µM), Spt4/5 or Spt4/5-P at 0.085 mg/ml (703 µM) was added to 6 µl of the assembled complexes. Following 10 minutes of incubation at 30°C, transcription was started after addition of 0.1 mM cold rNTPs. Transcription was stopped at 0s and 60s by addition of stop solution. Lane 1: buffer at 0s; 2: Cdk9 at 0s; 3: Spt4/5 at 0s; 4: Spt4/5-P at 0s; 5: buffer at 60s; 6: Cdk9 at 60s; 7: Spt4/5 at 60s; 8: Spt4/5-P at 60s

2.1.6 How do Pol II mutations affect transcription *in vivo*?

To study how E1106G and N494D Pol II mutations affect transcription in living cells, Dr Takuya Kajitani has performed precision nuclear run-on sequencing (PRO-seq) analysis of transcription maps the location of active polymerases across *S. pombe* genome with nucleotide resolution in WT and Rpb1 E1106G and N494D mutants.

Strikingly, PRO-seq analyses revealed global loss of promoter proximal Pol II pausing in fast and slow mutants that can be observed on representative *rpl902* gene compared to the WT (Figure 2.1.6C). Previous studies reported a shift of Pol II upstream of the transcription start site (TSS) at *ADH1* in *rpb1-E1103G* in budding yeast, which might be a result of a cryptic initiation event (Braberg et al., 2013; Kaplan et al., 2012). As there is another gene on the same strand upstream, this suggests that either *rpb1-E1106G* is similar to *rpb1-E1103G* in budding yeast has a preference for upstream start site selection or there is termination defect. Previous studies in budding yeast showed that *rpb1-N488D* (analogous to "slow" mutation N494D in fission yeast) impaired transcription initiation as residue N445 is affected, particularly the selection of transcription start site (Archambault et al., 1998; Berroteran et al., 1994). Thus, the "slow" mutation we generated may have the similar effect on transcription initiation by affecting the basic residues close to the Pol II active site. I conclude that "slow" mutations showed difference in TSS selection compared to WT, leading to loss in promoter proximal pausing.

Interestingly, E1106G mutant ("fast") showed transcription beyond transcription termination site observed in WT on multiple genes as exemplified by *rps2602*, encoding 40S ribosomal protein S26 (Figure 2.1.6A, region after green line). Failure to terminate properly was reported for fast pol II mutant on sn/snoRNA genes in *S. cerevisiae* and in mammals due to defective

recruitment of termination factors Sen1 helicase and Xrn2 exonuclease (Fong et al., 2015; Hazelbaker et al., 2013), in budding yeast rpb1-*E1103G* showed transcription read-through for selected snoRNA genes, and in humans *rpb1-E1126G* showed preference to distal termination compared to WT (Fong et al., 2015, 2014). I conclude fission yeast *rpb1-E1106G* leads to global termination defects *in vivo*.

In addition, while 5' pausing is reduced, 3'end pausing appears to increase in "fast" mutant (Figure 2.1.6A). The increased 3' end pausing could reflect compromised 3'end processing, which is further supported by termination defect observed in this mutant.

Strikingly, N494D Pol II mutant ("slow") showed overall reduction in transcription compared to WT. Lower transcription is observed in gene regions of *rps2602* (Figure 2.1.6A). Our *in vitro* data provides an explanation for this suggesting that reduced transcription in this mutant is due to initiation and elongation defect.

Next, we assessed termination of housing keeping genes. Surprisingly, early termination in *rps2602* is observed in slow mutant (Figure 2.1.6A, region before green line). In budding yeast, it is known that *rpb1-N488D* leads to early termination and shorter distribution of pre-snR33 transcripts compared to WT (Hazelbaker et al., 2013) Our data suggests that the slower elongation rate moves the termination window upstream compared to WT, leading to early termination.

Several regions of the *S. pombe* genome such as telomeric and centromeric regions are known to be transcriptionally silenced. However, paradoxically, transcription is required for establishment and maintenance of heterochromatin within these regions (Holoch and Moazed,

2015). Although the mechanisms underpinning this are not well understood, it was proposed that these regions possibly differ in the way their transcription is regulated. Therefore, we have decided next to analyse effect of Pol II mutants on transcription within these regions. The snapshot of telomeric region of Chromosome 2 encompassing *tlh2* shows striking increase in transcription in slow Pol II mutant compared to WT that indicates loss of silencing in Figure 2.1.6B. Similar effect is observed at other transcriptionally repressed telomeric regions and centromeric regions (data not shown). Interestingly, loss of transcriptional silencing in telomeric and centromeric loci is observed in "slow" pol II mutant and to a lesser extend also for "fast" mutant compared to WT. This suggest that accurate establishment of proper transcription rate is critical for either propagation or maintenance of heterochromatin within these regions and its deregulation leads to loss of silencing.

Taken together, "fast" mutant showed loss of promoter proximal pausing, increase in 3' end pausing and transcription read-through; whereas "slow" mutant showed impairment in transcription initiation, elongation defects, early termination and a loss of silencing in normally silenced regions of the genome.

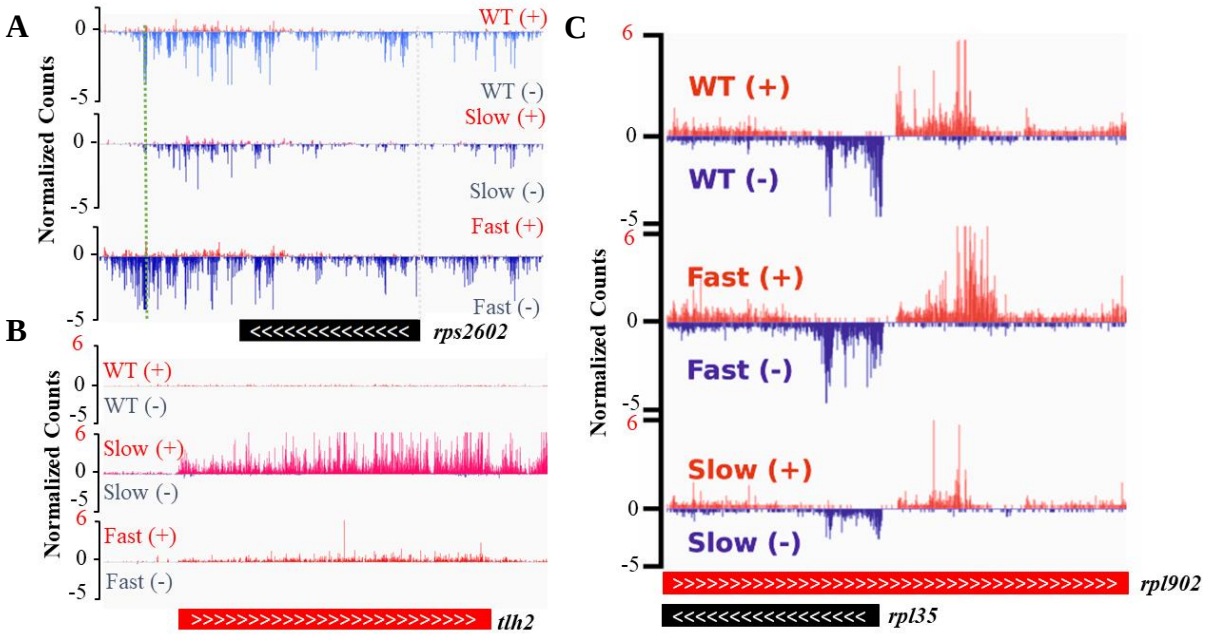


Figure 2.1.6 PRO-seq data on WT, fast and slow Pol II. PRO-seq experiment was carried out by Dr Takuya Kajitani using cells expressing WT (YP637), E1106G (YP1420) and N494D (YP1421) Pol II. Spike in *S. cerevisiae* (strain ATCC 204508/ S288c) control at 10% of *S. pombe* cell number was added to each sample to reduce sample processing bias. Sample preparation, nuclear run-on assays, adapter ligations, library generation and sequencing were performed according to (Mahat et al., 2016). The raw sequence data were processed by filtering out low quality reads and trimming adapter sequences. Sequences were then aligned to the genomic sequence and visualized in Integrative Genomics Viewer for downstream analysis. Selected gene regions *rps2602, tlh2, rpl35 and rpl902* are shown.

(A) Pol II WT, E1106G (fast) and N494D (slow) mutants PRO-seq data showing *rps2602* gene region. Arrows indicate the direction of transcription, and the coloured box indicates the location of the gene. Grey line: promoter region; green line: termination point of WT Pol II. Minus strand: blue; plus strand: red.

(B) Pol II WT, E1106G (fast) and N494D (slow) mutants PRO-seq data showing *tlh2* gene region. Arrows indicate the direction of transcription, and the coloured box indicates the location of the gene. Minus strand: blue; plus strand: red.

(C) Pol II WT, E1106G (fast) and N494D (slow) mutants PRO-seq data showing *rpl35* and *rpl902* gene region. Arrows indicate the direction of transcription, and the coloured box indicates the location of the gene. Minus strand: blue; plus strand: red.

2.1.7 Chapter conclusion

In this chapter, I have demonstrated the different processivity of Pol II mutants I constructed in *S. pombe*, through the established *in vitro* systems that allow the study of different aspects of Pol II processivity including nucleotide incorporation, pausing frequency and the ability to initiate transcription. E1106G is a gain-of-function mutation, which showed faster elongation rate (Figure 2.1.3C), loss of promoter proximal pausing, increase in 3' end processing pausing and termination defects (Figure 2.1.6). N494D is a loss-of-function mutation, which showed slower elongation rate compared to WT (Figure 2.1.3C), impairment in transcription initiation (section 2.1.3), less promoter proximal pausing, elongation defects and early termination (Figure 2.1.6). These mutants did not lose the ability to interact with elongation factor Spt4/5 (Figure 2.1.4C), however, it is not clear if the stimulatory effect observed for Pol II mutants is comparable to the WT. Moreover, Spt5 phosphorylation does not affect the elongation rate of Pol II at least *in vitro* suggesting that phosphorylation of Spt5 does not change its properties to stimulate Pol II transcription at least in the defined *in vitro* system with minimal number of components. Taken together, I have established a system that will provide a useful tool to understand how transcription rate affect processes such as mRNA 3' end processing that are coupled to transcription.

2.2. Investigating the role of Seb1 in transcription

2.2.1 Purification of Seb1

Previous studies from the Vasiljeva lab demonstrated that Seb1 is recruited at the end of transcription cycle and co-purifies with the CPF machinery (Wittmann et al., 2017), which is further confirmed by my data (Figure 2.2.1B).

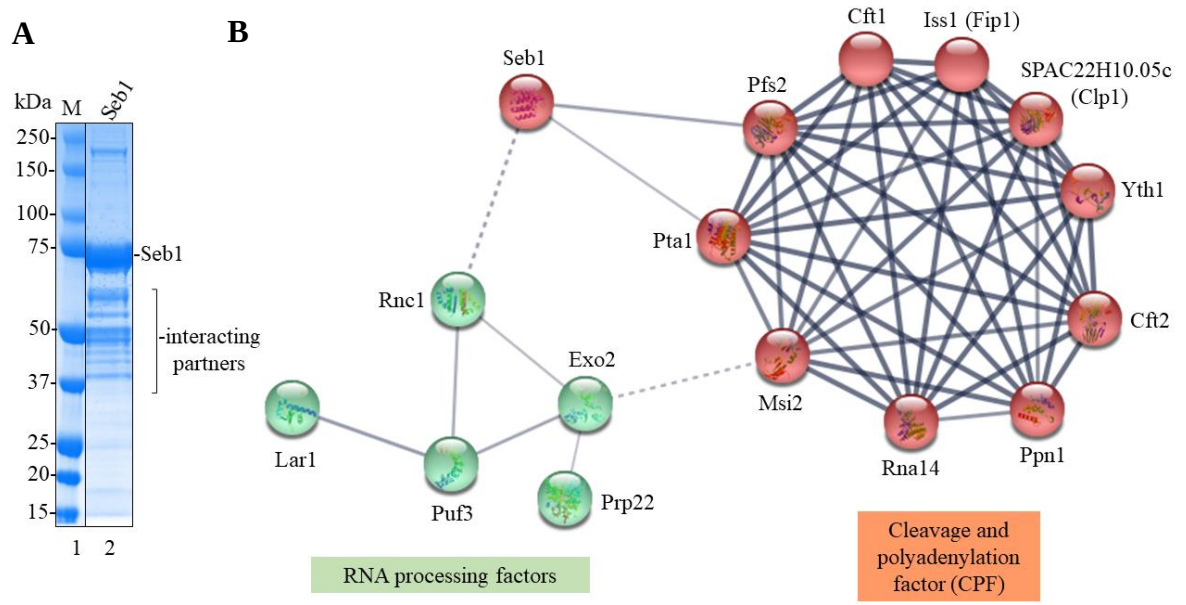


Figure 2.2.1 Seb1 interacts with cleavage and polyadenylation factor (CPF) complex and pre-mRNA-processing factors.

(A) A Coomassie-stained gel showing purified Seb1 and its potential interacting partners. Native Seb1 tagged with flag was purified from *S. pombe* through affinity chromatography and send to Advanced Proteomics Facility, Department of Biochemistry, University of Oxford for mass spectrometry (MS) and MS data analysis. Molecular weight (MW) of Seb1: 68 kDa (migrated to around 75 kDa). Lane 1: marker; 2: Seb1

(B) A protein association network showing potential interacting partners of Seb1. Abundant proteins such as metabolic, ribosomal and heat shock proteins were filtered, and intensity cut off 1E+8 was applied to the MS data of purified native Seb1. The data was then used to generate the network using STRING database. Disconnected nodes were hidden and interacting proteins were clustered into RNA processing factors and cleavage and polyadenylation factor. The brackets corresponding to the orthologs in budding yeast. Line thickness between protein nodes indicates the strength of data support (i.e. confidence based on interaction score calculated by STRING). Black line: highest confidence (interaction score 0.900); dark grey: high confidence (0.700); light grey: medium confidence (0.400); dotted light grey: low confidence (0.150).

2.2.2 Does Seb1 play a role in transcription elongation?

As mentioned in the introduction, Seb1 mutations in CID and RRM domains showed reduced Pol II occupancy within gene bodies suggesting elongation defects. This led me to ask whether Seb1 also has a role in transcription beyond termination and 3'end processing, hinted by its mammalian orthologue SCAF8, which was proposed to play a role in transcription elongation. To explore whether Seb1 could contribute to transcription elongation, I have performed transcription elongation assays in the presence and absence of Seb1 using the 3' labelling setup which only looks at the transcriptionally active complexes (Figure 2.2.2A). Strikingly, strong stimulation of transcription elongation can be observed upon addition of Seb1 as evidenced by the increased signal strength in runoff product (Figure 2.2.2B, lanes 3-4, 5-6). This suggests that Seb1 can stimulate processivity of Pol II.

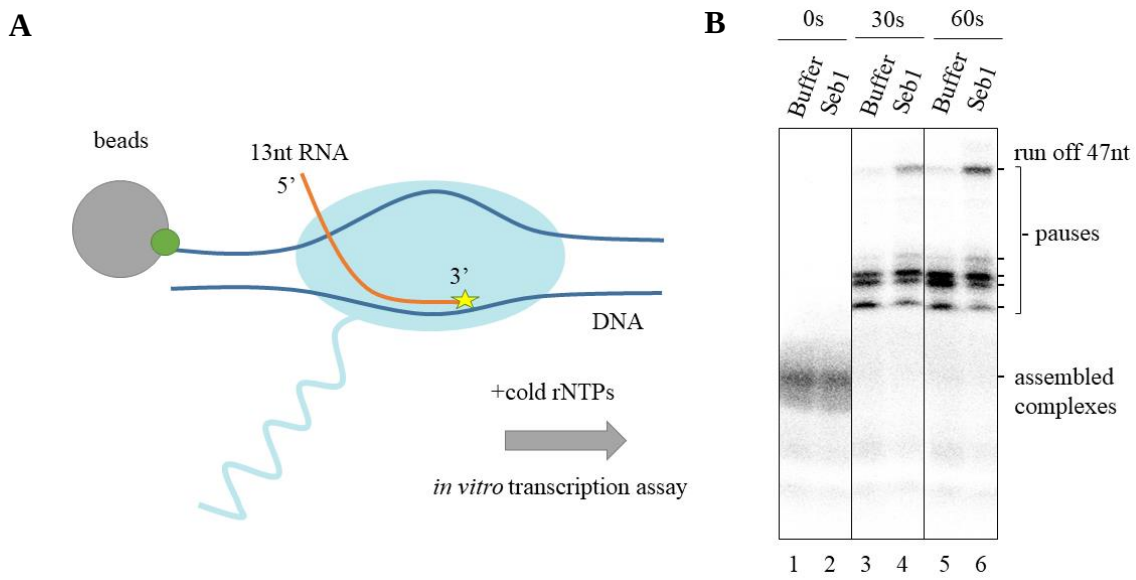


Figure 2.2.2 Seb1 stimulates transcription *in vitro*.

(A) **Setup of *in vitro* transcription assay.** Complexes composed of Pol II, RNA, template and biotin labelled non-template DNA are assembled and bound to streptavidin beads by 10 minutes incubation at RT. For 3' labelling of nascent RNA in the assembled complexes, 2.5 µl of [α-^{32}P]-GTP for HIV RNA (37MBq 1 mCi in 100 µl) was added to the assembled complexes. Following 5 minutes of incubation 30°C to incorporate radioactivity at 3'end of nascent RNA, excessive [α-^{32}P]-GTP were washed away. 3 µl of buffer or Seb1 at 0.5 mg/ml was added to 6 µl of assembled complexes and transcription was started after addition of 0.1 mM cold rNTPs. Transcription was stopped at 0s, 10s, 30s and 60s by addition of stop solution.

(B) **A gel showing transcription assay performed in the presence of with buffer only or 0.5 mg/ml Seb1.** Transcription was performed in the presence and absence of Seb1 and transcription stopped at 30 s and 60 s of incubation with 0.1 mM rNTPs following addition and incubation of seb1. Lane 1: buffer at 0 s, 2: Seb1 at 0 s, 3: buffer at 30 s, 4: Seb1 at 30; 5: buffer at 60 s, 6: Seb1 at 60s

2.2.3 Does Seb1 affect elongation rate to same extent in Pol II WT and its mutants?

To further investigate whether Seb1 can stimulate Pol II processivity to the same extent in Pol II WT and mutants, I have performed transcription elongation assay with and without Seb1 using the GT1 scaffold (Figure 2.2.3A) to look at the general transcription. The results revealed that Seb1 removed intermediate pauses in both WT and "fast" Pol II as evidenced from reduction of the pauses that Pol II enter at position 4 ("fast" Pol II) and 6 (WT Pol II) of the template (Figure 2.2.3B, lanes 7-8, 9-10). Seb1 has stimulatory effect on all 3 polymerases as shown by the increase in signal strength of runoff products (Figure 2.2.3B, lanes 7-12) as well as an increase of run off product (Figure 2.2.2.3C, position 2500). When looking at the intensity of runoff products (Figure 2.2.3C, position 2500), upon addition of Seb1, slow Pol II showed an increase in normalized photo-stimulated luminescent from 2 to 3.5 $x10^{-3}$ (blue line), whereas fast Pol II showed an increase from 4 to 5.5 $x10^{-3}$ (orange line) and WT from 2 to 5 $x10^{-3}$ (yellow line). We conclude that Seb1 affects Pol II processivity *in vitro* by removal of intermediate pauses. WT Pol II was stimulated better compared to Pol II mutants, this may be due to the intrinsic changes of Pol II properties in "fast" and "slow" mutant, affecting the interaction with Seb1. Similarly, one of the two Seb1 orthologs, SCAF8, was proposed to act as positive Pol II elongation factor based on *in vivo* data promoting transcriptional readthrough in the absence of the homologues factor SCAF4 (which similar to Seb1 co-purifies with CPF components) (Gregersen et al., 2019).

A

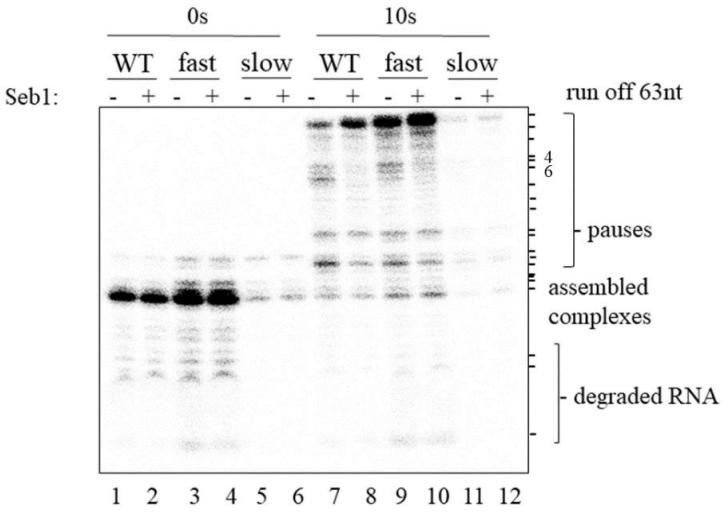
GT1_NT non-template DNA 5' - biotinylated - GGTCAAGGCAGTACTAGTAATGACCAGGCTCAATTGAGCTTGGAGTCAGTCGACGATGACTGG -3'

GT1_T template DNA 3'-CCAGTTCCGTCATGATCATTACTGGTCCGAGTTAACTCGAACCTCAGTCAGCTGCTACTGACC -5'

GACCAGGC - 3'

RNA1 5'- UGCUGUAAUAAA

B



C

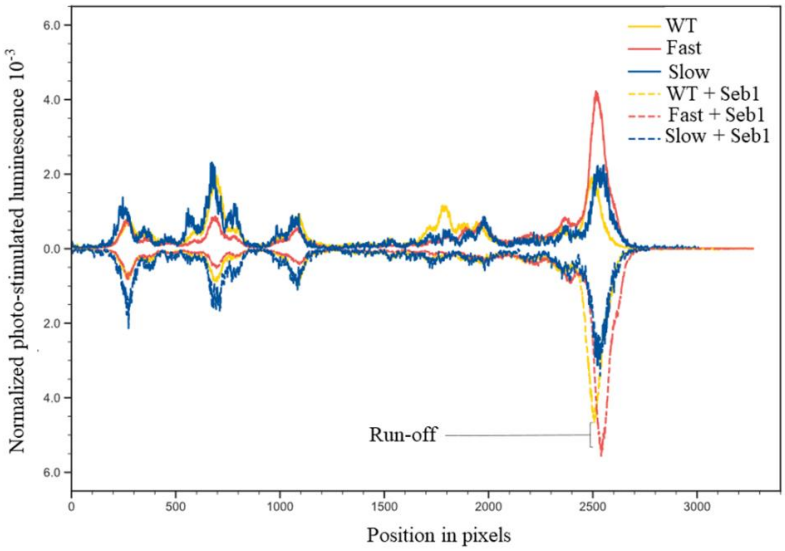


Figure 2.2.3 Seb1 remove pauses in transcription.

(A) Sequences of RNA1, GT1 template and non-template DNA. The location of where RNA1 was annealed to GT1 template DNA was shown.

(B) A gel showing effect of Seb1 on WT, fast and slow Pol II transcription at 0s and 10s. Lane1: WT Pol II without Seb1 at 0s, 2: WT Pol II with Seb1 at 0s; 3: fast Pol II without Seb1 at 0s; 4: fast Pol II with Seb1 at 0s; 5: slow Pol II without Seb1 at 0s; 6: slow Pol II with Seb1 at 0s; 7: WT Pol II without Seb1 at 10s, 8: WT Pol II with Seb1 at 10s; 9: fast Pol II without Seb1 at 10s; 10: fast Pol II with Seb1 at 10s; 11: slow Pol II without Seb1 at 10s; 12: slow Pol II with Seb1 at 10s

(C) A graph showing the quantification of runoff products from transcription assay with Seb1 at 10s. The bands were quantified in pixels and the background was subtracted from the measurements. The signal has been normalized to the total lane signal. The quantification was done by Dr Krzysztof Kuś with custom Python script.

2.2.4 Seb1 role in 3’ end processing and termination

To decipher the linkage between eukaryotic transcription, 3’ end processing and termination, I undertook a biochemical approach to study the role of Seb1 in these processes.

2.2.4.1 Is Seb1 a termination factor?

As discussed in the introduction, Seb1 CID and RRM mutants showed defects in 3’ end processing and termination *in vivo* (Wittmann et al., 2017), showing that Seb1 is important in 3’ end processing and termination. To understand if Seb1 directly contributes to dislodgement of Pol II from DNA, we have performed *in vitro* termination assay. This was designed and performed by Dr Krzysztof Kuś in collaboration with Nikolay Zenkin’s lab (University of New Castle). The idea behind this is to allow transcription of Pol II till the end of the template (transcription for 1 min), deactivate Pol II by chelating the magnesium ion in Pol II active site, and incubate Seb1 with the transcription complexes to see if Seb1 can release Pol II and mRNA from the template DNA (Supplementary Figure 3). Transcription is first carried out with the same procedure as 5’ labelled elongation assay (Supplementary Figure 1, Steps 1-4/ Supplementary Figure 3, Step 1). Then, transcription was started by addition of cold rNTPs and stopped after 1 minute by addition of EDTA (Supplementary Figure 3, Step 2). Seb1 was added to the transcription complexes and incubated for 10 minutes (Supplementary Figure 3, Step 3), the beads were centrifuged (Supplementary Figure 3, Step 4) and resulting in two fractions – supernatant and beads fraction. There are two scenarios for the results of termination assay: Scenario 1: Seb1 releases Pol II and mRNA, and Seb1, Pol II and mRNA will be found in supernatant fraction; Scenario 2: Seb1 does not release Pol II and mRNA, and Pol II and mRNA will remain bound in the transcription complex and found in the beads fraction (Figure 2.2.4.1A). The analysis of supernatant fractions with and without Seb1 (Figure 2.2.4.1B, lanes 6, 8) revealed that there is no obvious difference in mRNA product between the two supernatant

fractions. This data suggested that Seb1 is not a "bona fide" termination factor, rather it contributes to mRNA 3'end processing via CPF, which explains why Seb1 mutants show 3'end processing defect and as a result failure to terminate via torpedo model.

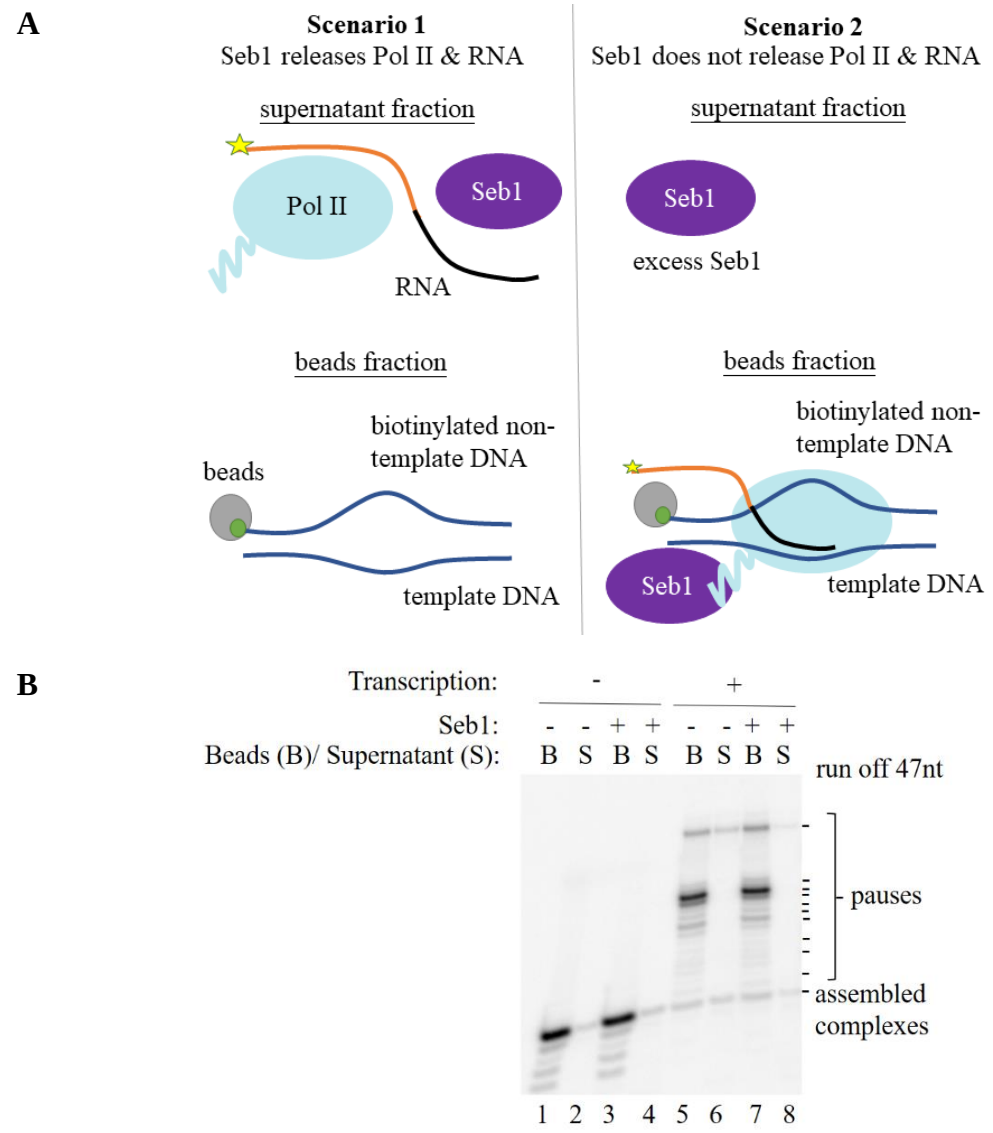


Figure 2.2.4.1 Seb1 is not sufficient to act as a termination factor *in vitro*.

(A) A diagram showing the rationale behind termination assay (Supplementary Figure3). Scenario 1: Seb1 releases Pol II and mRNA, Seb1, Pol II and mRNA will be found in supernatant fraction. Scenario 2: Seb1 does not release Pol II and mRNA, Pol II and mRNA will remain bound in the transcription complex and found in the beads fraction.

(B) A gel showing results of termination assay. Lane1: beads without Seb1 and transcription; 2: supernatant without Seb1 and transcription; 3: beads with Seb1 and without transcription; 4: supernatant with Seb1 and without transcription; 5: beads without Seb1 and with transcription; 6: supernatant without Seb1 and with transcription; 7: beads with Seb1 and transcription; 4: supernatant with Seb1 and transcription.

2.2.4.2 Reconstitution of 3’ end machinery *in vitro*

To further understand relationship between transcription and RNA processing we aim to reconstitute the whole 3’ end processing machinery *in vitro*; and establish mRNA 3’end processing coupled *in vitro* systems that combines transcription and mRNA 3’end processing. To reconstitute the fission yeast 3’end processing machinery, we have generated constructs for expression and purification of individual functional modules using the MultiBac system (Bieniossek et al., 2008). This system has been successfully employed in the lab to reconstitute polyadenylation module, phosphatase module of the CPF, 10 subunit RNA exosome complex and work is being in progress to reconstitute the nuclease module of the CPF.

To reconstitute Cleavage Factors IA and IB (CFIA and CFIB), I have cloned Msi2 (CFIB) tagged with Strep and optimized sequences of Pcf11, Rna14, Rna15 and Ctf1 and non-optimized Clp1 (components of CFIA) to pACEBac1 (Supplementary Table 2), and generated P1 and P2 viruses for recombinant protein expression in Sf9 insect cells. Our previous work suggest that Seb1 co-purifies with CPF subunits, however it is not known what module and subunits Seb1 interacts as a part of CPF. To find out how Seb1 interacts with CPF, I have also created a construct for expression of Seb1 in insect cells to be able to co-express Seb1 with other components. Recombinant Seb1 tagged with 6xHis was purified from bacteria (Figure 2.2.4.2IA, lane 2) by Ni-NTA affinity chromatography followed by size exclusion chromatography. Both Pcf11-8xHis-Clp1-Strep heterodimer, components of CFIA (Figure 2.2.4.2IB, lane 2), co-infected in the ratio of 49:1 in Sf9 cells; and Msi2, Cleavage Factor IB (Figure 2.2.4.2IC, lane 2) were purified by 3-step purification: Step affinity chromatography, ion exchange chromatography and size exclusion chromatography. The full CPF polymerase module (Figure 2.2.4.2ID, lane 2 by Dr Krzysztof Kuś); and the full CPF phosphatase module except Dis2 (Figure 2.2.4.2IE, lane 2 by a rotation student, Mr. Ho Alex Au) were also purified

by 3-step purification: Step affinity chromatography, ion exchange chromatography and size exclusion chromatography.

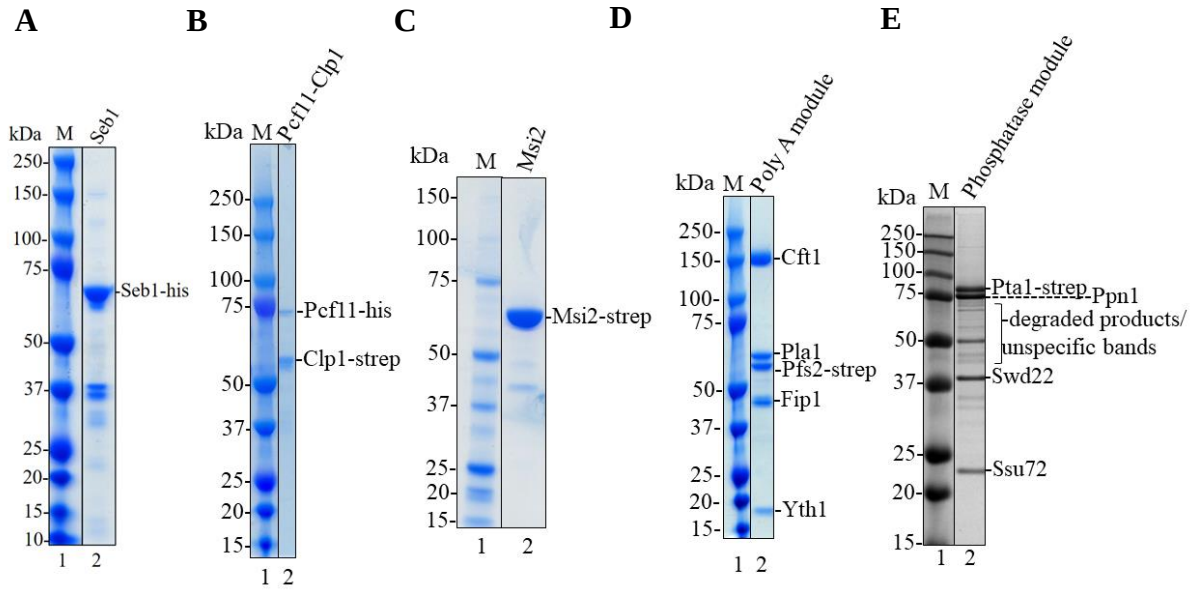


Figure 2.2.4.2I Reconstitution of 3' end processing machinery.

(A) **A Coomassie-stained SDS gel showing the recombinant Seb1 tagged with his.** The gel showed Seb1 after 2-step purification (his affinity chromatography and size exclusion chromatography) from bacteria. Lane 1: marker, 2: Seb1-his

(B) **A Coomassie-stained SDS gel showing the recombinant Pcf11-Clp1 tagged with his and strep respectively.** The gel showed the Pcf11-Clp1 hterodimer after 3-step purification (step affinity chromatography, ion exchange chromatography and size exclusion chromatography). Pcf11 and Clp1 was expressed together by co-infection in a ratio of 49:1 in Sf9 cells. MW of Pcf11-his: 71 kDa; Clp1-strep: 51 kDa. Lane 1: marker, 2: Pcf11-his Clp1-step

(C) **A Coomassie-stained SDS gel showing the recombinant Msi2 tagged with strep.** The gel showed Msi2 after 3-step purification (strep affinity chromatography, ion exchange chromatography and size exclusion chromatography). Msi2 was expressed in Sf9 cells. MW of Msi2-sterp: 53 kDa. Lane 1: marker, 2: Msi2-step

(D) **A Coomassie-stained SDS gel showing the recombinant Poly A module.** Pfs2 was tagged with strep. Poly A module was purified by Dr Krzysztof Kuś. MW of Pfs2-strep: 59 kDa; Cft1: 160 kDa; Pla1: 64 kDa; Fip1: 44 kDa; Yth1:19 kDa. Lane 1: marker, 2: Poly A module of CPF

(E) **A Coomassie-stained SDS gel showing the recombinant Poly A module.** Pta1 was tagged with strep. Phosphatase module was purified by Mr Alex Au. MW of Pta1-strep: 82 kDa; Ppn1: 77 kDa; Swd22: 39 kDa; Ssu72: 23kDa. Lane 1: marker, 2: Phosphatase module of CPF.

To verify if the protein complexes we purified are active functional complexes, I assessed RNA binding activity of the complexes by performing the electrophoretic mobility shift assay (EMSA) with Msi2 and CPF polymerase module using *rps2* WT and *rps2* WT F2 RNA (Figure 2.2.4.2IIA/ Supplementary Table 5). The results showed a band shift, indicating that the purified polymerase module (Figure 2.2.4.2IIB, lanes 3-6) and Msi2 (Figure 2.2.4.2IIC, lanes 6-9) bind to RNA and were active.

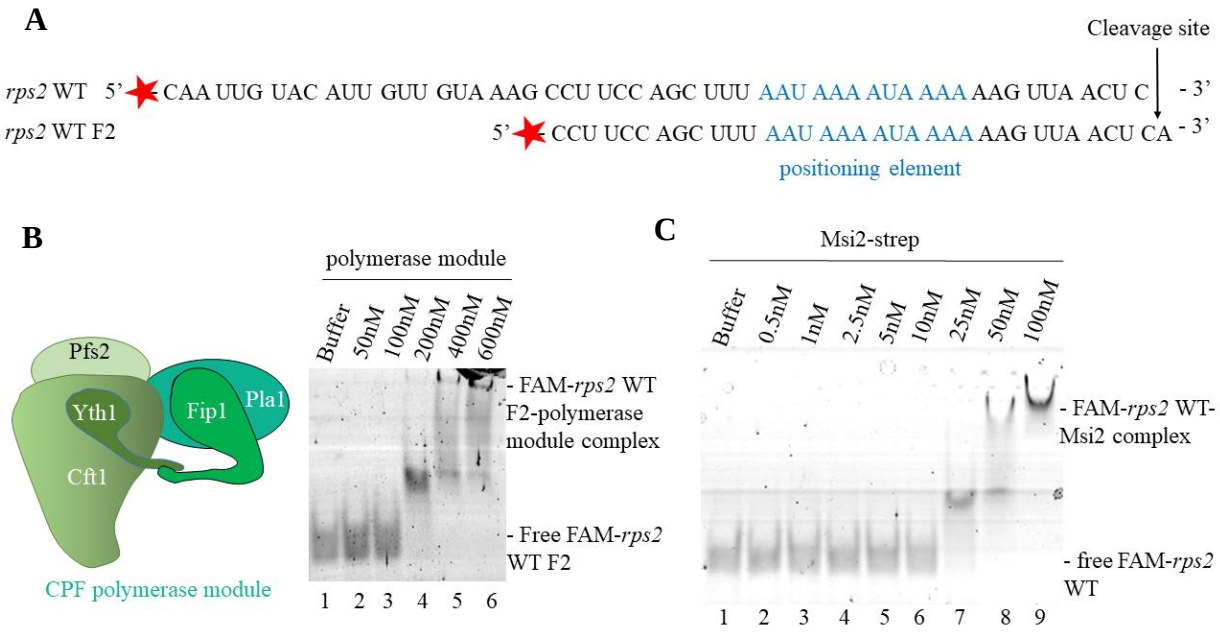


Figure 2.2.4.2II Purified Msi2 and Poly A module bind to PAS *in vitro*

(A) Sequence of FAM-labelled *rps2* WT and *rps2* WT F2. 5' FAM labelled RNA oligo spanning 3' UTR of *rps2* WT containing PAS (*rps2* WT, *rps2* WT F2 RNA, Supplementary Table 6).

(B) A gel showing Electrophoretic Mobility Shift Assay (EMSA) of polymerase module of CPF. 5' FAM labelled RNA oligo spanning 3' UTR of *rps2* containing PAS (*rps2* WT F2 RNA, Supplementary Table 6) in a final concentration of 50 nM was used. Lane 1: buffer; 2: 50 nM Poly A module; 3: 100 nM Poly A module; 4: 200 nM Poly A module; 5: 400 nM Poly A module; 6: 600 nM Poly A module

(C) A gel showing EMSA of Msi2 (Cleavage Factor IB). 5' FAM labelled *rps2* WT RNA in a final concentration of 50 nM was used. Lane 1: buffer; 2: 0.5 nM Msi2; 3: 1 nM Msi2; 4: 2.5 nM Msi2; 5: 5 nM Msi2; 6: 10 nM Msi2; 7: 25 nM Msi2; 8: 50 nM Msi2; 9: 100 nM Msi2

2.2.4.3 Which CPF subunit does Seb1 interacts with?

Our previous work suggested that Seb1 co-purifies with CPF subunits, however it is not known which module/ subunit Seb1 interacts as a part of CPF. To find out how Seb1 interacts with CPF, I performed pull down assays using Sf9 cells co-infected with Seb1 and Pcf11-Clp1, and Seb1 and Msi2 (Figure 2.2.4.3A).

The analysis of different fractions of the pulldown assay has revealed that Seb1 was presented in the flow through and wash 1 (Figure 2.2.4.3B, lanes 2-3), indicating that Seb1 does not interact with Pcf11-Clp1 heterodimer. Similarly, Seb1 was found in the flow through and wash 1 (Figure 2.2.3.3C, lanes 2-3) in the pull-down with Msi2, indicating that Seb1 does not interact with Msi2. Clp1 and Msi2 were found as potential interacting partners in my MS data (Figure 2.2.1B), however, they may be interacting with the presence of RNA, which forms a scaffold for binding and facilitating the interactions between components of 3' end machinery; or Msi2/ Pcf11-Clp1 co-purified with Seb1 interacting partners. I conclude that Seb1 does not interact directly with Pcf11-Clp1 nor Msi2 without the presence of RNA.

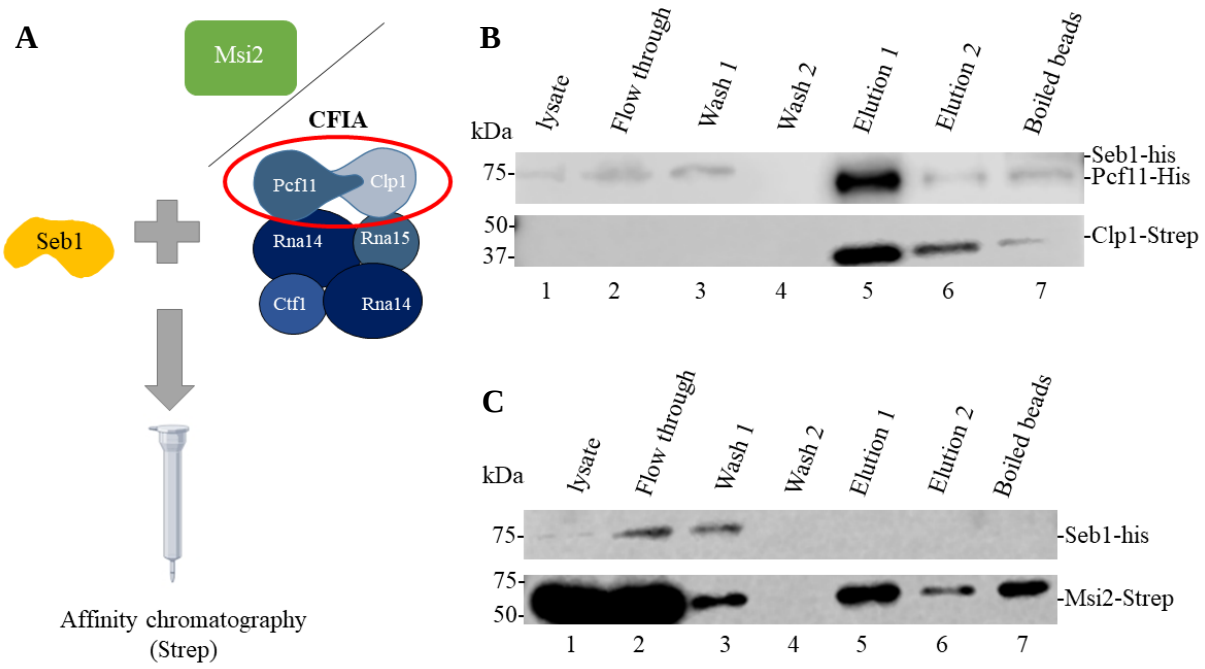


Figure 2.2.4.3 Seb1 does not interact with Pcf11-Clp1 nor Msi2.

(A) **A diagram showing the setup of the pull-down experiment.** Seb1-his was co-infected with Msi2-strep or Pcf11-his Clp1-strep, and affinity chromatography (strep) was performed using the lysate of co-infected cells.

(B) **A Western blot image showing the results of the pull-down experiment of Seb1 and Pcf11-Clp1 heterodimer.** Lane 1: lysate; 2: flow through; 3: wash 1; 4: wash 2; 5: elution 1; 6: elution 2; 7: boiled beads.

(C) **A Western blot image showing the results of the pull-down experiment of Seb1 and Msi2**. Lane 1: lysate; 2: flow through; 3: wash 1; 4: wash 2; 5: elution 1; 6: elution 2; 7: boiled beads.

2.2.4.4 Does Seb1 affect polyadenylation?

To understand what roles Seb1 plays in 3' end processing, I collaborated with Dr Krzysztof Kuś who performed *in vitro* polyadenylation assay in the presence of Seb1 to understand if Seb1 have an effect on polyadenylation by the polyadenylation module of CPF. Purified polymerase module was incubated with *rps2* WT RNA (Supplementary Table 5) for 30 minutes at room temperature with and without Seb1. Interestingly, the analysis of the assay results revealed that Seb1 suppresses polyadenylation as evidenced by disappearance of smear (polyadenylated RNA) at position 1, and a band is observed at position 35 nt (starting position for polymerase module with RNA at 0 min). This data suggests that Seb1 hinders

polyadenylation alone with polyadenylation module. Previous studies of Seb1 mammalian orthologs SCAF4 and SCAF8 showed that SCAF4 and SCAF8 are known to bind nascent RNA upstream of early poly(A) sites and act as anti-terminators to prevent early cleavage and polyadenylation of mRNA transcript. Our data showed that Seb1 may have a similar role preventing early polyadenylation. I conclude that Seb1 plays a role in suppressing polyadenylation *in vitro*. However, it is also possible that Seb1 competes for RNA binding with the polyadenylation module, therefore further studies are necessary in a context of fully reconstituted cleavage and polyadenylation system to understand role of Seb1 in mRNA 3'end formation.

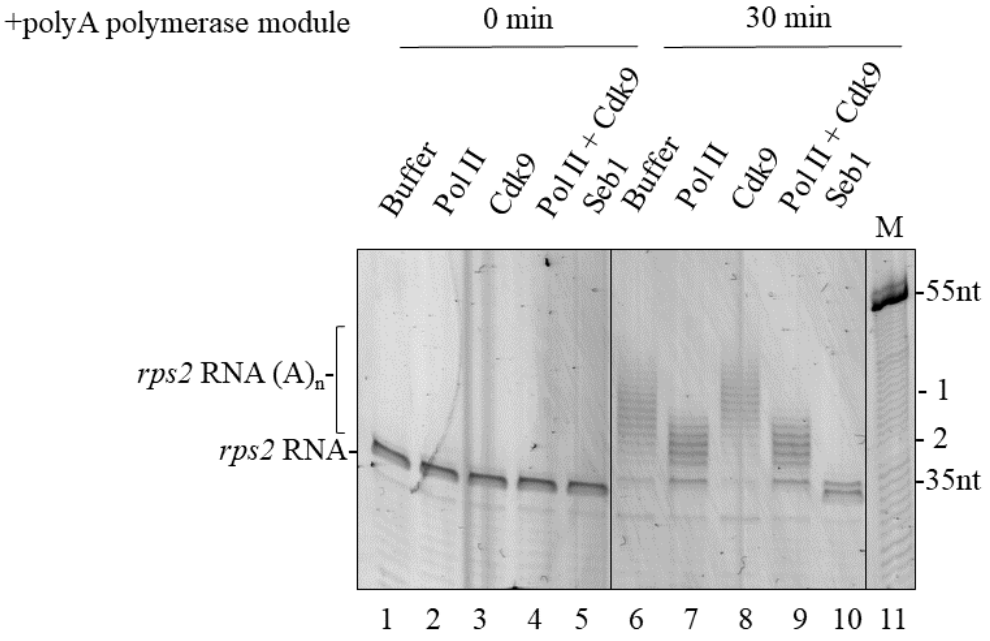


Figure 2.2.4.4 Seb1 suppresses polyadenylation. Polyadenylation assay was conducted by Dr Krzysztof Kuś. 2.5 µl of FAM-labelled *rps2* WT F2 RNA (Supplementary Table 5) at 1 µM was incubated at room temperature with 3µl of Poly A module at 1.6 µM and 0.2 µl of 100 mM ATP in 4.3 µl Buffer AS and kept in dark. The reaction at 0, 30 and 60 minutes was stopped by adding the 1 µl of stop solution (130 mM EDTA, 5% (w/v) SDS, 12 mg/ ml proteinase K in Buffer AS) to 10 µl of reaction mixture. The reaction mixture with stop solution was incubated for 15 mins at RT and 2 µl 6xLDE buffer was added to the mixture. The samples were loaded onto a Novex™ TBE Gel 4-20% and gel electrophoresis was carried out at 200V for 30 mins in 0.5x TBE. Lane 1: buffer (with polymerase module and RNA) at 0 min; 2: Pol II at 0 min; 3: Cdk9 at 0 min; 4: Pol II and Cdk9 at 0 min; 5: Seb1 at 0 min; 6: buffer at 30 min; 7: Pol II at 30 min; 8: Cdk9 at 30 min; 9: Pol II and Cdk9 at 30 min; 10: Seb1 at 30 min; 11: 55 nt marker

2.2.5 Chapter conclusion

In this chapter, I have demonstrated that Seb1 is an elongation factor and stimulates Pol II elongation rate (Figure 2.2.2B) but not a "bona fide" termination factor *in vitro* as it does not release Pol II and mRNA from the template DNA (Figure 2.2.4.1B). Seb1 plays a role in 3' end processing and termination via the CPF and appears to supress polyadenylation of the polymerase module (Figure 2.2.4.4). Thus, Seb1 may act as an anti-terminator like its orthologs in mammals SCAF4 and SCAF8 to suppress early cleavage and polyadenylation. The next step is to reconstitute the CPF nuclease module *in vitro* and investigate if Seb1 have an effect in cleavage and how it affects polyadenylation in a context of fully reconstituted system. Seb1 does not interact with Msi2 (CFIB) and Pcf11-Clp1 heterodimer (components of CFIA) (Figure 2.2.4.3B&C), however, Seb1 may interact with the Msi2 and Pcf11-Clp1 via RNA binding. Future work is needed to find out whether Seb1 interacts with fully reconstituted CFIA, nuclease and the phosphatase modules.

3. Discussion

The primary aim of my thesis is to understand the links between transcription and 3' end processing. To study how transcription rate affects the function of the Pol II associated factors such as elongation and 3' end processing factors, our ultimate goal is to reconstitute the whole 3' end processing machinery and transcription process *in vitro*. We have made progress towards this by establishing an *in vitro* transcription system as well as *in vitro* polyadenylation assay and have reconstituted and purified CPF polymerase and phosphatase modules, CFIB and are in the process of reconstituting CFIA and nuclease module. In this thesis, I was able to generate Pol II mutants and demonstrate that they show different processivity and interactions with trans-interacting factors involved in transcription and 3' end processing (section 2.1). Data presented here were supported by extensive research in Pol II mutants in budding yeast (Dangkulwanich et al., 2013; Kaplan et al., 2012; Malagon et al., 2006b; Qiu et al., 2016). Also, I was able to show that Seb1 plays a role as elongation factor and stimulate Pol II processivity *in vitro*, which is similar to its human ortholog SCAF8 (Gregersen et al., 2019) in addition to its reported role in 3'end formation (Lemay et al., 2016; Wittmann et al., 2017).

3.1 How does transcription rate affect the function of trans-acting factors that control transcription and 3' end processing?

The kinetic model was widely used to describe the effect of transcription rate on co-transcriptional events. Transcription rate affects the "window of opportunity" for trans-acting factors to bind to Pol II CTD or DNA/ RNA to perform their functions. "Fast" Pol II with a faster elongation compresses the "window of opportunity" upstream, whereas "slow" Pol II expanding the "window of opportunity" upstream, leading to different site selections and in turn generating transcripts with different isoforms, localization, functionality. Several studies

have proposed that transcription rates can affect co-transcriptional events (Bentley, 2014; Fong et al., 2014; Hazelbaker et al., 2013).

In this thesis, I have successfully generated the two Pol II Rpb1 mutants that showed different processivity in fission yeast (section 2.1) to further understand how transcription speed contributes to function of trans-acting factors such as elongation and 3'end processing factors. There is no previous studies in Pol II mutants in fission yeast, thus the construction was based on extensive studies on *S. cerevisiae* Pol II mutants (Malagon et al., 2006a; Qiu et al., 2016).

Using a genetic approach, I demonstrated that the Pol II mutants E1106G and N494D had an effect on transcription as evidenced by the additive growth defect when combined with mutations in the elongation factor Spt5. Through *in vitro* assays, we assessed Pol II pausing frequency, ability to incorporate nucleotides and initiate transcription by looking at transcriptionally active complex only. Pol II mutants E1106G and N494D showed faster and slower elongation rate compared to WT. Assessing the nucleotide incorporation efficiency of the two Pol II mutants, E1106G showed slightly more efficient nucleotide incorporation. N494D, however, was less efficient in incorporation of nucleotides. This is in line with the previous studies in budding yeast where E1103G and N488D mutations were reported to show faster and slow synthesis rates compared to WT Pol II (Irvin et al., 2014; Malagon et al., 2006a, p. 1; Qiu et al., 2016).

Looking at the ability to initiate transcription, E1106G showed similar ability to WT, whereas N494D mutant showed impairment in initiating transcription as evidenced by the number of active complexes assembled on DNA. This is documented in budding yeast, where transcription initiation or early elongation phases is impaired by the *rpb1-N488D* mutation as

it affects the N445 residue. This is further confirmed by genetic studies, in which the budding yeast N488D mutation is synthetically lethal with *soh1* deletion, which encodes for a subunit in the Mediator transcription complex (Malagon et al., 2006b). Combining the data from genomic studies using PRO-seq, both E1106G and N494D mutations showed global loss of promoter proximal Pol II pausing compared to the WT (Figure 2.1.6C, *rpl902* gene). The loss of promoter proximal pausing in N494D Pol II could be explained by the impairment in transcription initiation. Therefore, the N494D Pol II mutation we generated in fission yeast may have the similar effect on transcription initiation by affecting the basic residues close to the Pol II active site. E1106G in contrast, the loss in promoter proximal pausing might be due to termination defect as there a gene upstream of *rpl902* gene.

Using the HIV pause-induced scaffold, we showed that E1106G and N494D Pol II mutants have different pausing frequency or timing or pausing by looking at the intensity & position of intermediate products (pause sites) between WT and mutants. The pause sites in Pol II mutants are similar to that of WT, however, there is a minor difference in the strength of intermediate bands, suggesting the different processivity of Pol II mutants may influence the timing of each pause. This is also evidenced by our PRO-seq data. Interestingly, we demonstrate that in E1106G Pol II promoter proximal pausing is reduced, however polymerase tend to pause more in the end of transcription cycle. This suggest that proper regulation of pausing is compromised in this mutant. I speculate that the reason for this might be that faster rates of Pol II alter its interactions with the factors that regulate pausing such as Cdk9, Spt5 and Dis2. Previous studies in mammals have proposed that 3'end pausing might be important for proper regulation of events at the end of transcription, which might explain why termination defect is observed in this mutants (Glover-Cutter et al., 2008; Gromak et al., 2006; Nojima et al., 2018).

When investigating the effect of Pol II mutants in termination, we found that Pol II E1106G and N494D showed opposite effects. E1106G showed transcription readthrough compared to WT, whereas early termination is observed in N494D. Termination defects in E1106G was reported in budding yeast E1103G and in mammals due to the failure of recruiting termination factors Sen1 helicase and Xrn2 exonuclease (Fong et al., 2015; Hazelbaker et al., 2013a). In contrast, N494D showed early termination on multiple genes in fission yeast, which is consistent with data reported for *rpb1-N488D* in budding yeast that showed early termination and shorter distribution of pre-snR33 transcripts compared to WT (Hazelbaker et al., 2013). This suggested that the kinetic model of "window of opportunity" could potentially explain how the interactions between Pol II and termination factors are being affected by the transcription rate. The "window of opportunity" that allow termination factors to bind in upstream pA sites was expanded in N494D and was compressed in E1106G, leading to early termination and transcription readthrough respectively.

Additionally, N494D mutant show weaker PRO-seq signal across transcription units, suggesting the reduced overall transcription level compared to the WT Pol II. This could be explained by the impairment of transcription initiation as shown by our *in vitro* data. In contrast to euchromatic genes, striking loss of transcriptional silencing is observed in this mutant at telomeric and centromeric loci. This suggest that proper rate of transcription is important for recruitment of the silencing factors to heterochromatin. Further studies will test this possibility and identify which factors might be affected in this mutant.

Then, we assessed whether the interaction with transcription factors was affected due to the intrinsic change of Pol II properties in Pol II mutants. My data showed Pol II WT and mutants are stimulated by Spt4/5 (section 2.1.4), suggesting that Pol II mutants did not lose the ability

to interact with elongation factor Spt4/5. However, the signal strength of all three polymerases are uneven and the signal for N494D Pol II mutant is sub-optimal, thus I am unable to draw conclusion on to if the interaction is being affected and to what degree. For future work, I would need to enhance the signal strength of slow Pol II through optimizing the protocol.

To address the question on whether Spt5 phosphorylation plays a role in affect Pol II processivity, I have assessed the effects of Spt5 phosphorylation on transcription elongation (section 2.1.5). My data have revealed that Spt5 phosphorylation does not changes its properties in stimulating Pol II transcription with minimal components *in vitro*. However, Spt5 phosphorylation may play a more prominent role in *vivo*. In agreement with this, previous study in in budding yeast showed that Spt5 phosphorylation may be important for recruitment of PAF complex (Liu et al., 2009). Taken together, I speculate that effect of Spt5 phosphorylation on transcription inside the cell is likely to be mediated via trans-acting factors interacting with the CTR *in vivo*.

I conclude that the *in vitro* system established would provide a useful tool to understand how transcription rate affect co-transcriptional processes such as mRNA 3' end processing and that proper regulation of transcription rate is important for transcription associated processes.

3.2 What is the role of Seb1 in transcription and 3' end processing?

Function of the Pol II CTD is mediated by conserved CTD-interacting proteins. Seb1 is a fission yeast ortholog of *S. cerevisiae* Nrd1 and mammalian SCAF4 and SCAF8 and is known bind to CPF and Pol II CTD in previous study (Wittmann et al., 2017). Seb1 mutations in either CID or RRM domains resulted in 3' extended RNA species, lowered global gene expression as evidenced by a decrease in the total mRNA level, and reduced Pol II occupancy within gene

bodies, which suggest that transcription elongation might also be compromised in this mutant (Wittmann et al., 2017). To understand to which if these phenotypes Seb1 contributes directly, I aimed to study role of Seb1 in transcription and 3'end processing using reconstituted *in vitro* systems.

My *in vitro* data showed that Seb1 stimulates transcription strongly by removal of intermediate products (section 2.2.2), and stimulates all three polymerases to different extent, which is evidenced by the quantification of runoff product (section 2.2.3). We found that Pol II mutants were stimulated to a less extent compared to WT Pol II. This may be explained by the intrinsic change of Pol II processivity, affecting their interactions with Seb1 as mentioned in the previous section. My results concluded that Seb1 has a stimulatory effect on Pol II and Seb1 acts as a positive elongation factor. This suggests that Seb1 may have a similar role as SCAF8 (Gregersen et al., 2019).

Our data showed that Seb1 does not have a direct effect on transcription termination of Pol II (Figure 2.1.4.1). This demonstrates that Seb1 is not a "bona fide" termination factor and suggests that inefficient transcription termination reported in Seb1 is likely to result from failed 3'end processing. Therefore, I propose that Seb1 is likely to be a core component of the 3'end processing machinery.

Using the MultiBac system, we have successfully reconstituted part of the 3' end machinery including the CPF polymerase module and phosphatase module (except Dis2), part of cleavage factor IA (CFIA) and cleavage factor 1B (CFIB) *in vitro* and tested their activity using EMSA (section 2.2.4.2). My data showed that Seb1 does not interact with Pcf11-Clp1 nor Msi2 without the presence of RNA. As Clp1 and Msi2 are found in the MS data of purified Seb1,

this could be explained by the co-purification of Clp1 and Msi2 with CPF components that interact with Seb1 or they interact with Seb1 in the presence of RNA. Thus, our future work is to perform pull down assays of Seb1 with the three modules of CPF and CFs with and without the presence of RNA.

To further investigate if Seb1 affect polyadenylation, we performed polyadenylation assay with and without Seb1 (section 2.2.3.4). Our data revealed that Seb1 suppresses polyadenylation alone or competes with polyadenylation module for RNA. This suggests that Seb1 may have a similar role as its mammalian orthologs SCAF 4 and SCAF8, which both are known to function as anti-terminators and prevent early cleavage and polyadenylation of mRNA transcript. By comparing the aligned sequences of Seb1 and its orthologs Nrd1 in budding yeasts and SCAF4 and SCAF8 in mammals (Figure 3.2), CID and RRM domains are conserved across species and Seb1 sequence resembles those of SCAF4 and SCAF8. This suggested that Seb1 may have a similar role as SCAF4 and SCAF8 in transcription and 3' end processing. However, further studies are necessary in a context of fully reconstituted cleavage and polyadenylation system to understand role of Seb1 in mRNA 3'end formation.

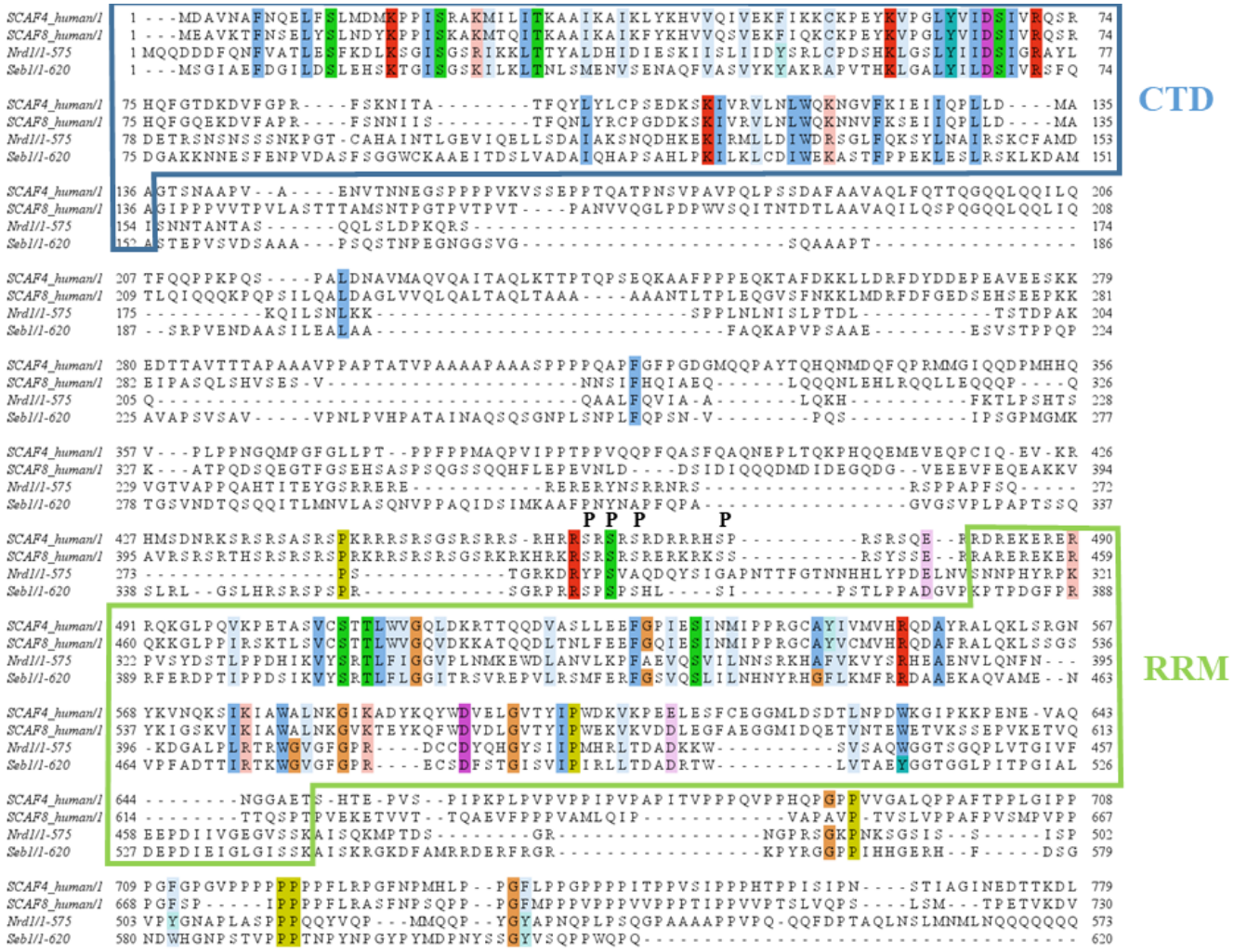


Figure 3.2 CID and RRM domains are conserved across species. The full-length sequences of Seb1, SCAF4, SCAF8 and Nrd1 were aligned using Clustal Omega and coloured using Jalview. The CID is indicated by a blue box and the RRM domain by a green box. Conserved amino acids (conservation threshold 70%) were coloured. Seb1 phosphorylation sites 360, 362, 364, 367 are marked above the residue by the letter P. The sites were determined by purification of FLAG-tagged Seb1 from S. pombe and followed by mass spectrometry (MS). MS was done by the Advance Proteomics Facility, Department of Biochemistry, Oxford.

For future work, our goal is to reconstitute the whole 3' end processing machinery and transcription process *in vitro*, and investigate how transcription rate and Seb1 contribute to each and every step of 3' end processing and termination by coupling and uncoupling transcription, cleavage, polyadenylation and termination using our established an *in vitro* system in lab.

4. Materials and Methods

4.1 *In vitro* biochemical methods

4.1.1 *S. pombe* genomic DNA extraction

5 ml of yeast culture was grown to saturation and cells were collected by centrifugation at 4000 g for 5 minutes. Cells were resuspended in 0.5 ml of water and transferred to a 1.5 ml microcentrifuge tube. The supernatant was removed, and cells were resuspended in residual liquid. 0.2 ml of Buffer A (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0), 200 µl of glass beads and 0.2 ml of phenol:chloroform:isoamyl alcohol (25:24:1) were added to the cells and the mixture was vortexed for 3 minutes. The supernatant was transferred to a new tube after centrifugation at 12,500 g for 5 minutes. The DNA was precipitated with 1 ml of absolute ethanol at RT, followed by 70% ethanol wash. The pellet was dried and resuspended in 100 µl of water.

4.1.2 Recombinant expression and purification in *E. coli*

4.1.2.1 Expression of full-length Seb1

Plasmid pET41a containing C-terminal 6xHis-tagged full-length Seb1 (3932/ LV107) (Wittmann et al., 2017) was transformed into bacterial strain (BL21 DE3) for recombinant expression. A single colony was picked and grown over-night at 37°C in 2xTY media with kanamycin, shaking at 190 rpm. The next day, cells were diluted in 1:1000 and allowed to grow at 37°C for an additional 3 hours. When the OD_{600} reached 0.3, the cells were moved to 18°C. Induction was done with 1 mM IPTG when cells reached 0.7 OD_{600}. Bacteria were grown for 14 hours before they were harvested by centrifugation at 5000 g for 10 minutes. Cell pellets were stored at -80°C.

4.1.2.2 Affinity purification of full-length Seb1

Cell pellets were resuspended in 5x excess (weight/volume) of lysis buffer (Buffer A (50 mM HEPES, pH 8.0, 150 mM NaCl, 1 mM 2-mercaptoethanol, 0.5 mM $MgCl_2$, 1 mM of DTT) supplemented with cOmplete EDTA-free protease inhibitor (Roche, manufacture recommendation), supernuclease (SinoBiological: SSNP01, 5mU/ ml). The lysate was incubated at 4°C with stirring for 20 minutes. The cells were lysed in a French Press and phenylmethylsulfonyl fluoride (PMSF) was added to 1 mM. Cell debris was removed by centrifugation at 40,000 g for 20 minutes at 4°C and cleared lysate was passed through 0.22 µm filter. The supernatant was incubated with Ni-NTA beads (Qiagen, 1 ml of slurry per 2L cell culture) equilibrated with Buffer A for 30 minutes at 4°C. The beads were washed with 50 column volumes (cv) of Buffer A + 25 mM Imidazole, followed by 50 cv of Buffer A + 50 mM Imidazole. For elution, 2 ml of Buffer B (Buffer A + 500 mM Imidazole) was added and incubated for 5 minutes. The elution was carried out twice. The lysate, flow-through, washes and elution fractions were analysed by Coomassie staining of SDS-PAGE.

4.1.2.3 Purification of full-length Seb1 by size exclusion chromatography

The fractions containing the desired proteins were mixed and injected to a Superdex 200 column (HiLoad 16/600 or 10/300 GE Healthcare) on the ÄKTA system. The proteins eluted with 1.5 cv Buffer A.

4.1.3 Recombinant expression and purification in Sf9 cells

4.1.3.1 Media and growth conditions for Sf9 cells

Sf9 cells are cultured in Insect-XPRESS (Lonza: BE12-730Q) at 27°C at 190 rpm. Cells are maintained at a concentration of 0.5-4 x 10^6/ ml unless indicated. All Sf9 manipulations were performed in a sterile laminar flow-hood.

4.1.3.2 Bacmid preparation, P1 and P2 virus generation

To generate constructs for expression of *S. pombe* SPAC22H10.05c (Clp1), Seb1 and Msi2 proteins in insect cells, corresponding genomic cDNAs were cloned into pACEBac1 (4196/ LV354) vector (a kind gift from Kim Nasmyth lab). To do that genomic cDNAs were generated using a standard protocol (Invitrogen SuperScript^{TM} III Reverse Transcriptase: 18080-093). A 20 µl reaction was set up with 1 µl oligo(dT)$_{20}$ at 50 µM (5702, Supplementary Table 1), *S. pombe* total RNA at 10 pg – 5 µg (provided by Dr. Dong-hyuk Heo, Vasilieva's Lab), 1 µl dNTP mix (dATP, dGTP, dCTP, dTTP) at 10 mM and 13 µl sterile water. The reaction mixture was incubated at 65°C for 5 minutes followed by 1 minute on ice. The reaction mixture was then added with 4 µl 5x First-Strand Buffer, 1 µl 0.1 M DTT, 1 µl RNasin Ribonuclease inhibitor (Promega N2111), and 1 µl SuperScript III RT and incubated at 50°C for 1 hour. The reaction was inactivated by incubation at 70°C for 15 minutes. To remove the RNA complementary to cDNA, 1 µl of RNase H was added to the reaction mixture and incubated at 37°C for 20 minutes. *S. pombe* Pcf11, Rna14, Rna15 and Ctf1 were constructed using codon-optimized gene fragments for Sf9 expression (Integrated DNA Technologies (IDT)). Assembled genes were amplified by PCR to generate DNA fragments containing upstream SalI and downstream NotI. 8xhis or twin-strep tags were inserted to the 3' end of Pcf11 (8xhis), Seb1 (8xhis), Msi2 (8xhis, strep) and Clp1 (8xhis, strep) respectively using 8xhis tag or strep tag containing primers for another round of PCR. The genes were cloned to pACEBac1 (4196/ LV354) and the sequences were confirmed by sequencing. The plasmids pACEBac1-Seb1-8xHis (4189), pACEBac1-Pcf11 opt-8xHis (4253), pACEBac1-Rna14 opt-8xHis (4254), pACEBac1-Rna15 opt-8xHis (4255), pACEBac1-Ctf1 opt-His (4256), pACEBac1-Msi2-Strep (4250), pACEBac1-Clp1-Strep (4252) (a kind gift from Kim Nasmyth lab) and blue-white colony screening was carried out to select for plasmids. The colonies were cultured overnight

in 2xTY with kanamycin, gentamycin and tetracycline and the bacmids were extracted using a modified DNA extraction protocol (Bieniossek et al., 2008).

To generate the P1 virus, 2 ml of 0.5 x 10^6 cells/ ml per well of Sf9 cells were added to a 6 well plate and incubated at RT for 1 hr. Bacmid DNA was diluted to 1000 ng/ µl and 2µl of bacmid DNA was mixed with 198µl of Insect Xpress medium and 6µl of FuGENE HD Transfection reagent (Promega E2311), and incubated at RT for 15 minutes. This mixture and subsequently 0.8 ml of Insect Xpress medium were added to each well. The 6-well plate was incubated at 27°C for 5 days. For harvesting the virus, the supernatant in the wells was collected and centrifuged at 4000 g for 5 minutes. The supernatant was transferred to a new 15 ml Falcon tube with 500 µl FBS (Corning 35010CV), the virus was filtered by 0.1 µm filter.

To generated P2 virus, 500 µl of P1 virus was added to 50 ml of Sf9 cells at the concentration of 1 x 10^6 cells/ ml (Pcf11) or 2 x 10^6 cells/ ml (Clp1, Msi2). The cells were incubated at 27°C for 3 days and the P2 virus was harvested and filtered.

4.1.3.3 Protein expression

For protein expression, the P2 virus was added to 2 x 10^6 cells/ ml in the ratio of 1:100 and incubated at 27°C for 3 days. For pull-downs, the cell cultures were co-infected with the P2 viruses in 1:1 ratio, except Pcf11 and Clp1 were co-infected in 49:1 ratio. When harvesting the proteins, the cultures were centrifuged at 1500 g for 10 minutes at 4°C, washed with 30 ml of cold 1x PBS, and transferred to 50 ml Falcon tubes. After another round of centrifugation, the supernatant was removed, and the cell pellet was frozen using liquid nitrogen and stored at -80°C.

4.1.3.4 Affinity chromatography for purification of strep-tagged Msi2, Pcf11-Clp1, poly A and phosphatase module of the CPF

Cells infected with the P2 viruses generated from pACEBac1-Msi2-Strep (4250), co-infected with the P2 viruses generated from pACEBac1-Pcf11-8xHis (4190), pACEBac1-Clp1-Strep (4252) were harvested. Cell pellet was thawed in and resuspended in 5x excess (weight/volume) of the lysis buffer (Buffer AS (50 mM HEPES, pH 8.0, 150 mM NaCl, 1 mM 2-mercaptoethanol, 0.5 mM $MgCl_2$, 1 mM of DTT) supplemented with cOmplete EDTA-free protease inhibitor (Roche, manufacture recommendation), supernuclease (SinoBiological: SSNP01, 5mU/ml), 1 mM PMSF. The lysate was incubated at 4°C with stirring for 20 minutes. The cells were lysed by sonication at 80% amplitude 5s on 10s off, and centrifuged at 45,000 g for 20 minutes at 4°C. The lysate was filtered using a 0.22 µm filter and incubated with pre-washed MagStep "type 3" XT Beads (IBA Lifesciences 2-4090-002) for 30 minutes. The reaction tube was placed in the magnetic separator and the supernatant was removed. For washing, 100 µl of Buffer AS/ µl of beads was added to the tube, the tube was vortexed shortly and the supernatant was removed. This step was repeated three times before elution. For elution, 25 µl Buffer BXT (IBA Lifesciences 2-1042-025)/ µl beads were added and vortexed. The mixture was incubated at 4°C with rotation for 10 minutes.

To optimize Pcf11-Clp1 heterodimer purification, Buffer AS with a higher salt concentration (250 mM NaCl) was used to remove excess Clp1.

For large scale purification using the ÄKTA system, the lysate was loaded onto a regenerated StrepTrap column 5 ml (GE Healthcare GE28-9075-48). The column was washed with 50 ml Buffer AS and the proteins were eluted with 50 ml Buffer BS (Buffer AS with 5 mM

desthiobiotin (IBA Life Sciences 2-1000-005)). The column was regenerated by washing with 25 ml 0.5 M NaOH, followed by water.

4.1.3.5 Purification of 8xhis-tagged Seb1 and Msi2 by Ni-NTA affinity chromatography

Cells infected with P2 viruses generated from pACEBac1-Seb1-8xHis (4189), pACEBac1-Msi2-8xHis (4251) were harvested. Cell lysis procedure as for Strep-tagged protein affinity chromatography was used. After filtering the lysate, the lysate was incubated with Ni-NTA agarose (Invitrogen R90110) for 20 minutes at 4°C. The beads were then washed with 5 cv Buffer AS with 5 mM Imidazole and transferred to 5 ml Bio-rad gravity column. The beads were further washed 5 times with 100 µl Buffer AS with 15 mM Imidazole and 100 µl Buffer AS with 40 mM Imidazole. For elution, 100 µl of Buffer AS with 200 mM Imidazole was added and incubated for 3 minutes. For the second elution, the beads were incubated for 2 minutes.

4.1.3.6 Purification of strep-tagged Msi2, Pcf11-Clp1, poly A and phosphatase module of the CPF and 8xhis-tagged Seb1 and Msi2 by ion exchange chromatography

Samples containing the strep-tagged Msi2, Pcf11-Clp1, poly A and phosphatase module of the CPF and 8xhis-tagged Seb1 and Msi2 from affinity chromatography were topped to 50 ml with Buffer QAS (50mM HEPES pH 8.0, 50 mM NaCl, 1 mM 2-mercaptoethanol, 0.5 mM $MgCl_2$). The sample was pumped to HiTrap Q HP 5 ml column (GE Healthcare 17115401), and washed with 10 cv Buffer QAS, followed by 5 cv of 8% Buffer QBS (as Buffer QAS but 1M salt (100%)). Elution was done in 10 cv with linear 8-40% gradient of Buffer QBS.

4.1.3.7 Purification of strep-tagged Msi2, Pcf11-Clp1, poly A and phosphatase module of the CPF and 8xhis-tagged Seb1 and Msi2 by size exclusion chromatography

The elution fractions containing the strep-tagged Msi2, Pcf11-Clp1, poly A and phosphatase module of the CPF and 8xhis-tagged Seb1 and Msi2 from ion-exchange chromatography were concentrated by centrifugation at 4000 g 4°C for 20 minutes. The sample was then transferred to a clean microcentrifuge tube and centrifuged for 2 minutes at 10,000 g at 4°C to spin down protein aggregates. The sample was injected to the ÄKTA system with Superdex 200 10/300 column. The proteins were eluted with 1.5 cv Buffer AS.

4.1.4 Purification of native Pol II from *S. pombe* cells

4.1.4.1 Affinity chromatography for FLAG-tagged Rpb9 Pol II

YP637 (WT), YP1420 (*rpb1-E1106G*), and YP1421 (*rpb1-N494D*) were harvested at OD_{600} =1. Cells were disrupted in a freezer mill (SPEX SamplePrep). Cell extract powder was resuspended in 5x excess (weight/ volume) of lysis buffer (L1 buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.7, 10% glycerol, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$, 0.5% Triton X-100) supplemented with cOmplete EDTA-free protease inhibitor (Roche, manufacture recommendation), supernuclease (SinoBiological: SSNP01, 5mU/ml), 1 mM PMFS. Extract was incubated at 4°C for 30 minutes followed by centrifugation at 40,000 g for 20 minutes at 4°C. Cleared lysate was incubated with L1 buffer pre-equilibrated anti-flag M2 resin (0.2 ml/ L of culture) 1 hr with rotation at 4°C. Beads were washed 3 times with 50 ml W1 buffer (1 M NaCl, 50 mM Tris-HCl, pH 7.7, 10% glycerol, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$, 0.5% Triton X-100), 1 M urea and 0.5 mM 2-mercaptoethanol, and washed 2 times with 50 ml W2 buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.7, 10% glycerol, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$, 0.05% Triton X-100) with 0.5 mM 2-mercaptoethanol. For elution, 2.5 ml of 2.5 mg/ml flag peptide in W2 buffer was added to beads and incubated for 10 minutes. The beads

were incubated for 2 minutes for the second elution. The beads were washed with 20ml of Buffer QA (20 mM NaCl, 50 mM Tris-HCl, pH 7.7, 10% glycerol, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$, 2 mM 2-mercaptoethanol) plus the two elution. The elution was further diluted to 50 ml with Buffer QA.

4.1.4.2 Ion exchange chromatography for FLAG-tagged Rpb9 Pol II

The elution fractions containing Pol II from affinity chromatography were combined and diluted in Buffer QA (20 mM NaCl, 50 mM Tris-HCl, pH 7.7, 10% glycerol, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$, 2 mM 2-mercaptoethanol) to a total volume of 50 ml and loaded onto the HiTrap Q HP 5ml column using the ÄKTA system. The column was washed with 10 cv of Buffer QA followed by 5 cv 8% of Buffer QB (2M NaCl (100%), 50mM Tris-HCl, pH 7.7, 10% glycerol, 0.5mM $MgCl_2$, 0.5mM $Mg(OAc)_2$, 2 mM 2-mercaptoethanol). The proteins were eluted by 10 cv in a linear 8-40% gradient of Buffer QB. The proteins were buffer exchanged to Buffer GF (20mM HEPES, pH 7.5, 150mM NaCl, 0.5mM $MgCl_2$, 0.5mM $Mg(OAc)_2$, 1mM 2-mercaptoethanol) for storage. Protein was snap-frozen in liquid nitrogen and stored at -80°C.

4.1.5 EMSA

Serial dilutions of Msi2 (final concentration of 0.5, 1, 2.5, 5, 10, 25, 50, 100 nM) and polymerase module (final concentration of 50, 100, 200, 400, 600 nM) were prepared. 0.5 µl of FAM-labelled *rps2-WT* RNA (IDT) for Msi2 or *rps2-WT* F2 RNA for polymerase module (Supplementary Table 5) was mixed with 5.5 µl Buffer S (50 mM HEPES pH 8.0, 1 mM 2-mercaptoethanol, 0.5 mM $MgCl_2$, 1 mM of DTT) to a final concentration of 0.05 µM, and the mixture was boiled at 70°C for 5 minutes to remove any secondary structures. 6 µl of RNA/ buffer mix was added to 4 µl of protein and incubated on ice for 20 minutes in the dark. 1.5 µl

of 6xLDE buffer (30% glycerol, 0.2 mg/ml bromophenol) was added immediately after the incubation and 10 µl of the mixture was loaded to 6% DNA retardation gel (Invitrogen EC6365BOX). Gel electrophoresis was performed in the cold room with 0.5xTBE at 100 V for 40 minutes in the dark, and visualization of the fluorescent signal was achieved using Image Analyser (FLA-600 Fuji) with semi-wet gel.

4.1.6 SDS-PAGE

Samples were boiled at 95°C for 5 minutes in a 1x SDS loading buffer (5x SDS buffer: 10% SDS, 500 mM DTT, 50% Glycerol, 250 mM Tris-HCl, pH 6.8, 0.5% bromophenol blue dye). The Mini-PROTEAN gel pouring and electrophoresis system (Bio-Rad) was utilized for Coomassie stained and Western blot gels. 8% to 12% polyacrylamide resolving gel was used depending on the protein size with a 4% stacking gel. Gels were run at 100 V for 15 minutes followed by 120 V for 1 hr. The Mini Gel Tank (Thermo Fisher Scientific) was used for EMSA gels and silver stain gels. Gels were run at 100 V for 40 minutes (EMSA) or 200 V for 40 minutes (all other gels).

4.1.7 SDS-PAGE Coomassie and Silver staining

Staining was performed using InstantBlue (Expedeon ISB01L) according to the manufacturer's instructions, followed by detaining with MQ water until the background was clear. Silver staining of gels was carried out using the Pierce Silver Stain Kit (Thermo Fisher Scientific 24612) according to the standard basic staining protocol.

4.1.8 Western blotting

After SDS-PAGE, proteins were transferred onto a PVDF or nitrocellulose membrane through a wet transfer at 25V 4°C for 16 hours or 400 mA 4°C for 3 hours using the transfer buffer (25 mM Tris-HCl, pH 8.3, 192 mM glycine, 1 mM EDTA and 20% (v/v) methanol). For membrane blocking, TBST (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% Tween-20) + 5% (w/v) skim milk powder or 5% (w/v) bovine serum albumin (BSA) was added and blocked for 30 minutes at RT. A horseradish peroxidase (HRP)-coupled antibody (mouse 6xHis (Takara Bio 631210) diluted 1:10,000; mouse Anti-Flag M2 (Sigma Aldrich F3165) diluted 1:5,000 were diluted in TBST + 5% skim milk powder or Streptavidin-HRP (Thermo Scientific 11896814) diluted 1:4000 was diluted in TBST + 5% BSA and incubated with the membrane overnight at 4°C/ 3 hours at RT. The membrane was washed three times for 5 minutes with TBST. Proteins were visualized using Clarity Western ECL (Bio-Rad 1705061) according to the manufacturer's instructions.

4.1.9 Phosphorylation of Spt4/5 and phospho-tag gel

3.4 µg of Spt4/5 (pRSFDuet-Spt5wt-Spt4, 4164/ LV323) (Kecman et al., 2018) with mixed with 0.3 µg Cdk9 (pACEBac1-Cdk9 FL, 4206/ LV364) (Kilchert et al., 2019) and 0.5 mM ATP in Buffer AS (50 mM HEPES, pH 8.0, 150 mM NaCl, 1 mM 2-mercaptoethanol, 0.5 mM $MgCl_2$, 1 mM of DTT) with 50 mM $MnCl_2$ in a total volume of 40 µl. The mixture was incubated at 30°C for 30 minutes. 1 µg of the phosphorylated Spt4/5 (11.76 µl) was boiled at 95°C for 5 minutes in 5x SDS loading buffer (2-mercaptoethanol (5%), bromophenol blue (0.02%), glycerol (30%), SDS (Sodium dodecyl sulphate) (10%), Tris-HCl (250 mM, pH 6.8)). Samples were loaded onto 7.5% acrylamide gel with 7.5 µM phospho-tag (FujiFilm AAL-107) and 50 µM $MnCl_2$. Bands were visualised with InstantBlue Coomassie staining (Expeadon).

4.1.10 5’ labelling of RNA oligos with ^{32}P

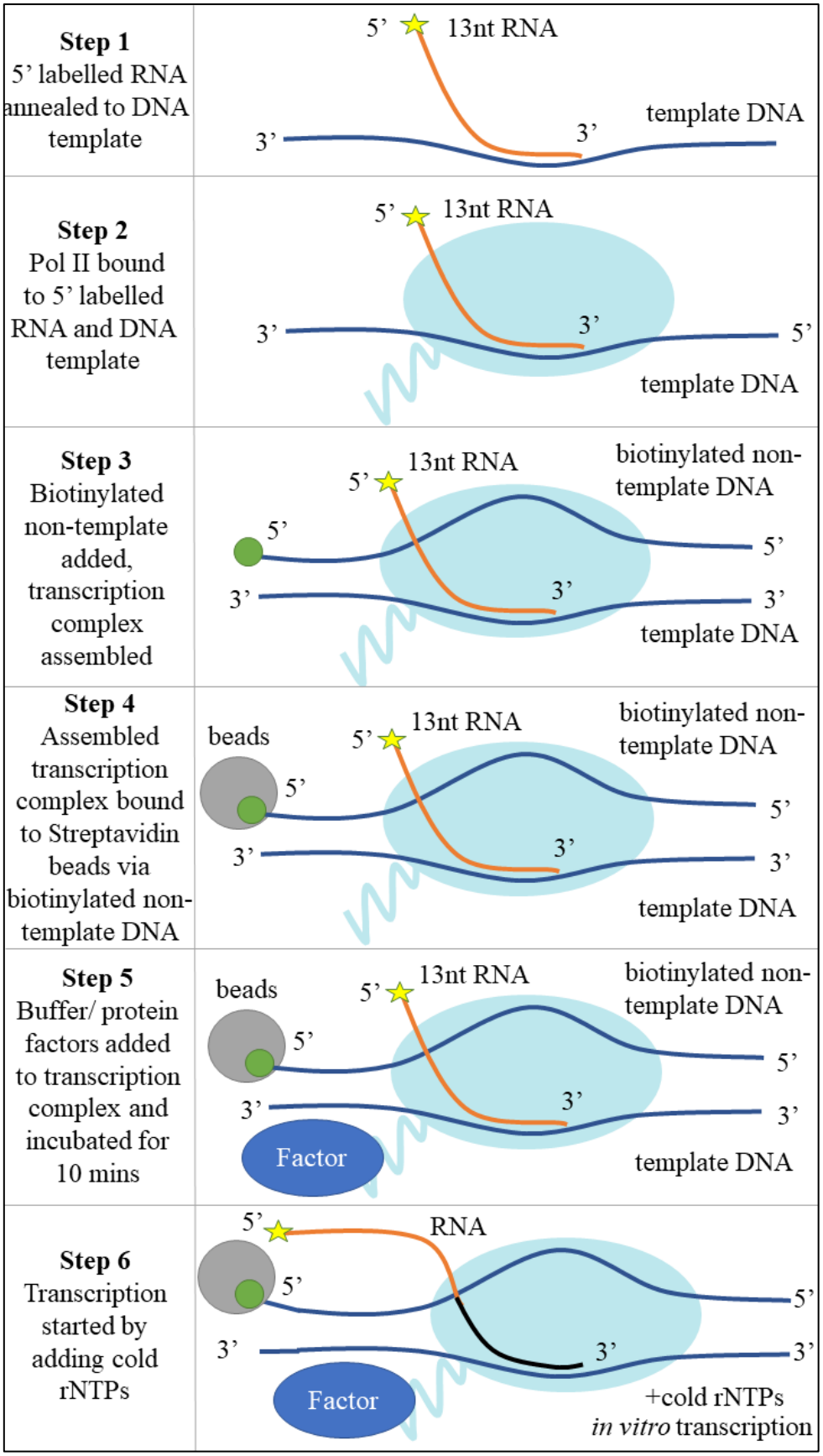
In one round, 700 pmol of RNA (HIV_R or RNA1, Supplementary Table 6) was 5’-labelled with ^{32}P using 2 µl PNK (NEB) and 4 µl $\gamma^{32}P$ ATP (Hartmann Analytic) in the total volume of 34 µl of 1X PNK buffer. The mixture was incubated at 37°C for 1 hour and an additional 1 µl of PNK was added with another 20 minutes interval. Then, 6 µl of 100 mM EDTA was added and the reaction mix was incubated at 72°C for 10 minutes. Labelled RNA was purified with water-equilibrated Micro Bio-Spin Chromatography Column P-6 (Bio-rad 732-6221) following manufacturer recommendations. The RNA was stored at -20°C until used.

4.1.11 Transcription elongation assay

4.1.11.1 Using radioactively 5’-labelled RNA

The assay was done using pre-assembled transcription complexes immobilised on Streptavidin beads via biotin attached to the 5’-end of non-template DNA strand. To prepare transcriptionally competent Pol II complexes, 5’-radioactively labelled RNA (20 pmol) was annealed to the template DNA strand in 1XTB buffer (40 mM KCl, 20 mM Tris-HCl, 200 µM EDTA, pH 7.9) at 45°C for 5 min, followed by 10 min at room temperature (Supplementary Figure 1, Step 1). Then, DNA-RNA hybrid was incubated with Pol II (or its mutants, usually 4-8 pmol) at 30°C for 10 min (Supplementary Figure 1, Step 2), followed by addition of an excess of 5’-biotinylated non-template DNA strand (100 pmol) and kept at 30°C for 10 min (Supplementary Figure 1, Step 3). For pause-inducing scaffold following nucleic acids were used: HIV_R (RNA, 13nt), HIV_T (template strand), HIV_NT (non-template strand), whereas for general transcription template RNA1 (20nt), GT1_T, GT1_NT were assembled, respectively (Supplementary Table 6). Pol II-nucleic acids complexes were immobilized on pre-equilibrated in 1XTB Streptavidin beads (Sepharose High Performance, GE Healthcare, 100 µl slurry per complex) with 10 min incubation at RT (Supplementary Figure 1, Step 4).

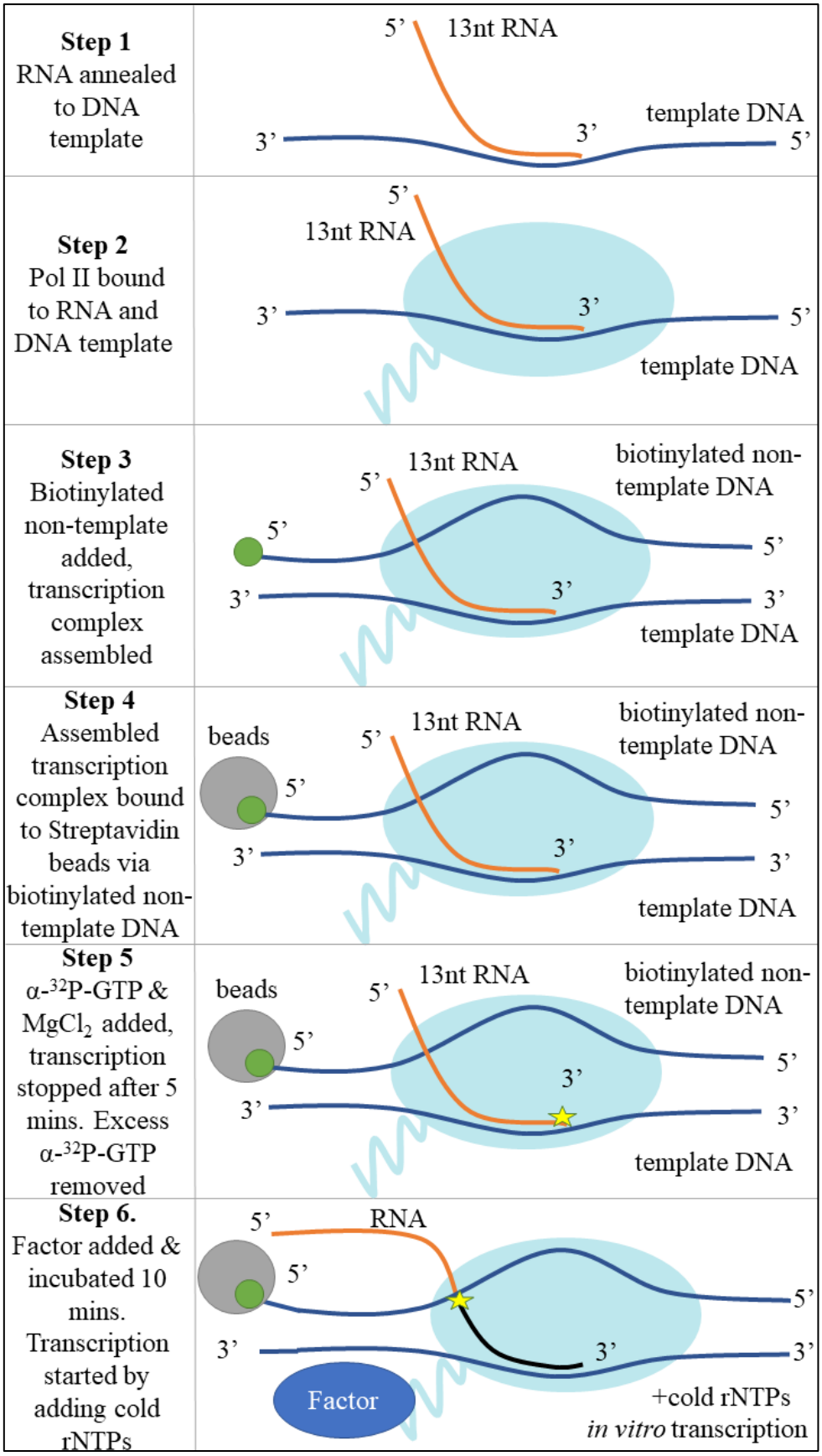
Next, to remove the excess of nucleic acids or not properly assembled complexes, beads were washed with 1 ml of W1000 buffer (20 mM Tris-HCl, pH7.9, 1 M NaCl) and twice with 1X TB buffer. To test the impact of Seb1 or Spt4/5 on transcription, complexes were mixed with buffer or proteins at a final concentration of 2.2 or 0.4 µM, respectively (Supplementary Figure 1, Step 5). Transcription was initiated with a final concentration of 10 µM rNTPs and 5 mM $MgCl_2$ (Supplementary Figure 1, Step 6). The reaction was carried out at 30°C for times specified in the figures (usually 0, 10, 30, 60, 300 s) and stopped immediately by addition of an equal volume of Stop Buffer (20 mM EDTA, 7 M urea, 100 ug/ml heparin, 0.02% bromophenol blue, 0.03% xylene cyanol in formamide). Products were boiled at 95°C for 2 minutes and resolved on 15% 19:1 acrylamide:bisacrylamide (19:1), 8 M urea gel at 30 W in 1xTBE buffer for 45 minutes. The gel was exposed to phosho-screen and the signal was scanned on the FLA-7000 Fuji instrument. Total intensity normalized lane profiles were calculated and plotted using Python scripts written by Dr Krzysztof Kuś.



Supplementary Figure 1. A diagram showing the setup of the transcription assay for 5' labelled RNA

4.1.11.2 Using 3’ labelled RNA

As during every protein preparation, a fraction of molecules might be inactive due to many reasons. One of the ways to look only at transcriptionally competent Pol II is to allow the enzyme to incorporate α-^{32}P-labeled rNTP which labels 3’-end of the RNA in the assembled complex. The steps of assembling Pol II-nucleic acids complex were performed essentially as in the paragraph above except unlabelled RNA and 50 ul of streptavidin slurry were used (Supplementary Figure 2, Steps 1-4). Immobilized, washed and transcriptionally competent Pol II complexes were walked on corresponding templates for 5 min using 2.5 µl of α-^{32}P-GTP (Hartmann Analytic) or α-^{32}P-UTP (Hartmann Analytic) with 5 mM $MgCl_2$ for HIV_RNA or RNA1 containing assemblies, respectively (Supplementary Figure 2, Step 5). Unincorporated nucleotides were removed during washes with 500 µl 1X TB with 20 mM EDTA, followed with 3 additional washes with 1X TB buffer only. These complexes with one radioactive nucleotide incorporated were used for the elongation assay as described in the section above (Supplementary Figure 2, Step 6).



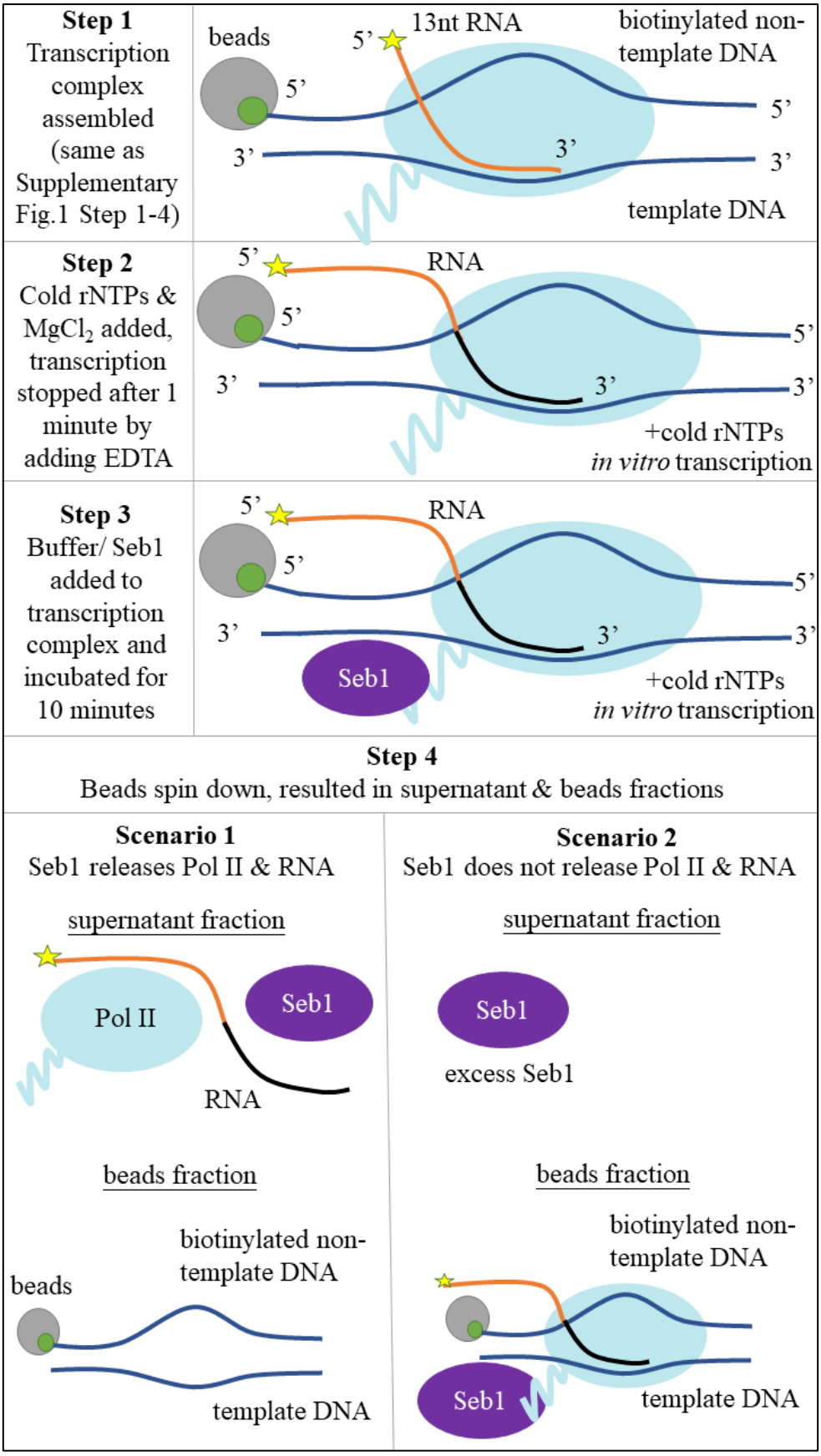
Supplementary Figure 2 A diagram showing the setup of the transcription assay for 3' labelled RNA

4.1.12 Pol II termination assay

Essentially, these assays were carried out following the 5'-labeled RNA elongation protocol except 50 µl of streptavidin beads slurry, W500 buffer (20 mM Tris-HCl, pH7.9, 0.5 M NaCl) instead of W1000 were used (Supplementary Figure 3, Step 1). Washed Pol II complex was divided into two fractions. For one sample Pol II transcription was carried out in the presence of 0.9 mM rNTPs and 5mM $MgCl_2$ whereas another one was treated with only $MgCl_2$. The reaction was halted with 10 mM EDTA (final concentration) (Supplementary Figure 3, Step 2) and each of the samples was divided into two tubes treated with a buffer or Seb1 (Supplementary Figure 3, Step 3). Mixtures were incubated at 30°C for 10 min, spin-down, and the supernatant was mixed with an equal volume of Stop Buffer. Beads were washed with 300 µl of 1xTB buffer and resuspended in 0.5X Stop Buffer (Supplementary Figure 3, Step 4). Samples were boiled at 95°C and processed as before.

4.1.13 Polyadenylation assay

The polyadenylation assay was performed as described in (Casañal et al., 2017) with modifications. A master mix solution containing 0.5 µM of FAM-*rps2* WT F2 RNA (Supplementary Table 5), 0.5 µM polymerase module in AS buffer was split in different tubes and supplemented with either buffer, Cdk9 (0.2 µM), Pol II (0.5 µM), Seb1 (1 µM) or their indicated combinations. Aliquots (6 µl), were taken at 0 or 30 min and the reaction was stopped by adding the 1 µl of Digestion Solution (130 mM EDTA, 2% (w/v) SDS, 2 mg/ ml proteinase K) and incubated 15 min at RT. The samples were mixed with an equal amount of Stop Buffer and loaded on 12.5% 19:1 acrylamide:bisacrylamide (19:1), 8 M urea gel and run at 30 W in 1xTBE buffer for 30 minutes. The gel was scanned on the FLA-7000 Fuji instrument.



Supplementary Figure 3 A diagram showing the setup of the termination assay coupled to transcription assay for 5' labelled RNA

4.2 *S. pombe* manipulation

4.2.1 Yeast growth and culture condition

S. pombe strains were grown in YES medium to OD_{600} 1 at 30°C unless indicated. Strains used in this study are listed in Supplementary Table 3.

4.2.2 Genetic manipulation and transformation

4.2.2.1 Yeast transformation

S. pombe strains were grown in YES medium to OD_{600} 0.5-1, cells were centrifuged at 4000 g for 5 minutes to remove the medium. Cells were then washed twice with LiAc to 10^9 cells/ ml, distributed to 150 µl aliquots and incubated for 45 minutes at 30°C. 2µg DNA and 370 µl of pre-warmed 50% (w/v) PEG4000 with 10 mM Tris-HCl, pH 7.5 were added to the cells and mixed by vortexing. The cells were then incubated for 1 hour at 30°C. Cells were heat-shocked at 42°C for 15 minutes, followed by 10 minutes RT. The supernatant was removed after centrifugation and the cell pellet was washed twice with YES medium. Cells were resuspended in YES and cultured overnight at 25°C 190 rpm. Cells were plated on the next day on appropriate selection media after washing with a fresh YES medium.

4.2.2.2 Tetrad dissection

Yeast strains were streaked on YES plates from freezing stocks and incubated for 2 days. Strains were then crossed on an EMMG plate, and the cells were checked under the microscope for tetrads every day afterwards. Once a considerable number of tetrads were observed under the microscope, the cells are streaked on the left side of a YES plate. Tetrads were selected and picked up using the dissection microscope (Singer Instruments) and were incubate at 30°C for 2-3 hours. Then, 4 spores of a complete tetrad were organized in a row and the plate was incubated at 30°C until colonies were visible. The plates were scanned and replica plating on

appropriate selection plates was carried out to determine the genotype of the spores. For Pol II Rpb1 mutant strains, colonies are confirmed by sequencing.

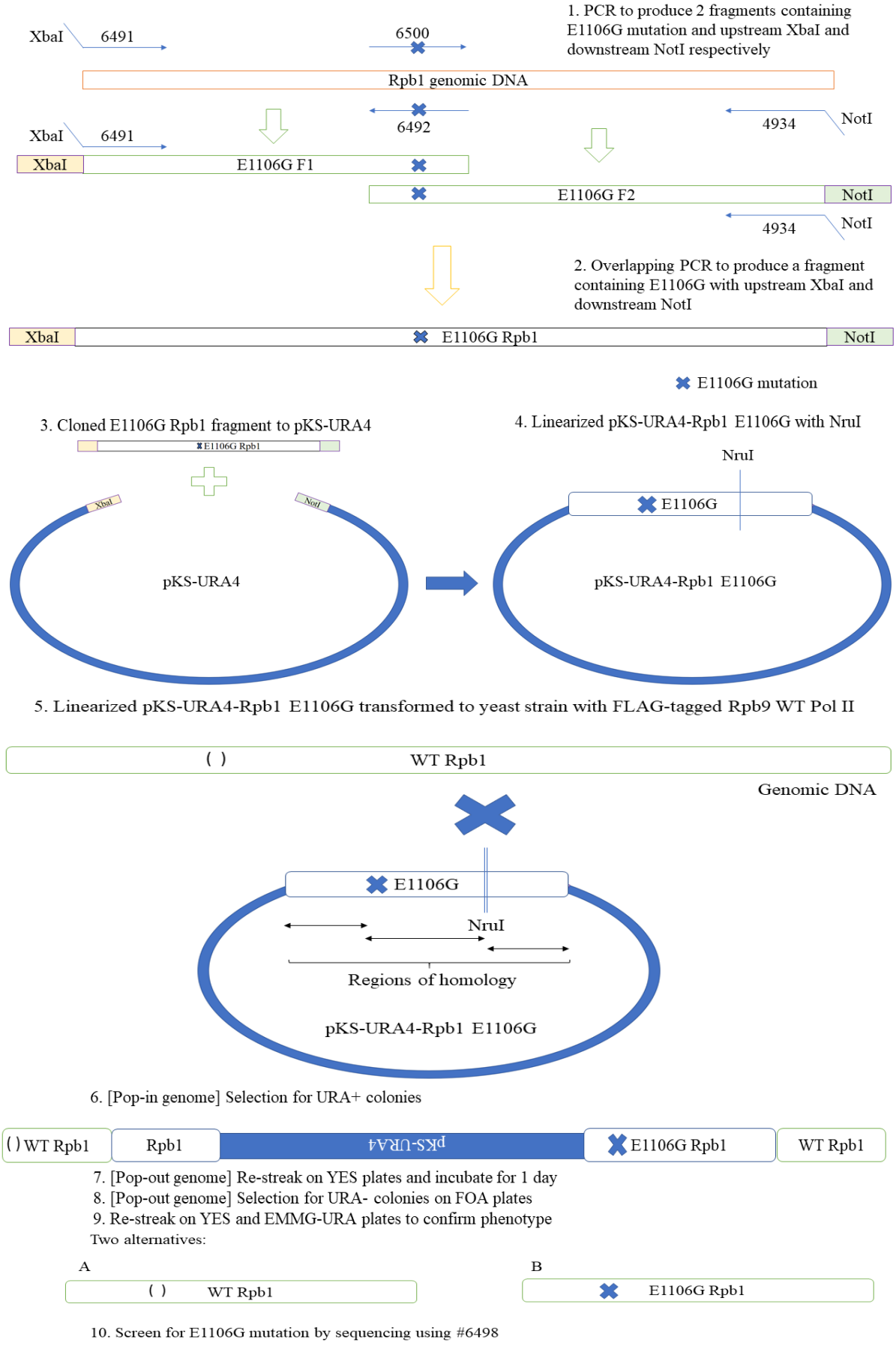
4.2.2.3 Construction of Rpb1 mutants

For the construction of Pol II Rpb1 E1106G and N494D mutations, the pop-in pop-out allele replacement method was used as described in (Gao et al., 2014). Rpb1 fragments containing the mutations *rpb1-E1106G* and *rpb1-N494D* were generated by PCR using the genomic DNA of WT *S. pombe* and primers containing desired mutations. Primer set 6491+6492 (Supplementary Table 1) were used to generate *rpb1-E1106G* fragment 1, 6500+4934 (Supplementary Table 1) to generate *rpb1-E1106G* fragment 2; 6494+6502 (Supplementary Table 1) to generate *rpb1-N494D* fragment 1 and 6495+6501 (Supplementary Table 1) to generate *rpb1-N494D* fragment 2 (Supplementary Figure 4, Step 1). Rpb1 fragments containing upstream XbaI and downstream NotI sequences and mutations were generated by overlapping PCR using *rpb1-E1106G/ rpb1-N494D* fragments 1 and 2 with primer set 6491+4934 for *rpb1-E1106G* and 6501+6502 for *rpb1-N494D* (Supplementary Figure 4, Step 2). The fragments were cloned to pKS-URA4 digested with XbaI and NotI to generated plasmids pKS-URA4 *rpb1-E1106G* (4170/ LV329) and pKS-URA4 *rpb1-N494D* (4171/ LV330) (Supplementary Table 2) (Supplementary Figure 4, Step 3). The NruI linearized plasmids 4170 and 4171 carrying E1106G and N494D Rpb1 mutations are transformed to a Rpb9-3xflag tagged *S. pombe* strain (YP637, Supplementary Table 3) and selected for URA+ on EMMG-URA plates (pop-in) (Supplementary Figure 4, Steps 4-6). After the colonies became visible on plate, the colonies are re-streaked on YES plates for 1 day to allow the loss of URA and re-streaked on 5-FOA for 2-3 days for selection of URA- (pop-out) (Supplementary Figure 4, Steps 7-8). The URA-colonies that grew on 5-FOA plates were re-streaked on YES and EMMG-URA plates to ensure pop-out was complete (Supplementary

Figure 4, Step 9). The colonies were screened and confirmed by sequencing using 6497 (N494D) or 6498 (E1106G) (Supplementary Figure 4, Step 10).

4.2.3 Spot test

Yeast cultures were grown overnight at 30°C. To count the cells, 1 ml of culture was collected, sonicated at 20% A for 20 s and vortexed. 10 µl of the culture was pipetted to the cell counter. The OD of cells was calculated assuming that 0.1 OD_{600} equals 1.5 x 10^6 cells. The required volume of cell culture was pipetted to a new tube and centrifuged at max speed for 1s. Cells were resuspended with 500 µl of sterile YES medium or water (for yeast strains with pNMT promoter) and serial dilutions (10x) were carried out for 5 repeats. For the spot test, 5 µl of each dilution was pipetted to YES or EMMG plates and left to dry. Plates were incubated at 25°C, 30°C and 37°C and were photographed every day for a week.



Supplementary Figure 4. A diagram showing the construction of Pol II Rpb1 mutants

4.3 PRO-seq

The PRO-seq experiment was performed according to (Mahat et al., 2016). *S. pombe* strains YP637 (WT), YP1420 (*rpb1-E1106G*), YP1421 (*rpb1-N494D*) (Supplementary Table 4) were grown to 0.5 OD_{600} at 30°C. 25 million cells were harvested and washed with ice cold PBS. Spike in *S. cerevisiae* (strain ATCC 204508/ S288c) control at 10% of *S. pombe* cell number was added to each sample to reduce sample processing bias. Cell pellet was lysed in ice cold permeabilization buffer (0.5% Sarkosyl, 0.5 mM DTT, one tablet of protease inhibitors cocktail per 50 ml and 4 units per ml RNase inhibitor) in concentration of 1 x 10^6 cells/ ml, and resuspended in storage buffer (10 mM Tris-HCl, pH 8.0, 25% (vol/vol) glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA and 5 mM DTT). Nuclear run-on (NRO) assays were performed with biotin-11-A/G/C/UTP (rNTPs). 100 µl of permeabilized cells was mixed with 100 µl of the pre-heated 2x master mix (20 mM Tris-HCl pH 7.7, 32 mM $MgCl_2$, 0.5 mM DTT, 200 mM KCl), and mixed 15 times at 30°C heating block by gentle pipetting. Extracted and purified nascent RNA was fragmented using base hydrolysis. Fragmented RNA was bound to Streptavidin M280 beads (Invitrogen cat. number: K1585-01) and incubated for 20 min at 4 °C. Beads were washed twice with high-salt wash buffer (50 mM Tris-HCl pH 7.4, 2 M NaCl, 0.5%(vol/vol) Triton X-100 in DEPC H_2O), twice with binding buffer (10 mM Tris-HCl pH 7.4, 300 mM NaCl, 0.1%(vol/vol) Triton X-100 in DEPC H_2O) and once with low-salt wash buffer (5 mM Tris-HCl pH 7.4, 0.1%(vol/vol) Triton X-100 in DEPC H_2O). Biotinylated RNA was extracted, and ethanol precipitated. Ligation of 3′ RNA adaptors to biotinylated RNA was performed followed by second round of biotin-streptavidin purification. mRNA cap was degraded followed by the ligation of 5’ RNA adaptor to the nascent RNA and third round of biotin-streptavidin purification. Complementary DNA (cDNA) of the adaptor ligated nascent RNA was produced by reverse transcription (RT) using RP1 reverse transcription primer (100 µM). cDNA was then amplified with different RPI-n (25 µM) primer to generate barcodes for

different libraries, and library fractions at equimolar concentrations were pooled together. Libraries were sequenced using Illlumina NextSeq 500 platform on a high-output flow cell. The raw sequence data was processed by filtering out low quality reads and trimming adapter sequences. Sequences were then aligned to the genomic sequence and visualized in a genome browser for downstream analysis.

4.4 Databases, miscellaneous software and online services

S. pombe gene names and DNA sequences were extracted from Pombase (www.pombase.org). UniProt (www.uniprot.org) was used to translate nucleotides sequences into protein sequences and determine the isoelectric point (pI) values of protein complexes. Multiple sequence alignment was done using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and conserved amino acids were coloured using Jalview (https://www.jalview.org/). Interacting partners of Seb1 were visualized using STRING (https://string-db.org/). Analysis of protein structures was conducted using Pymol software. PRO-seq data was analysed and exported using Integrative Genomics Viewer.

Supplementary Information

Supplementary Table 1. List of oligonucleotides used in this study

Oligo No.	Oligo Name	Sequence	Purpose
4934	Rpb1_rev_XbaI	AACTTTtctagaCGAGAACTTAGACACTACACTTC	Creating point mutants (E1106G) for rbp1
5702	Oligo-dT (rooted)	ttttttttttttttttttttttttttv	RT-PCR
6491	E1106G_F	GACATAGCGGCCGCGATGAGGAGTATACACAGCTGGTTG	Creating point mutant (E1106G) for rbp1, colony PCR E1106G
6492	E1106G_R	CTTAATATTTTTAGCGACATTCAAAATACCTTTCAAACGAGGAACACC	Creating point mutant (E1106G) for rbp1, colony PCR E1106G
6493	N494D_F	GACATAGCGGCCGCCTTAGGTCTTAATGAGCAGTATGCTAGAC	Creating point mutant (N494D) for rbp1
6494	N494D_F2	GCTGACTTTGATGGTGATGAAATGGATATGCATGTCCCTCAGTCAG	Creating point mutant (N494D) for rbp1
6495	N494D_R	CTGACTGAGGGACATGCATATCCATTTCATCACCATCAAAGTCAG	Creating point mutant (N494D) for rbp1
6496	494_seq_F	CAGCTCGTACTGTGATTACCGG	Sequencing N494D
6497	494_seq_R	CAACAGTCTTCTTGTCAACCAC	Sequencing N494D
6498	1106_seq_F	CAACGCTACCTTGTTATTCCAG	Sequencing E1106G
6499	1106_seq_R	CTTCTGCCTTACGGTCATCATC	Sequencing E1106G
6500	E1106G_F2	GTGTTCCTCGTTTGAAAGGTATTTTGAATGTCGCTAAAAATATTAAG	Creating point mutant (E1106G) for rbp1
6501	N494D_Fprom	GACATAGCGGCCGCGTGATTGTTGTATGCACATGTTTC	Creating point mutant (N494D) for rbp1
6502	N494D_RProm	AACTTTTCTAGAGTGCTGAGCATCCTGAATACATTC	Creating point mutant (N494D) for rbp1
6721	Seb1_SalI_F	aatacGTCGACATGTCGGGAATCGCTGAATTC	Making baculovirus construct with Seb1-TEV
6722	Seb1_AGA_TEV_R	GCTCTGAAAGTACAGATCCTCTTGGGGTTGCCAAGGAGGTTG	Making baculovirus construct with Seb1-TEV
6723	Pcf11_SalI_F	acataaGTCGACATGGATTTAGTGGAATTGGATTAC	Making baculovirus construct with Pcf11-TEV

6724	Pcf11_AGA_TEV_R	GCTCTGAAAGTACAGATCCTCCTTGCTTTCTTGTTTAATAACTTTAG	Making baculovirus construct with Pcf11-TEV
6725	Clp1_SalI_F	acaataGTCGACATGGATTACCAAGATGATGG	Cloning of recombinant Clp1
6726	Clp1_AGA_TEV_R	GCTCTGAAAGTACAGATCCTCAGAAATTAGCCGGCTTTTAG	Cloning of recombinant Clp1
6727	Tev_His8x_R	aataacGCGGCCGCttaGTGGTGGTGATGGTGATGATGATGGCTCTGAAAGTACAGATCCTC	Adding 8xHis tag to baculovirus construct
6728	TEV_Strep_F	GAGGATCTGTACTTTCAGAGCTGGTCACATCCGCAGTTTGAAAAAGG	Adding Strep tag to baculovirus construct
6729	SP_strep_NotI_R	aataacGCGGCCGCTTATTTCTCAAACTGAGGATGGC	Adding Strep tag to baculovirus construct
7025	728_seq_pACEBac1	TAAAATGATAACCATCTCGC	Colony PCR for baculovirus constructs
7038	A_gib_F	GGTCCTAAGGTAGCGAGTTTTAACGGTCCTAAGGTAGCGAGTTTAAAC	Cloning of genes to pACEBac1 using Gibson Assembly
7039	A_gib_R	CACGTCATCATCTATCGTGCATAGAATCGATACTAGTTTAGATGGGTATACCCTAGGG	Cloning of genes to pACEBac1 using Gibson Assembly
7040	B_gib_F	TATGCACGATAGATGATGACGTGTATACCCATCTAAACTAGTATCGATTCGCGACCTACT	Cloning of genes to pACEBac1 using Gibson Assembly
7041	B_gib_R	TCGCGAATCGATACTAGTGTTTCAATTAGATGGGTATACCCTAGGGGTTATGA	Cloning of genes to pACEBac1 using Gibson Assembly
7046	pACEBac1_seq_R	GTTTGGACAAACCACAACTAG	Colony PCR for baculovirus constructs
7315	Pcf11_F_opt	CAATAACTATcacactacacagctagc	Cloning of optimized Pcf11 sequence to pACEBac1 to make CFIA, use with 7316
7316	Pcf11_R_opt	TAGGTCGCGAATCGATACTAGTG	Cloning of optimized Pcf11 sequence to pACEBac1 to make CFIA, use with 7315
7317	Rna14_F_opt	atcagccagtcagacagcaagGTTG	Cloning of optimized Rna14 sequence to pACEBac1, use with 7318
7318	Rna14_R_opt	ctgtgctagctgtgtagtgtgATAG	Cloning of optimized Rna14 sequence to pACEBac1, use with 7317
7319	Rna15_F_opt	CTATAACGGTCCTAAGGTAGCGAG	Cloning of optimized Rna15 sequence to

			pACEBac1, use with 7320
7320	Rna15_R_opt	CCAACcttgctgtctgactggc	Cloning of optimized Rna15 sequence to pACEBac1, use with 7319
7321	Ctf1_F_opt	CTGTAACTATAACGGTCCTAAGG	Cloning of optimized Ctf1 sequence to pACEBac1, use with 7322
7322	Ctf1_R_opt	CGCGAgatgctgcatcgactgtG	Cloning of optimized Ctf1 sequence to pACEBac1, use with 7321
7323	Pcf11_gib_F	TATAACGGTCCTAAGGTAGCGAGTTTAAACCAATAACTATcacactacacagctagcacag	Cloning of optimized Pcf11 sequence to pACEBac1, use with 7324
7324	Pcf11_gib_R	TAGGTCGCGAATCGATACTAGTGTTTTAGGTCGCGAATCGATACTAGTGTTTg	Cloning of optimized Pcf11 sequence to pACEBac1, use with 7323
7325	Pcf11_seq	cacactacacagctagcacag	Sequencing of optimized sequence Pcf11
7326	Clp1_seq	gcgcatcagcagcatgcaacc	Sequencing of Clp1 in pACEBac1
7327	Rna14_seq	atcagccagtcagacagcaag	Sequencing of optimized sequence Rna14
7328	Rna15_seq	acacagtcgatgcagcatc	Sequencing of optimized sequence Rna15
7329	Ctf1_seq	agtacaggcatgcatgcattg	Sequencing of optimized sequence Ctf1

Supplementary Table 2. List of plasmids used in this study

Freezer Number	Plasmid Name	Purpose
3392	pKS-URA4	Cloning of Rpb1 mutations
3932	pET41a-Seb1-6xHis full-length	Recombinant expression of Seb1
4164	pRSFDuet-Spt5wt-Spt4	Recombinant expression of Sp4/5
4170	pKS-URA4 Rpb1-E1106G	*S. pombe* strain construction

4171	pKS-URA4 Rpb1-N494D	*S. pombe* strain construction
4189	pACEBac1-Seb1-8xHis	For recombinant expression of Seb1 in Sf9 cells
4190	pACEBac1-Pcf11-8xHis	For recombinant expression of Pcf11 in Sf9 cells
4191	pACEBac1-Clp1-8xHis	For recombinant expression of Clp1 in Sf9 cells
4196	pACEBac1	For cloning of recombinant genes
4206	pACEBac1-Cdk9 FL	For recombinant expression of Cdk9 in Sf9 cells
4250	pACEBac1-Msi2-Strep	For recombinant expression of Msi2 in Sf9 cells
4251	pACEBac1-Msi2-8xHis	For recombinant expression of Msi2 in Sf9 cells
4252	pACEBac1-Clp1-Strep	For recombinant expression of Clp1 in Sf9 cells
4253	pACEBac1-Pcf11 opt-8xHis	For recombinant expression of codon optimized Pcf11 in Sf9 cells
4254	pACEBac1-Rna14 opt	For recombinant expression of codon optimized Rna14 in Sf9 cells
4255	pACEBac1-Rna15 opt	For recombinant expression of codon optimized Rna15 in Sf9 cells
4256	pACEBac1-Ctf1 opt	For recombinant expression of codon optimized Ctf1 in Sf9 cells

Supplementary Table 3. List of *S. pombe* strains used in this study

Strain number	**Genotype**	**Source/ Constructed by**
YP71	h-	Erik Boye via Beata Grallert
YP144	h+, leu1-32, ura4-Δ18, ade6-M216, his3Δ::1	Francois Bachand
YP602	h+, leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; rpb9-3xflag::kanMX; Rpb1-TEV cleavage site	Tea Kecman
YP637	h+, leu1-32, ura4Δ18, ade6-M216, his3Δ::1, rpb9-3xflag::kanMX	Cornelia Kilchert
YP839	h+, leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Seb1-3xFlag::nat	Katie Duckett
YP1349	h-, leu1-32; ade6-M216; his3Δ::1; Spt5-wt-3xFlag::natMX	Tea Kecman
YP1350	h+, leu1-32; ade6-M216; his3Δ::1; Spt5-wt-3xFlag::natMX	Tea Kecman
YP1351	h-, leu1-32; ade6-M216; his3Δ::1; Spt5-CTR(T1A)-3xFlag::natMX	Tea Kecman
YP1352	h+, leu1-32; ade6-M216; his3Δ::1; Spt5-CTR(T1A)-3xFlag::natMX	Tea Kecman
YP1353	h+, leu1-32; ade6-M216; his3Δ::1; Spt5-CTR(T1E)-3xFlag::natMX	Tea Kecman
YP1362	h+, leu1-32; ade6-M216; his3Δ::1	Tea Kecman

YP1420	h+, leu1-32, ura4-Δ18, ade6-M216, his3Δ::1, rpb9-3xflag::kanMX, Rpb1-E1106G	This study
YP1421	h+, leu1-32, ura4-Δ18, ade6-M216, his3Δ::1, rpb9-3xflag::kanMX, Rpb1-N494D	This study
YP1482	h-, ade6-M216; his3Δ::1 Spt5-CTR(T1E)-3xFlag::natMX	This study
YP1504	h+, leu1-32, ade6-M216, his3Δ::1, ura4-Δ18, Rpb1 E1106G, Spt5-CTR(T1A)-3xFlag::natMX	This study
YP1507	h+, leu1-32, ade6-M216, his3Δ::1, ura4-Δ18, Rpb1 N494D, Spt5-CTR(T1E)-3xFlag::natMX	This study
YP1532	h+, leu1-32, ade6-M216, his3Δ::1, Rpb1 E1106G, Spt5-WT-3xFlag::natMX, rpb9-3xflag::kan	This study
YP1533	h-, leu1-32, ade6-M216, his3Δ::1, Rpb1 N494D, Spt5-WT-3xFlag::natMX, rpb9-3xflag::kan	This study
YP1534	h+, leu1-32, ade6-M216, his3Δ::1, ura4-Δ18, Rpb1 E1106G	This study
YP1567	h+, leu1-32, ade6-M216, his3Δ::1, E1106G Rpb1, Spt5-WT-3xFlag::natMX	This study
YP1568	h+, leu1-32, ade6-M216, his3Δ::1, N494D Rpb1, Spt5-WT-3xFlag::natMX	This study
YP1613	h+, leu1-32, ade6-M216, his3Δ::1, E1106G Rpb1	This study
YP1656	h+, leu1-32, ade6-M216, his3Δ::1, E1106G Rpb1, Spt5-CTR(T1E)-3xFlag::natMX	This study
YP1657	h+, leu1-32, ade6-M216, his3Δ::1, E1106G Rpb1, Spt5-CTR(T1A)-3xFlag::natMX	This study
YP1662	h+, leu1-32, ade6-M216, his3Δ::1, N494D Rpb1, Spt5-CTR(T1E)-3xFlag::natMX	This study
YP1663	h+, leu1-32, ade6-M216, his3Δ::1, N494D Rpb1	This study
YP1706	h+, leu1-32, ade6-M216, his3Δ::1, N494D Rpb1, Spt5-CTR(T1A)-3xFlag::natMX	This study

Supplementary Table 4. Antibodies used in this study

Description	**Company cat number**
mouse 6xHis-HRP	Takara Bio 631210
Anti-Flag M2-HRP	Sigma Aldrich F3165
Streptavidin-HRP	Thermo Scientific 11896814

Supplementary Table 5. RNAs used in this study

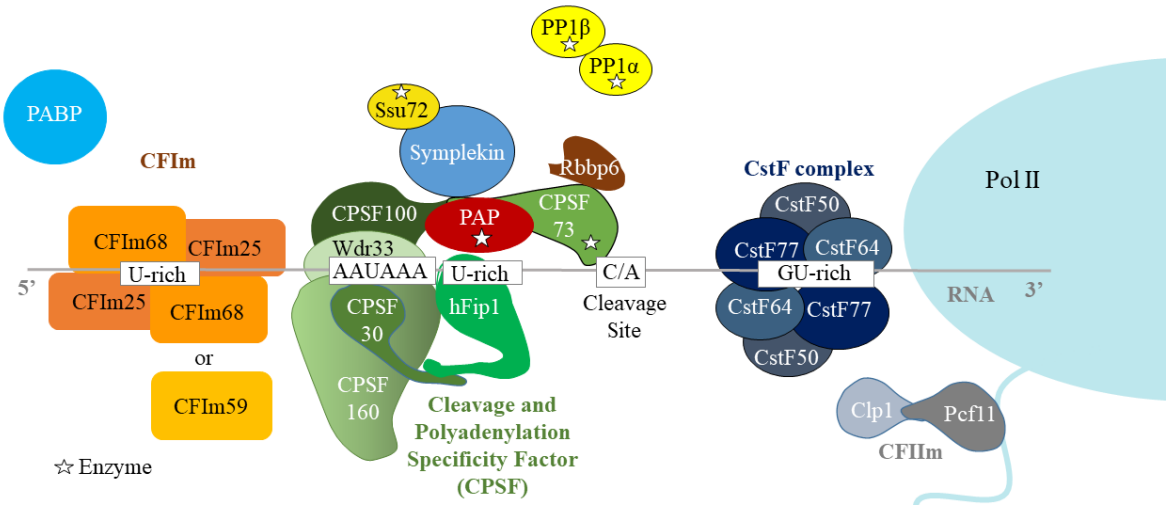
Oligo number	**RNA**	**Sequence**	**Purpose**
7572	*rps2* WT	CAAUUGUACAUUGUUGUAAAGCCUUCCAG CUUUAAUAAAAUAAAAAAGUUAACUCAAA GCUAGUUUAUCUUUUGUUCAGUAACGUGU AUAUUGGUAUCAGUCAAUUUUUCAUAGU CACUGUAAGUA	EMSA

7573	*rps2* WT F1	CAAUUGUACAUUGUUGUAAAGCCUUCCAG C	EMSA
7574	*rps2* WT F2	CCUUCCAGCUUUAAUAAAAUAAAAAAGUUA ACUCA	Polyadenylation assay/ EMSA
7575	*rps2* WT F3	CAAAGCUAGUUUAUCUUUUGUUCAGUAACG UGUAUAUUGGUAUCAGUCAAUUUUUC	EMSA
7576	PolyU	UUUUUUUUUU	EMSA

Supplementary Table 6. RNA oligos, template and non-template DNA oligos used in this study (*in vitro* assay)

Oligo number	**Template (T)/ Non-template (NT) DNA**	**Sequence**	**Purpose**
7577	RNA1	UGCUGUAAUAAAGACCAGGC	*In vitro* assays, use with 7564 and 7565
7578	GT1_NT	5’ - biot-GGTCAAGGCAGTACTAGTAATGACCAGG CTCAATTGAGCTTGGAGTCAGTCGACGA TG ACTGG -3’	*In vitro* assays, use with 7562 and 7565
7579	GT1_T	3’ - CCAGTTCCGTCATGATCATTACTGGTCCG AGTTAACTCGAACCTCAGTCAGCTGCTA CTGACC - 5’	*In vitro* assays, use with 7562 and 7564
7580	HIV RNA	AAUAAUCGAGAGG	*In vitro* assays, use with 7566 and 7567
7581	HIV_NT	5’ - biot-ACTTACAGCCATCGAGAGGGAACCCACT GCTTAA GCCTCAATAAAGC -3’	*In vitro* assays, use with 7563 and 7567
7582	HIV_T	3’-TGAATGTCGGTAGCTCTCCCTTGGGTGA CGAATTCGGAG TTATTTCG - 5’	*In vitro* assays, use with 7563 and 7566

Supplementary Figures



Supplementary Figure 5. Architecture of the 3'end processing complex in humans. A diagram showing the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF) complex, cleavage factor Im and IIm, other 3' end processing factors and their recognized RNA sequences. This figure is based on (Clerici et al., 2018, 2017; Kaufmann et al., 2004; Mandel et al., 2006; Schönemann et al., 2014; Takagaki et al., 1990; Xiang et al., 2014)

References

Ahn, S.H., Kim, M., Buratowski, S., 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Mol. Cell 13, 67–76. https://doi.org/10.1016/s1097-2765(03)00492-1

Archambault, J., Jansma, D.B., Kawasoe, J.H., Arndt, K.T., Greenblatt, J., Friesen, J.D., 1998. Stimulation of Transcription by Mutations Affecting Conserved Regions of RNA Polymerase II. J. Bacteriol. 180, 2590–2598. https://doi.org/10.1128/JB.180.10.2590-2598.1998

Armache, K.-J., Mitterweger, S., Meinhart, A., Cramer, P., 2005. Structures of Complete RNA Polymerase II and Its Subcomplex, Rpb4/7. J. Biol. Chem. 280, 7131–7134. https://doi.org/10.1074/jbc.M413038200

Aslanzadeh, V., Huang, Y., Sanguinetti, G., Beggs, J.D., 2018. Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. Genome Res. 28, 203–213. https://doi.org/10.1101/gr.225615.117

Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., Söding, J., Cramer, P., 2017. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Mol. Cell 66, 38-49.e6. https://doi.org/10.1016/j.molcel.2017.02.009

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., Nudler, E., 2005. A ratchet mechanism of transcription elongation and its control. Cell 120, 183–193. https://doi.org/10.1016/j.cell.2004.11.045

Bataille, A.R., Jeronimo, C., Jacques, P.-É., Laramée, L., Fortin, M.-È., Forest, A., Bergeron, M., Hanes, S.D., Robert, F., 2012. A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. Mol. Cell 45, 158–170. https://doi.org/10.1016/j.molcel.2011.11.024

Bentley, D.L., 2014. Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. 15, 163–175. https://doi.org/10.1038/nrg3662

Berroteran, R.W., Ware, D.E., Hampsey, M., 1994. The sua8 suppressors of Saccharomyces cerevisiae encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to sua7 (TFIIB) mutations. Mol. Cell. Biol. 14, 226–237. https://doi.org/10.1128/MCB.14.1.226

Bieniossek, C., Richmond, T.J., Berger, I., 2008. MultiBac: Multigene Baculovirus-Based Eukaryotic Protein Complex Production. Curr. Protoc. Protein Sci. 51, 5.20.1-5.20.26. https://doi.org/10.1002/0471140864.ps0520s51

Blythe, A.J., Yazar-Klosinski, B., Webster, M.W., Chen, E., Vandevenne, M., Bendak, K., Mackay, J.P., Hartzog, G.A., Vrielink, A., 2016. The yeast transcription elongation factor Spt4/5 is a sequence-specific RNA binding protein. Protein Sci. Publ. Protein Soc. 25, 1710–1721. https://doi.org/10.1002/pro.2976

Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F., Tang, L.K., Fraser, J.S., Holstege, F.C.P., Hieter, P., Guthrie, C., Kaplan, C.D., Krogan, N.J., 2013. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell 154, 775–788. https://doi.org/10.1016/j.cell.2013.07.033

Brueckner, F., Cramer, P., 2008. Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. Nat. Struct. Mol. Biol. 15, 811–818. https://doi.org/10.1038/nsmb.1458

Buratowski, S., 2009. Progression through the RNA polymerase II CTD cycle. Mol. Cell 36, 541–546. https://doi.org/10.1016/j.molcel.2009.10.019
Buratowski, S., 2003. The CTD code. Nat. Struct. Mol. Biol. 10, 679–680. https://doi.org/10.1038/nsb0903-679
Butler, J.S., Platt, T., 1988. RNA processing generates the mature 3' end of yeast CYC1 messenger RNA in vitro. Science 242, 1270–1274. https://doi.org/10.1126/science.2848317
Casañal, A., Kumar, A., Hill, C.H., Easter, A.D., Emsley, P., Degliesposti, G., Gordiyenko, Y., Santhanam, B., Wolf, J., Wiederhold, K., Dornan, G.L., Skehel, M., Robinson, C.V., Passmore, L.A., 2017. Architecture of eukaryotic mRNA 3′-end processing machinery. Science 358, 1056–1059. https://doi.org/10.1126/science.aao6535
Chan, S., Choi, E.-A., Shi, Y., 2011. Pre-mRNA 3'-end processing complex assembly and function. Wiley Interdiscip. Rev. RNA 2, 321–335. https://doi.org/10.1002/wrna.54
Chang, C.-H., Luse, D.S., 1997. The H3/H4 tetramer blocks transcript elongation by RNA polymerasem II in vitro. J. Biol. Chem. 272, 23427–23434. https://doi.org/10.1074/jbc.272.37.23427
Chen, F., MacDonald, C.C., Wilusz, J., 1995. Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acids Res. 23, 2614–2620.
Chen, X., Müller, U., Sundling, K.E., Brow, D.A., 2014. Saccharomyces cerevisiae Sen1 as a Model for the Study of Mutations in Human Senataxin That Elicit Cerebellar Ataxia. Genetics 198, 577–590. https://doi.org/10.1534/genetics.114.167585
Cheng, H., He, X., 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. https://doi.org/10.1128/MCB.24.7.2932-2943.2004
Cheung, A.C.M., Cramer, P., 2011. Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature 471, 249–253. https://doi.org/10.1038/nature09785
Clerici, M., Faini, M., Aebersold, R., Jinek, M., 2017. Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. eLife 6, e33111. https://doi.org/10.7554/eLife.33111
Clerici, M., Faini, M., Muckenfuss, L.M., Aebersold, R., Jinek, M., 2018. Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. Nat. Struct. Mol. Biol. 25, 135–138. https://doi.org/10.1038/s41594-017-0020-6
Connelly, S., Manley, J.L., 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. 2, 440–452. https://doi.org/10.1101/gad.2.4.440
Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., Bentley, D.L., 2019. Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. Mol. Cell 76, 896-908.e4. https://doi.org/10.1016/j.molcel.2019.09.031
Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., Pelechano, V., Styles, E.B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J.S., Srikumar, T., Bahr, S., Chen, Y., Deshpande, R., Kurat, C.F., Li, S.C., Li, Z., Usaj, M.M., Okada, H., Pascoe, N., Luis, B.-J.S., Sharifpoor, S., Shuteriqi, E., Simpkins, S.W., Snider, J., Suresh, H.G., Tan, Y., Zhu, H., Malod-Dognin, N., Janjic, V., Przulj, N., Troyanskaya, O.G., Stagljar, I., Xia, T., Ohya, Y., Gingras, A.-C., Raught, B., Boutros, M., Steinmetz, L.M., Moore, C.L., Rosebrock, A.P., Caudy, A.A., Myers, C.L., Andrews, B., Boone, C., 2016. A global interaction network maps a wiring diagram of cellular function. Science 353. https://doi.org/10.1126/science.aaf1420

Cramer, P., Bushnell, D.A., Kornberg, R.D., 2001. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. Science 292, 1863–1876. https://doi.org/10.1126/science.1059493

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., Corden, J.L., 2011. Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. 7, e1002329. https://doi.org/10.1371/journal.pgen.1002329

Dangkulwanich, M., Ishibashi, T., Liu, S., Kireeva, M.L., Lubkowska, L., Kashlev, M., Bustamante, C.J., 2013. Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. eLife 2, e00971. https://doi.org/10.7554/eLife.00971

Darby, M.M., Serebreni, L., Pan, X., Boeke, J.D., Corden, J.L., 2012. The Saccharomyces cerevisiae Nrd1-Nab3 Transcription Termination Pathway Acts in Opposition to Ras Signaling and Mediates Response to Nutrient Depletion. Mol. Cell. Biol. 32, 1762–1775. https://doi.org/10.1128/MCB.00050-12

Davidson, L., Francis, L., Cordiner, R.A., Eaton, J.D., Estell, C., Macias, S., Cáceres, J.F., West, S., 2019. Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. Cell Rep. 26, 2779-2791.e5. https://doi.org/10.1016/j.celrep.2019.02.012

Dheur, S., Vo, L.T.A., Voisinet-Hakil, F., Minet, M., Schmitter, J.-M., Lacroute, F., Wyers, F., 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. https://doi.org/10.1093/emboj/cdg253

Dupin, A.F., Fribourg, S., 2014. Structural basis for ATP loss by Clp1p in a G135R mutant protein. Biochimie 101, 203–207. https://doi.org/10.1016/j.biochi.2014.01.017

Eaton, J.D., West, S., 2018. An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. Transcription 9, 321–326. https://doi.org/10.1080/21541264.2018.1498708

Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., Sekine, S., 2017. Structure of the complete elongation complex of RNA polymerase II with basal factors. Science 357, 921–924. https://doi.org/10.1126/science.aan8552

Eick, D., Geyer, M., 2013. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. Chem. Rev. 113, 8456–8490. https://doi.org/10.1021/cr400071f

Elkon, R., Ugalde, A.P., Agami, R., 2013. Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. 14, 496–506. https://doi.org/10.1038/nrg3482

Exinger, F., Lacroute, F., 1992. 6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae. Curr. Genet. 22, 9–11. https://doi.org/10.1007/BF00351735

Fabrega, C., Shen, V., Shuman, S., Lima, C.D., 2003. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. Mol. Cell 11, 1549–1561. https://doi.org/10.1016/S1097-2765(03)00187-4

Fitz, J., Neumann, T., Pavri, R., 2018. Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. EMBO J. 37. https://doi.org/10.15252/embj.201797965

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M., Sheridan, R.M., Nguyen, T., Karp, S., Bentley, D.L., 2015. Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. Mol. Cell 60, 256–267. https://doi.org/10.1016/j.molcel.2015.09.026

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.-D., Bentley, D.L., 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. 28, 2663–2676. https://doi.org/10.1101/gad.252106.114

Fuda, N.J., Ardehali, M.B., Lis, J.T., 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461, 186–192. https://doi.org/10.1038/nature08449

Fusby, B., Kim, S., Erickson, B., Kim, H., Peterson, M.L., Bentley, D.L., 2016. Coordination of RNA Polymerase II Pausing and 3′ End Processing Factor Recruitment with Alternative Polyadenylation. Mol. Cell. Biol. 36, 295–303. https://doi.org/10.1128/MCB.00898-15

Gao, J., Kan, F., Wagnon, J.L., Storey, A.J., Protacio, R.M., Davidson, M.K., Wahls, W.P., 2014. Rapid, efficient and precise allele replacement in the fission yeast Schizosaccharomyces pombe. Curr. Genet. 60, 109–119. https://doi.org/10.1007/s00294-013-0406-x

Ghazy, M.A., Gordon, J.M.B., Lee, S.D., Singh, B.N., Bohm, A., Hampsey, M., Moore, C., 2012. The interaction of Pcf11 and Clp1 is needed for mRNA 3′-end formation and is modulated by amino acids in the ATP-binding site. Nucleic Acids Res. 40, 1214–1225. https://doi.org/10.1093/nar/gkr801

Glover-Cutter, K., Kim, S., Espinosa, J., Bentley, D.L., 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. 15, 71–78. https://doi.org/10.1038/nsmb1352

Gordon, J., Shikov, S., Kuehner, J.N., Liriano, M., Lee, E., Stafford, W., Poulsen, M.B., Harrison, C., Moore, C., Bohm, A., 2011. Reconstitution of CF IA from overexpressed subunits reveals stoichiometry and provides insights into molecular topology. Biochemistry 50, 10203–10214. https://doi.org/10.1021/bi200964p

Graber, J.H., Cantor, C.R., Mohr, S.C., Smith, T.F., 1999. In silico detection of control signals: mRNA 3′-end-processing sequences in diverse species. Proc. Natl. Acad. Sci. U. S. A. 96, 14055–14060.

Gregersen, L.H., Mitter, R., Ugalde, A.P., Nojima, T., Proudfoot, N.J., Agami, R., Stewart, A., Svejstrup, J.Q., 2019. SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. Cell 177, 1797-1813.e18. https://doi.org/10.1016/j.cell.2019.04.038

Gromak, N., West, S., Proudfoot, N.J., 2006. Pause Sites Promote Transcriptional Termination of Mammalian RNA Polymerase II. Mol. Cell. Biol. 26, 3986–3996. https://doi.org/10.1128/MCB.26.10.3986-3996.2006

Gross, S., Moore, C., 2001. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. Proc. Natl. Acad. Sci. U. S. A. 98, 6080–6085. https://doi.org/10.1073/pnas.101046598

Gruber, A.J., Schmidt, R., Gruber, A.R., Martin, G., Ghosh, S., Belmadani, M., Keller, W., Zavolan, M., 2016. A comprehensive analysis of 3′ end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. Genome Res. 26, 1145–1159. https://doi.org/10.1101/gr.202432.115

Grzechnik, P., Gdula, M.R., Proudfoot, N.J., 2015. Pcf11 orchestrates transcription termination pathways in yeast. Genes Dev. 29, 849–861. https://doi.org/10.1101/gad.251470.114

Haberle, V., Stark, A., 2018. Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. 19, 621–637. https://doi.org/10.1038/s41580-018-0028-8

Hampsey, M., Reinberg, D., 1999. RNA polymerase II as a control panel for multiple coactivator complexes. Curr. Opin. Genet. Dev. 9, 132–139. https://doi.org/10.1016/S0959-437X(99)80020-3
Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., Buratowski, S., 2013. Kinetic Competition between RNA Polymerase II and Sen1-dependent Transcription Termination. Mol. Cell 49, 55–66. https://doi.org/10.1016/j.molcel.2012.10.014
Hill, C.H., Boreikaitė, V., Kumar, A., Casañal, A., Kubík, P., Degliesposti, G., Maslen, S., Mariani, A., von Loeffelholz, O., Girbig, M., Skehel, M., Passmore, L.A., 2019. Activation of the Endonuclease that Defines mRNA 3′ Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex. Mol. Cell 73, 1217-1231.e11. https://doi.org/10.1016/j.molcel.2018.12.023
Hirtreiter, A., Damsma, G.E., Cheung, A.C.M., Klose, D., Grohmann, D., Vojnic, E., Martin, A.C.R., Cramer, P., Werner, F., 2010. Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. Nucleic Acids Res. 38, 4040–4051. https://doi.org/10.1093/nar/gkq135
Holbein, S., Scola, S., Loll, B., Dichtl, B.S., Hübner, W., Meinhart, A., Dichtl, B., 2011. The P-Loop Domain of Yeast Clp1 Mediates Interactions Between CF IA and CPF Factors in Pre-mRNA 3′ End Formation. PLoS ONE 6. https://doi.org/10.1371/journal.pone.0029139
Holoch, D., Moazed, D., 2015. RNA-mediated epigenetic regulation of gene expression. Nat. Rev. Genet. 16, 71–84. https://doi.org/10.1038/nrg3863
Hsin, J.-P., Manley, J.L., 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. 26, 2119–2137. https://doi.org/10.1101/gad.200303.112
Huang, X., Wang, D., Weiss, D.R., Bushnell, D.A., Kornberg, R.D., Levitt, M., 2010. RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. Proc. Natl. Acad. Sci. 107, 15745–15750. https://doi.org/10.1073/pnas.1009898107
Irvin, J.D., Kireeva, M.L., Gotte, D.R., Shafer, B.K., Huang, I., Kashlev, M., Strathern, J.N., 2014. A Genetic Assay for Transcription Errors Reveals Multilayer Control of RNA Polymerase II Fidelity. PLOS Genet. 10, e1004532. https://doi.org/10.1371/journal.pgen.1004532
Izumi, K., 2016. Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes. Mol. Syndromol. 7, 262–273. https://doi.org/10.1159/000448747
Jasnovidova, O., Stefl, R., 2013. The CTD code of RNA polymerase II: a structural view. WIREs RNA 4, 1–16. https://doi.org/10.1002/wrna.1138
Jimeno-González, S., Haaning, L.L., Malagon, F., Jensen, T.H., 2010. The Yeast 5′-3′ Exonuclease Rat1p Functions during Transcription Elongation by RNA Polymerase II. Mol. Cell 37, 580–587. https://doi.org/10.1016/j.molcel.2010.01.019
Jonkers, I., Lis, J.T., 2015. Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 16, 167–177. https://doi.org/10.1038/nrm3953
Kachaev, Z.M., Lebedeva, L.A., Kozlov, E.N., Shidlovskii, Y.V., 2020. Interplay of mRNA capping and transcription machineries. Biosci. Rep. 40. https://doi.org/10.1042/BSR20192825
Kaplan, C.D., Jin, H., Zhang, I.L., Belyanin, A., 2012. Dissection of Pol II Trigger Loop Function and Pol II Activity–Dependent Control of Start Site Selection In Vivo. PLOS Genet. 8, e1002627. https://doi.org/10.1371/journal.pgen.1002627

Kaplan, C.D., Larsson, K.-M., Kornberg, R.D., 2008. The RNA Polymerase II Trigger Loop Functions in Substrate Selection and is Directly Targeted by α-amanitin. Mol. Cell 30, 547–556. https://doi.org/10.1016/j.molcel.2008.04.023

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., Keller, W., 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO J. 23, 616–626. https://doi.org/10.1038/sj.emboj.7600070

Kecman, T., Kuś, K., Heo, D.-H., Duckett, K., Birot, A., Liberatori, S., Mohammed, S., Geis-Asteggiante, L., Robinson, C.V., Vasiljeva, L., 2018. Elongation/Termination Factor Exchange Mediated by PP1 Phosphatase Orchestrates Transcription Termination. Cell Rep. 25, 259-269.e5. https://doi.org/10.1016/j.celrep.2018.09.007

Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A., 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. https://doi.org/10.1101/gad.11.19.2545

Kilchert, C., Kecman, T., Priest, E., Hester, S., Kus, K., Castello, A., Mohammed, S., Vasiljeva, L., 2019. System-wide analyses of the fission yeast poly(A)+ RNA interactome reveal insights into organisation and function of RNA-protein complexes. bioRxiv 748194. https://doi.org/10.1101/748194

Kim, M., Ahn, S.-H., Krogan, N.J., Greenblatt, J.F., Buratowski, S., 2004a. Transitions in RNA polymerase II elongation complexes at the 3’ ends of genes. EMBO J. 23, 354–364. https://doi.org/10.1038/sj.emboj.7600053

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., Buratowski, S., 2004b. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature 432, 517–522. https://doi.org/10.1038/nature03041

Kireeva, M.L., Komissarova, N., Waugh, D.S., Kashlev, M., 2000. The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. J. Biol. Chem. 275, 6530–6536. https://doi.org/10.1074/jbc.275.9.6530

Kireeva, M.L., Nedialkov, Y.A., Cremona, G.H., Purtov, Y.A., Lubkowska, L., Malagon, F., Burton, Z.F., Strathern, J.N., Kashlev, M., 2008. Transient Reversal of RNA Polymerase II Active Site Closing Controls Fidelity of Transcription Elongation. Mol. Cell 30, 557–566. https://doi.org/10.1016/j.molcel.2008.04.017

Klein, B.J., Bose, D., Baker, K.J., Yusoff, Z.M., Zhang, X., Murakami, K.S., 2011. RNA polymerase and transcription elongation factor Spt4/5 complex structure. Proc. Natl. Acad. Sci. U. S. A. 108, 546–550. https://doi.org/10.1073/pnas.1013828108

Kobor, M.S., Greenblatt, J., 2002. Regulation of transcription elongation by phosphorylation. Biochim. Biophys. Acta BBA - Gene Struct. Expr., Transcription Elongation Control-2002 1577, 261–275. https://doi.org/10.1016/S0167-4781(02)00457-8

Kulish, D., Struhl, K., 2001. TFIIS enhances transcriptional elongation through an artificial arrest site in vivo. Mol. Cell. Biol. 21, 4162–4168. https://doi.org/10.1128/MCB.21.13.4162-4168.2001

Kumar, A., Clerici, M., Muckenfuss, L.M., Passmore, L.A., Jinek, M., 2019. Mechanistic insights into mRNA 3′-end processing. Curr. Opin. Struct. Biol., Catalysis and Regulation ● Protein Nucleic Interactions 59, 143–150. https://doi.org/10.1016/j.sbi.2019.08.001

Kwak, H., Lis, J.T., 2013. Control of Transcriptional Elongation. Annu. Rev. Genet. 47, 483–508. https://doi.org/10.1146/annurev-genet-110711-155440

Kyburz, A., Sadowski, M., Dichtl, B., 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.

Larson, M.H., Zhou, J., Kaplan, C.D., Palangat, M., Kornberg, R.D., Landick, R., Block, S.M., 2012. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. Proc. Natl. Acad. Sci. 109, 6555–6560. https://doi.org/10.1073/pnas.1200939109

Latchman, D.S., 1996. Transcription-Factor Mutations and Disease. N. Engl. J. Med. 334, 28–33. https://doi.org/10.1056/NEJM199601043340108

Lee, T.I., Young, R.A., 2013. Transcriptional Regulation and its Misregulation in Disease. Cell 152, 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Leeper, T.C., Qu, X., Lu, C., Moore, C., Varani, G., 2010. Novel protein-protein contacts facilitate mRNA 3′-processing signal recognition by Rna15 and Hrp1. J. Mol. Biol. 401, 334–349. https://doi.org/10.1016/j.jmb.2010.06.032

Lemay, J.-F., Marguerat, S., Larochelle, M., Liu, X., Nues, R. van, Hunyadkürti, J., Hoque, M., Tian, B., Granneman, S., Bähler, J., Bachand, F., 2016. The Nrd1-like protein Seb1 coordinates cotranscriptional 3′ end processing and polyadenylation site selection. Genes Dev. 30, 1558–1572. https://doi.org/10.1101/gad.280222.116

Lewis, B.A., Reinberg, D., 2003. The mediator coactivator complex: functional and physical roles in transcriptional regulation. J. Cell Sci. 116, 3667–3675. https://doi.org/10.1242/jcs.00734

Liu, X., Hoque, M., Larochelle, M., Lemay, J.-F., Yurko, N., Manley, J.L., Bachand, F., Tian, B., 2017. Comparative analysis of alternative polyadenylation in S. cerevisiae and S. pombe. Genome Res. 27, 1685–1695. https://doi.org/10.1101/gr.222331.117

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., Hahn, S., 2009. Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. Mol. Cell. Biol. 29, 4852–4863. https://doi.org/10.1128/MCB.00609-09

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

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., Lis, J.T., 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat. Protoc. 11, 1455–1476. https://doi.org/10.1038/nprot.2016.086

Maiuri, P., Knezevich, A., De Marco, A., Mazza, D., Kula, A., McNally, J.G., Marcello, A., 2011. Fast transcription rates of RNA polymerase II in human cells. EMBO Rep. 12, 1280–1285. https://doi.org/10.1038/embor.2011.196

Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M., Strathern, J.N., 2006a. Mutations in the Saccharomyces cerevisiae RPB1 Gene Conferring Hypersensitivity to 6-Azauracil. Genetics 172, 2201–2209. https://doi.org/10.1534/genetics.105.052415

Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M., Strathern, J.N., 2006b. Mutations in the Saccharomyces cerevisiae RPB1 Gene Conferring Hypersensitivity to 6-Azauracil. Genetics 172, 2201–2209. https://doi.org/10.1534/genetics.105.052415

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., Tong, L., 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end processing endonuclease. Nature 444. https://doi.org/10.1038/nature05363

Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C., Cramer, P., 2011. Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. EMBO J. 30, 1302–1310. https://doi.org/10.1038/emboj.2011.64

Martin-Tumasz, S., Brow, D.A., 2015. Saccharomyces cerevisiae Sen1 Helicase Domain Exhibits 5'- to 3'-Helicase Activity with a Preference for Translocation on DNA Rather than RNA. J. Biol. Chem. 290, 22880–22889. https://doi.org/10.1074/jbc.M115.674002

Mayer, A., Landry, H.M., Churchman, L.S., 2017. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. Curr. Opin. Cell Biol. 46, 72–80. https://doi.org/10.1016/j.ceb.2017.03.002

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Program, A.E., Shuman, S., Bentley, D.L., 1997. 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. https://doi.org/10.1101/gad.11.24.3306

Meyer, P.A., Li, S., Zhang, M., Yamada, K., Takagi, Y., Hartzog, G.A., Fu, J., 2015. Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. Mol. Cell. Biol. 35, 3354–3369. https://doi.org/10.1128/MCB.00520-15

Mischo, H.E., Proudfoot, N.J., 2013. Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. Biochim. Biophys. Acta BBA - Gene Regul. Mech., RNA polymerase II Transcript Elongation 1829, 174–185. https://doi.org/10.1016/j.bbagrm.2012.10.003

Misteli, T., Spector, D.L., 1999. RNA Polymerase II Targets Pre-mRNA Splicing Factors to Transcription Sites In Vivo. Mol. Cell 3, 697–705. https://doi.org/10.1016/S1097-2765(01)80002-2

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L., 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. https://doi.org/10.1074/jbc.M304454200

Neve, J., Patel, R., Wang, Z., Louey, A., Furger, A.M., 2017. Cleavage and polyadenylation: Ending the message expands gene regulation. RNA Biol. 14, 865–890. https://doi.org/10.1080/15476286.2017.1306171

Ni, Z., Olsen, J.B., Guo, X., Zhong, G., Ruan, E.D., Marcon, E., Young, P., Guo, H., Li, J., Moffat, J., Emili, A., Greenblatt, J.F., 2011. Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B. Transcription 2, 237–242. https://doi.org/10.4161/trns.2.5.17803

Ni, Z., Xu, C., Guo, X., Hunter, G.O., Kuznetsova, O.V., Tempel, W., Marcon, E., Zhong, G., Guo, H., Kuo, W.-H.W., Li, J., Young, P., Olsen, J.B., Wan, C., Loppnau, P., El Bakkouri, M., Senisterra, G.A., He, H., Huang, H., Sidhu, S.S., Emili, A., Murphy, S., Mosley, A.L., Arrowsmith, C.H., Min, J., Greenblatt, J.F., 2014. RPRD1A and RPRD1B Are Human RNA Polymerase II C-Terminal Domain Scaffolds for Ser5 Dephosphorylation. Nat. Struct. Mol. Biol. 21, 686–695. https://doi.org/10.1038/nsmb.2853

Noble, C.G., Beuth, B., Taylor, I.A., 2007. Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic Acids Res. 35, 87–99. https://doi.org/10.1093/nar/gkl1010

Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S.M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., Proudfoot, N.J., 2018. Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. Mol. Cell 72, 970-984.e7. https://doi.org/10.1016/j.molcel.2018.10.011

Ohnacker, M., Barabino, S.M., Preker, P.J., 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. https://doi.org/10.1093/emboj/19.1.37

Palo, M.Z., Mishanina, T.V., Landick, R., 2019. Conserved mechanisms of transcriptional pausing regulate diverse RNA polymerases. FASEB J. 33, 624.2-624.2. https://doi.org/10.1096/fasebj.2019.33.1_supplement.624.2

Parua, P.K., Booth, G.T., Sansó, M., Benjamin, B., Tanny, J.C., Lis, J.T., Fisher, R.P., 2018. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. Nature 558, 460–464. https://doi.org/10.1038/s41586-018-0214-z

Patturajan, M., Wei, X., Berezney, R., Corden, J.L., 1998. A Nuclear Matrix Protein Interacts with the Phosphorylated C-Terminal Domain of RNA Polymerase II. Mol. Cell. Biol. 18, 2406–2415. https://doi.org/10.1128/MCB.18.4.2406

Pérez-Cañadillas, J.M., 2006. Grabbing the message: structural basis of mRNA 3′UTR recognition by Hrp1. EMBO J. 25, 3167–3178. https://doi.org/10.1038/sj.emboj.7601190

Phatnani, H.P., Greenleaf, A.L., 2006. Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev. 20, 2922–2936. https://doi.org/10.1101/gad.1477006

Porrua, O., Libri, D., 2013. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nat. Struct. Mol. Biol. 20, 884–891. https://doi.org/10.1038/nsmb.2592

Proudfoot, N.J., 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science 352, aad9926. https://doi.org/10.1126/science.aad9926

Proudfoot, N.J., 1989. How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem. Sci. 14, 105–110. https://doi.org/10.1016/0968-0004(89)90132-1

Qiu, C., Erinne, O.C., Dave, J.M., Cui, P., Jin, H., Muthukrishnan, N., Tang, L.K., Babu, S.G., Lam, K.C., Vandeventer, P.J., Strohner, R., Van den Brulle, J., Sze, S.-H., Kaplan, C.D., 2016. High-Resolution Phenotypic Landscape of the RNA Polymerase II Trigger Loop. PLoS Genet. 12. https://doi.org/10.1371/journal.pgen.1006321

Rosado-Lugo, J.D., Hampsey, M., 2014. The Ssu72 Phosphatase Mediates the RNA Polymerase II Initiation-Elongation Transition. J. Biol. Chem. 289, 33916–33926. https://doi.org/10.1074/jbc.M114.608695

Saba, J., Chua, X.Y., Mishanina, T.V., Nayak, D., Windgassen, T.A., Mooney, R.A., Landick, R., 2019. The elemental mechanism of transcriptional pausing. eLife 8, e40981. https://doi.org/10.7554/eLife.40981

Sainsbury, S., Bernecky, C., Cramer, P., 2015. Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 16, 129–143. https://doi.org/10.1038/nrm3952

Schlackow, M., Marguerat, S., Proudfoot, N.J., Bähler, J., Erban, R., Gullerova, M., 2013. Genome-wide analysis of poly(A) site selection in Schizosaccharomyces pombe. RNA 19, 1617–1631. https://doi.org/10.1261/rna.040675.113

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A.R., Keller, W., Zavolan, M., Wahle, E., 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes Dev. 28, 2381–2393. https://doi.org/10.1101/gad.250985.114

Schreieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., Passmore, L.A., 2014. RNA polymerase II termination involves CTD tyrosine dephosphorylation by CPF subunit Glc7. Nat. Struct. Mol. Biol. 21, 175–179. https://doi.org/10.1038/nsmb.2753

Schwer, B., Sanchez, A.M., Shuman, S., 2012. Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. Proc. Natl. Acad. Sci. 109, 18024–18029. https://doi.org/10.1073/pnas.1208995109
Schwer, B., Schneider, S., Pei, Y., Aronova, A., Shuman, S., 2009. Characterization of the Schizosaccharomyces pombe Spt5-Spt4 complex. RNA 15, 1241–1250. https://doi.org/10.1261/rna.1572709
Schwer, B., Shuman, S., 2011. Deciphering the RNA Polymerase II CTD code in fission yeast. Mol. Cell 43, 311–318. https://doi.org/10.1016/j.molcel.2011.05.024
Shatkin, A.J., 1976. Capping of eucaryotic mRNAs. Cell 9, 645–653. https://doi.org/10.1016/0092-8674(76)90128-8
Sheets, M.D., Ogg, S.C., 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.
Shetty, A., Kallgren, S.P., Demel, C., Maier, K.C., Spatt, D., Alver, B.H., Cramer, P., Park, P.J., Winston, F., 2017. Spt5 plays vital roles in the control of sense and antisense transcription elongation. Mol. Cell 66, 77-88.e5. https://doi.org/10.1016/j.molcel.2017.02.023
Shuman, S., 2001. Structure, mechanism, and evolution of the mRNA capping apparatus. Prog. Nucleic Acid Res. Mol. Biol. 66, 1–40. https://doi.org/10.1016/s0079-6603(00)66025-7
Soares, L.M., Buratowski, S., 2012. Yeast Swd2 Is Essential Because of Antagonism between Set1 Histone Methyltransferase Complex and APT (Associated with Pta1) Termination Factor. J. Biol. Chem. 287, 15219–15231. https://doi.org/10.1074/jbc.M112.341412
Sonkar, A., Gaurav, S., Ahmed, S., 2017. Fission yeast Ctf1, a cleavage and polyadenylation factor subunit is required for the maintenance of genomic integrity. Mol. Genet. Genomics 292, 1027–1036. https://doi.org/10.1007/s00438-017-1329-x
Sonkar, A., Yadav, S., Ahmed, S., 2016. Cleavage and polyadenylation factor, Rna14 is an essential protein required for the maintenance of genomic integrity in fission yeast Schizosaccharomyces pombe. Biochim. Biophys. Acta BBA - Mol. Cell Res. 1863, 189–197. https://doi.org/10.1016/j.bbamcr.2015.11.007
Spåhr, H., Calero, G., Bushnell, D.A., Kornberg, R.D., 2009. Schizosacharomyces pombe RNA polymerase II at 3.6-Å resolution. Proc. Natl. Acad. Sci. 106, 9185–9190. https://doi.org/10.1073/pnas.0903361106
Steinmetz, E.J., 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. https://doi.org/10.1128/mcb.16.12.6993
Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., Bartel, D.P., 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature 508, 66–71. https://doi.org/10.1038/nature13007
Sugiyama, T., Sugioka-Sugiyama, R., Hada, K., Niwa, R., 2012. Rhn1, a Nuclear Protein, Is Required for Suppression of Meiotic mRNAs in Mitotically Dividing Fission Yeast. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0042962
Sun, Y., Zhang, Y., Hamilton, K., Manley, J.L., Shi, Y., Walz, T., Tong, L., 2018. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc. Natl. Acad. Sci. 115, E1419–E1428. https://doi.org/10.1073/pnas.1718723115
Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J., Shenk, T., 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. Genes Dev. 4, 2112–2120. https://doi.org/10.1101/gad.4.12a.2112

Tian, B., Graber, J.H., 2012. Signals for pre-mRNA cleavage and polyadenylation. Wiley Interdiscip. Rev. RNA 3, 385–396. https://doi.org/10.1002/wrna.116
Toulokhonov, I., Zhang, J., Palangat, M., Landick, R., 2007. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. Mol. Cell 27, 406–419. https://doi.org/10.1016/j.molcel.2007.06.008
Valentini, S.R., Weiss, V.H., Silver, P.A., 1999. Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. RNA 5, 272–280.
Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., Meinhart, A., 2008. The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. 15, 795–804. https://doi.org/10.1038/nsmb.1468
Vermulst, M., Denney, A.S., Lang, M.J., Hung, C.-W., Moore, S., Moseley, M.A., Thompson, J.W., Madden, V., Gauer, J., Wolfe, K.J., Summers, D.W., Schleit, J., Sutphin, G.L., Haroon, S., Holczbauer, A., Caine, J., Jorgenson, J., Cyr, D., Kaeberlein, M., Strathern, J.N., Duncan, M.C., Erie, D.A., 2015. Transcription errors induce proteotoxic stress and shorten cellular lifespan. Nat. Commun. 6, 1–11. https://doi.org/10.1038/ncomms9065
Viktorovskaya, O.V., Engel, K.L., French, S.L., Cui, P., Vandeventer, P.J., Pavlovic, E.M., Beyer, A.L., Kaplan, C.D., Schneider, D.A., 2013. Divergent contributions of conserved active site residues to transcription by eukaryotic RNA polymerases I and II. Cell Rep. 4, 974–984. https://doi.org/10.1016/j.celrep.2013.07.044
von Hippel, P.H., Bear, D.G., Morgan, W.D., McSwiggen, J.A., 1984. Protein-nucleic acid interactions in transcription: a molecular analysis. Annu. Rev. Biochem. 53, 389–446. https://doi.org/10.1146/annurev.bi.53.070184.002133
Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., Cramer, P., 2018a. Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. Nature 560, 607–612. https://doi.org/10.1038/s41586-018-0440-4
Vos, S.M., Farnung, L., Urlaub, H., Cramer, P., 2018b. Structure of paused transcription complex Pol II–DSIF–NELF. Nature 560, 601–606. https://doi.org/10.1038/s41586-018-0442-2
Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J., 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.
Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., Kornberg, R.D., 2006. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell 127, 941–954. https://doi.org/10.1016/j.cell.2006.11.023
Watts, J.A., Burdick, J., Daigneault, J., Zhu, Z., Grunseich, C., Bruzel, A., Cheung, V.G., 2019. cis Elements that Mediate RNA Polymerase II Pausing Regulate Human Gene Expression. Am. J. Hum. Genet. 105, 677–688. https://doi.org/10.1016/j.ajhg.2019.08.003
Webb, S., Hector, R.D., Kudla, G., Granneman, S., 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. Genome Biol. 15, R8. https://doi.org/10.1186/gb-2014-15-1-r8
West, S., Gromak, N., Proudfoot, N.J., 2004. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432, 522–525. https://doi.org/10.1038/nature03035
Wittmann, S., Renner, M., Watts, B.R., Adams, O., Huseyin, M., Baejen, C., El Omari, K., Kilchert, C., Heo, D.-H., Kecman, T., Cramer, P., Grimes, J.M., Vasiljeva, L., 2017. The conserved protein Seb1 drives transcription termination by binding RNA

polymerase II and nascent RNA. Nat. Commun. 8, 1–15. https://doi.org/10.1038/ncomms14861
Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M.M., Manley, J.L., Tong, L., 2010. Crystal structure of the human symplekin–Ssu72–CTD phosphopeptide complex. Nature 467, 729–733. https://doi.org/10.1038/nature09391
Xiang, K., Tong, L., Manley, J.L., 2014. Delineating the Structural Blueprint of the Pre-mRNA 3′-End Processing Machinery. Mol. Cell. Biol. 34, 1894–1910. https://doi.org/10.1128/MCB.00084-14
Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., Handa, H., 2006. P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. Mol. Cell 21, 227–237. https://doi.org/10.1016/j.molcel.2005.11.024
Yang, F., Hsu, P., Lee, S.D., Yang, W., Hoskinson, D., Xu, W., Moore, C., Varani, G., 2017. The C terminus of Pcf11 forms a novel zinc-finger structure that plays an essential role in mRNA 3′-end processing. RNA 23, 98–107. https://doi.org/10.1261/rna.058354.116
Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., Corden, J.L., 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proc. Natl. Acad. Sci. 93, 6975–6980. https://doi.org/10.1073/pnas.93.14.6975
Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., Ansari, A.Z., 2012. Ssu72 Phosphatase-dependent Erasure of Phospho-Ser7 Marks on the RNA Polymerase II C-terminal Domain Is Essential for Viability and Transcription Termination. J. Biol. Chem. 287, 8541–8551. https://doi.org/10.1074/jbc.M111.335687
Zhang, Z., Gilmour, D.S., 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. https://doi.org/10.1016/j.molcel.2005.11.002
Zhao, J., Hyman, L., Moore, C., 1999a. Formation of mRNA 3′ Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. Microbiol. Mol. Biol. Rev. 63, 405–445.
Zhao, J., Kessler, M., Helmling, S., O'Connor, J.P., Moore, C., 1999b. Pta1, a Component of Yeast CF II, Is Required for Both Cleavage and Poly(A) Addition of mRNA Precursor. Mol. Cell. Biol. 19, 7733–7740.
Zhou, K., Kuo, W.H.W., Fillingham, J., Greenblatt, J.F., 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. Proc. Natl. Acad. Sci. 106, 6956–6961. https://doi.org/10.1073/pnas.0806302106