Investigating the role of Dis2 phosphatase in RNA Polymerase II transcription termination in fission yeast



Tea Kecman

Hertford College

University of Oxford

A dissertation submitted for the degree of

Doctor of Philosophy

Trinity term 2018

To my parents

Acknowledgements

First and foremost I would like to express my gratitude to my supervisor Lidia. Thank you for the support and guidance throughout my DPhil, your enthusiasm for science is restless and contagious.

I am also grateful to all my fellow labmates for the kindness, the help and the extraordinary work environment. A special thank you to Katie and Kris who were involved in my project (I will always be grateful for all your protein preps!). To Dong-Hyuk who shared with me his knowledge on ChIP and was kind enough to provide me with his Ser5P data. Thanks to Doğukan, Sina, Cornelia, Adrien and Kris for discussions about any aspect of science and research. Also, Monica and Beth for many nice chats.

Furthermore, I would like to thank Hamish, Sina, Guifeng, and Naomi for their help with bioinformatics. I am particularly grateful to Hamish who shared with me his ChIP-seq alignment script.

I am grateful to my thesis committee, Rob Klose and Alfredo Castello, for their critical comments and feedback.

My gratitude goes also to Hertford College for giving me the opportunity to be here.

To Francesco, for always being there and believing in me. My time in Oxford wouldn't have been as happy without you.

To my parents and my sister - thank you for your unconditional love and support. My achievements are yours.

Abstract

Investigating the role of Dis2 phosphatase in RNA Polymerase II transcription termination in fission yeast

Tea Kecman, Hertford College, *DPhil* in Biochemistry, Trinity term 2018

The end of the RNA polymerase II (Pol II) transcription cycle is a tightly regulated process that requires: (i) dissociation of elongation factors, (ii) recruitment of the 3' end processing machinery that is essential for pre-mRNA cleavage and polyadenylation, and (iii) association of termination factors to dislodge Pol II from the DNA template. A characteristic feature of termination factors is that they directly interact with the phosphorylated C-terminal domain (CTD) of the largest subunit of Pol II (Rpb1). The CTD consists of heptad repeats ($Y_1S_2P_3T_4S_5P_6S_7$) where Tyr1, Ser2, Ser5, Ser7 and Thr4 can be reversibly phosphorylated during the transcription cycle. The work presented here reveals that Thr4 phosphorylation levels are enriched at the 3' ends of protein-coding genes in fission yeast and identifies Dis2 as a Pol II-CTD phosphatase that dephosphorylates Thr4 both *in vitro* and *in vivo*. Dis2 activity is required for Thr4 dephosphorylation at the polyadenylation site, for recruitment of termination factors Pcf11, Rhn1 and Seb1 and for normal Pol II termination. Interestingly, Seb1 co-transcriptional recruitment is shown to be dependent not on Dis2 activity on CTD but on its ability to dephosphorylate the elongation factor Spt5, suggesting that Dis2 regulates the elongation-to-termination transition by targeting multiple substrates.

Contents

1 Introduction **1**

1.1 The RNA Polymerase II carboxy-terminal domain (CTD) 2

1.1.1 The CTD code 2

1.1.2 CTD phosphorylation 4

1.1.3 CTD writers and erasers 6

1.1.4 Coordination of transcription and RNA processing by CTD phosphorylation 8

1.2 Transcription initiation and elongation 9

1.2.1 Initiation. 9

1.2.2 Elongation 10

1.2.3 Spt5 complex: a multi-faceted regulator of transcription elongation. 12

1.3 Transcription termination 14

1.3.1 3' end processing machinery 15

1.3.2 Termination: allosteric vs torpedo model 17

1.3.3 The role of CID-containing proteins in transcription termination 19

1.3.4 The role of CTD phosphatases in transcription termination . . . 21

1.4 The PP1 phosphatase 21

1.5 Objectives of this work 24

2 Thr4 phosphorylated CTD in fission yeast **26**

2.1 Introduction 26

2.2 Tyr1 and Thr4 CTD phospho-marks are enriched at the 3' ends of protein-coding genes 27

2.3 A constitutively phosphorylated CTD on Thr4 is lethal for *S. pombe* . . 29

2.4 Termination factors bind Thr4P *in vitro* 30

2.4.1 *S. pombe* termination factors bind Thr4P *in vitro* 30

2.4.2 *S. cerevisiae* termination factors bind Thr4P *in vitro* 32

2.5 Discussion 34

3 Dis2 is a phosphatase that targets Thr4P-CTD in fission yeast and is important for transcription termination 37
3.1 Introduction . . . 37
3.2 Dis2 mutants show an increase in Thr4P-CTD levels in the whole cell extract . . . 38
3.3 Dis2 specifically dephosphorylates Thr4 *in vitro* . . . 39
3.4 Deletion of *dis2* leads to impaired transcription termination and higher levels of transcriptionally engaged CTD-Thr4P . . . 41
3.5 Recruitment of termination factors is impaired upon *dis2* deletion . . . 43
3.6 Discussion . . . 46

4 Dis2 regulates Seb1 recruitment by dephosphorylating Spt5 48
4.1 Introduction . . . 48
4.2 Spt5 phosphorylation disrupts its ability to bind RNA . . . 49
4.3 Hyper-phosphorylation of Spt5 leads to partial loss of Seb1 recruitment 51
4.4 Seb1 competes with Spt5 for binding to Pol II *in vitro* . . . 53
4.5 Hyper-phosphorylation of Spt5 is not the sole cause of the termination defect in *dis2*Δ . . . 54
4.6 Discussion . . . 58

5 Discussion 61
5.1 What is the role of Thr4P-CTD at the gene 3' ends? . . . 61
5.2 What is Dis2 role in transcription termination? . . . 64
5.3 What is the importance of Spt5 dephosphorylation at the end of transcription? . . . 66
5.4 How could phosphorylation regulate transcription termination? . . . 68

6 Material and methods 71
6.1 *S.pombe* media and growth conditions . . . 71
6.2 Yeast strain construction . . . 71
6.3 Yeast transformation . . . 72
6.4 Yeast whole-cell protein extraction for WB . . . 73
6.5 Yeast protein purification . . . 73
6.6 Phosphatase assay . . . 74
6.7 Cloning of recombinant proteins . . . 75
6.8 Transformation of *E.coli* cells . . . 76
6.9 Protein expression and purification from *E.coli* . . . 76

6.10 Fluorescent anisotropy (FA) assay 78
6.11 Electrophoretic mobility shift assay (EMSA) 78
6.12 SDS-PAGE 79
6.13 Immunoblotting 79
6.14 ChIP-sequencing 80
6.15 Yeast strains used in this study 81
6.16 Plasmids used in this study 84
6.17 Oligonucleotides used in this study 84
6.18 Antibodies used in this study 86

List of Figures

1.1 Sequence of the CTD in mammals and yeast. 3
1.2 Schematic representation of Spt5 protein domains 13
1.3 Alignment of human and yeast PP1 isoforms 23

2.1 Differential distribution of phospho-marks on protein-coding genes. . . 28
2.2 Effect of Thr4 mutations on fission yeast viability 29
2.3 Purification of recombinant Seb1-CID and SUMO Pcf11-CID. 31
2.4 Fluorescence anisotropy (FA) assay for *S. pombe* termination factors. . 32
2.5 Fluorescence anisotropy (FA) assay for *S. cerevisiae* termination factors. 33

3.1 Deletion of *dis2* affects Pol II-CTD Thr4 phosphorylation *in vivo*. . . . 38
3.2 Dis2 is Thr4P phosphatase *in vitro*. 40
3.3 Increased Thr4P downstream of PAS upon loss of Dis2 coincides with transcription termination defect. 42
3.4 Lack of Dis2 activity leads to increased recruitment of the termination factors Pcf11 and Rhn1. 44
3.5 Lack of Dis2 activity leads to decreased recruitment of the termination factor Seb1. 45

4.1 Spt5 T1E does not bind RNA. 50
4.2 Spt5 protein levels in *spt5* mutants. 51
4.3 *spt5* T1A rescues Seb1 recruitment 52
4.4 Purification of CTD-less Pol II, recombinant Spt4/5 and Seb1 for binding assay. 53
4.5 Seb1 competes with Spt5 for binding to Pol II. 54
4.6 *spt5* T1A partially rescues 3' extended transcription on *rpl1202*. 55
4.7 Pol II metagene profile in *spt5* mutants. 56
4.8 *spt5* T1A partially rescues Pol II termination defects. 58

5.1 Sequence alignments of CIDs from yeast termination factors. 63

5.2 Model. 66

List of Tables

1.1 CTD kinases in mammals and yeast 7
1.2 CTD phosphatases in mammals and yeast 8
1.3 Components of the 3' end processing machinery in humans and yeast. . 16
1.4 CID-containing proteins in yeast and humans 20

3.1 Dis2, Ppn1 and associated CPF subunits identified by MS 39

6.1 Yeast strains used in this study 81
6.2 Plasmids used in this study 84
6.3 Oligonucelotides used in this study 84
6.4 Antibodies used in this study 86

Declaration

All of the data described in this thesis were generated by myself, except as detailed below:

• Ser5P ChIP-seq (both practical part and data alignment) was entirely done by Dr Dong-Hyuk Heo, Vasilieva lab (section 2.2).

• CTD T4A and T4E diploid strains with subsequent tetrad analysis were made by Dr Adrien Birot, Vasilieva lab (section 2.3).

• Recombinant expression and purifications of S. cerevisiae termination factors (Rtt103, Pcf11, Nrd1) with subsequent fluorescence anisotropy assays were carried out by Ms Katie Duckett, Vasilieva lab, under my supervision (section 2.4.2).

• Purification of S. pombe Pcf11 in section 2.4.1 was carried out by Dr Krzysztof Kus, Vasilieva lab.

• Trypsin digestion and mass spectrometric analysis of the Dis2/Dis2 R245Q purification was performed by the TDI Mass Spectrometry Laboratory, Kessler lab , University of Oxford (section 3.3).

• The practical part of the first Pol II and Thr4P ChIP-seq experiment in section 3.4 was done by Dr Dong-Hyuk Heo, Vasilieva lab. The practical part of the second replica was done by me together with Dr Dong-Hyuk Heo (section 3.4).

• Purification of the CTD-less Pol II, Spt5 and Seb1 in section 4.4 was done by Dr Krzysztof Kus, Vasilieva lab.

• Some strains used in this thesis were obtained from outside the lab (section 6.14) and some plasmids were generated by Ms Katie Duckett (section 6.15).

• Most of the data presented in this thesis, both text and figures, are based on a manuscript which has been submitted for publication and was co-written by Dr Lidia Vasilieva

Abbreviations

E. coli	*Escherichia coli*
H. sapiens	*Homo sapiens*
S. cerevisiae	*Saccharomyces cerevisiae* (budding yeast)
S. pombe	*Schizosaccharomyces pombe* (fission yeast)
aa	amino acid
CF	cleavage factor
ChIP	Chromatin Immunoprecitation
CID	CTD-interacting domain
CPF	cleavage and polyadenylation factor
CPSF	cleavage and polyadenylation specificity factor
CstF	cleavage stimulation factor
CTD	C-terminal domain of Rpb1 in Pol II
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSIF	DRB-sensitivity factor
DTT	dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
EMMG	Edinburgh minimal medium + glutamate

EMSA	electrophoretic mobility shift assay
FAM	5'-fluorescein amidite
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	horse radish peroxidase
IP	immunoprecipitation
K_d	dissociation constant
kDa	kilo Dalton
LB	Lysogeny broth
mRNA	messenger RNA
MS	mass spectrometry
ncRNA	non-coding RNA
NELF	negative elongation factor
nt	nucleotide(s)
OD_{600}	optical density measured at 600 nm
ORF	open reading frame
P-TEFb	positive elongation factro b
pA	polyadenylation
PAGE	polyacrylamide gel electrophoresis
PAS	poly(A) site
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
Pol II	RNA polymerase II
RNA	ribonucleic acid
RRM	RNA recognition motif

rRNA	ribosomal RNA
RT	reverse transcription
RT-PCR	reverse transcriptase PCR
SDS	sodium dodecyl sulfate
Seq	sequencing (refers to next-generation sequencing methods)
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SUMO	small ubiquitin-like modifier
TBE	Tris-buffered EDTA
TBST	Tris-buffered saline + Tween
TCA	trichloroacetic acid
TEV	tobacco etch virus
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
ts	temperature sensitive
TSS	transcription start site
U	unit of enzymatic activity
WT	wild-type
YES	yeast extract broth + supplements

Chapter 1

Introduction

Transcription of DNA into RNA is an essential process and is carried out by RNA Polymerases. While prokaryotes rely on one RNA Polymerase, eukaryotes have evolved multiple specialized forms of RNA polymerases, each transcribing a distinct subset of RNAs. RNA Polymerase I (Pol I) is responsible for the synthesis of most ribosomal RNAs (rRNAs), while RNA Polymerase III (Pol III) produces the 5S rRNA, transfer (t)RNAs, and the U6 small nuclear (sn)RNA. RNA Polymerase II (Pol II) also transcribes many non-coding RNAs (ncRNAs) including small nucleolar (sno) and four snRNAs, but is the sole polymerase responsible for the transcription of all protein-coding genes into messenger RNAs (mRNAs). Plants have two additional RNA polymerases (Pol IV and V) that evolved as specialized forms of Pol II (Luo and Hall, 2007) and orchestrate non-coding RNA-mediated gene silencing (Haag and Pikaard, 2011).

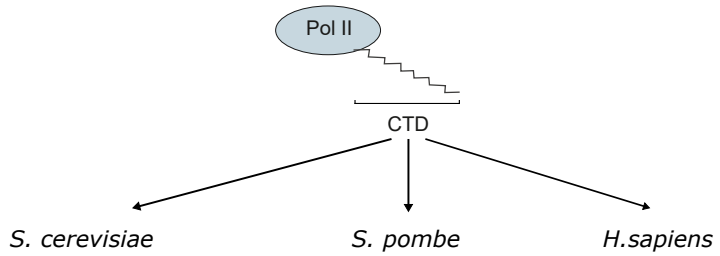
Transcription of mRNAs is a highly coordinated process that requires: recruitment of Pol II to promoters, transcription initiation, elongation and termination. Processing of mRNA precursor to form mature and functional transcripts occurs simultaneously with their synthesis by Pol II and includes 5' capping, RNA splicing and 3' cleavage/polyadenylation of the transcript. Both transcription and RNA processing are tightly regulated. This regulation is crucial for the production of functional transcripts.

1.1 The RNA Polymerase II carboxy-terminal domain (CTD)

1.1.1 The CTD code

Pol II is a multi-subunit complex (12 subunits in humans and yeast) and one of the key players in regulating the different stages of transcription is the carboxy-terminal domain (CTD) of Rpb1, Pol II largest subunit. The CTD is an unstructured domain that is composed of tandemly repeated heptapeptides of consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. The number of heptad repeats varies from 26 in budding yeast (*Saccharomyces cerevisiae/S.cerevisiae*), 29 in fission yeast (*Schizosaccharomyces pombe/ S.pombe*) to 52 in mammals (Figure 1.1). While in yeast almost all the heptapeptides perfectly match the consensus sequence, the mammalian CTD counts 21 consensus and 31 non-consensus repeats (reviewed in Eick and Geyer, 2013; Corden, 2013; Jeronimo et al., 2016; Zaborowska et al., 2016). The CTD is a unique feature of Pol II. No similar domains are found in Pol I or Pol II, nor in the prokaryotic RNA Polymerase. Plant-specific Pol IV and Pol V also lack the CTD-like structure, although Pol V largest subunit harbours a unique repetitive sequence that binds to ARGONAUTE 4 (Haag and Pikaard, 2011).

The CTD is subject to extensive post-translational modifications, with the serine (Ser), tyrosine (Tyr) and threonine (Thr) being phosphorylated *in vivo*, and the two prolines existing in either cis- or trans-configuration. Glycosylation also occurs throughout the CTD *in vitro*, but whether this is the case *in vivo* is still unclear. In mammals, the divergent nature of the CTD gives space to additional modifications (acetylation, ubiquitinylation, methylation). The patterns of CTD post-translational modifications change during the transcription cycle, generating a CTD code (Buratowski, 2003) that orchestrates transcription and RNA processing (reviewed in Eick and Geyer, 2013; Corden, 2013; Jeronimo et al., 2016; Zaborowska et al., 2016).



S. cerevisiae	*S. pombe*	*H.sapiens*
FSPTSPT	YGLTSPS	YSPTSPA
YSPTSPA	YSPSSPG	YEPRSPGG
YSPTSPS	YSTSPA	YTPQSPS
YSPTSPS	YMPSSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPN
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSATSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPA	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPS	YSPTSPS	YSPTSPS
YSPTSPN	YSPTSPS	YSPTSPN
YSPTSPS	YSPTSPS	YSPTSPN
YSPTSPG	YSPTSPS	YTPTSPS
YSPGSPA	YSPTSPS	YSPTSPS
YSPKQDEQKHNENENSR	YSPTSPS	YSPTSPN
	YSPTSPS	YTPTSPN
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
		YSPTSPS
		YSPSSPR
		YTPQSPT
		YTPSSPS
		YSPSSPS
		YSPASPK
		YTPTSPS
		YSPSSPE
		YTPTSPK
		YSPTSPK
		YSPTSPK
		YSPTSPT
		YSPTTPK
		YSPTSPT
		YSPTSPV
		YTPTSPK
		YSPTSPT
		YSPTSPK
		YSPTSPT
		YSPTSPKGST
		YSPTSPG
		YSPTSPT
		YSLTSPAISPDDSDEEN

Figure 1.1: Sequence of the CTD in mammals and yeast.
The CTDs from yeast (*S.cerevisiae and S.pombe*) and mammals (*H.sapiens*) are shown. The heptapeptides that perfectly match the consensus sequence are highlighted in red.

1.1.2 CTD phosphorylation

Patterns during the transcription cycle. The phosphorylation changes of the CTD during the transcription cycle have been analyzed genome-wide in *S. cerevisiae* and mammals using chromatin immunoprecipitation coupled to DNA microarray (ChIP-chip) first, and later to next generation sequencing (ChIP-seq). ChIP has been for many years the best technique to study the binding of proteins to genomic DNA. Recently, mammalian native elongating transcript sequencing (mNET-seq) has been also used in human cells to map the location of transcribing Pol II relative to the nascent transcript (Nojima et al., 2015).

ChIP-seq has shown that the phosphorylation of CTD at Ser2 and Ser5 is similar in budding yeast and mammals. Ser5P is enriched close to the transcription start sites (TSS), while Ser2P peaks at the 3' end of protein-coding genes (Hintermair et al., 2012; Kim et al., 2010; Bataille et al., 2012; Mayer et al., 2010; Milligan et al., 2016; Tietjen et al., 2010). mNET-seq has also detected an enrichment of Ser2P at the 3' end of human genes, but showed high levels of Ser5P over spliced exons, rather than an enrichment at the gene 5' ends (Nojima et al., 2015). mNET-seq detects the nascent RNA in the active site of immunoprecipitated Pol II. This RNA needs to be at least 35 nt long to be captured. Thus, it is possible that at the beginning of the transcription cycle Ser5-phosphorylated Pol II has not yet synthesized RNAs long enough to be detected by mNET-seq (Nojima et al., 2015). Ser7P phospho-mark is also placed early during the transcription cycle, but while in mammals the Ser7P signal drops downstream the TSS (Brookes et al., 2012; Descostes et al., 2014), in budding yeast it is retained until the end of genes (Kim et al., 2010; Tietjen et al., 2010).

While the patterns of serine phosphorylation are similar in budding yeast and mammals, the distributions of Tyr1P and Thr4P are not. Tyr1P is enriched at the beginning

and end of mammalian protein-coding genes (Descostes et al., 2014; Schlackow et al., 2017), while in budding yeast Tyr1P is enriched in gene bodies and depleted at the 5' and 3' ends of genes (Mayer et al., 2012). Thr4P is highest at the end of genes in mammals (Hintermair et al., 2012; Schlackow et al., 2017), while in *S. cerevisiae* Thr4P is present both within the gene body and at the gene 3' ends (Mayer et al., 2012; Harlen et al., 2016).

Relative abundance of phosphorylation sites. Analysis of CTD phosphorylation by ChIP or NET-seq relies on monoclonal antibodies that recognize the specific CTD phospho-marks. Antibodies are a powerful tool to determine the phosphorylation patterns, but they can't give a measurement of the relative levels of the different phospho-marks. Therefore for many years the *in vivo* stoichiometry of CTD phosphorylation remained unknown. Mass spectrometry would have solved the problem if it wasn't for the repetitive nature of the CTD and the lack of protease cleavage sites within the repeats. Only recently, a combined genetic and mass spectrometry (MS) approach partially overcame this problem. By genetically modifying the CTD sequence in budding yeast (Schüller et al., 2016; Suh et al., 2016) and mammals (Schüller et al., 2016), it was possible to introduce multiple trypsin cleavage sites that could be used to generate unique peptides. Mass spectrometry of this engineered CTD has showed that only one-third of the repeats are phosphorylated, mainly with one phosphate per heptad. Ser2P and Ser5P are much more abundant than phosphorylated Tyr1 and Ser7 in both mammals and budding yeast. Thr4P levels are higher that Tyr1P/Ser7P but lower than Ser2P/Ser5P in mammals. For what concerns the stoichiometry of Thr4P in budding yeast, there is some discrepancy between the published works. Schüller et al. (2016) found that Thr4P levels are similar to those of Ser2P, while Suh et al. (2016) showed low Thr4 phosphorylation. All these values were averaged over a mixed population of Pol II and there may be subsets of Pol II where the relative abundance

of the phospho-marks differs from the average. It is also important to note that this MS-based quantification of CTD phosphorylation may be incomplete since peptides for which the phospho-site could not be assigned univocally were discarded from the analysis (Schüller et al., 2016) and that low Ser7P levels may reflect the multiple mutations made at this position (Schüller et al., 2016; Suh et al., 2016).

1.1.3 CTD writers and erasers

The dynamic phosphorylation of the CTD that is so important to coordinate transcription is regulated by the combined action of specific kinases ("writers") and phosphatases ("erasers") that are timely recruited at different stages of transcription. Many of the enzymes acting on the CTD have been identified (Tables 1.1, 1.2), although for some there are discrepancies between different studies and further investigation is needed to determine their specificity and role *in vivo*. For many years, the main interest was to find and characterize the enzymes acting on Ser2, Ser5 and Ser7. Only recently, more effort has been made to determine which phosphatases and kinases target Tyr1 and Thr4. Thus, the enzymes acting on these two phospho-marks are often unknown or need to be further investigated (reviewed in Eick and Geyer, 2013; Corden, 2013; Jeronimo et al., 2016; Zaborowska et al., 2016).

CTD kinases. The metazoan Cdk7 subunit of the transcription initiation factor TFIIH phosphorylates Ser5 and Ser7 early during transcription, as do its orthologues, budding yeast Kin28 and fission yeast Mcs6. In addition to Cdk7, the Cdk8 subunit of the the Mediator complex targets Ser5 *in vitro*, but its role *in vivo* is still unclear (reviewed in Eick and Geyer, 2013; Corden, 2013; Jeronimo et al., 2016; Zaborowska et al., 2016).

Ctk1 in *S. cerevisiae* and Lsk1 in *S. pombe* are the major Ser2 kinases. In both yeast

there is an additional Ser2 kinase, Bur1 in *S.cerevisiae* and Cdk9 in *S.pombe*. Despite the progress in yeast, the identity of the major Ser2 kinase in higher eukaryotes remains unclear. For years Ser2 was considered to be the main target of the mammalian Cdk9, subunit of P-TEFb. Although inhibition of Cdk9 decreases Ser2P level as detected by western blot, Cdk9 preferentially targets Ser5 *in vitro*. Recently, two additional Ser2 kinases have been identified in higher eukaryotes: Cdk12 and Cdk13. These metazoan homologues of yeast Lsk1/Ctk1, have Ser2, Ser5 and Ser7 phosphorylation activity *in vitro*, while their role *in vivo* needs further investigation. Additionally, the bromodomain protein Brd4 has recently being characterized as a kinase for Ser2 *in vitro* and *in vivo* (reviewed in Heidemann et al., 2013; Zaborowska et al., 2016; Jeronimo et al., 2016).

In human cells, Plk3 has been characterized as a Thr4-specific kinase (Hintermair et al., 2012), while c-Abl has been reported to phosphorylate Tyr1 *in vitro*, but its role *in vivo* is still unclear (Baskaran et al., 1993). Recently, Hintermair et al. (2016) showed that mammalian Plk1 phosphorylates Thr4 in mitotic cells. To date no Thr4 kinase have been identified in yeast, while Slt2 has been recently reported to phosphorylate Tyr1 in *S.cerevisiae* (Yurko et al., 2017).

Table 1.1: CTD kinases in mammals and yeast
? indicates an enzyme inferred from homology, whose ability to target the CTD has not been tested directly.

Mammals	*S. cerevisiae*	*S. pombe*	CTD specificity
Cdk7	Kin28	Msc6	Ser5, Ser7
Cdk8	Srb10	Cdk8	Ser2, Ser5
Cdk9	Bur1	Cdk9	Ser2, Ser5, Ser7
Cdk12,13	Ctk1	Lsk1	Ser2, Ser5, Ser7
Brd4	Bdf1/2?	Bdf1/2?	Ser2
c-Abl	-	-	Tyr1
Mapk7?	Slt2	Pmk1?	Tyr1
Plk3,Plk1	Cdc5?	Plo1?	Thr4

CTD phosphatases. Ser5 phospho-mark is erased within promoter-proximal regions by yeast Rtr1 (RPAP2 in humans) and to a lesser extent by Ssu72. Ssu72 mainly acts at the end of the transcription cycle where it dephosphorylates both Ser5P and Ser7P (reviewed in Zaborowska et al., 2016; Jeronimo et al., 2016). Fcp1 is the Ser2-specific phosphatase and is associated with both ends of protein-coding genes, although it is more active at the 3' end. In *S. cerevisiae* a strain with a catalytically dead Fcp1 shows increased levels of Thr4P on western blots, suggesting that Fcp1 may also target Thr4P in addition to Ser2P (Allepuz-Fuster et al., 2014). The protein phosphatase 1 (PP1) Glc7, a component of the 3' end processing machinery in budding yeast, dephosphorylates Tyr1 close to the polyadenylation site (PAS) (Schreieck et al., 2014). In *Kluyveromyces lactis* (yeast), Rtr1 can also dephosphorylate Tyr1 *in vitro* (Hsu et al., 2014).

Table 1.2: CTD phosphatases in mammals and yeast
? indicates an enzyme inferred from homology, whose ability to target the CTD has not been tested directly.

Mammals	*S. cerevisiae*	*S. pombe*	CTD specificity
RPAP2	Rtr1	Rtr1?	Ser5
Ssu72	Ssu72	Ssu72?	Ser5, Ser7
Fcp1	Fcp1	Fcp1	Ser2, Ser5, Thr4
PP1?	Glc7	Dis2/Sds21?	Tyr1

1.1.4 Coordination of transcription and RNA processing by CTD phosphorylation

Changes in CTD phosphorylation are believed to coordinate the association of transcription and RNA-processing factors to polymerase (discussed below). This fine-tuned recruitment of transcription factors is crucial for Pol II transitions between the different

stages of transcription: initiation, elongation and termination.

1.2 Transcription initiation and elongation

1.2.1 Initiation.

Transcription initiation begins with the assembly at gene promoters of a pre-initiation complex (PIC), which consists of Pol II and the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIH. TFIID is recruited to promoters through interaction between the TBP (TATA box binding protein) subunit of TFIID and the TATA box motif in gene promoters. In addition, TFIID-associated proteins interact with elements downstream of the TSS and contribute to TFIID recruitment in many promoters. TFIIA and TFIIB also associate with TFIID to facilitate promoter binding and Pol II recruitment. Pol II together with TFIIF binds to the promoter followed by TFIIE and TFIIH. Subsequent to the PIC formation, the DNA strands are separated by the TFIIH helicase subunit to form the transcription "bubble", which enables RNA synthesis to begin from the transcription start site (reviewed in Kwak and Lis, 2013; Sainsbury et al., 2015; Sikorski and Buratowski, 2009).

The Pol II that enters the PIC harbours a hypo-phosphorylated CTD. The Mediator complex, a general transcription factor, binds to this CTD. Mediator can act both as a positive and negative regulator of gene expression (Allen and Taatjes, 2015). It first supports the recruitment of Pol II to the promoter. Then, following PIC assembly, Mediator stimulates CTD phosphorylation by the TFIIH subunit Cdk7, at least *in vitro* (Kim et al., 1994). The current initiation model suggests that phosphorylation by Cdk7 of Pol II on Ser5 and Ser7 disrupts the CTD-Mediator interaction (Søgaard et al., 2007), which triggers promoter escape and formation of the elongation complex (EC). Recently, Shah et al. (2018) showed that Tyr-to-Phe mutation within the 39 more distal

repeats of the mammalian CTD impairs Mediator association with Pol II, suggesting that Tyr1 may also regulate Mediator recruitment.

1.2.2 Elongation

In higher eukaryotes, after transcribing the initial 20–60 nt of RNA, Pol II becomes paused. Particularly important for this promoter-proximal pausing is the association to Pol II of two negative transcription elongation factors: NELF (Negative Elongating Factor) and the Spt4/Spt5 complex, also known as DSIF (DRB Sensitivity Inducing Factor) (reviewed in Kwak and Lis, 2013; Guo and Price, 2013). Biochemical and structural data have suggested that NELF and DSIF can bind the EC only upon promoter clearance because their binding sites are occupied in the PIC by initiation factors (Cheng and Price, 2007; Grohmann et al., 2011; He et al., 2016; Vos et al., 2018b). Recently, it was shown that NELF prevents binding of the anti-pausing transcription elongation factor IIS (TFIIS) to Pol II (Vos et al., 2018b). In addition, the structure of Pol II-NELF-DSIF revealed a tilted DNA-RNA hybrid in the active site that prevents binding of the nucleoside triphosphate (Vos et al., 2018b), providing the molecular mechanism of Pol II promoter-proximal pausing. This pausing is thought to create a "window of opportunity" for the recruitment of RNA processing and elongation factors. Particularly important is the recruitment to Pol II of the capping enzyme, which is dependent on Ser5 phosphorylated CTD. Substitution of Ser5 to alanine is lethal in both yeast and mammals, but the viability is rescued in *S. pombe* by tethering the capping enzyme to the CTD (Schwer and Shuman, 2011). Ser5 phosphorylation is also involved in the recruitment of other factors including the histone modifying enzymes Set1 and Rpd3S.

This early elongation checkpoint is overcome by the positive elongation factor P-TEFb and by the association with Pol II of the elongation factor Paf1 complex (PAF)

and Spt6 (Vos et al., 2018a). The P-TEFb kinase subunit Cdk9 phosphorylates Spt5, converting it into a positive elongation factor, and NELF, causing its dissociation from the EC (reviewed in Guo and Price, 2013). NELF release seems to free space on Pol II for PAF to bind (Vos et al., 2018a). Cdk9 not only phosphorylates the negative elongation factors, but also Rpb1 (i) on the linker that tethers the CTD to Pol II core, which then recruits Spt6 (Vos et al., 2018a) and (ii) on Ser2 of the CTD, which helps recruit RNA processing factors to emerging transcripts in particular at the end of elongation, when the mRNA is cleaved and polyadenylated. In addition, Ser2P, together with Ser5P, is involved in the recruitment of the spliceosome and the histone methyltransferase Set2 (reviewed in Eick and Geyer, 2013; Corden, 2013; Jeronimo et al., 2016; Zaborowska et al., 2016).

Unlike polymerase in metazoans, yeast Pol II does not accumulate downstream of the TSS as measured by ChIP-seq. Therefore, for a long time, it has been thought that early transcriptional pausing is specific to higher eukaryotes and is absent in yeast, partly because NELF is not conserved (Guo and Price, 2013). However, recent studies have challenged this view. In *S. cerevisiae*, strand-specific map of Pol II occupancy relative to nascent transcripts have detected an accumulation of polymerase shortly after the TSS (Churchman and Weissman, 2011; Milligan et al., 2016) and an early elongation check-point has been described that does not seem to be controlled by P-TEFb but by the Kin28 kinase (Rodríguez-Molina et al., 2016) . In *S. pombe*, precision Run-On sequencing (PRO-seq) to map the positions of Pol II active sites revealed an accumulation of polymerase immediately downstream of the TSS (Booth et al., 2016). In addition, Cdk9 was found to regulate this promoter-proximal checkpoint, although through a different mechanism than in higher eukaryotes (Booth et al., 2018). Cdk9 inhibition in mammals leads to an increase in promoter-proximal Pol II and loss of signal on the gene body (Laitem et al., 2015), indicating that Cdk9 is needed for promoter-

proximal escape. On the other hand, inhibition of *S.pombe* Cdk9 has a little effect on promoter-proximal regions, but results in accumulation of Pol II just downstream of these regions (TSS to +1kb) as a consequence of reduced elongation rate of Pol II (Booth et al., 2018). Like in higher eukaryotes, *S. pombe* Cdk9 phosphorylates Spt5 (Pei and Shuman, 2003). If and to which extent Spt5 phosphorylation contributes to regulate this early check point in fission yeast is still not clear.

1.2.3 Spt5 complex: a multi-faceted regulator of transcription elongation.

Spt5 is an essential multi-domain protein, the only known transcription factor that is conserved in all three kingdoms of life (Werner and Grohmann, 2011). The archeal Spt5 and the bacterial homologue NusG consist of an N-terminal (NGN) domain and a C-terminal Kyrpides-Ouzounis-Woese (KOW) domain. Eukaryotic Spt5 has additional domains and consists of an acidic N-terminal region followed by a NGN domain, multiple KOW domains (five in yeast), and a C-terminal region (CTR). This C-terminal region of Spt5, similarly to Rpb1 CTD, is unstructured and repetitive. The number of repeats and their sequence vary between species (Figure 1.2).

The majority of Spt5 protein is found in complex with Spt4, a small zinc finger protein present in eukaryotes and archea, but not in bacteria. Spt4 binds to Spt5 NGN domain and by doing so stabilizes the Spt5-Pol II interaction. In addition, Spt4 enhances the thermal stability of archeal Spt5 and similarly stabilizes Spt5 protein levels in budding yeast (reviewed in Hartzog and Fu, 2013).

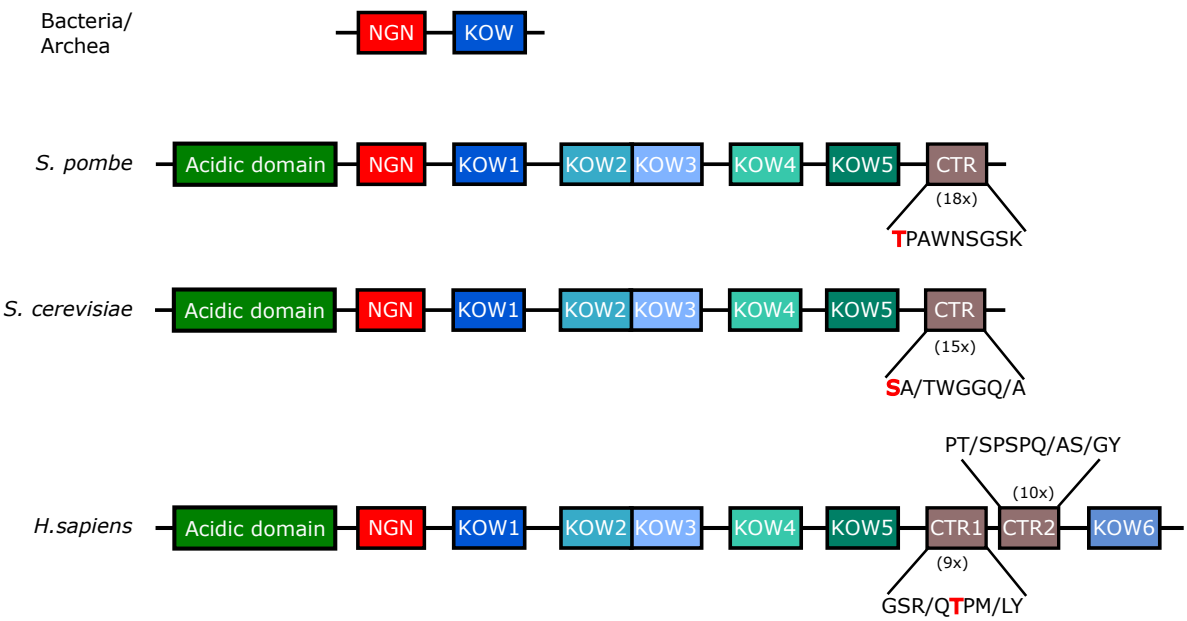


Figure 1.2: Schematic representation of Spt5 protein domains.
The domain organisation is based on published studies (Stachora et al., 1997; Zhou et al., 2009). For the CTR, the number of repeats and their consensus sequence are shown, with in red the residue that is phosphorylated by Cdk9/Bur1.

The Spt4-Spt5 complex promotes elongation by reducing the frequency of transcription pausing or arrest. Structural analysis of the elongation complex gave insights on how Spt4-Spt5 may fulfil its function (Ehara et al., 2017; Bernecky et al., 2017; Vos et al., 2018a). Pol II core binding involves Spt5 NGN domain and multiple KOW domains. In addition, the Spt5 NGN-KOW1 domains together with Spt4 contact the DNA coming out of the transcription bubble, establishing a DNA exit tunnel and thus enabling proper DNA rewinding. KOW1 also physically separates the DNA and RNA exit channels, preventing potential DNA-RNA interactions. KOW4 and KOW5 domains contact Pol II close to the the RNA exit channel, directing the exiting transcript. This result is in line with work from Qiu and Gilmour (2017) showing that KOW5 directly contacts the nascent RNA. Spt5-CTR structure was not resolved due to the flexibility of this domain, but, being downstream of KOW5, it is very likely that this repetitive region is also in close proximity to the RNA exit channel (Ehara et al., 2017; Bernecky et al., 2017). Spt4-Spt5 ability to restrict RNA/DNA orientation and interac-

tion gives to the elongation complex the structure required for processive transcription, while counteracting pausing and arrest (Ehara et al., 2017).

Spt5 can have both a positive and inhibitory function in elongation. In higher eukaryotes, in fact, the Spt5-Spt4 complex together with NELF halts Pol II soon after initiation. Pol II is released from this checkpoint upon phosphorylation of NELF and Spt5-CTR by pTEFb subunit Cdk9. *In vitro*, phosphorylation of human Spt5 is in fact critical to the positive activity of DSIF on elongation, but not to its elongation repression activity (Yamada et al., 2006). The phosphorylated Spt5 facilitates the association to Pol II of elongation factors (Chen et al., 2009). As noted above, NELF is absent in yeast. In addition, of the two opposing roles for metazoan DSIF, only the positive role seems to be conserved in yeast, where phosphorylation of Spt5-CTR stimulates the recruitment of the PAF complex in *S. cerevisiae* (Liu et al., 2009). Mammalian and fungal mRNA capping enzyme also binds to the Spt5-CTR, in addition to Pol II CTD. While the Pol II-capping enzyme interaction is Ser5P-dependent, phosphorylation of Spt5-CTR antagonizes its association with the capping enzyme in fission yeast (Doamekpor et al., 2014). Interestingly, in contrast to null mutations, deletion of the Spt5 CTR is not a lethal mutation in budding and fission yeast (Liu et al., 2009; Schneider et al., 2010; Zhou et al., 2009) and does not affect genome-wide Pol II distribution as measured by ChIP-seq (Mayer et al., 2012; Shetty et al., 2017).

1.3 Transcription termination

Termination of transcription by RNA polymerase II requires two distinct processes: (i) co-transcriptional pre-mRNA cleavage and polyadenylation by the 3' end processing machinery, and (ii) dissociation of Pol II from the DNA template.

1.3.1 3' end processing machinery

3' end processing of pre-mRNA consists of cleavage followed by polyadenylation of the upstream cleaved product and is dependant on two sub-complexes in budding yeast, the CPF (cleavage and polyadenylation factor) and the CF (cleavage factor, consisting of CFIA and CFIB) (reviewed in Mischo and Proudfoot, 2013; Porrua and Libri, 2015; Loya and Reines, 2016). Although CFIA and CFIB contribute to the specificity and efficiency of 3' end processing reactions, cleavage and polyadenylation are catalysed by subunits of the CPF complex (Table 1.3). Structural analysis of the latter has revealed its subdivision into three functionally distinct catalytic modules: the polyadenylation, nuclease and phosphatase modules (Casañal et al., 2017). The polyadenylation module is involved in the recognition of the PAS and in the polyadenylation of the mRNA 3' end following cleavage by the nuclease module. The phosphatase module is known to contain two phosphatases, Ssu72 and Glc7, that target the CTD (see section 1.3.4).

Homologues of the majority of the 3' end processing subunits from *S. cerevisiae* are also involved in 3' end formation in mammals (Table 1.3). However, the organisation into sub-complexes is different from yeast and consists of the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulation factor (CstF) and the cleavage factors CFIm and CFIIm. Overall, all proteins that are conserved in both mammals and budding yeast have been identified in fission yeast by purification of homologues of 3' end processing factors (Table 1.3). However, for most of them there is no evidence for their involvement in 3' end formation.

Table 1.3: Components of the 3' end processing machinery in humans and yeast.
Components of the polyadenylation, nuclease and phosphatase modules are highlighted in red, blue and green. Proteins with catalytic activity are indicated with asterisks (*). The tables is based on Roguev et al., 2004; Shi et al., 2009; Chan et al., 2011; Vanoosthuyse et al., 2014

S. cerevisiae	*H. sapiens*	*S. pombe*
Pap1* (CPF)	PAP	Pla1
Pfs2 (CPF)	WDR33	Pfs2
Cft1 (CPF)	CPSF-160	Cft1
Yth1 (CPF)	CPSF-30	Yth1
Fip1 (CPF)	Fip1 (CPSF)	Iss1
Ysh1* (CPF)	CPSF-73	Ysh1
Cft2 (CPF)	CPSF-100	Cft2
Mpe1 (CPF)	Rbbp6	Mpe1
Pta1 (CPF)	Symplekin	Pta1
Pti1 (CPF)	-	-
Syc1 (CPF)	-	-
Glc7* (CPF)	PP1CA/CB	Dis2
Ref2 (CPF)	-	-
Swd2 (CPF)	WDR82	Swd2.2
Ssu72* (CPF)	Ssu72	Ssu72
Hrp1 (CFIB)	-	Msi2
Rna14 (CFIA)	CstF-77	Rna14
Rna15 (CFIA)	CstF-64	Rna15
-	CstF-50	-
Clp1(CFIA)	Clp1 ($CFII_m$)	SPAC22H10.05c
Pcf11 (CFIA)	Pcf11 ($CFII_m$)	Pcf11
-	CFI-68kDa	-
-	CFI-25kDa	-

The CPF-CF complex is recruited to the nascent RNA once the elongating Pol II transcribes the polyadenylation signal (PAS). While the human PAS consists of a highly conserved AAUAAA hexamer, in addition to other less conserved sequence recognition elements, the RNA sequences directing 3' end processing are more degenerate in *S.cerevisiae* (reviewed in Mischo and Proudfoot, 2013; Porrua and Libri, 2015; Loya and Reines, 2016). They include (i) an AU-rich efficiency element EE (UAUAUA being the

optimal sequence) ∼ 25-40 nt upstream of the cleavage site, (ii) an A-rich positioning element PE (with AAUAAA being the most effective sequence) ∼ 10-30 nt upstream of the cleavage position, (iii) an U-rich element at the cleavage site. In *S. pombe*, the polyadenylation signal consists of the AAUAAA motif and a UGUA motif, both upstream of the cleavage site (Mata, 2013; Schlackow et al., 2013).

Many components of the CPF-CF complex are RNA-binding proteins that recognize the specific polyadenylation signals of the nascent transcript and bind to them as soon as they emerge from the polymerase. Several factors additionally bind the Rpb1 CTD. In particular, the CFIA subunit Pcf11 preferentially interacts with CTD repeats phosphorylated on Ser2 (Barilla et al., 2001; Licatalosi et al., 2002; Sadowski et al., 2003). Once the CPF-CF complex is recruited, the nascent transcript is cleaved by the CPF endonuclease (Ysh1 in *S.cerevisiae*, Brr5 in *S.pombe* or CPSF-73 in mammals) at the PAS and a pA tail of about 70 nt in budding yeast (200 nt in mammals) is added to the 5' cleavage product by Pap1/PAP polymerase. Polyadenylation protects the mRNA from 3' degradation and promotes its export to the cytoplasm (reviewed in Mischo and Proudfoot, 2013; Porrua and Libri, 2015; Loya and Reines, 2016).

1.3.2 Termination: allosteric vs torpedo model

After release of the cleaved mRNA, Pol II keeps transcribing the downstream DNA. The polymerase is then removed from the DNA template, but how this happens is still under debate and two models have been proposed, the torpedo and the allosteric model. The two models are not mutually exclusive, and a combined allosteric/torpedo mechanism has been proposed. The allosteric model postulates that dissociation of elongation factors and binding of the termination complex after transcription of the polyadenylation site cause a conformational change in the elongation complex with consequent

decrease in processivity and ultimately termination. This model is supported by the observation that specific elongation factors dissociate from Pol II before termination occurs which could destabilise the transcribing complex (Kim et al., 2004b) and that the termination factor Pcf11 alone can dislodge an elongation complex *in vitro* by simultaneous binding to the Pol II CTD and to the nascent RNA (Zhang et al., 2005). The torpedo model posits that cleavage of the pre-mRNA by Ysh1/CPSF-73 creates an entry point for a 5'-3' exonuclease (Rat1 in *S. cerevisiae*, Xrn2 in mammals) that degrades the downstream RNA until it catches up with the transcribing Pol II, triggering the release of the elongation complex from the DNA. Supporting this model, inactivation/depletion of Rat1 or Xrn2 causes read-through transcription and increased Pol II occupancy downstream of the polyadenylation site (Kim et al., 2004b; West et al., 2004; Eaton et al., 2018; Baejen et al., 2017). Termination defects were also observed upon depletion of CPSF-73 in human cells (Nojima et al., 2015). However both Osheim et al. (2002) and Zhang et al. (2015) showed that a polyadenylation signal is sufficient for termination to happen, independently from cleavage. Furthermore, it seems that addition of Rat1/Rai is insufficient to displace Pol II from its template *in vitro* (Luo et al., 2006; Dengl and Cramer, 2009). Recently, Fong et al. (2015) showed that termination is slowed down rather than prevented in mammalian cells upon depletion of Xrn2 by shRNA and simultaneously expression of a catalytically dead version of the protein. In addition, introducing Pol II mutants that transcribe slower or faster than the wild-type enzyme leads, respectively, to shorter and longer terminated transcripts (Fong et al., 2015), supporting the idea of kinetic competition between the transcribing elongation complex and Rat1. Thus, although cleavage and the subsequent degradation of the nascent transcript by Rat1/Xrn2 are not absolutely required for termination *in vitro*, they affect the efficiency and speed of termination *in vivo*. This gives rise to the idea that both an allosteric change resulting from crossing the poly(A) site and the torpedo effects of Rat1 are necessary for efficient dismantling of the elongation complex from

DNA (Luo et al., 2006; Zhang et al., 2015). Supporting this idea, (Schaughency et al., 2014) showed that upon nuclear depletion of Ysh1 in *S.cerevisiae*, Pol II accumulates, but does not extend further downstream of the poly(A) site, suggesting that Pol II is deficient in the release step, presumably because the lack of cleavage prevents a nuclease torpedo from gaining access to the elongation complex.

1.3.3 The role of CID-containing proteins in transcription termination

Several RNA-processing factors recognise the CTD by means of a conserved CTD-interacting domain (CID). There are three CID-containing factors in *S. cerevisiae* that participate in Pol II transcription termination: Pcf11, Rtt103 and Nrd1. While the latter binds Ser5 phosphorylated CTD (Vasiljeva et al., 2008), Pcf11 and Rtt103 associate with Ser2P-CTD (Meinhart and Cramer, 2004; Lunde et al., 2010). This different specificity for Pol II isoforms may dictate which termination pathway is used. Sections 1.3.1-1.3.2 described the mechanism to terminate transcription of protein-coding genes. Pol II however also transcribes short non-coding RNAs, whose termination and processing in *S. cerevisiae* relies on a different mechanism that utilises the Nrd1-Nab3-Sen1 (NNS) complex (Steinmetz et al., 2001; Arigo et al., 2006; Thiebaut et al., 2006; Kim et al., 2006; Vasiljeva et al., 2008). Thus, for the Nrd1-dependent pathway, high Ser5P close to the TSS recruits the NNS complex to act on short RNA. On the contrary, for termination of mRNA-coding genes, high Ser2P near the poly(A) site provides a mechanism to recruit Rtt103 and Pcf11 at the 3' end of protein-coding genes. Recently, Thr4P was also implicated in aiding recruitment of Rtt103 (Nemec et al., 2017; Jasnovidova et al., 2017; Harlen et al., 2016). While Pcf11 is an essential factor for both 3' end processing and transcription termination (Sadowski et al., 2003; Ghazy et al., 2012), the function of Rtt103 is unclear. It is not essential for viability and its main function seems to be the recruitment of Rat1 to Ser2P Pol II after transcript cleavage

(Schönemann et al., 2014). However, deletion of *rtt103* does not impair termination (Kim et al., 2004b; Luo et al., 2006). In fact Pcf11 is also able to aid recruitment of Rat1, providing a possible explanation of why Rtt103 is dispensable for termination (Luo et al., 2006).

In *S. pombe* and humans, one or more homologues can be found for each CID-containing protein (Table 1.4). It's not clear, however, if they have any function in transcription termination, with the following exceptions. Pcf11 plays an essential role in termination in higher eukaryotes (Zhang and Gilmour, 2006; West et al., 2008) and this seems to be conserved in *S. pombe* (Wittmann et al., 2017). In fission yeast, the Nrd1 homologue Seb1 is also a termination factor, but, in contrast to Nrd1, binds Ser2P-CTD, is recruited downstream of the poly(A) site and is involved in 3' end processing/termination of both protein-coding and non-coding genes (Wittmann et al., 2017).

Table 1.4: CID-containing proteins in yeast and humans
CTD binding preferences, if known, are indicated in brackets and based on published work (Vasiljeva et al., 2008; Lunde et al., 2010; Noble et al., 2005; Becker et al., 2008; Ni et al., 2014; Pineda et al., 2015)

S. cerevisiae	*S. pombe*	*H. sapiens*
Nrd1(Ser5P)	Seb1 (Ser2P)	Scaf4 (Ser2P), Scaf8 (Ser2P)
Pcf11 (Ser2P)	Pcf11	Pcf11 (Ser2P)
Rtt103 (Ser2P, Thr4P)	Rhn1	Rprd1A (Ser2P), Rprd1B (Ser2P), Rprd2 (Ser2P)

Several CID-containing factors bind RNA, in addition to CTD. Nrd1 and its homologues have an RNA recognition motif (RRM) that mediates the interaction with the nascent RNA. Concomitant binding of Nrd1 to both CTD and RNA transcripts promotes efficient Pol II termination (Heo et al., 2013; Tudek et al., 2014). Similarly

both RNA- and CTD-binding are essential for Seb1 function in transcription termination (Wittmann et al., 2017). Interestingly, Pcf11 can also bind RNA through its CID domain, however the RNA and CTD interactions are mutually exclusive (Hollingworth et al., 2006).

1.3.4 The role of CTD phosphatases in transcription termination

There are three phosphatases that are recruited at the 3' end of protein coding genes and target the CTD: Fcp1, Ssu72 and PP1/Glc7/Dis2 (Table 1.3). Compromising the enzymatic activity of Fcp1 and Ssu72 does not affect termination genome-wide (Bataille et al., 2012), but their ability to dephosphorylate the CTD at the end of the transcription cycle is thought to be important to recycle Pol II after transcription termination and to provide unphosphorylated enzyme available for assembly of new PICs. In fact, depletion of Fcp1 leads to an increase in the amount of non-chromatin bound phosphorylated Pol II (Fuda et al., 2012).

Unlike Ssu72 and Fcp1, nuclear depletion of the PP1 phosphatase Glc7 impairs transcription termination genome-wide in budding yeast (Schreieck et al., 2014). Because PP1 phosphatase is a major focus of this thesis, it is discussed in more detail in the following section.

1.4 The PP1 phosphatase

PP1 is a major protein Ser/Thr phosphatase that targets many different substrates, thus regulating a wide range of cellular processes (reviewed in Cohen, 2002). Specificity is achieved through association with regulatory subunits, also known as PP1 interacting proteins (PIPs). The PP1 phosphatase is highly conserved among all eukaryotes (Figure

1.3), the variability being localized almost exclusively at the C-terminus of the proteins. Interestingly, with the exception of *S. cerevisiae*, which has one PP1 gene (*glc7*), all eukaryotes encode for multiple PP1 genes (Moorhead et al., 2007).

PP1 is a component of the 3' end processing machinery in humans and yeast (Table 1.3). In *S. pombe*, there are two PP1 homologues, Dis2 and Sds21. However, only Dis2 is part of the CPF where it associates with Swd2.2 (Table 1.3) through interaction with Ppn1, one of the PIPs in fission yeast (Vanoosthuyse et al., 2014). The type 1 protein phosphatases are encoded by three genes in humans (*PP1CA*, *CB* and *CC*). Of these, only PP1CA/CB co-purify with the 3' end processing machinery, together with the mammalian Ppn1 homologue, PNUTS (Shi et al., 2009). The *S. cerevisiae* PP1 phosphatase, Glc7, interacts with the Ref2 and Swd2 subunits of the CPF (Table 1.3) (Nedea et al., 2008). There is no Ppn1/PNUTS homologue in budding yeast and whether and which PIP is needed for Glc7 recruitment to the 3' end processing machinery is unknown.

As discussed above, *S. cerevisiae* Glc7 is a CTD phosphatase that targets Tyr1P both *in vivo* and *in vitro* (Schreieck et al., 2014). Tyr1P inhibits binding of CID-containing proteins and was proposed to exclude the termination factors Pcf11 and Rtt103 from the gene body of protein-coding genes in budding yeast (Mayer et al., 2012). Through the action of Glc7, Tyr1P levels drop at the poly(A) site in *S. cerevisiae*, allowing the correct recruitment of Pcf11 and Rtt103 at the 3' end of protein-coding genes (Mayer et al., 2012; Schreieck et al., 2014). Whether this ability of Glc7 to target the CTD is conserved between species is unclear. The only other PP1 that has been tested for dephosphorylating Pol II is PP1CC in humans. PP1CC does not co-purify with the CPF (Shi et al., 2009), but is chromatin associated and dephosphorylates Ser2P and Ser5P *in vitro* (Washington et al., 2002). Therefore, despite the extremely high

amino acid similarity, the PP1 phosphatases seem to target different CTD phospho-marks in different species.

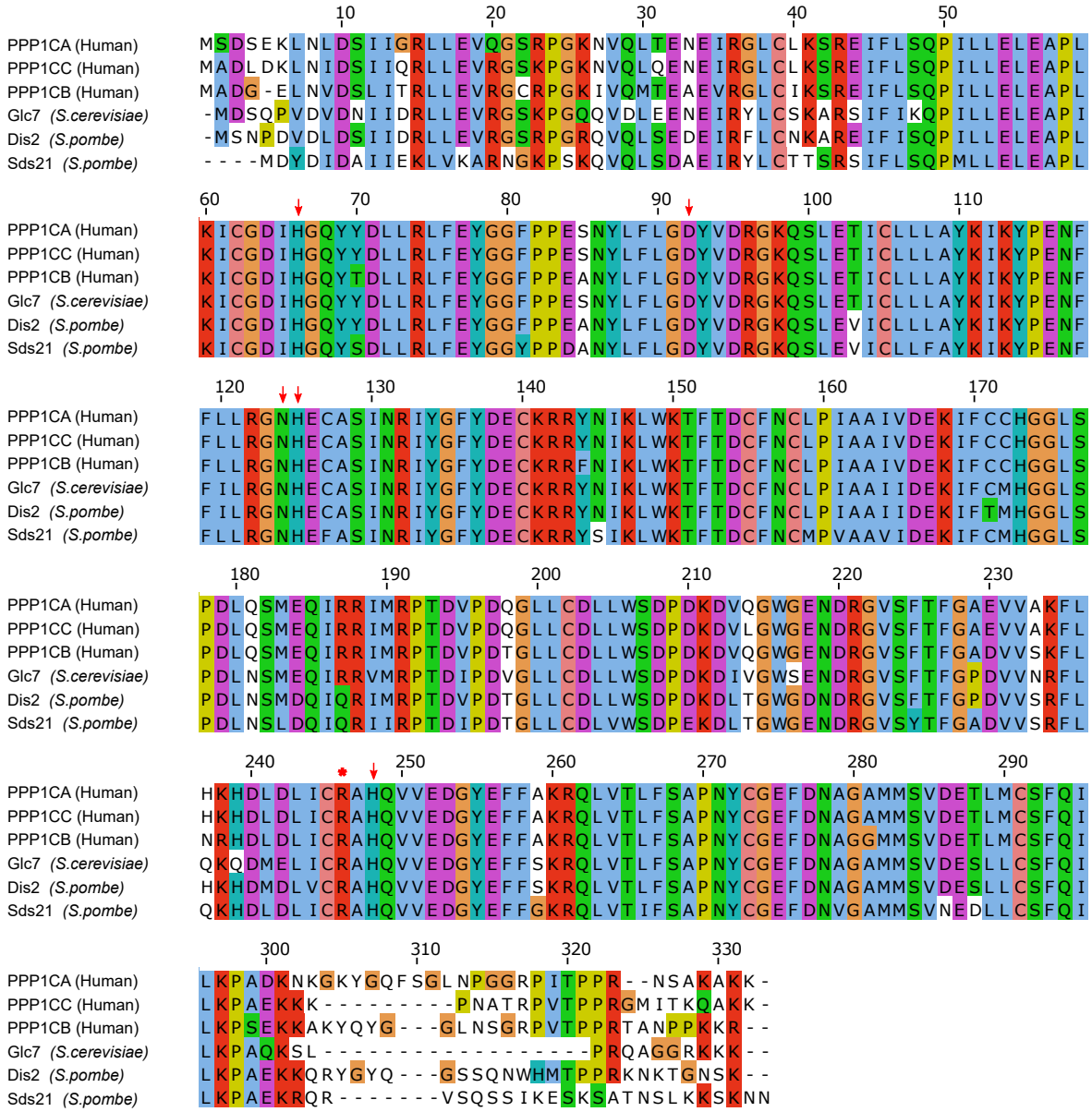


Figure 1.3: Alignment of human and yeast PP1 isoforms.
Red arrows indicate the amino acids of the catalytic pocket. Red star marks an arginine whose mutation results in catalytically inactive Dis2 used in this study (Kinoshita et al., 1990).

1.5 Objectives of this work

The role of Ser2 phosphorylation in recruiting 3' processing factors to the transcription machinery is well established. Interestingly, recent genome-wide studies have revealed the co-occurrence of multiple phospho-marks at gene 3' ends. In addition to Ser2P, there is an enrichment of Thr4P in *S. cerevisiae*, of Thr4P and Tyr1P in mammals. Despite progress in many model systems, the CTD code in *S. pombe* is poorly understood. ChIP-chip experiment by Coudreuse et al. (2010) showed a peak of Ser2P at the 3' ends of protein-coding genes. The following chapter proves that there is also an enrichment of Tyr1P and Thr4P at the end of the transcription cycle in *S. pombe*, resembling the patterns observed in mammalian cells. Chapter 1 also shows that phosphorylated Thr4 may contribute to the CTD code by recruiting termination factors and that a constitutively phosphorylated CTD on Thr4 is lethal for fission yeast. This result suggests that placing and removing the Thr4 phospho-mark at the correct time during transcription is critical.

Prior to the work described herein, the identity of the enzyme that can dephosphorylate Thr4P remained unclear. Chapter 2 demonstrates that the PP1 phosphatase Dis2 in fission yeast targets Thr4 phosphorylated CTD both *in vitro* and *in vivo*. Furthermore, it is shown that Dis2 is important for transcription termination. Deletion of *dis2*, in fact, causes delayed Pol II pausing at the gene 3' ends and impaired recruitment of termination factors. The major effect was seen for Seb1, whose association with chromatin was partially lost in absence of Dis2.

While the work described in this thesis was in progress, Parua et al. (2018) discovered that Dis2 dephosphorylates Spt5 downstream of the polyadenylation site. Chapter 3 shows that Seb1 and Spt5 compete for interaction with Pol II *in vitro* and that there is an antagonistic relationship between Spt5 and Seb1 that depends on Spt5

phosphorylation and is regulated by Dis2.

Chapter 2

Thr4 phosphorylated CTD in fission yeast

2.1 Introduction

As discussed above, CTD phosphorylated at serines is recognized by specific transcription and RNA processing factors that act on Pol II and has been extensively investigated. The CTD is also phosphorylated on Thr4, however the role of this phospho-mark is not fully understood. Thr4P is required for efficient transcription elongation and proper M-phase progression in human cells (Hintermair et al., 2012, 2016), while it's crucial for processing of the replication-activated histone genes in chicken DT40 cells (Hsin et al., 2011). In budding yeast, Thr4P facilitates the recruitment of the termination factor Rtt103 and regulates Pol II pausing after the poly(A) site (Harlen et al., 2016).

Despite progress in many model systems, the role of Thr4-phosphorylated CTD in *S. pombe* has not been addressed. This chapter shows that this phospho-mark is enriched at the 3' end of genes in fission yeast and may contribute to the recruitment of termination factors at the end of the transcription cycle.

2.2 Tyr1 and Thr4 CTD phospho-marks are enriched at the 3' ends of protein-coding genes

To understand how different Pol II phospho-marks are distributed genome-wide during transcription in fission yeast, chromatin immunoprecipitation experiments followed by high-throughput sequencing (ChIP-seq) were performed. Metagene profiles, heatmaps and representative individual protein-coding genes are shown (Figure 2.1). Pol II occupancy was monitored using an antibody specific for the CTD (8WG16). The 8WG16 antibody is known to recognize the unphosphorylated form of the CTD. However, metagene analysis over protein-coding genes revealed 8WG16 enrichment toward the 3' end of genes (Figure 2.1E), following a pattern similar to ChIP-chip experiments where an antibody against the tagged Rbp3 subunit of Pol II was used (Coudreuse et al., 2010). Similarly, in budding yeast, Pol II profiles using 8WG16, antibodies against Rpb3 or against tagged Rpb1 follow the same pattern (Bataille et al., 2012). This result may seem surprising, but could be explained by the fact that only 30% of the CTD repeats are phosphorylated (Schüller et al., 2016; Suh et al., 2016).

Localization of Pol II phosphoisoforms was mapped using monoclonal antibodies against Tyr1P, Ser2P, Thr4P, Ser5P (Figure 2.1). Ser5P is enriched across the gene body, whereas Ser2P increases towards the 3' end of genes consistent with the reported distribution of this phospho-marks in fission yeast (Coudreuse et al., 2010). While the ChIP-chip experiment by Coudreuse et al. (2010) also detected a peak of Ser5P at the poly(A) site (PAS), Figure 2.1D shows a depletion of this phospho-mark at the 3' end of genes. This could be explained by the fact that the present study employed the monoclonal antibody 3E8 instead of the polyclonal H14 used by Coudreuse et al. (2010). As reported previously, ChIPs for the same mark sometimes give different profiles using different antibodies (Bataille et al., 2012).

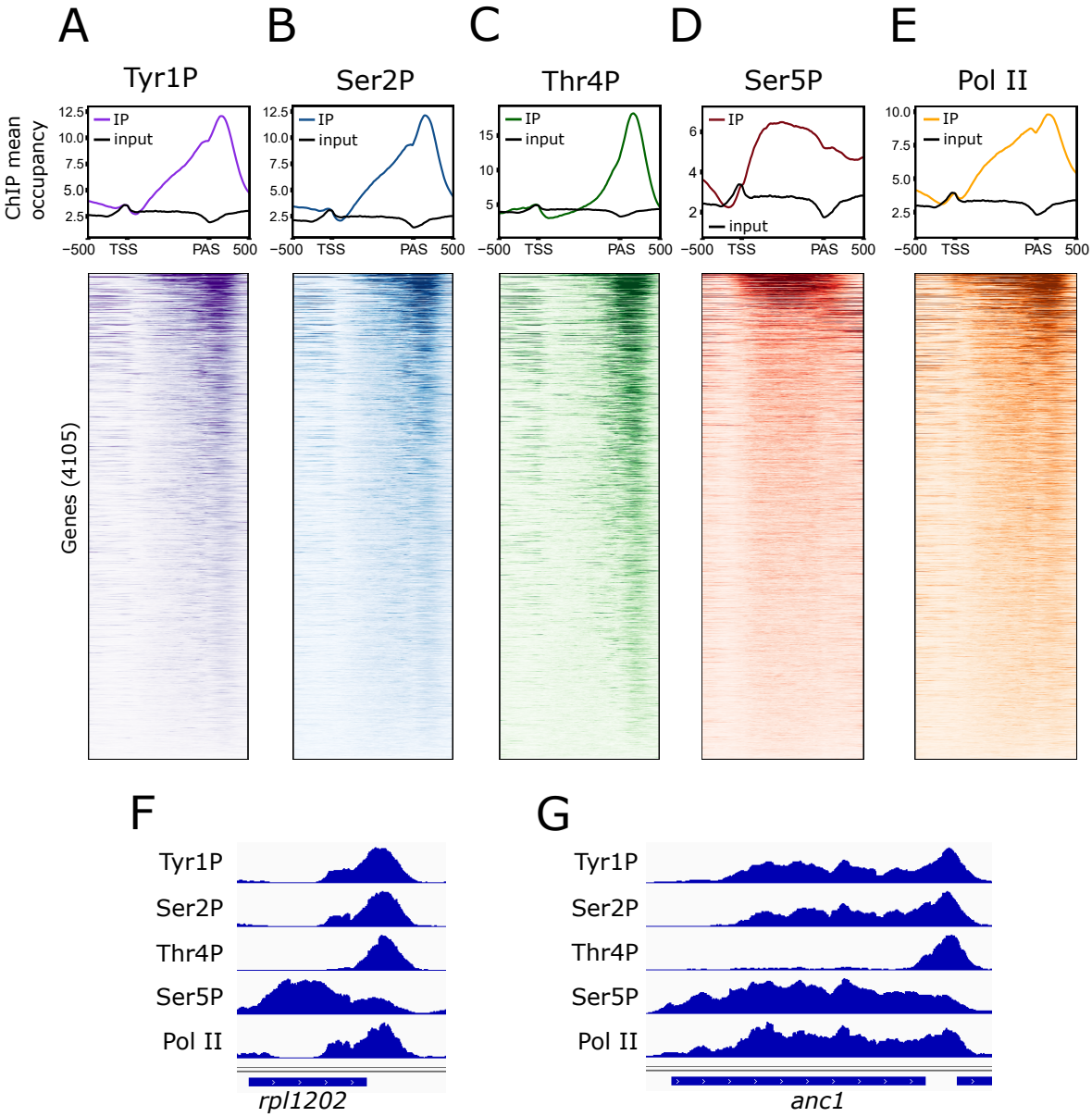


Figure 2.1: Differential distribution of phospho-marks on protein-coding genes.
(A-E) Averaged occupancy profiles of Pol II (8WG16), phospho-Pol II and input from ChIP-seq on all 4105 yeast protein coding-genes (upper panel). Color-coded heatmaps illustrating the phospho-Pol II profiles across single genes (lower panel). Profiles are aligned to the TSS and PAS as indicated. Ser5P ChIP-seq experiment was done by Dr Heo.
(F-G) ChIP-seq profiles across *rpl1202* (60S ribosomal protein L12.1) and *anc1* (adenine nucleotide carrier).

Tyr1P levels are also increased towards gene 3' ends, peaking after the PAS similar to Ser2P, suggesting that both marks are likely to contribute to the regulation of events

during 3' end formation. Interestingly, Thr4P gives an even more 3' end distal profile, peaking after the PAS and with almost no signal in the body of genes. This may be consistent with the possible global role for Thr4P late during transcription. Notably, the Tyr1P and Thr4P profiles described here for fission yeast resemble the distribution of these two phospho-marks in mammals (Hintermair et al., 2012; Nojima et al., 2015; Schlackow et al., 2017)

2.3 A constitutively phosphorylated CTD on Thr4 is lethal for *S. pombe*

To determine the importance of Thr4 phosphorylation during the transcription cycle, Thr4 residues were replaced either with an alanine (T4A), preventing phosphorylation at this position *in vivo*, or with a glutamic acid (T4E), mimicking a constitutively phosphorylated Thr4 (Figure 2.2). In the *S. pombe* CTD, the first four repeats deviate from the consensus heptad (YSPTSPS) (Figure 1.1). T4A and T4E mutations were introduced within the distal 25 heptads that perfectly match to the consensus sequence (with the exception of an alanine instead of Pro3 in the ninth heptad).

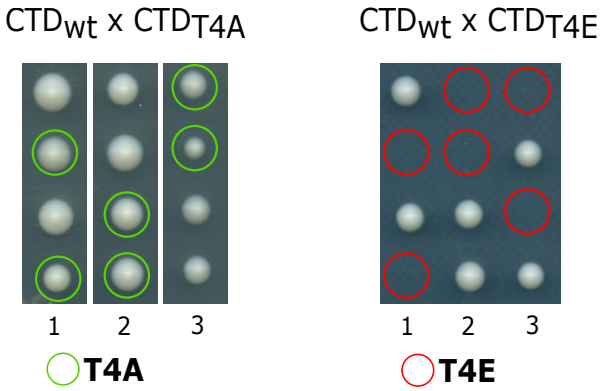


Figure 2.2: Effect of Thr4 mutations on fission yeast viability.
Results from a tetrad dissection of heterozygous T4A-CTD (left) and T4E-CTD (right) mutants. Three dissections are shown. This experiment was done by Dr Birot.

T4A mutant is viable in agreement with previous work (Schwer and Shuman, 2011),

where a truncated CTD with 14 canonical repeats carrying the T4A mutation was also shown to be viable. T4E mutant is lethal, suggesting that tight control of where and when the Thr4 phospho-mark is placed and kept is important for cell survival.

2.4 Termination factors bind Thr4P *in vitro*

While it is known that termination factors bind Ser2P, whether Thr4P contributes to their recruitment is not fully understood. Recently, it was shown that Rtt103 binds Thr4P and that this interaction may be important for recruitment to the CTD *in vivo* (Harlen et al., 2016; Nemec et al., 2017; Jasnovidova et al., 2017). To test if this is the case for other CID-containing termination factors, their CIDs were recombinantly expressed and tested for binding to FAM-labelled, differently modified two-repeat CTD peptides using fluorescence anisotropy (FA).

2.4.1 *S. pombe* termination factors bind Thr4P *in vitro*

There are three CID-containing termination factors in *S. pombe*: Seb1, Pcf11 and Rhn1. Recombinant *S. pombe* Seb1-CID and Pcf11-CID were successfully purified from bacteria (Figure 2.3). It was not possible to purify Rhn1-CID: different constructs were tested, but the protein was always insoluble. For Pcf11-CID, expression levels were low and the protein was insoluble. This was resolved by adding N-terminally a small ubiquitin-like modifier (SUMO) tag which is known to increase protein solubility (Marblestone et al., 2006).

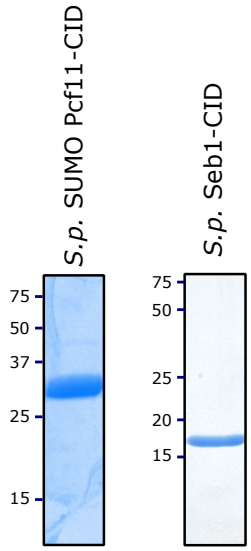
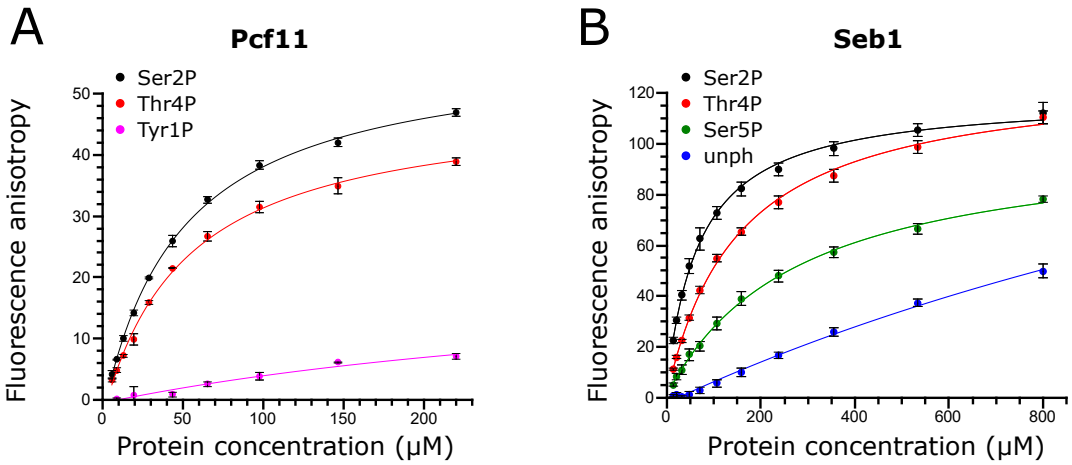


Figure 2.3: Purification of recombinant Seb1-CID and SUMO Pcf11-CID. SDS-PAGE showing the preparations for *S.pombe* SUMO Pcf11-CID and Seb1-CID. Pcf11 purification was done by Ms Duckett.

Purified Seb1 and Pcf11 CIDs were used for fluorescence anisotropy measurements using unmodified or differently phosphorylated CTD peptides (Figure 2.4A-B). Both Seb1 and Pcf11 can bind Thr4 phosphorylated CTD. However, while Pcf11 has similar affinity to Ser2P and Thr4P peptides, Seb1 shows preference to Ser2P and lower affinity for Thr4P peptides (Figure 2.4C).



C

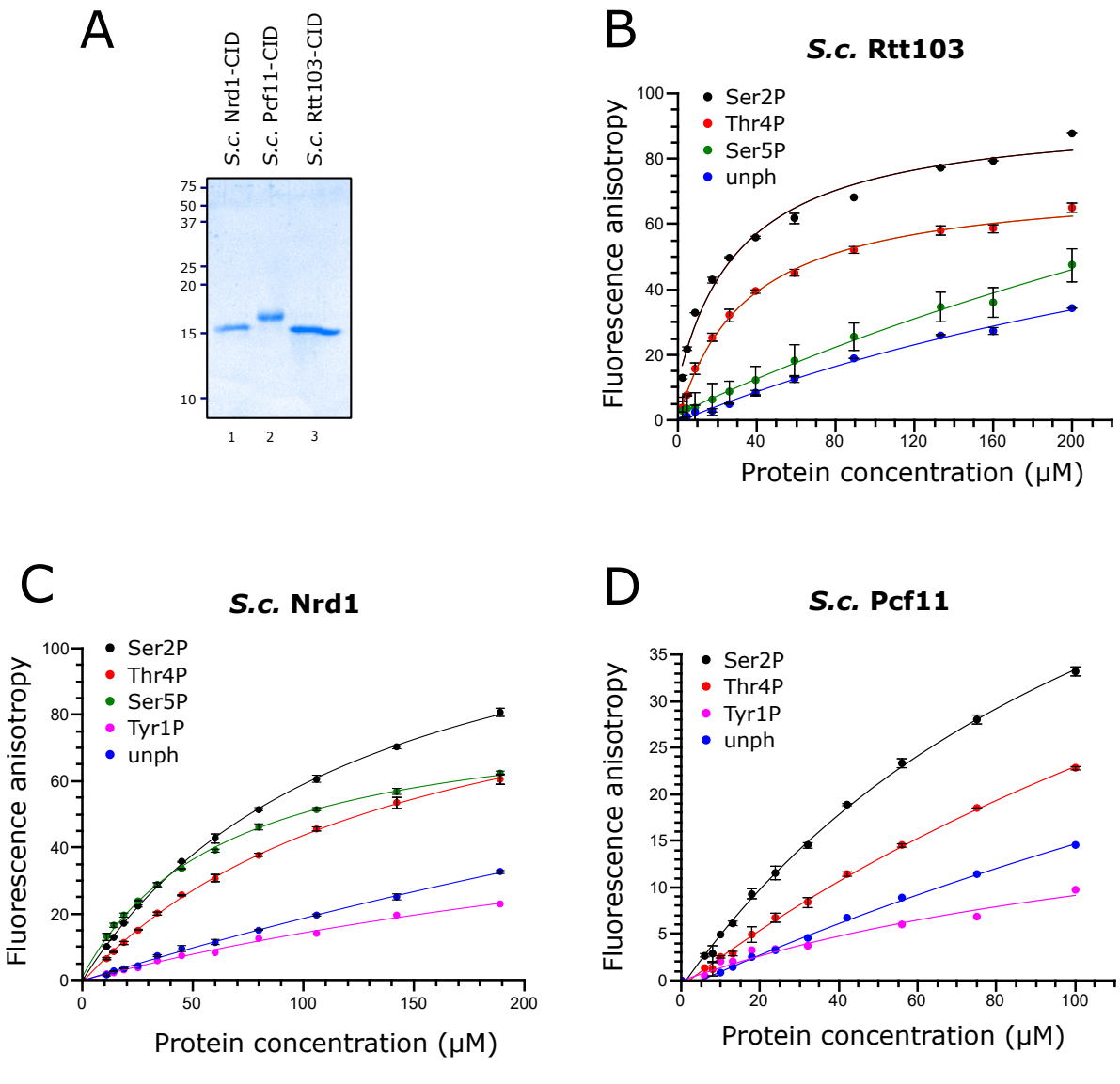
K_d (μM)

	Pcf11	Seb1
Ser2P	48.1 ± 2.7	72.3 ± 6.4
Thr4P	51.1 ± 5.2	153.2 ± 12.1
Ser5P	-	280.9 ± 23.1

Figure 2.4: Fluorescence anisotropy (FA) assay for *S. pombe* termination factors.
(A) Binding of SUMO-Pcf11-CID to the FAM-tagged two-repeat non-phosphorylated or phosphorylated CTD peptides measured by FA. Error bars show the standard deviation of two technical repeats. This experiment was done by Ms Duckett.
(B) FA assay for Seb1-CID. Error bars show the standard deviation of two technical repeats.
(C) K_d values of CID–CTD interactions determined from plots in (A-B).

2.4.2 *S. cerevisiae* termination factors bind Thr4P *in vitro*

The affinities to phosphorylated CTD peptides were also measured for *S. cerevisiae* termination factors: Pcf11, Rtt103 and Nrd1 (Figure 2.5).



E

K_d (μM)

	Rtt103	**Nrd1**	**Pcf11**
Ser2P	32.3 ± 6.8	126 ± 4	127 ± 15
Thr4P	34.8 ± 3.2	148 ± 7	269 ± 57
Ser5P	>1000	65.8 ± 3	-

Figure 2.5: Fluorescence anisotropy (FA) assay for *S. cerevisiae* termination factors. These experiments were done by Ms Duckett.
(A) SDS-PAGE showing the preparations for *S. cerevisiae* Rtt103-CID, Nrd1-CID and Pcf11-CID. Done by Ms Duckett.
(B-D) FA assay for *S. cerevisiae* CID containing factors from (A). Error bars show the standard deviation of two technical repeats. Done by Ms Duckett.
(E) K_d values of CID–CTD interactions determined from plots in (B-D).

In agreement with previous studies, Rtt103 (Rhn1 in *S. pombe*) has the strongest affinity to Ser2P among *S. cerevisiae* termination factors. In addition, Rtt103 interacts equally well with Ser2P and Thr4P. Consistent with previous results (Mayer et al., 2012; Lunde et al., 2010), Nrd1 preferentially binds Ser5P, while Pcf11 has the highest affinity to Ser2P. All proteins show no binding to Tyr1P in support of the proposed role for this residue in negatively affecting recruitment of termination factors (Mayer et al., 2012).

2.5 Discussion

The ChIP-seq analysis of *S. pombe* Pol II phosphoisoforms presented in this chapter shows enrichment of Ser2P, Tyr1P and Thr4P at 3' ends of protein-coding genes, resembling the patterns observed in mammalian cells (Hintermair et al., 2012; Nojima et al., 2015; Schlackow et al., 2017). The presence of Tyr1P downstream of the poly(A) site is surprising since this phospho-mark is known to impair recruitment of termination factors (Mayer et al., 2012). In mammals, the fact that Tyr1P and Ser2P rarely coexist in the same repeat and that Tyr1P-containing heptad repeats represent approximately 5% of the total phosphorylated CTD (Schüller et al., 2016) may explain how termination factors can be still recruited at gene 3' ends when Tyr1P is present. If this could also be the case for fission yeast is hard to say since the stoichiometry of its phosphorylated CTD is unknown. Unlike mammals, fission yeast does not show widespread promoter-proximal Pol II pausing nor enrichment of Tyr1P at the 5' end of genes.

This study together with published work (Hintermair et al., 2012; Schlackow et al., 2017) shows that Thr4P is enriched at gene 3' ends in both fission yeast and mammals, while in *S. cerevisiae* Thr4P is present both within the gene body and after the PAS (Mayer et al., 2012; Harlen et al., 2016; Milligan et al., 2016). The presence of

a Thr4P peak at the gene 3' end that is conserved between species poses the question whether, in addition to Ser2P, Thr4-phosphorylated CTD also contributes to the recruitment of termination factors. Here it is shown that termination factors can indeed bind Thr4P *in vitro* but with different affinities. As expected, Nrd1 has the highest affinity for Ser5P in agreement with its function in the early stages of transcription in budding yeast (Vasiljeva and Buratowski, 2006). Of the remaining two termination factors in *S. cerevisiae*, only Rtt103 can bind equally well Ser2P and Thr4P. This is in agreement with recent work that showed similar Rtt103 affinity for Ser2 and Thr4-phosphorylated CTD (Jasnovidova et al., 2017; Nemec et al., 2017). In addition, mass spectrometry analysis of the Thr4P-CTD interactome revealed enrichment of Rtt103, but not of *S. cerevisiae* Pcf11 (Harlen et al., 2016) in line with our results showing differential binding to Thr4P among termination factors. Interestingly, *S. pombe* Pcf11 differs from the *S. cerevisiae* homologue as it binds both Ser2P and Thr4P with similar affinity. Seb1, on the other hand, has the highest affinity for Ser2P. The fission yeast Rtt103 homologue, Rhn1, could not be tested as the recombinant protein was insoluble.

What is the importance of Thr4P? The question can be addressed genetically by mutating Thr4 residues to Ala or Glu. Here, it is shown that T4A substitution within the canonical repeats of CTD does not significantly affect cell growth in fission yeast. This is consistent with work in budding yeast, where T4V and T4A mutants are also viable (Harlen et al., 2016; Nemec et al., 2017). Despite viability, NET-seq analysis of *S. cerevisiae* T4V mutant revealed an increased Pol II occupancy after the PAS. Unlike yeast, T4V and T4A substitutions are lethal for chicken and human cells (Hintermair et al., 2012; Hsin et al., 2011). ChIP-seq experiment in human T4A mutant showed accumulation of Pol II downstream of the TSS and a decrease at the 3' end of genes, indicating elongation defect in the mutant. Analysis of the transcriptome of T4V chicken cells revealed a specific defect in processing of non-polyadenylated histone mRNAs,

while abundance and processing of other mRNAs were not affected. In *S. cerevisiae*, histone genes are polyadenylated (Fahrner et al., 1980) and no defects in their 3' end processing were observed in T4A mutants (Nemec et al., 2017).

Removal of a phospho-mark at the right time is likely to be at least as important as the addition of that phospho-mark in the first place. Dephosphorylation of the CTD residues is essential to produce the classic pattern of CTD phosphorylation across genes, and previous work has shown that replacement of serines 2 and 5 by a phosphomimic is not tolerated in both mammals and yeast (reviewed in Eick and Geyer (2013); Corden (2013)). Here it is shown that substitution of threonine 4 to phosphomimetic glutamic acid (T4E) in all canonical repeats of *S. pombe* CTD leads to cell death. This is consistent with work in budding yeast, where T4E is also lethal (Harlen et al., 2016), suggesting that removing Thr4P is critical in both yeast species. Whether this is also the case for mammals is unknown. Despite the emerging importance of Thr4P, very little is known about the kinases and phosphatases that regulate this phospho-mark.

Chapter 3

Dis2 is a phosphatase that targets Thr4P-CTD in fission yeast and is important for transcription termination

3.1 Introduction

Both phosphorylation and dephosphorylation are important to determine specific CTD patterns during the different stages of transcription. Many of the CTD kinases and phosphatase have been identified, but not all. In fact, very little is known about the enzymes that target Thr4. Polo-like kinase 3 (PLK3) is a Thr4-specific kinase in human cells (Hintermair et al., 2012), while the identity of the phosphatase that target Thr4P remains unclear. As discussed above, Thr4P is enriched at the 3' end of protein-coding genes in *S. pombe*. There are three known phosphatases that are recruited at the end of the transcription cycle: Fcp1, Ssu72 and Dis2 (Glc7 in *S. cerevisiae*). Dis2 is one of the two protein phosphatase 1 (PP1) homologues in fission yeast, the other being Sds21 (Ohkura et al., 1989). Of the two, only Dis2 is part of the 3' end processing machinery (Vanoosthuyse et al., 2014). This chapter shows that Dis2 is a phosphatase that targets Pol II CTD on Thr4 and is important for transcription termination.

3.2 Dis2 mutants show an increase in Thr4P-CTD levels in the whole cell extract

To understand whether Dis2 or Sds21 affect CTD dephosphorylation, levels of phosphorylated CTD in whole-cell extract were analysed in the absence of Dis2 or Sds21 (Figure 3.1).

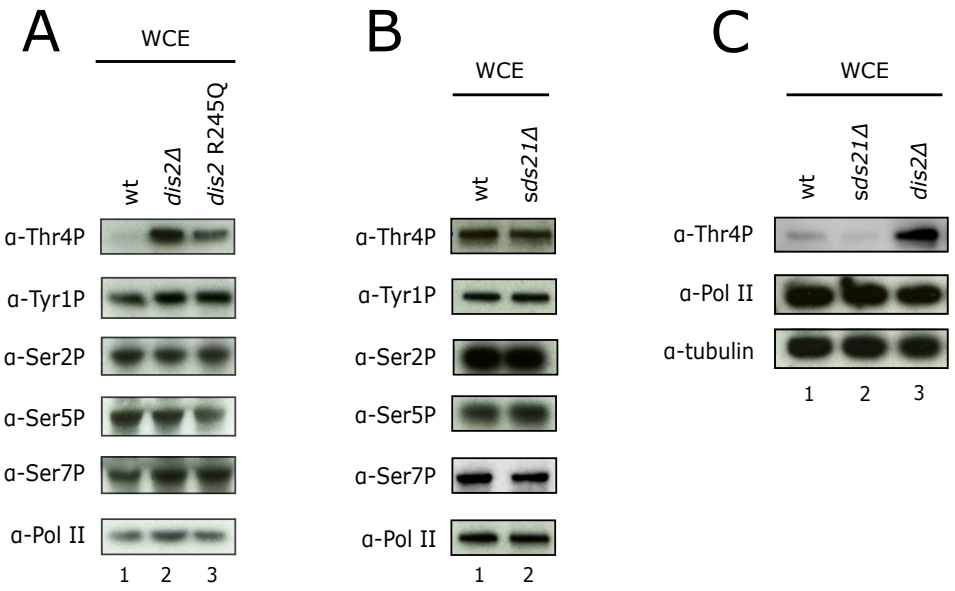


Figure 3.1: Deletion of *dis2* affects Pol II-CTD Thr4 phosphorylation *in vivo*.
(A) Whole cell extracts were prepared from cells with *dis2* deleted (*dis2*Δ) or expressing the phosphatase-defective *dis2* R245Q allele. Extracts were analysed by Western blot using antibodies specific for Thr4P, Tyr1P, Ser2P, Ser5P, Ser7P and Pol II (8WG16).
(B) Whole cell extract levels of phospho-Pol II were analysed by Western blot using the indicated antibodies in a wild-type (wt) and *sds21*-deleted strain (*sds21*Δ).
(C) Whole cell extract levels of phospho-Pol II were analysed by Western blot using the indicated antibodies in wt, *sds21*Δ and *dis2*Δ strains.

Deletion of *dis2* resulted in an increase in Thr4P-CTD, whereas other phospho-marks were not affected (Figure 3.1A, lines 1-2). The same phenotype was observed in a strain carrying a *dis2* R245Q allele (Figure 3.1A, lines 1-3), in line with previous studies, showing that this point mutant has reduced catalytic activity (Kinoshita et al., 1990). Deletion of *sds21* had no effect on phosphorylated CTD (Figure 3.1B), suggesting that Dis2 is the only PP1 isoform targeting Pol II in *S. pombe*. Levels of Thr4P in the

wild-type (wt) and *sds21*Δ strains were barely detectable when compared to *dis2*Δ (Figure 3.1C); however, Thr4P could be detected in these strains using longer exposure (Figure 3.1B).

3.3 Dis2 specifically dephosphorylates Thr4 *in vitro*

To investigate whether Dis2 is directly responsible for dephosphorylating Pol II on Thr4-CTD, Dis2 activity was tested *in vitro*. Since phosphatases rely on regulatory co-factors that provide specificity in targeting their substrate (Shi, 2009), both Dis2 and the catalytically inactive variant (Dis2 R245Q) in complex with associated proteins were purified from fission yeast (Figure 3.2A). In line with previous studies (Vanoosthuyse et al., 2014), Dis2 co-purified with the CPF subunits, including Ssu72 (Table 3.1). The Dis2 R245Q purification also contained some CPF subunits, although the association seemed to be negatively affected by the arginine-to-glutamine mutation.

Table 3.1: Dis2, Ppn1 and associated CPF subunits identified by MS.
The number of peptides and the MS intensity score for each protein are indicated. MS intensities were centred to median = zero for each individual sample.

Protein name	Peptide counts		Intensity score	
	Dis2	Dis2 R245Q	Dis2	Dis2 R245Q
Dis2	213	99	2.48E+09	4.91E+08
Swd2.2	38	0	1.08E+08	-
Cft1	64	3	5.56E+07	1.08E+06
Cft2	54	1	4.71E+07	6.40E+05
Ysh1	40	3	2.98E+07	3.47E+05
Pta1	47	0	2.45E+07	-
Fip1	11	2	1.14E+07	1.75E+05
Pfs2	7	0	6.97E+06	-
Yth1	10	0	6.70E+06	-
Ssu72	8	0	6.69E+06	-
Pla1	6	0	5.35E+05	-
Ppn1	48	0	7.31E+07	-

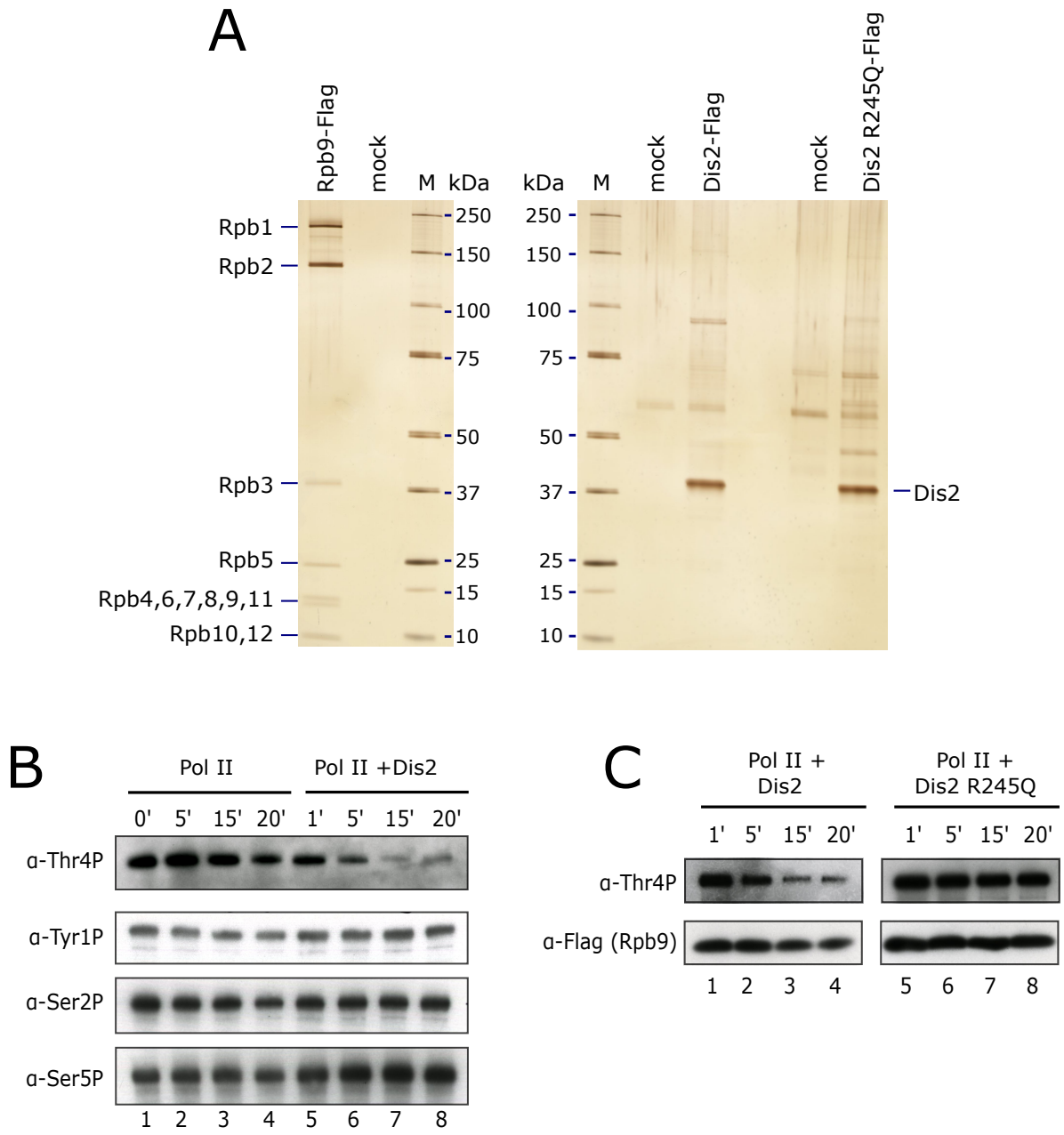


Figure 3.2: Dis2 is a Thr4P phosphatase *in vitro*.
(A) Silver-stained SDS-PAGE analysis of Pol II, Dis2 and catalytically inactive Dis2 R245Q purified from yeast.
(B-D) Purified Pol II was incubated with Dis2 or Dis2 R245Q and levels of Pol II were assessed by Western Blot with anti-CTD and Flag antibodies.

Next, Dis2 phosphatase activity was tested using native Pol II purified from yeast as a substrate (3.2A). In agreement with the *in vivo* data, addition of Dis2 resulted in dephosphorylation of Thr4P Pol II, but not Tyr1P or Ser2P (Figure 3.2B). Ser5P

levels did not decrease despite the presence of Ssu72, but actually increased. Ssu72 was much less abundant than Dis2 (Table 3.1) and might not have been enough to dephosphorylate the CTD in the time-scale used for the assay. Several kinases (Cka1, Ckb1, Sty1, Hhp1, Win1) were present at low levels in the Dis2 purification that could explain the increase in Ser5P levels. None of these kinases is known to target the CTD and their specificity would need to be tested *in vivo*. Addition of the inactive R245Q enzyme did not affect Thr4P levels (Figure 3.2C), further proving that Dis2 is a CTD phosphatase that specifically targets Thr4P on Pol II.

3.4 Deletion of *dis2* leads to impaired transcription termination and higher levels of transcriptionally engaged CTD-Thr4P

To further understand the function of Dis2 on transcription, ChIP-seq was carried out to assess the consequences of Dis2 loss on the genome-wide distribution of Pol II and Thr4P-CTD. ChIP-seq experiments were done by mixing defined numbers of *S. cerevisiae* cells (the spike-in genome) with *S. pombe* samples prior to cross-linking (spike-in ChIP-seq). This "spike" chromatin then served as an internal control to which the experimental signals were adjusted (Hu et al., 2015; Bonhoure et al., 2014; Orlando et al., 2014). Bioinformatic analysis focused on a subset of 1735 protein-coding genes that do not overlap with any other transcription unit in a region from 250 nt upstream to 500 nt downstream of the PAS. Metagene analysis around poly(A) sites (Figure 3.3A) as well as analysis of single genes (Figure 3.3D-E) revealed an accumulation of Pol II further downstream of the PAS in the Dis2 mutant (*dis2*Δ) compared to wt strain, indicating that lack of Dis2 may delay Pol II release from the DNA template. Figure 3.3B-E shows that, in addition to impaired transcription termination, deletion of *dis2* caused a pronounced increase of Thr4P-CTD (p-value<0.001,two-sided Wilcoxon

rank-sum test) in line with the role of Dis2 as a CTD phosphatase.

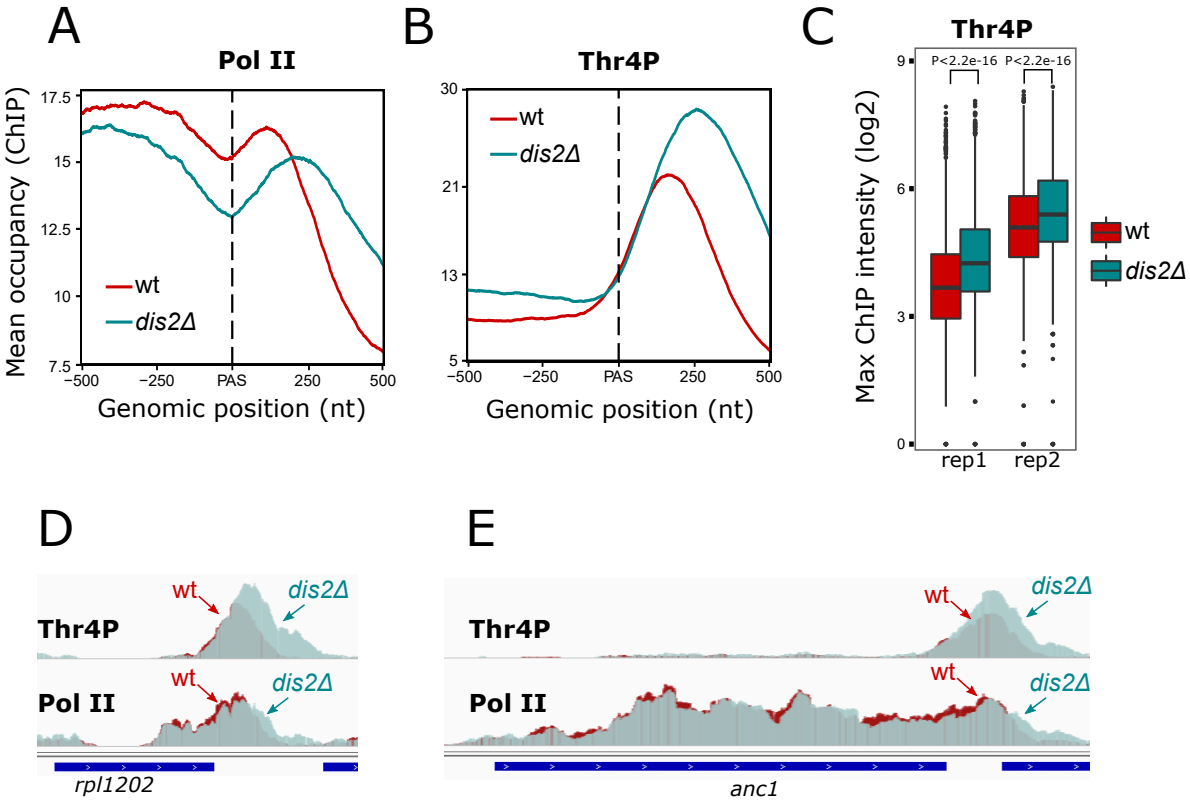


Figure 3.3: Increased Thr4P downstream of PAS upon loss of Dis2 coincides with transcription termination defect. These experiments were done together with Dr Heo.
(A-B) Spike-in normalized ChIP-seq meta-profiles of total Pol II (8WG16) and Thr4P around the PAS, averaged from two biological replicas. Protein-coding genes that do not overlap with any other transcription unit in a region from 250 nt upstream to 500 nt downstream of the PAS were selected for the analysis (n=1735).
(C) Quantitation (boxplots) of spike-in normalized maximum Thr4P ChIP signal intensity downstream of the PAS. Protein-coding genes for analysis were selected as above (n=1735). Duplicate data is presented. p-values (two-sided Wilcoxon rank-sum test) were calculated to test the significance of the difference between wt and *dis2*Δ.
(D-E) Thr4P and Pol II profiles on individual genes.

3.5 Recruitment of termination factors is impaired upon *dis2* deletion

Since termination factors are known CTD-interactors, whether their recruitment was impaired upon deletion of *dis2s* was tested next. Spike-in ChIP-seq was carried out to compare the occupancy of Pcf11, Rhn1 and Seb1 in wt and *dis2*Δ strains. Deletion of *dis2* caused a shift in recruitment of all three termination factors further downstream of the PAS compared to wt (Figure 3.4, 3.5), most likely as a consequence of delayed Pol II pausing in *dis2*Δ (Figure 3.3A). In addition, bioinformatic analysis revealed a modest, but significant increase in the recruitment of Pcf11 and Rhn1 upon *dis2* deletion (Figure 3.4A-F, two-sided Wilcoxon rank-sum test p-value$\leq$0.001). On the contrary, Seb1 recruitment to chromatin was decreased in *dis2*Δ (Figure 3.5A-D, two-sided Wilcoxon rank-sum test p-value$<$0.001). Similarly, Pol II was found to pull down much less Seb1 II when immunoprecipitated from *dis2*Δ extract (Figure 3.5E, lanes 3-4) while levels of Seb1 in the whole cell extract were not affected in *dis2*Δ (Figure 3.5E, lanes 1-2), suggesting that loss of Seb1 recruitment was not a result of decreased cellular levels of Seb1 .

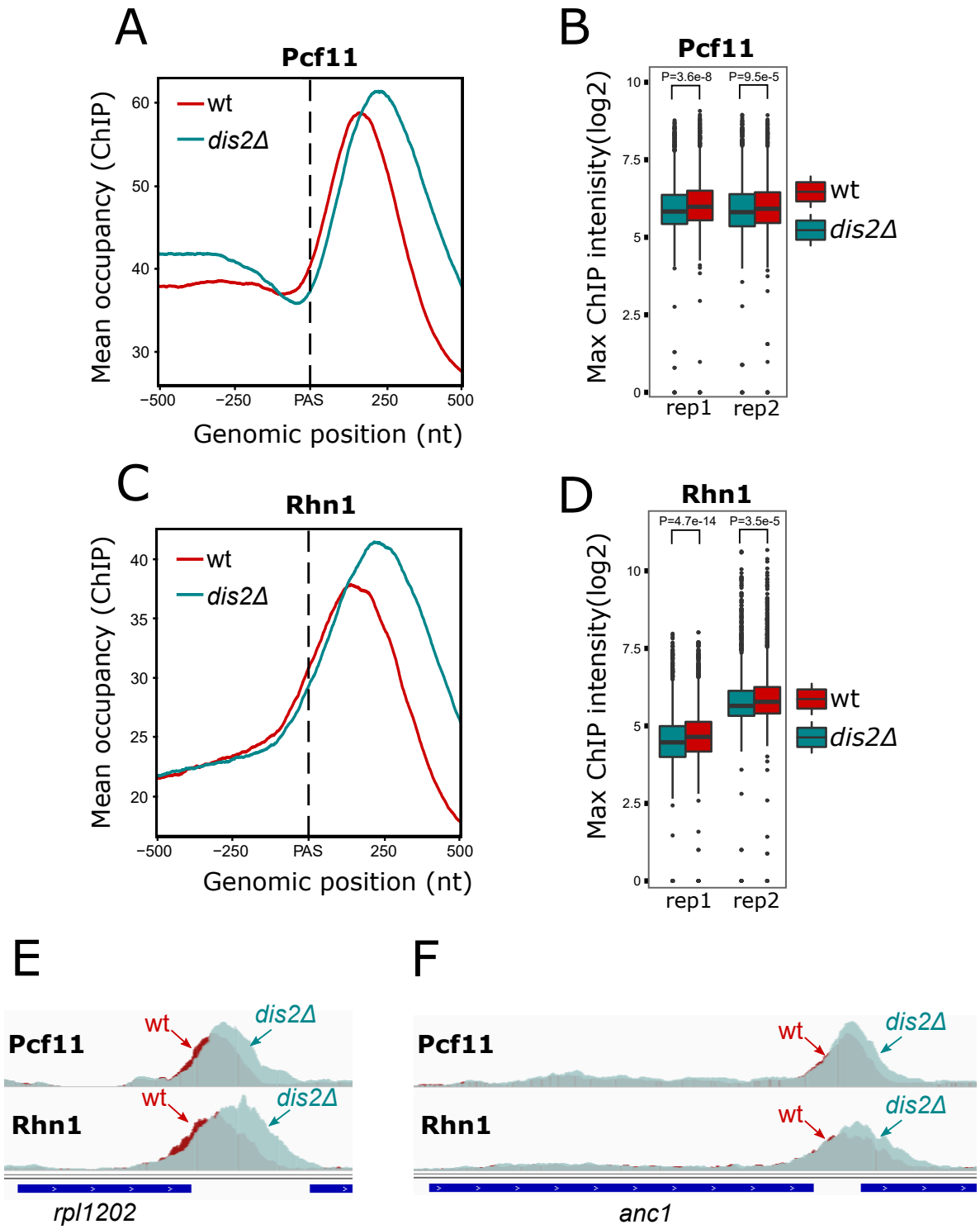


Figure 3.4: Lack of Dis2 activity leads to increased recruitment of the termination factors Pcf11 and Rhn1.

(A-B) Spike-in normalized ChIP-seq meta-profiles of Pcf11 and Rhn1 around the PAS, averaged from two biological replicas. Protein-coding genes for analysis were selected as above (n=1735).

(C-D) Quantitation (boxplots) of spike-in normalized maximum Pcf11 and Rhn1 ChIP signal intensity downstream of the PAS. Protein-coding genes for analysis were selected as above (n=1735). Duplicate data is presented.

(E-F) Pcf11 and Rhn1 profiles on individual genes.

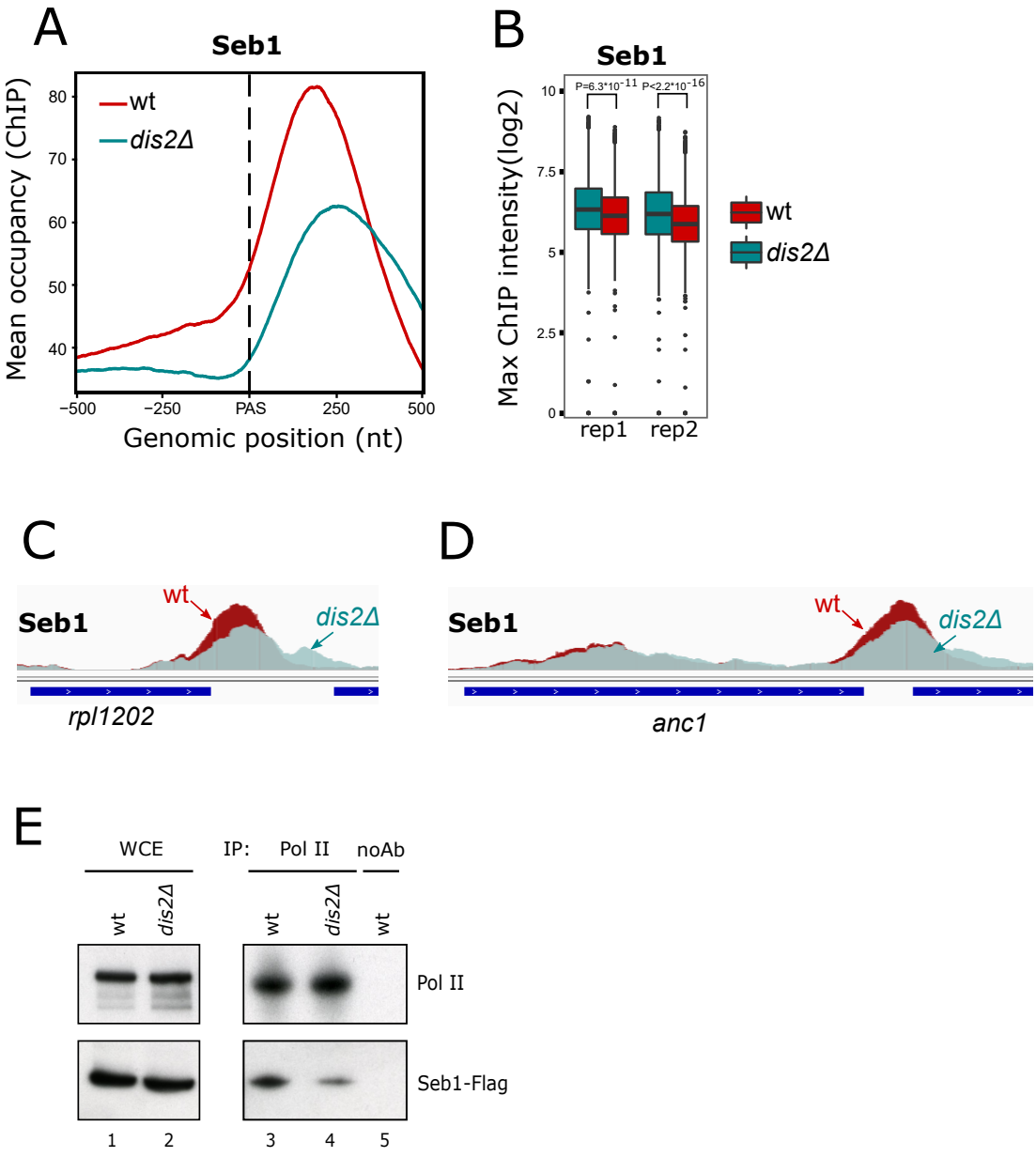


Figure 3.5: Lack of Dis2 activity leads to decreased recruitment of the termination factor Seb1.

(A) Spike-in normalized ChIP-seq meta-profiles of Seb1 around the PAS, averaged from two biological replicas. Protein-coding genes for analysis were selected as above (n=1735).

(B) Quantitation (boxplots) of spike-in normalized maximum Seb1 ChIP signal intensity downstream of the PAS.

(C-D) Seb1 profiles on individual genes.

(E) Association of Seb1-Flag to Pol II in wt and *dis2*Δ was determined by immunoprecipitation of Pol II (8WG16), followed by Western blotting.

3.6 Discussion

Whereas previous work showed that the Pol II CTD is phosphorylated at Thr4, it remained unknown which enzyme dephosphorylates Thr4. Here, it is shown that Dis2 is a Pol II CTD phosphatase that acts at the gene 3' ends with specificity for Thr4P and is important for transcription termination in fission yeast.

Dis2 belongs to the PP1 phosphatase family and is one of the two PP1 homologues in fission yeast, the other being Sds21. Interestingly, we show that while Dis2 specifically targets Thr4P-CTD, Sds21 does not dephosphorylate the CTD despite the high degree of sequence identity (78.75%) with Dis2. This result may seem surprising, but could be explained by the fact that phosphatases need regulatory co-factors to provide substrate specificity (Shi, 2009) and while Dis2 co-purifies with the CPF complex, Sds21 does not (Vanoosthuyse et al., 2014). Notably, *dis2*Δ and *sds21*Δ single mutants are viable, but the double mutant is lethal (Ohkura et al., 1989). PP1 phosphatases play a role in a wide range of cellular processes (reviewed in Cohen (2002)), thus it's likely that although Dis2 and Sds21 differ in their ability to target the CTD, their functions may overlap to regulate a different substrate.

Dis2 homologues, Glc7 in budding yeast and PP1 in metazoans (Figure 1.3), also target the CTD. The similarities do not end there. Here it is shown that Dis2 is important for transcription termination. The same has been found for Glc7 and PP1. However, despite the analogies, the CTD phospho-isoform they target and the mechanism by which these phosphatase regulate transcription termination are different. In humans, only PP1CC has been tested *in vitro* and showed specificity for Ser2P and Ser5P. In *Drosophila melanogaster*, loss of function of the PP1-cofactor PNUTS causes increased levels of Ser5P and, to less extent, of Ser2P, but not of Thr4P (Ciurciu et al., 2013). In mouse, silencing of PNUTS causes read-through transcription at the 3' ends. However,

whether PP1 or its catalytic activity are involved and how is not clear (Austenaa et al., 2015). Glc7 phosphatase activity on Tyr1P has been proved both *in vivo* and *in vitro* (Schreieck et al., 2014). If Glc7 can also target Thr4P has not been tested, but in either case its role in recruiting termination factors differs from Dis2. In fact, upon depletion of Glc7, high levels of Tyr1P at gene 3' ends inhibit CID binding to the CTD and cause termination defects by loss of Pcf11 and Rtt103 recruitment. This is just the opposite of what happens in *dis2*Δ where Pcf11 and Rhn1 occupancy are increased, mimicking the increase in Thr4P-CTD. Unlike budding yeast, *S. pombe* has a third CID-containing protein important for transcription termination of protein-coding genes, Seb1, whose association to Pol II is partially lost in *dis2*Δ. Thus, deletion of *dis2* affects differently the occupancy on chromatin of the different termination factors. Increased levels of Pcf11 and Rhn1 may affect the Seb1-CTD interaction, thus the partial loss of Seb1. However, it is also possible that other factors contribute to reduced recruitment of Seb1 in the absence of Dis2. Impaired recruitment of termination factors could explain the delayed Pol II pause in *dis2*Δ.

Chapter 4

Dis2 regulates Seb1 recruitment by dephosphorylating Spt5

4.1 Introduction

The previous chapter showed that Dis2 is important for correct recruitment of termination factors and transcription termination, since its loss delays Pol II release from the DNA template as measured by ChIP-seq. Similar results were recently obtained by Parua et al. (2018), where it was shown that Dis2 inactivation promotes transcription beyond the normal termination zone as detected by precision run-on transcription and sequencing. While work described here has demonstrated that Dis2 dephosphorylates Pol II CTD on Thr4, Parua et al. (2018) found that Dis2 dephosphorylates Spt5. Spt5, like Rpb1, has a C-terminal region (CTR) that is composed of 18 tandem repeats of the consensus $T_1P_2A_3W_4N_5S_6G_7S_8K_9$, where Thr1 is phosphorylated by the Pol II CTD kinase Cdk9 early during the transcription cycle (Booth et al., 2018; Pei and Shuman, 2003; Swanson et al., 1991) (Figure 1.2). How much this phosphorylation contributes to promoting elongation in yeast remains unclear. On the contrary, the role of Spt5 phosphorylation in regulating the transition from initiation to elongation has been widely investigated in metazoans. The current model suggests that Spt5, in its unphosphorylated form, displaces the initiation factor TFIIE from Pol II (Grohmann et al., 2011; Schulz et al., 2016), and mediates Pol II promoter-proximal pausing via recruitment

of the negative elongation factor (NELF). Then, upon phosphorylation by Cdk9, Spt5 becomes a positive elongation factor and releases the paused Pol II that can enter the elongation phase of the transcription cycle. Parua et al. (2018) showed that Spt5 CTR is dephosphorylated in fission yeast by Dis2 at the PAS and that there is an accumulation of hyper-phosphorylated Spt5 at the genes 3' end in the catalytically inactive *dis2* R245Q strain. In addition, Parua et al. (2018) showed that Dis2 is a target of negative regulation by Cdk9 and proposed that Dis2 becomes active near the PAS owing to a drop in Cdk9 activity. Parua et al. (2018) then speculated that Dis2-dependent dephosphorylation of Spt5 reverses elongation-rate enhancement at the end of the transcription cycle and facilitates Pol II termination. How this would happen remains unclear. One possibility is that Spt5 dephosphorylation at the end of transcription could affect the ability of Spt5 to interact with Pol II, RNA or DNA. It is also possible that Dis2 facilitates termination of Pol II by mediating an interplay between elongation (Spt5) and termination (Seb1) factors during Pol II transition through the poly(A) site. This chapter investigates the relationship between Spt5 and Seb1 and demonstrates that Seb1 recruitment is dependent on dephosphorylation of Spt5 by Dis2. In addition, we show that Spt5 dephosphorylation may be important to regulate Spt5 binding to RNA.

4.2 Spt5 phosphorylation disrupts its ability to bind RNA

It was recently shown in budding yeast that Spt5 binds to nascent RNA once Pol II transcribes past the poly(A) site (Baejen et al., 2017). This is the exact same time Spt5-CTR is dephosphorylated by Dis2 in fission yeast. From the structural analysis of Pol II-Spt5 complex, it's very likely that Spt5-CTR is in close proximity to the RNA exit channel (Ehara et al., 2017). Thus, dephosphorylating the CTR would mean removing negative charges that could interfere with RNA binding. To test this hypothesis, con-

structs were cloned into bacteria encoding for Spt5 wt or a Spt5 variant where Thr1 was replaced on all the repeats within CTR harbouring this residue (17 out of the 18) with a glutamic acid (T1E), mimicking a hyper-phosphorylated Thr1. How much this resembles the phosphorylation of Spt5 *in vivo* is a question that can't be answered since the *in vivo* stoichiometry of CTR phosphorylation is unknown. Recombinant Spt5 wt and Spt5 T1E were purified from bacteria (Figure 4.1A) and their ability to bind RNA was assessed by electrophoretic mobility shift assays (EMSA).

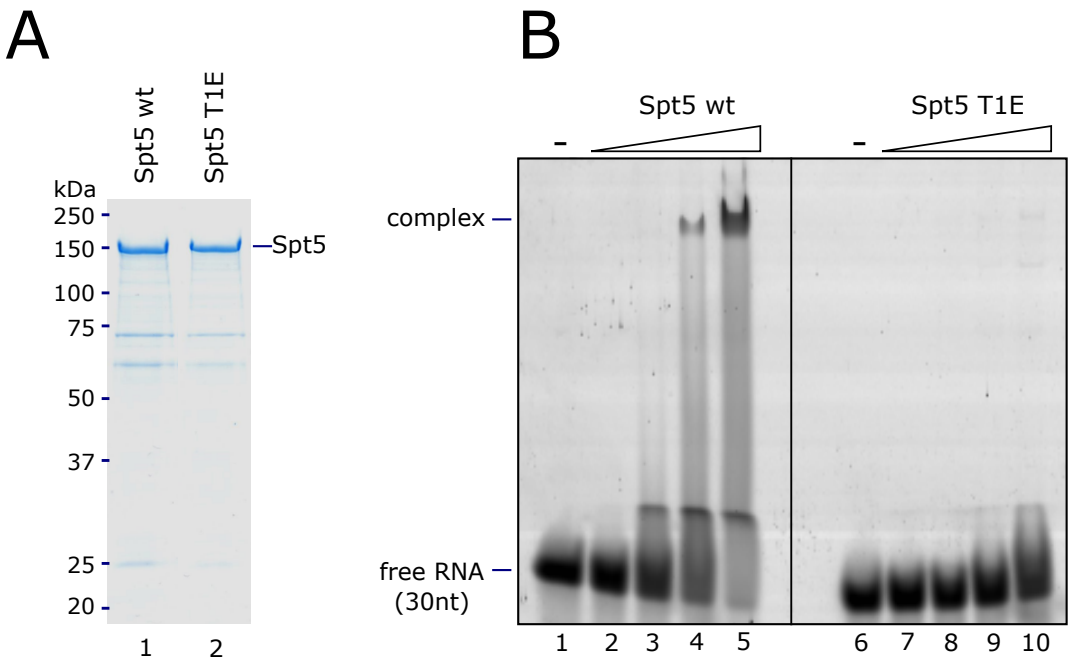


Figure 4.1: Spt5 T1E does not bind RNA.
(A) SDS-PAGE of Spt5 WT and T1E used for electrophoretic mobility shift assays (EMSAs).
(B) Binding of proteins shown in (A) to RNA was assessed by EMSA. A FAM-labelled RNA bearing the random 30-nt sequence AUUAGUAAAAUAUAUG-CAUAAAGACCAGGC was incubated with increasing protein concentrations and RNA binding was observed by a shift in RNA mobility on native gels which was detected using a phosphorimager.

Figure 4.1B shows that while addition of Spt5 wt caused RNA to shift (lanes 1-5), no such change in mobility was observed when RNA was incubated with Spt5 T1E (lanes 6-10). This result suggests that indeed Spt5 dephosphorylation is critical for the ability of the protein to bind nascent RNA and possibly to regulate transcription.

4.3 Hyper-phosphorylation of Spt5 leads to partial loss of Seb1 recruitment

To investigate the relationship between Spt5 and Seb1 at the end of transcription, yeast strains were generated where Thr1 was replaced on all the repeats harbouring this residue (17 out of the 18) within the Spt5 CTR with either an alanine (T1A), preventing phosphorylation at this position i*n vivo*, or a phosphomimetic glutamic acid (T1E, as in Figure 4.1). All *spt5* mutants were stable at the protein level with amounts comparable to the Spt5 wt (Figure 4.2).

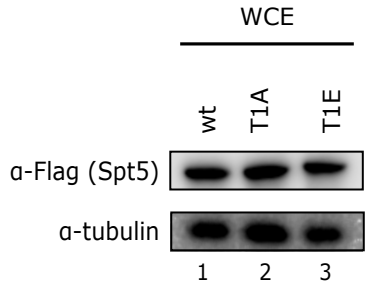


Figure 4.2: Spt5 protein levels in *spt5* mutants.
Western blot analysis of Spt5 wt/T1A/T1E in whole cell extract.

The effect of these mutations on Seb1 recruitment was then analysed by ChIP-seq. Similar to the decrease observed in *dis2*Δ mutant, Seb1 occupancy was strongly reduced in *dis2+, spt5* T1E and *dis2*Δ*, spt5* T1E mutants (Figure 4.3A-C). Reciprocally, levels of Seb1 reverted to wt when phosphorylation of Spt5 was prevented in T1A (Figure 4.3A-C). Scatterplot analysis in Figure 4.3D shows Seb1 ChIP intensity for each gene in *dis2*Δ (upper panel) and *dis2*Δ*, spt5* T1A (lower panel) compared to wt and shows that rescue of Seb1 recruitment in *spt5* T1A affected most of the genes analysed. This suggests that Spt5 phosphorylation controls Seb1 recruitment during late stages of transcription and loss of Seb1 in the *dis2* mutant is primarily due to hyper-phosphorylation of Spt5.

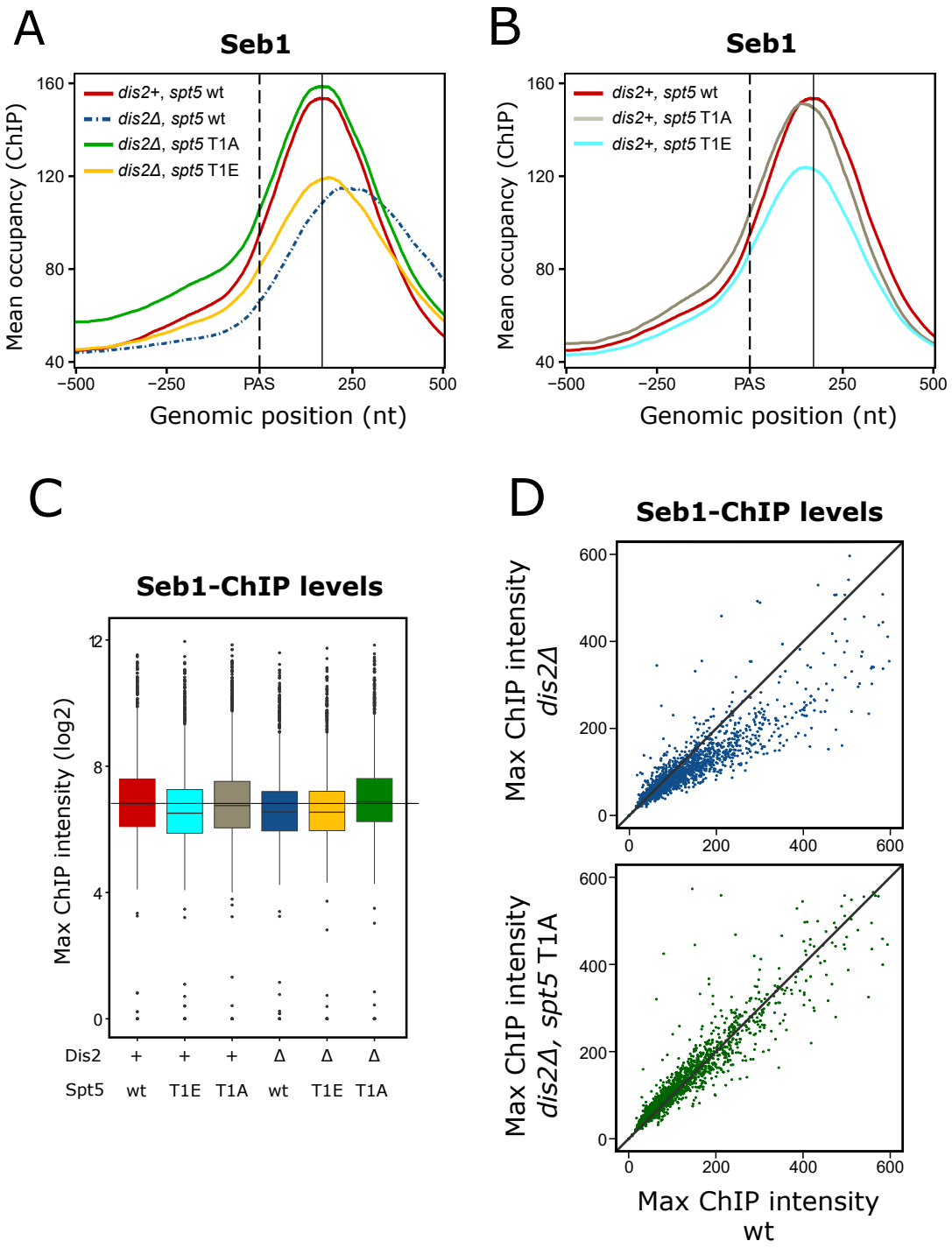


Figure 4.3: *spt5* T1A rescues Seb1 recruitment.
(A-B) Spike-in normalized ChIP-seq meta-profiles of Seb1 around the PAS in the indicated mutants from four biological replicas (*dis2+, spt5* wt/T1E and *dis2*Δ*, spt5* wt/T1A/T1E) or three biological replicas (*dis2+, spt5* T1A). Protein-coding genes for analyses were selected as above (n=1735). The black solid line defines the position of Seb1 peak in the wt (*dis2+, spt5* wt) strain.
(C) Quantitation (box plots) of Seb1 ChIP signal from A and B.
(D) Scatterplot of Seb1 ChIP signal intensities in wt versus *dis2*Δ (upper panel) or *dis2*Δ*, spt5* T1A (lower panel). Each point corresponds to the spike-in normalized maximum ChIP signal intensity for each gene downstream of the PAS.

4.4 Seb1 competes with Spt5 for binding to Pol II *in vitro*

Seb1 is known to interact with Pol II CTD (Wittmann et al., 2017). In addition, Seb1-Pol II cross-linking experiments done in our lab by Dr.Krzysztof Kus have revealed that Seb1 may also interact with the Pol II core in proximity to the regions where Spt4/5 has been shown to bind (Ehara et al., 2017). To test whether Spt5 and Seb1 binding sites on Pol II core overlap, binding assays were carried out using Pol II purified from a *S. pombe* strain where CTD can be cleaved off (Figure 4.4A-B). Using this CTD-less Pol II has the advantage to confine Seb1 binding exclusively to the core of Pol II. Full-length recombinant Spt4/5 heterodimer (Figure 4.4C) and recombinant Seb1 (Figure 4.4D) were purified from bacteria.

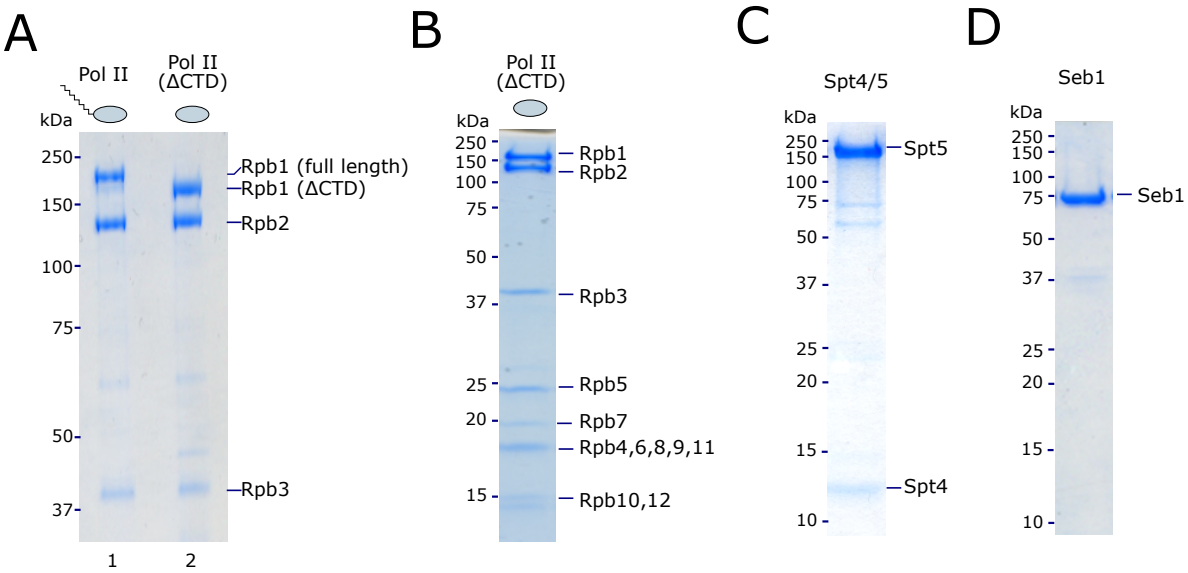


Figure 4.4: Purification of CTD-less Pol II, recombinant Spt4/5 and Seb1 for binding assay. These experiments were done by Dr Kus.
(A) SDS-PAGE analysis of Pol II before and after TEV cleavage.
(B) Ion-exchange purification of CTD-less Pol II.
(C) SDS-PAGE analysis of the purified recombinant Spt4/5.
(D) SDS-PAGE analysis of the purified recombinant Seb1.

The Pol II (ΔCTD) – Spt4/5 complex was then reconstituted *in vitro* (Figure 4.5A,

lane 1). Interestingly, addition of increasing amounts of Seb1 to the Pol II (ΔCTD) – Spt4/5 complex reduced the amount of Spt5 on Pol II, suggesting that Seb1 and Spt5 compete for binding to Pol II, at least *in vitro* (Figure 4.5A-B, lanes 2-5). Another possible explanation is that Spt5 binds to Seb1 and this complex no more interacts with Pol II.

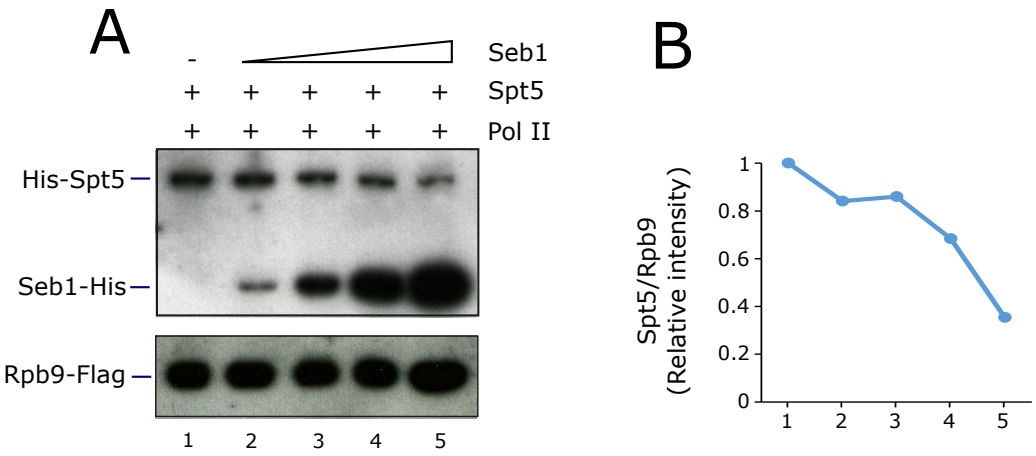


Figure 4.5: Seb1 competes with Spt5 for binding to Pol II.
(A)The Pol II-Spt4/5 complex was incubated with increasing amount of recombinant purified Seb1 (25, 75, 225 and 680 ng). The competition between Seb1 and Spt5 for Pol II binding was analysed by Western Blot with antibodies specific for Seb1 (His), Spt5 (His) and Pol II subunit Rpb9 (Flag).
(B) Spt5 levels normalized to Rpb9 from (A) were quantified by ImageJ.

4.5 Hyper-phosphorylation of Spt5 is not the sole cause of the termination defect in *dis2Δ*

Next, it was tested whether reduced Seb1 recruitment in the *dis2* mutant is the reason for delayed termination of Pol II transcription in this strain. Therefore, levels of the 3' extended read-through RNA from *rpl1202* were compared in the different *spt5* mutants by RT-PCR. RT-PCR was performed using either primers downstream or spanning the PAS (Figure 4.6).

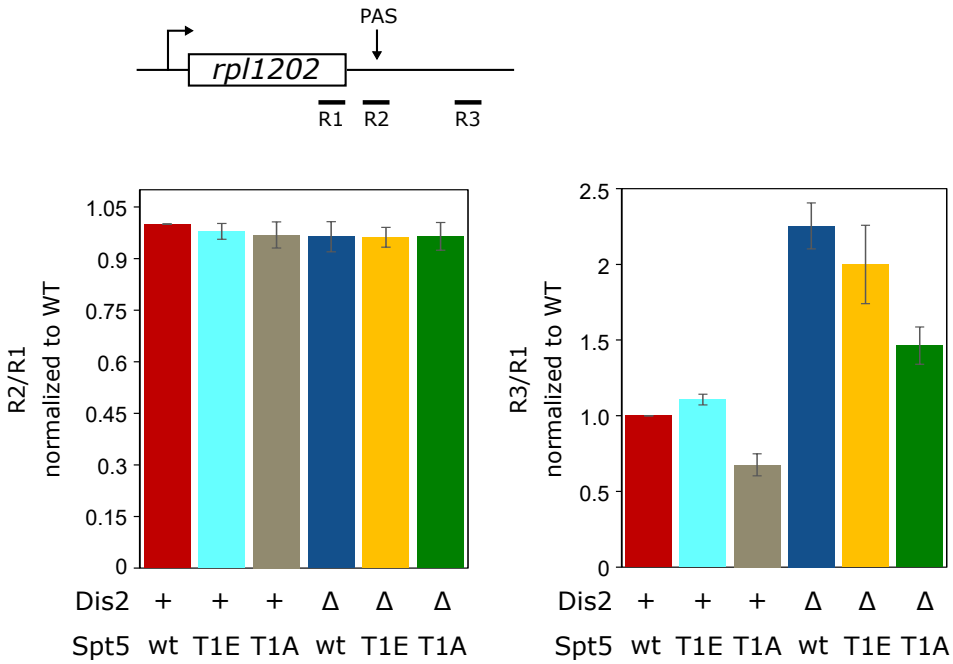


Figure 4.6: *spt5* T1A partially rescues 3' extended transcription on *rpl1202*. Analysis of termination defects of the *rpl1202* gene. Position of the RT–PCR products is shown in the schematic. Error bars indicate the standard error of the mean of biological triplicates.

Accumulation of the 3' extended read-through RNA in *dis2*Δ was detected using the downstream primers, whereas the intensity of PCR bands using primers spanning the PAS was the same in both *dis2*Δ and wt strain. This suggests that Dis2 regulates transcription termination after completion of the *rpl1202* mRNA cleavage.

If the termination defect in *dis2*Δ is only due to the reduction in Seb1 recruitment, a suppression of this defect is to be expected when Seb1 levels go back to normal in *dis2*Δ, *spt5* T1A (Figure 4.3A). Surprisingly, only partial suppression of the termination defect is observed (Figure 4.6), suggesting that although reduced Seb1 recruitment may be partly responsible, it is not the sole reason for the delayed release of Pol II in *dis2*Δ. On the other hand, loss of Seb1 due to hyper-phosphorylated Spt5 in the presence of functional Dis2 (Figure 4.3B), was not sufficient to cause read-through transcription (Figure 4.6).

To understand the degree by which *spt5* T1A can suppress the termination defect seen in *dis2*Δ, the analysis was extended to the whole genome by Pol II ChIP-seq (Figure 4.7).

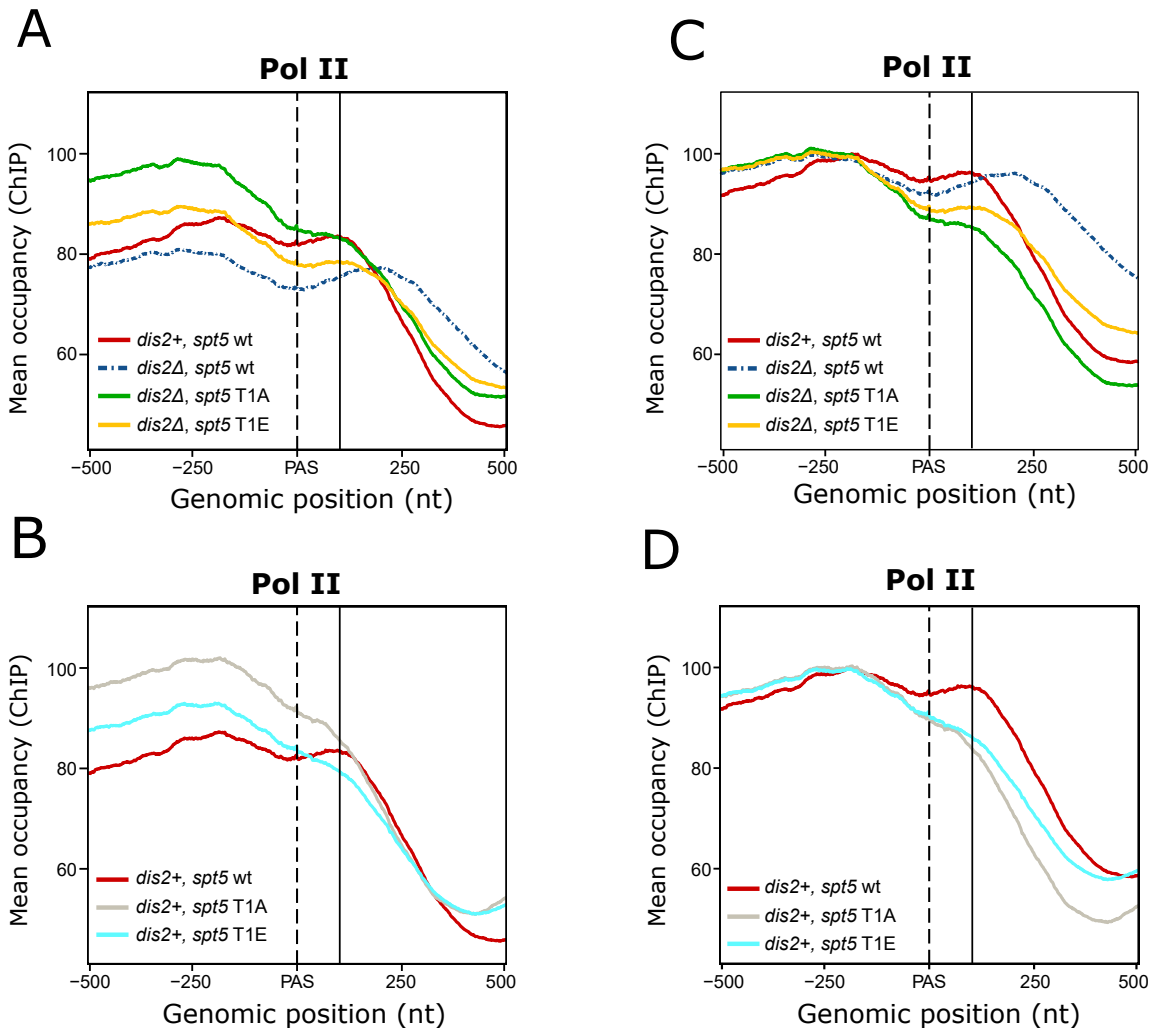


Figure 4.7: Pol II metagene profile in *spt5* mutants.
(A-B) Spike-in normalized ChIP-seq meta-profiles of total Pol II (8WG16) around the PAS, averaged from two biological replicas. Protein-coding genes for analysis were selected as above (n=1735).
(C-D) Profiles from (A-B) were horizontally translated to have approximately equal occupancy levels upstream of the PAS. The black solid line defines the position of Pol II peak downstream of the PAS in the wt (*dis2+, spt5* wt) strain.

Metagene analysis was done by normalizing the ChIP signal to the spike-in (Figure

4.7A-B). In addition, to better visualize termination defects, the analysis was repeated by horizontally translating the Pol II profiles to have approximately equal Pol II occupancy upstream of the PAS (Figure 4.7C-D). In both cases, metagene analysis of Pol II occupancy showed a rescue of the termination defect upon loss of Spt5 phosphorylation (T1A mutant) in *dis2*Δ strain (Figure 4.7A and C). From Figure 4.7B and D it is also obvious that introducing *spt5* T1A or T1E mutations in presence of functional Dis2 caused a change in Pol II profile, with a less pronounced peak after the PAS compared to the wt strain. This change in Pol II profile made it harder to compare termination defects by metagene analysis. Thus, for each strain a Pol II termination index was calculated by dividing reads downstream of the poly(A) site to the reads mapping within 500 nt before and after the PAS (Figure 4.8A). Consistent with the RT-PCR data, the global analysis of Pol II termination also revealed a partial rescue of the termination defect upon loss of Spt5 phosphorylation (T1A mutant) in *dis2*Δ strain that affected most of the genes (Figure 4.8A-B). Interestingly both RT-PCR and genome-wide analysis revealed that the threonine-to-alanine mutation in Spt5 speeds up transcription termination even when Dis2 is functional. This result seems to be in line with the idea by Parua et al. (2018) of Spt5 dephosphorylation slowing elongation, so that termination can happen. However, introducing phosphomimetic mutation in Spt5 (*spt5* T1E) did not lead to delayed Pol II termination in the presence of active Dis2 (Figure 4.8A) and correlates with lack of the 3' read-through observed in *rpl1202* (Figure 4.6C). The *dis2*Δ*, spt5* T1E mutant recapitulated the transcription termination defect observed in the strain lacking functional PP1 phosphatase (*dis2*Δ), however the defect was somewhat less pronounced. This could be explained by the two-fold difference in charge between phospho-threonine and glutamic acid (Craft and Legge, 2005).

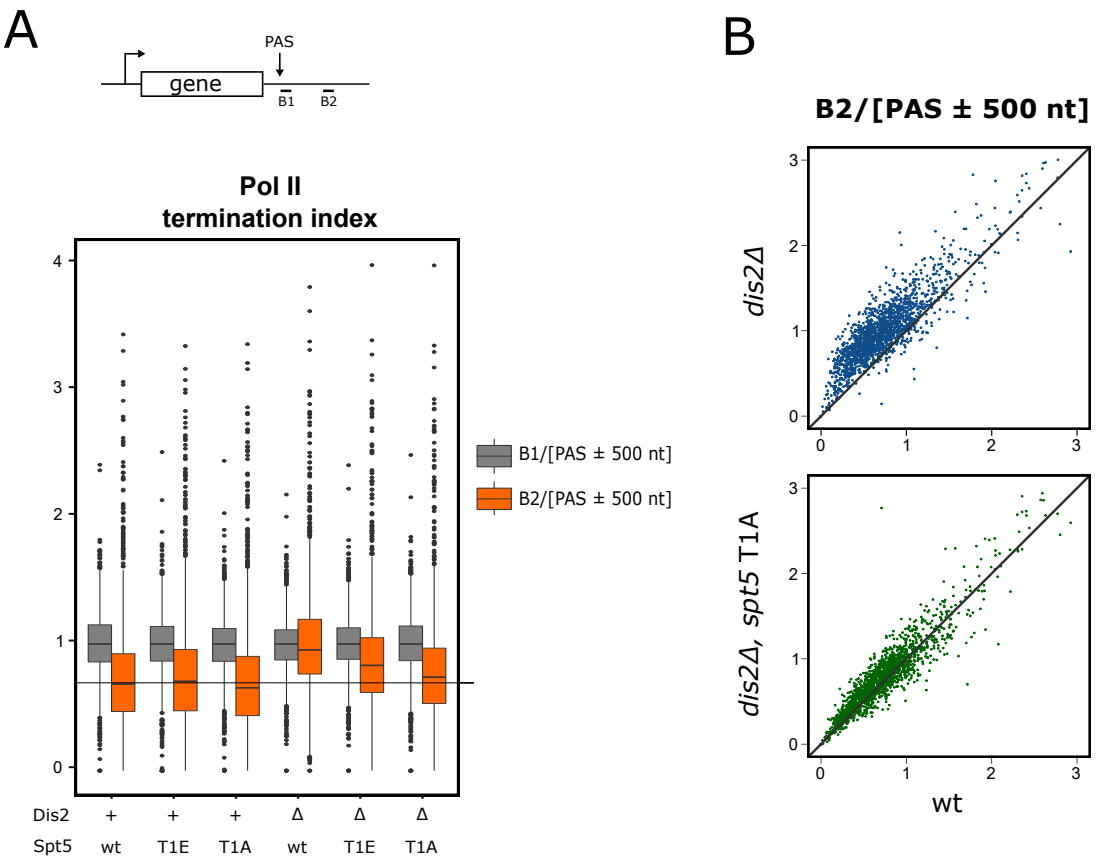


Figure 4.8: *spt5* T1A partially rescues Pol II termination defects.
(A) Pol II termination index was determined by dividing mapped reads in window B1 or B2 by the read count within PAS±500 nt. Protein-coding genes for analysis were selected as above (n=1735). B1 corresponds to the region from PAS to PAS+50 nt. B2 corresponds to the region from PAS+350 nt to PAS+400 nt.
(B) Scatterplot of Pol II termination index in wt versus *dis2*Δ (upper panel) or *dis2*Δ, *spt5* T1A (lower panel).

4.6 Discussion

The ability of Spt5 to bind RNA has been shown *in vitro* (Cheng and Price, 2008; Viktorovskaya et al., 2011; Qiu and Gilmour, 2017), but only recently Baejen et al. (2017) found that interaction with RNA *in vivo* is localized at the TSS and downstream of the PAS, just where Spt5 exists on Pol II in its unphosphorylated form. This chapter shows that indeed hyper-phosphorylated Spt5 does not bind RNA *in vitro*, suggesting

that Spt5 dephosphorylation may be important for its ability to bind nascent RNA once Pol II transcribes past the poly(A) site. What is the importance of the Spt5-RNA interaction at the end of the transcription cycle is an open question. Binding to RNA is required for Spt5 stable association with Pol II *in vitro* in *Drosophila* (Missra and Gilmour, 2010), suggesting that binding to the nascent RNA recruits Spt5 at the beginning of the transcription cycle. If this was the case in fission yeast, one would expect T1E mutation of the CTR to be lethal since Spt5 is an essential transcription factor (Pei and Shuman, 2002; Shetty et al., 2017). However, *spt5* T1E strain is viable. In line with this result, Spt5 association with Pol II in fission yeast does not require RNA (Figure 4.5), although it can't be excluded that RNA may enhance Spt5 recruitment.

The fact that phosphorylation of Spt5 changes the biochemical properties of the protein, in particular its ability to interact with other transcription factors, has been widely proven in the initiation to elongation transition (Grohmann et al., 2011; Schulz et al., 2016; Missra and Gilmour, 2010; Werner, 2012; Doamekpor et al., 2014, 2015; Schneider et al., 2010). Here it is shown that change in Spt5 phosphorylation status is also important to regulate the recruitment of termination factors (Seb1). Seb1 is an essential factor for mRNA 3' processing and transcription termination in fission yeast (Wittmann et al., 2017). Seb1 interacts with Pol II via its CTD-interacting domain (CID), but also has an RNA recognition motif (RRM) that binds nascent RNA as it emerges from the RNA exit channel on Pol II (Wittmann et al., 2017). Work in the lab by Dr Kus showed that Seb1 also forms extensive contacts with the Pol II core via its RRM (data not shown). Interestingly, this chapter shows that Seb1 and Spt5 compete for binding to Pol II *in vitro*, suggesting that, at the transition from elongation to termination, Spt5 needs to dissociate from the Pol II core for Seb1 to bind. In addition, this chapter reveals an antagonistic relationship between Spt5 and Seb1 that depends on Spt5 phosphorylation and is regulated by Dis2. Seb1 occupancy on chromatin in

*dis2*Δ reverts to wt when Spt5 phosphorylation is prevented by T1A mutation, while Seb1 levels are reduced in all the mutants where Spt5 is hyper-phosphorylated. Interestingly, it seems that there is no direct correlation between partial loss of Seb1 and transcription termination defects, suggesting that even low levels of Seb1 are enough for this termination factor to fulfil its function in *spt5* T1E. *spt5* T1A is able only partially to suppress the delayed Pol II 3' pause and lower levels of Seb1 in *dis2+, spt5* T1E are not sufficient to cause read-through. These results suggest that the termination defect seen in *dis2*Δ is not only caused by Spt5 hyper-phosphorylation, but there most be another Dis2 substrate that contributes to the termination phenotype.

Chapter 5

Discussion

5.1 What is the role of Thr4P-CTD at the gene 3' ends?

Termination factors recognize the CTD by means of the CTD-interacting domain (CID). CTD phosphorylation at Ser2 is enriched at the gene 3' ends, where it recruits CID-containing proteins that are important for 3' end processing and transcription termination. In addition to Ser2P, there is an enrichment of Thr4 phosphorylated CTD downstream of the poly(A) site in both yeast and mammals, raising the possibility that both these phospho-marks may be important to recruit termination factors at the end of the transcription cycle.

Structures from several CID-containing proteins bound to peptides have revealed that CIDs interact with the consensus sequence via a highly conserved binding pocket. The NMR structure of Rtt103-Ser2P by Lunde et al. (2010) showed that there is one important arginine (Arg108) in the CID that makes direct contacts with the phosphate moiety of Ser2P (Figure 5.1). Interestingly, the corresponding residue in *S.cerevisiae* Pcf11 is an asparagine (Asn107) and this is the reason why Pcf11-CID binds Ser2P with a lower affinity than Rtt103 (Lunde et al., 2010). In fact, mutating Asn107 to Arg increases Pcf11 affinity for Ser2P to values comparable to Rtt103 (Lunde et al., 2010). Reciprocally, Arg-to-Asn mutation in Rtt103 causes a decrease in affinity for

Ser2P leading to lower occupancy of Rtt103 on chromatin (Lunde et al., 2010). Indeed, the X-ray structure of *S. cerevisiae* Pcf11-CID showed that the CID does not form direct contacts with the Ser2P moiety, but seems to recognize the phosphate indirectly (Meinhart and Cramer, 2004). This indirect Ser2P recognition may result from stabilization of the CTD peptide by an additional intramolecular hydrogen bond between the Ser2 phosphate and the side-chain hydroxyl group of Thr4 on the same CTD repeat (Meinhart and Cramer, 2004). Structure of *S.pombe* Pcf11-CID has not been solved and prior to the work described herein, the affinity to the different CTD isoforms were unknown. Like its homologue in budding yeast, *S. pombe* Pcf11-CID lacks the arginine that is important for Ser2P binding by Rtt103 and has a Gln instead (Figure 5.1). In spite of this, it is shown in the present work that the affinity of *S. pombe* Pcf11 to Ser2P is comparable to that of Rtt103, thus considerably higher than the affinity observed for *S. cerevisiae* Pcf11. This suggests that Pcf11-Ser2P interaction in fission yeast is different from any other known factor. Like *S. pombe* Pcf11, Rhn1, the Rtt103 homologue in fission yeast, also lacks the arginine corresponding to Arg108 in Rtt103 and has a Tyr instead (Figure 5.1). Rhn1-CID was insoluble in our hands and couldn't be tested for its affinity to the different CTD phosphoisoforms. *S. pombe* Seb1 is the only other yeast termination factor that has a basic amino acid (Lys124) corresponding to Arg108 in Rtt103 (Figure 5.1). Biochemical studies have shown that not only Lys124, but also Lys121 is critical for binding to Ser2P (Wittmann et al., 2017), although how these two residues engage the phosphorylated CTD is not known. Interestingly, Lys121 is highly conserved in all the CIDs (Figure 5.1). Unlike for Seb1, this residue is not critical for binding of *S. cerevisiae* CIDs. Its importance in *S. pombe* Pcf11 and Rhn1 needs to be tested.

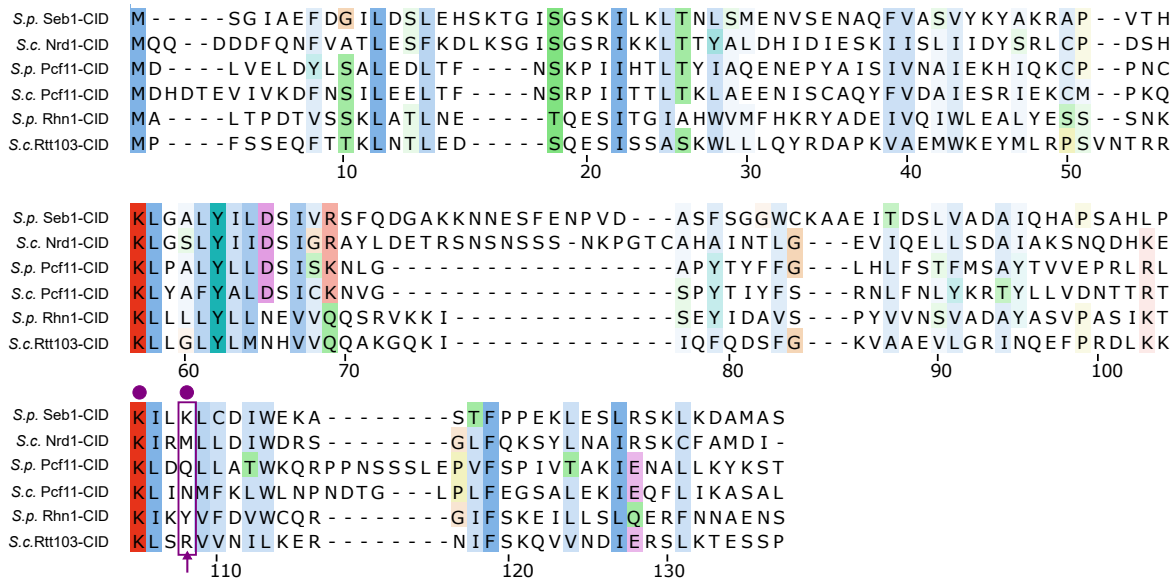


Figure 5.1: Sequence alignments of CIDs from yeast termination factors. The colour indicates amino acid conservation. Purple box and arrow marks the Arg previously shown to contact Ser2P in Rtt103 and the corresponding residues in the other CIDs. The purple circles highlight the amino acids important for Seb1 binding to Ser2P. The alignment was done using Promals3D.

This study has tested the binding of all termination factors for Thr4P in both budding and fission yeast. What emerged is that indeed these factors bind Thr4P-CTD peptide *in vitro*, but not with the same affinity, suggesting that Thr4P may help recruiting some factors, but not all. In fact, only *S. cerevisiae* Rtt103 and *S.pombe* Pcf11 can bind equally well Ser2P and Thr4P. Interestingly, Jasnovidova et al. (2017); Nemec et al. (2017) recently found that the structure of Rtt103-Thr4P is similar to that of Rtt103-Ser2P. In fact, the same Arg108 that binds Ser2P forms an hydrogen bond with the phosphate moiety of Thr4P. As discussed above, structure of fission yeast Pcf11-CID has not been solved, so whether the binding to Ser2P and Thr4P involves the same residues on the CID is unknown. What is clear, however, is that the biochemical properties of Pcf11 differ between fission and budding yeast. Not only the work presented here demonstrates that *S.pombe* Pcf11 has a higher affinity for Ser2P compared to its *S. cerevisiae* homologue, but it also shows that *S.p.* Pcf11 binds equally well

Ser2P and Thr4P, while the budding yeast Pcf11 has a preference specifically for Ser2P. These results suggest that Pcf11 function *in vivo* may differ between the two organisms, although further work needs to be done to prove this point.

A role of Thr4P-CTD in recruiting termination factors would explain why this phospho-mark is exclusively enriched at the 3' end of protein-coding genes in fission yeast and mammals. High levels of Thr4P on the gene body would recruit termination factors early during transcription, leading to premature termination. Indeed, high levels of Thr4P within the gene body of long non-coding transcripts have been shown to correlate with abortive transcription in mammals (Schlackow et al., 2017). In addition, the present study shows that higher levels of Thr4P at gene 3' ends upon deletion of *dis2* correspond to increased occupancy of Pcf11 and Rhn1. Interestingly, while loss of Dis2 is not lethal for fission yeast, expressing a T4E mutated CTD leads to cell death. This may be explained by the fact that while increase in Thr4P is prominent downstream of the poly(A) sites in *dis2*Δ, the T4E mutated CTD is likely to recruit termination factors early during transcription, causing premature termination. Further research is necessary to test this hypothesis, even more so in light of the observation that Thr4P-CTD is present, although low, on the gene body of protein-coding genes in *S. cerevisiae*.

5.2 What is Dis2 role in transcription termination?

The C-terminal part of Rpb1 and Spt5 are unique in their repetitive and unstructured nature. Genetic interactions suggest a functional overlap between Spt5 CTR and the Pol II CTD (Schneider et al., 2010). Interestingly, both CTR and CTD are phosphorylated by the same kinase Cdk9 and dephosphorylated by the same phosphatase Dis2. Loss of Dis2 or expression of a catalytically inactive variant causes: (i) hyper-phosphorylation

of Spt5 (Parua et al., 2018) with (ii) consequent decrease in Seb1 occupancy on chromatin, (iii) higher levels of Thr4P-CTD and (iv) delayed release of Pol II from the DNA template. Although depleting Seb1 from fission yeast also causes a shift in Pol II ChIP-seq signal further downstream of the poly(A) site (Lemay et al., 2016), this thesis shows that partial loss of Seb1 as a consequence of hyper-phosphorylated Spt5 in the presence of functional Dis2 is not enough to cause delayed Pol II pausing. An explanation is that only full loss of chromatin-associated Seb1 causes termination defects. Since Seb1 associates with Pol II by different means, it is also tempting to speculate that *spt5* T1E mutation affects only one of them, leaving a pool of Seb1 protein that is functional and enough to fulfil its role in termination. The fact that introducing phosphomimetic mutation in Spt5 with consequent loss of Seb1 in the presence of active Dis2 does not cause termination defect leads to the hypothesis (Figure 5.2) that the defect seen in *dis2*Δ is due to hyper-phosphorylation of both Dis2 substrates: Spt5 and Thr4-CTD. To test whether Rpb1 CTD T1A in combination with Spt5 CTR T1A in *dis2*Δ fully rescues the termination defect would prove the validity of the model.

It is also possible that other substrates of Dis2 contribute to the phenotype. The number of phosphatases is much less than the number of kinases (Shi, 2009), that is phosphatases have many multiple targets. Interestingly, the PP1 phosphatase Glc7 in *S. cerevisiae* can dephosphorylate several CPF subunits *in vitro* (He and Moore, 2005). Whether this is relevant *in vivo* has been tested only for the Pta1 subunit. Glc7 dephosphorylates Pta1 and by doing so controls poly(A) synthesis (He and Moore, 2005). Mammalian PP1 is also specifically required for mRNA polyadenylation - the poly(A) tails are shorter upon PP1 inhibition (Shi, 2009). Whether this function is conserved in *S. pombe*, whether Dis2 can target any other CPF subunit and if this contributes to the delayed pause of Pol II in *dis2*Δ needs to be tested.

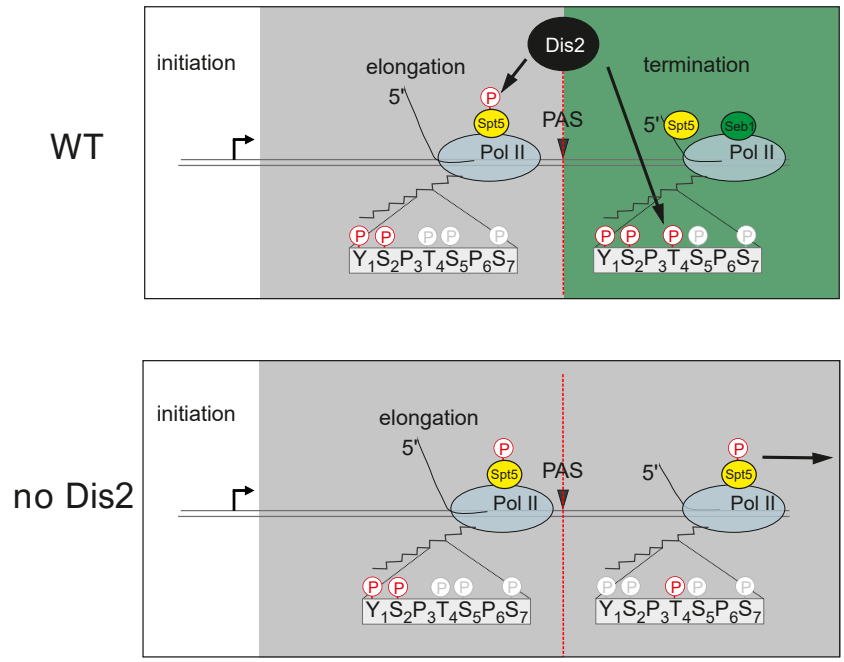


Figure 5.2: Model. PP1-Dis2 dephosphorylates Spt5 and Thr4P-CTD at the end of the transcription cycle (WT). Upon deletion of *dis2*, hyper-phosphorylation of both Spt5 and Thr4P causes defects in transcription termination.

5.3 What is the importance of Spt5 dephosphorylation at the end of transcription?

During the transcription cycle, Pol II progresses through three main stages: initiation, elongation and termination. Spt5 is an elongation factor that regulates the initiation to elongation transition. Interestingly, Spt5 stays associated to the chromatin past the poly(A) site (Baejen et al., 2017), suggesting that it may also have a function in transcription termination. This thesis together with work from Parua et al. (2018) proves that Spt5 plays a role in fine-tuning the elongation to termination transition and reveals parallels with the initiation to elongation transition.

Progression through the different stages of transcription requires a change of factors associated to Pol II. In the initiation to elongation transition, Spt4/5 can displace

TFIIE from the Pol II elongation complex and stimulate processivity (Grohmann et al., 2011). Similarly, interaction sites on Pol II of Spt5 and the termination factor Seb1 overlap (Dr Krzysztof Kus, personal communication), and this thesis shows that these two factors compete for binding to the core of polymerase *in vitro*. However, both Seb1 (Figure 3.5) and Spt5 (Parua et al., 2018) associate with chromatin downstream of the PAS as measured by ChIP-seq, suggesting that they co-exist on the transcribing machinery *in vivo*. This may seem surprising, but could be easily explained by the fact that proteins dynamically contact the chromatin-associated Pol II on multiple sites. Interestingly, Seb1 can interact with the Pol II core, with the Pol II CTD and the nascent RNA. Whereas both RNA and CTD-binding are essential for correct transcription termination, only RNA-binding is important for Seb1 recruitment (Wittmann et al., 2017). Spt5 also can bind RNA in addition to Pol II core. This thesis has shown that unphosphorylated Spt5, but not hyper-phosphorylated Spt5, can bind RNA. Spt5 is phosphorylated at gene 5' ends by Cdk9 and dephosphorylated by Dis2 at the poly(A) site (Parua et al., 2018). Interestingly, Spt5 interacts with the nascent RNA at the 5' and 3' end of genes (Baejen et al., 2017), that is when Spt5 is unphosphorylated. It is therefore possible that Spt5 interacts with the Pol II core until the elongation complex transcribes through the poly(A) site, when, following Dis2-dependent dephosphorylation, Spt5 may translocate to the nascent RNA and leave space on the polymerase core for Seb1 to bind (Figure 5.2). Further research is necessary to test this hypothesis, but this model would explain why correct recruitment of Seb1 depends on Spt5 dephosphorylation. Hyper-phosphorylated Spt5 at gene 3' ends in *dis2*Δ wouldn't bind the nascent RNA, wouldn't translocate from the core of polymerase and wouldn't leave space for Seb1 to bind Pol II core.

In agreement with a role of Spt5 in transcription termination, Mayer et al. (2012) showed that Spt5 CTR is important for the recruitment of pre-mRNA cleavage factor I

(CFI) to the 3' ends of *S. cerevisiae* genes. In addition, Spt5 CTR interacts with CFI *in vitro*. Similarly, in the early stages of transcription, Spt5-CTR recruits the capping enzyme and the negative elongation factor NELF and directly interacts with these factors - this happens exclusively when the CTR is unphosphorylated. (Doamekpor et al., 2014, 2015; Schneider et al., 2010; Missra and Gilmour, 2010). A similar mechanism may exist at the end of the transcription cycle where changes in the phosphorylation status of Spt5 could affect its ability to interact with 3' end processing factors.

Spt5 and its role in promoting elongation are conserved in all three kingdoms of life reflecting its importance in transcription regulation (Werner and Grohmann, 2011). Interestingly, bacterial Spt5 homologue NusG antagonizes the transcription termination activity of Rho during the synthesis of actively translated mRNAs, but helps Rho to terminate untranslated transcripts (Sullivan and Gottesman, 1992; Cardinale et al., 2008; Peters et al., 2012). Recently, Lawson et al. (2018) found that accessibility of NusG-KOW domain determines whether NusG will act to promote or prevent Rho-dependent termination. Whether a similar mechanism exists in higher eukaryotes is not known, but it is tempting to speculate that communication between elongation and termination factors to regulate properties of RNA polymerase is conserved across different domains of life.

5.4 How could phosphorylation regulate transcription termination?

At the end of the transcription cycle, Pol II is dislodged from the DNA template. How this happens is still object of debate. Two models have been proposed. The torpedo model posits that cleavage of the pre-mRNA creates an uncapped 5' end which is degraded by a 5'-3' exonuclease until it catches up with the transcribing Pol II, leading

to its dislodgement from the DNA. The allosteric model postulates that Pol II release is triggered by conformational changes of the elongation complex upon dissociation of elongation factors and binding of the termination complex once Pol II transcribes through the poly(A) site. Interestingly, not all elongation factors dissociate from the transcribing complex at gene 3' ends. Kim et al. (2004a) showed in budding yeast that while the Paf and TREX/THO elongation complexes do not crosslink to transcribed regions downstream of polyadenylation sites, a group of elongation factors (Spt4/5, Spt6/Iws1, and Spt16/Pob3) does, indicating that they probably have other roles during termination.

What does trigger association/dissociation of factors? Phosphorylation of the CTD is one of the best studied mechanism. At the 3' end of protein-coding genes high Ser2P signal is a well-known feature of poly(A) site recognition. This study shows that enrichment of Thr4P may act similarly to Ser2P in recruiting termination factors to Pol II. Phosphorylation of proteins other than Rpb1 might be equally important to regulate the last stages of transcription. For example, Xrn2 phosphorylation promotes its enzymatic activity and recruitment to chromatin (Sansó et al., 2016), whereas Pcf11-CID phosphorylation weakens its interaction with Pol II, promoting mRNA release from chromatin-associated Pol II (Volanakis et al., 2017). Removal of post-translational modifications is likely as important as the addition of the same modifications in the first place. Indeed, this thesis shows that Dis2-dependant Spt5 dephosphorylation as transcription complexes traverse the cleavage and polyadenylation signal fine-tunes recruitment of the termination factor Seb1 to Pol II. Interestingly, many components of the CPF complex are phosphorylated *in vivo* (Carpy et al., 2014; Ryan and Bauer, 2008) and the PP1 phosphatase Glc7 in *S. cerevisiae* can dephosphorylate many of them *in vitro* (He and Moore, 2005). Further work will help to understand to which extent cycles of phosphorylation and dephosphorylation of the 3' end processing/termination

machinery contribute to Pol II conformational changes that are important for transcription termination.

Chapter 6

Material and methods

6.1 *S.pombe* media and growth conditions

Yeast cells were grown in YES liquid media (0.5% yeast extract, 3% glucose, 225 mg/l adenine, 225 mg/l uracil, 225 mg/l histidine, 225 mg/l leucine, 225 mg/l lysine) at 30 °C and 200 rpm to $OD_{600} \sim$ 0.4-0.7.

6.2 Yeast strain construction

Standard PCR-based methodology was used for epitope tagging (Bähler et al., 1998). The pop-in, pop-out method allele replacement methods was used to introduce a TEV cleavage site (GAGAATTTGTATTTTCAGGGT) upstream the Rpb1-CTD. DNA fragments carrying the TEV cleavage site were generated by 2-step PCR (see Table 6.3 for primers). Fragments were cloned into pKS-URA4 vector by NotI/XbaI digestion and ligation. HpaI was used to linearise the plasmid before transformation into cells with depleted ura4 locus. Ura+ colonies were selected on EMMG media lacking uracil (3 g/l Potassium hydrogen phthalate, 2.2 g/l Na_2HPO_4, 3.7 g/l glutamic acid, 1.05 g/l $MgCl_2$ x 6 H_2O, 1 g/l KCl, 40 mg/l $NaSO_4$, 14.7 mg/l $CaCl_2$ x 2 H_2O, 10 mg/l nicotinic acid, 10 mg/l myo-inositol, 1 mg/l pantothenic acid, 10 μg/l biotin, 1 mg/l citric acid, 0.5 mg/l boric acid, 0.4 mg/l $MnSO_4$, 0.4 mg/l $ZnSO_4$ x 7 H_2O, 0.236 mg/l $FeCl_2$ x 4 H_2O, 0.1 mg/l KI, 40 μg/l $CuSO_4$ x 5 H_2O, 225 mg/l adenine, 225 mg/l histidine, 225 mg/l

leucine, 225 mg/l lysine) and genotyped by DNA sequence analysis of PCR products. Positive clones were plated onto 0.1% 5-FOA to select cells in which the ura4+ gene had been lost (pop-out).

Spt5-CTR mutants. *S. pombe* Spt5-CTR sequences of Thr1 mutants (T1A and T1E) with an optimized codon usage were synthesized by Eurofins Genomics. 17 out of the 18 repeats within the CTR have a Thr at position one and the T1A/T1E mutations were introduced in all of these 17 repeats.

Rbp1-CTD mutants. These strains were made by Dr Adrien Birot (postdoctoral researcher, Vasilieva lab). *S. pombe* Rpb1-CTD sequences of Thr4 mutants (T4A and T4E) with an optimized codon usage were synthesized by Eurofins Genomics.
In the *S.pombe* CTD, the first four repeats deviate from the consensus heptad (YSPT-SPS). The distal 25 heptads perfectly match to the consensus sequence (with the exception of an alanine instead of Pro3 in the ninth heptad). It is in these 25 repeats that T4A and T4E mutations were introduced. Rpb1 CTD T4A and T4E were transformed into diploid yeast cells. Tetrad dissection was performed following a 3 days incubation of diploid strains on EMMG media at 30 °C.

6.3 Yeast transformation

5 mL of yeast cells were grown in YES to $OD_{600} \sim 0.5$. Cells were harvested and washed with water and LiTE (0.1 M LiAc, 10 mM Tris-HCl pH 7.5, and 1 mM EDTA), then resuspended in 100 μL LiTE. After addition of 20 μg ssDNA and 1 μg of plasmid or PCR product, the reaction was incubated at room temperature for 10 min. 260 μL PLATE (40% PEG, 0.1 M LiAc, 10 mM Tris-HCl pH 7.5, and 1 mM EDTA) was added. Cells were incubated at 30 °C for 1h. After addition of 43 μL DMSO (dimethyl sulfoxide), cells were heat shocked at 42 °C for 5 min. Cells were harvested, washed

with water and resuspended in 650 μL YES or EMMG and left to recover at 30 °C overnight. Cells were spread on plates supplied with the appropriate selection marker.

6.4 Yeast whole-cell protein extraction for WB

3 OD of cells in exponential growth phase were harvested and lysed by addition of 100 μL 1.85 M NaOH. After 5 min incubation on ice, 100 μL TCA (trichloroactic acid) was added. The reaction was incubated for 30 min at 4 °C. Whole cell extract was recovered by centrifugation at 25,000 g for 15 min. The pellet was resuspended in 100 μL 2.5x Lammli buffer (Laemmli, 1970), 0.1 M Tris-HCl pH 8.0.

6.5 Yeast protein purification

Dis2 for phosphatase assay. Flagged-Dis2 *S. pombe* strain was grown at 30 °C to exponential phase and harvested by filtration (Churchman et al. (2012)). Pellets were frozen in liquid nitrogen and cells were disrupted by manual grinding with mortar and pestle. All steps from this point were performed on ice with precooled buffers. Cells were resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 0.5% TritonX-100, 0.5 mM DTT, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$) containing protease inhibitors (1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM benzamidine, 2.3 μM leupeptin (1 μg/ml), 1.5 μM pepstatin A (1 μg/ml), 81 μM bestatin (25 μg/ml), 1.5 μM aprotinin (10 μg/ml)). 5 ml of lysis buffer was used per 1l of yeast culture. The lysate was clarified by centrifugation at 16,000 g for 20 min and incubated with M2 agarose beads (Sigma) on a rotator for 2 h at 4 °C. 50 μl of beads (packed) were added per 1l of initial yeast culture. Beads were washed 4 times for 2 min with lysis buffer, 4 times for 2 min with buffer A (lysis buffer without TritonX-100), then Dis2 complexes were eluted with four volumes of 2.5 mg/ml Flag-peptide (Sigma) in buffer B (25 mM

Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 3 mM DTT, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$).

Pol II for phosphatase assay. For Pol II purification, the Rpb9 subunit was fused with a Flag-tag. Flagged-Rpb9 was purified as above with the following modifications. 100 μl of M2 beads (packed) were added per 1l of initial yeast culture. The M2 agarose gel was washed 4 times with buffer C (50 mM Tris-HCl pH 7.5, 1 M NaCl, 1 M urea, 10% glycerol, 0.5% TritonX-100, 0.5 mM DTT, 0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$), 4 times with buffer A, then eluted as above.

CTD-less Pol II for Spt5/Seb1 competition assay A TEV clevage site (GAGA-ATTTGTATTTTCAGGGT) was integrated in yeast upstream the CTD of Rpb1. The purification of this CTD-less polymerase (Rpb9-flagged) was carried by Dr K. Kus (postdoctoral researcher, Vasilieva lab) as above with the following modifications. After Flag-purification and elution, the sample was subjected to an ion exchange chromatography column (HiTrap Q HP, GE Healthcare) equilibrated with buffer QA (50 mM TRIS pH 7.7, 5 mM NaCl, 10% glycerol, 0.5 mM MgCl2, 0.5 Mg(OAc)2, 1 mM β-mercaptoethanol). The column was washed with 8% buffer QB (buffer QA with 2000 mM NaCl). The Pol II complex was eluted with a gradient of buffer QB (up to 40%). Pol II containing-fractions were pooled together and incubated with home-made Tev protease at room temperature for 1 h. The sample was then subjected to size-exclusion chromatography using a Superose 6 resin (GE Healthcare). The CTD-less Pol II complex was stored at -20 °C until usage.

6.6 Phosphatase assay

For the de-phosphorylation assay, Pol II (1l) and Dis2 (1l) complexes were mixed together and incubated at 30 °C for the indicated time in buffer B. In parallel, Pol II (1l)

was incubated without Dis2 as a negative control. Reactions were stopped by addition of Lammli buffer and phosphorylation status of Rpb1 was analysed by western blotting with phospho-specific antibodies (see Table 6.4).

6.7 Cloning of recombinant proteins

CID-containing proteins. The cloning of the CID-containing proteins was done by Ms Katie Duckett (PartII student, Vasilieva lab), except for Seb1 CID, cloned by Dr Sina Wittmann (postdoctoral researcher, Vasilieva lab). Gene-specific primers (see Table 6.3) were used to amplify the CID domain of CID-containing proteins from *S.pombe* cDNA. Primers contained restriction site for subsequent cloning into expression vectors (see Table 6.2 for all plasmids used in this study). The *S.cerevisiae* Nrd1 CID_{1-152} ,Pcf11 CID_{1-144} and Rtt103 CID_{1-152} and the *S.pombe* Seb1 CID_{1-148} constructs were cloned into pET41a(+), resulting in a C-terminal Hisx8-tag. The *S.pombe* Pcf11 CID_{1-139} with an N-terminal Hisx6-SUMO tag was expressed from pOPINS3C expression vector (Bird, 2011).

Full-length Seb1 Full-length Seb1 with a C-terminal Hisx8-tag was cloned into pET41a(+) by restriction enzyme digestion by Dr Sina Wittmann (Wittmann et al., 2017).

Spt5 and Spt4. The full-length *S.pombe* genes encoding Spt4 and Spt5 (WT or T1E) were cloned into a single pRSFDuet plasmid by Gibson assembly (Gibson et al., 2009). Spt5 contained an N-terminal thioredoxin-Hisx8-tag. Gene-specific primers were used to amplify Spt5 and Spt4 from *S.pombe* cDNA for WT proteins or from a synthesized vector by Eurofins Genomics for Spt5 T1E (see section 5.2). The primers introduced a second-ribosomal binding site between Spt5 and Spt4. A fragment encoding for the thioredoxin-Hisx8-tag was amplified from pET-TRX1a to be introduced upstream Spt5.

All the DNA fragments (thioredoxin-tag, Spt5, Spt4) had 20 pb overlapping ends and were cloned into the pRSFDuet vector digested with PacI/NcoI by Gibson Assembly (NEB Kit).

6.8 Transformation of *E.coli* cells

For cloning, chemically competent *E.coli* TOP10 were used. For protein expression, chemically competent BL21(DE3) with or without the Rosetta plasmid were transformed. 10 μl of a ligation mixture/Gibson assembly mixture (for cloning) or 25 ng of a purified vector (for protein expression) was used to transform 50 μl of cells. Cells were then incubated on ice for 30 min and subjected to heat shock at 42 °C for exactly 30 sec. Incubation in 250 μl of SOC medium at 37 °C for 1 h was followed by selection on LB-agar plates supplied with the appropriate bacterial selection marker.

6.9 Protein expression and purification from *E.coli*

CID-containing proteins. Purifications of *S. cervisiae* CIDs were done by Ms Katie Duckett (PartII student, Vasilieva Lab). Purification of *S. pombe* Pcf11 was done by Dr Krzysztof Kus (postdoctoral researcher, Vasilieva lab). *S.cerevisiae* Pcf11 was purified as in Mayer et al. (2012). *S.cerevisiae* Nrd1, Rtt103 and *S.pombe* Seb1, Pcf11 were purified as in Wittmann et al. (2017). Briefly, after transformation, *E.coli* Rosetta cells were grown in Terrific Broth (12 g/l tryptone, 24 g/l yeast extract, 3.5% (v/v) glycerol, 2.3 g/l KH2PO4, 12.3 g/l K2HPO4, 1% (w/v) glucose) at 37 °C to $OD_{600} \sim 0.8$ before protein expression with 1mM IPTG at 18 °C for 18h. Cells were harvested, resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 5 mM imidazole) and lysed by French press. After addition of 600 mM NaCl, 0.1% Tween-20 and 1 mM PMSF, the lysate was centrifuged at 100,000g for 1h at 4 °C. The supernatant was passed through a 45 μm

filter, then run through NiNTA agarose resin (Qiagen). Bound protein was eluted by addition of increasing amounts of imidazole (40 mM, 100 mM, 250 mM and 500 mM). Protein-containing fractions were pooled together and subjected to size-exclusion chromatography using a HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) on an AKTA purifier (GE Healthcare). *S. pombe* Pcf11 was subjected to size-exclusion chromatography using a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare) to eliminate soluble but high molecular species. For *S. cerevisiae* Pcf11, Ni-NTA elution fractions were loaded on a MonoQ anion exchange column (Amersham Bioscience) prior to size-exclusion chromatography.

Full-length Seb1 Full-length Seb1 was purified by Dr Krzysztof Kus (postdoctoral researcher, Vasilieva lab) as above with the following modifications. Protein expression was induced with 0.5 mM IPTG for 4h at 37 °C. After binding to the NiNTA resin, protein was eluted with 500 mM imidazole. Seb1-containing fractions were pooled together and run through a 10/300 Superdex 200 column (GE Healthcare). Fractions containing monomeric Seb1 were stored at -80 °C until usage.

Spt5 and Spt4. Spt4/5 purifcation was done by Dr Krzysztof Kus (postdoctoral researcher, Vasilieva lab). The same protocol was used to purify Spt5 WT and T1E. After transformation, *Escherichia coli* BL21 were grown at 37 °C to $OD_{600} \sim 0.8$ and induced with 1mM IPTG for 3h at 37 °C. Harvested cells were resuspended in buffer NiTA (50 mM Tris-HCl pH 7.8, 500 mM NaCl, 5 mM imidazole, 1 mM 2-mercaptoethanol, 2500 U of SuperNuclease (Sino Biological Inc.) and lysed by French Press. The lysate was then supplemented with 0.5% Tween-20 and 1 mM PMSF. After centirfugation at 30,000g for 20 min at 4 °C, cleared lysate was subjected to affinity chromatography using Ni-NTA resin. The resin was washed with 5 ml of buffer NiTA and 50 ml buffer NiTC (buffer NiTA with 1M NaCl and 1M urea). Proteins were eluted with 200 mM imidazole, then subjected to ion exchange chromatography (Q HP, GE Healthcare). Bound protein was

washed with 8% of buffer Q2 (buffer Q1 with 2M NaCl), before elution with a gradient of salt (up to 30% of Q2). Fractions containing Spt4/5 were pooled together and subjected to size-exclusion chromatography using a Superdex 200 10/300 (GE Healthcare) equilibrated with buffer GF (20 mM HEPES pH 7.5, 150 mM NaCl,0.5 mM $MgCl_2$, 0.5 mM $Mg(OAc)_2$, 1 mM 2-mercaptoethanol).

6.10 Fluorescent anisotropy (FA) assay

FA assays for all CID-containing proteins, except Seb1 and *S. pombe* Pcf11, were done by Ms Katie Duckett (PartII student, Vasilieva Lab). For the fluorescent anisotropy assay, binding to 75 nM of two-repeat CTD peptides containing an N-terminal 5'-fluorescein amidite (FAM)-tag (Peptides&elephants, Potsdam, Germany) was determined after incubation at room temperature for 20 min as in Wittmann et al. (2017). For *S.c.* Rtt103, *S.c.*Nrd1 and *S.p.* Pcf11 measurements were performed in the presence of 150 mM NaCl. For *S.p.* Seb1 measurements were done in the presence of 200 mM NaCl, and for *S.c.* Pcf11 in the presence of 50 mM NaCl. Excitation of the ligand was performed with linearly polarized light at 485 nm and emission was measured both parallel and perpendicular to plane at 520 nm at 25 °C using a FLUOstar-Omega microplate reader (BMG-Labtech). All measurements were performed in duplicates unless otherwise stated, results were plotted against the protein concentration and K_d values were determined via curve fitting as described in Reinstein et al. (1990).

6.11 Electrophoretic mobility shift assay (EMSA)

A FAM-labelled RNA of the sequence AUUAGUAAAAUAUAUGCAUAAAGACCAGGC (obtained from Eurofins) was used for EMSAs. The RNA was heated for 5 min at 95 °C prior incubation with proteins. 0, 3.5 pmol, 8.75 pmol, 17.5 pmol or 38.5 pmol of re-

combinant Spt5 WT/T1E was incubated with 5 pmol of the above RNA in buffer EB (20 mM Hepes pH 7.5, 30 mM NaCl, 1mM $MgCl_2$, 0.01% NP-40, 5% glycerol) for 40 min at room temperature. Samples were then analysed on a pre-run 4-20% TBE gel (Invitrogen) in 0.5x TBE buffer at 140 V at 4 °C. Visualisation of the fluorescent signal was achieved with an Image Analyser (FLA-600, Fujifilm, Japan).

6.12 SDS-PAGE

Proteins were analyzed by SDS denaturating polyacrylamide gel electrophoresis (SDS-PAGE). After addition of Lammli buffer (Laemmli, 1970), samples were heated at 95 °C for 5 min. Gels were run at 100 V for 15 min, followed by 175 V until the bromophenol blue tracking dye reached the bottom of the gel. Gels were stained with InstantBlue (Expedeon) or SilverQuest Silver Staining Kit (Invitrogen) according to manufacturer's instructions.

6.13 Immunoblotting

To detect proteins by Western blot, after SDS-PAGE, proteins were transferred onto an Immobilion-P polyvinylidene fluoride (PVDF) membrane via wet transfer. Protein transfer was performed at 25V for 18h at 4 °C in transfer buffer (25 mM Tris, 192 mM glycine, 1 mM EDTA and 20% (v/v) methanol). Membranes were blocked with TBST (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween-20) with 5% (w/v) skim milk powder at room temperature for 30 min, then incubated with a primary antibody or a horse radish peroxidase (HRP)-coupled antibody for 1h at room temperature. The membrane was then washed three times for 5 min. If required, membranes were incubated with a secondary antibody as above. After addition of Clarity Western ECL (Bio-Rad), membranes were exposed to Amersham Hyperfilm ECL (GE Healthcare)

and developed using a SRX-101A developer (Konica Minolta). Table 6.4 lists all the antibody used in this study.

6.14 ChIP-sequencing

All ChIP-seq experiments were performed in duplicate unless otherwise stated. Chromatin was prepared as previously described (Shah et al. (2014)). For spike-in control, *S. cerevisiae* cells were added to *S. pombe* cells before cross-linking. S. *cerevisiae* cells correspond to 15% of *S. pombe* cells. Briefly, 140 OD of cells at OD $\sim$ 0.6 were cross-linked with 11% formaldehyde for 20 min at room temperature. The reaction was then quenched with 3M glycine, 20 mM Tris for 5 min. Cells were collected by centrifugation, washed once with TBS and once with FA buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% TritonX-100) containing 0.1%SDS. Cells were lysed in FA buffer/0.5% SDS by vortexing thirty times for 30 s at max speed. Between each round of vortexing, samples were put on ice for 30 s. After centrifugation at 200,000 g for 20 min, the pellet was washed with FA buffer/0.1%SDS. Next, the chromatin was sheared by 15 s on/45 s off sonication cycles on high power for 80 min using a Biorupter sonicator (Diagenode). After centrifugation at 25,000 g for 20 min, the sheared chromatin was in the supernatant. NaCl concentration was adjusted to 275 mM, except when immunoprecipitating with Thr4P antibody, for which all buffers had NaCl concentration at 150 mM. Immunoprecipitations were conducted overnight at 4 °C with IgG agarose beads (Sigma, 60 μl per sample) or antibodies against Rpb1 (8WG16, Millipore)/phosphorylated CTD (see Table 6.4) coupled to protein G dynabeads (Life Technologies, per sample: 60 μl of beads conjugated with 15 μl of anitibody). After eluting bound material from the beads, crosslinks were reverted by incubation with 0.4 mg Pronase for 1 hr at 4 °C, followed by overnight incubation at 65 °C. RNA was degraded by incubating samples with 0.02 mg RNAse A (Roche) for 1 hr at 37 °C. DNA was purified using ChIP DNA

Clean & Concentrator kit (Zymo Research, USA). Libraries were constructed using NEBNext Fast DNA Library Prep Set for Ion TorrentTM Kit (for early ChIP-seq) or NEBNext Ultra DNA Library Prep Kit for Illumina (for late ChIP-seq). Libraries were sequenced on the Ion Torrent Proton or the Illumina NextSeq500.

Reads shorter than 20 nt and with average quality per base below 20 were discarded. Reads were aligned to a concatenated genome (*S. pombe* + *S. cerevisiae*) with Bowtie2 Langmead and Salzberg (2012). Reads that mapped more than once were discarded and PCR duplicates were removed using SAMtools Li et al. (2009). To assess changes in protein occupancy between strains, a spike-in normalization was used where *S. pombe* reads were adjusted using the number of *S. cerevisiae* reads in each IP sample. Genomic DNA sequencing of the input mixture of fission and budding yeast was also used to correct for any variation in cell mixture ratios. For metagene profiles and single gene analysis, *S. pombe* genome annotation from (Eser et al., 2016) was used. The analysis was done over a subset of 1735 protein-coding genes unless otherwise stated. These 1735 genes have a distance greater than 275 nt from the TSS of the downstream gene (on the same strand) and do not overlap with any other transcription unit (on the opposite strand) in a region from 250 nt upstream to 500 nt downstream of the PAS.

6.15 Yeast strains used in this study

Table 6.1: Yeast strains used in this study

Name	Organism	Genotype	Reference
YP144	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1	Lemieux et al. (2011)
YP602	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1; Rpb9-3xFLAG::kanMX; Rpb1-TEV-CTD (a TEV cleavage site was inserted upstream the CTD)	This study

YP606	*S.pombe*	h-, leu1; ura4; dis2::ura4+	National BioResource Project (NBRP), Japan
YP611	*S.pombe*	h-; leu1; dis2-11	National BioResource Project (NBRP), Japan
YP637	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Rpb9-3xFLAG::kanMX	This study
YP732	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1; dis2::ura4+	This study
YP781	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1; Seb1-3XHA-TAP::kanMX	This study
YP782	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1; dis2::ura4+; Seb1-3XHA-TAP::kanMX	This study
YP783	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1; Pcf11-TAP::kanMX	This study
YP784	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1; dis2::ura4+; Pcf11-TAP::kanMX	This study
YP822	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1; Rhn1-3XHA-TAP::kanMX	This study
YP823	*S.pombe*	h+, leu1-32, ura4-Δ18, ade6-M216; his3Δ::1; dis2::ura4+; Rhn1-3XHA-TAP::kanMX	This study
YP839	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Seb1-3xFlag::natMX	This study
YP840	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; dis2::ura4+; Seb1-3xFlag::natMX	This study
YP910	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Seb1-3XHA-TAP::kanMX; Spt5-wt-3xFlag::natMX	This study
YP911	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Seb1-3XHA-TAP::kanMX; Spt5-T1E-3xFlag::natMX	This study
YP912	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Seb1-3XHA-TAP::kanMX; Spt5-T1A-3xFlag::natMX	This study

YP913	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; dis2::ura4+; Seb1-3XHA-TAP::kanMX; Spt5-wt-3xFlag::natMX	This study
YP914	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; dis2::ura4+; Seb1-3XHA-TAP::kanMX; Spt5-T1E-3xFlag::natMX	This study
YP915	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; dis2::ura4+; Seb1-3XHA-TAP::kanMX; Spt5-T1A-3xFlag::natMX	This study
YP956	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Spt5-wt-3xFlag::natMX	This study
YP957	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Spt5-T1A-3xFlag::natMX	This study
YP958	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; Spt5-T1E-3xFlag::natMX	This study
YP959	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; dis2::ura4+ Spt5-wt-3xFlag::natMX	This study
YP960	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; dis2::ura4+; Spt5-T1A-3xFlag::natMX	This study
YP961	*S.pombe*	h+; leu1-32; ura4-Δ18; ade6-M216; his3Δ::1; dis2::ura4+; Spt5-T1E-3xFlag::natMX	This study
YP1209	*S.pombe*	h-/h+; ade6-M210/ade6-M216; arg3-Δ4/arg3-Δ4; ura4-Δ18/ura4-Δ18; his3-Δ1/his3-Δ1; rpb1-wt/rpb1-T4A-KanMX6	Adrien Birot (Vasilieva lab)
YP1083	*S.pombe*	h-/h+; ade6-M210/ade6-M216; arg3-Δ4/arg3-Δ4; ura4-Δ18/ura4-Δ18; his3-Δ1/his3-Δ1; rpb1-wt/rpb1-T4E-KanMX6	Adrien Birot (Vasilieva lab)
YF336	*S.cerevisae*	MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0	Kim et al. (2004b)
YF1429	*S.cerevisae*	MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; Pcf11-TAP::His3	Ghaemmaghami et al. (2003)

6.16 Plasmids used in this study

Table 6.2: Plasmids used in this study

Name	Description	Purpose	Reference
3932	pET41a(+)-Seb1(full-length)-His8	recombinant expression	Sina Wittmann (Vasilieva lab)
3933	pET41a(+)-SpSeb1-His8 CID(1-148)	recombinant expression	Sina Wittmann (Vasilieva lab)
4097	pET41a(+)-ScPcf11-His8 CID(1-144)	recombinant expression	Katie Duckett (Vasilieva lab)
4098	pET41a(+)-ScRtt103-His8 CID(1-162)	recombinant expression	Katie Duckett (Vasilieva lab)
4110	pOPINS3C-His6-SpPcf11 CID(1-139)	recombinant expression	Katie Duckett (Vasilieva lab)
4156	pEX-A128-Spt5-T1A	cloning	Eurofins Genomics
4157	pEX-A128-Spt5-T1A	cloning	Eurofins Genomics
4164	pRSFDuet-Trx-His8-Spt5wt-Spt4	recombinant expression	This study
4165	pRSFDuet-Trx-His8-Spt5(T1E)-Spt4	recombinant expression	This study

6.17 Oligonucleotides used in this study

Table 6.3: Oligonucelotides used in this study

N.	Name	Sequence	Purpose
4933	Rpb1-for-C	GAATTTGTATTTTCAGGGTggggta	cloning
4934	Rpb1-rev-XbaI	AACTTTtctagaCGAGAACTTAGACACTA-CACTTC	cloning
4936	Rpb1A-TEV-NotI	GACATAgcggccgcGCGGCAAATTGAAGA-TGGTAC	cloning

5057	Rpb1-E-rev	ACCCTGAAAATACAAATTCTCTTTATA-AGGACTAGCAGCCCCCAA	cloning
NA	Rpb9-3xFlag-L3	NA	cloning
4953	Rpb9-3xFlag-L4	gtctccacccccgcctccccCTGTTCTTCAAAA-GCAAATC	cloning
4954	Rpb9-3xFlag-L4	gtttaaacgagctcgaattcTATGCGACAATTCG-GTCTAA	cloning
NA	Rpb9-3xFlag-L6	NA	cloning
4980	Dis2_rev	TAGTTCAAACAACTATTGCTCGGT	cloning
5058	Dis2-L3	TGATCCTGACAAGGATCTTACAG	cloning
5059	Dis2-L4-3xFlag	gtctccacccccgcctccccCTTTGAATTTCCTG-TCTTATTCTTCCG	cloning
5060	Dis2-L5-3xFlag	TGATCCTGACAAGGATCTTACAG	cloning
5312	Dis2-afterRQ-for	GTTGTGGAAGACGGTTATGAG	cloning
5447	Pcf11-KpnI-for	TAATTAggtaccATGGATTTAGTGGAAT-TGGATTACTTAAGTG	cloning
5448	Pcf11-PvuII-139-rev	TTATATcagctgttaTGATTTGTATTTTAG-TAATGCGTTTTC	cloning
5494	Pcf11-CID-NdeI-for	TAATTAcatatgGATCACGACACAGAAGTT	cloning
5495	Pcf11-CID-NotI-140-r	TATATAgcggccgcAGCACTGGCCTTTAT-TAAAAACTG	cloning
5496	Rtt103-CID-NdeI-f	TAATTAcatatgCCTTTCTCTTCTGAGC-AAT	cloning
5497	Rtt103-CID-XhoI-131-r	TATATActcgagAAGACTTCTTTCTATG-TCATTGACTAC	cloning
6156	TEV-fwd	ctttaataaggagatataccATGAAACATCACC-ATCACC	cloning
6157	TEV-rev	tcgtatccatGGCGCCCTGAAAATAAAG	cloning
6158	spt5-aa1-fwd	tcagggcgccATGGATACGAATTCTCCG	cloning
6159	spt5-wt-rev	atttatatctccttcttaaagttaaacTTAAGGAGA-AGGAGGTACATATTTG	cloning

6160	spt4-fwd	gtttaactttaagaaggagatataaatATGGATAA-GCTAAATAGGACTAGATC	cloning
6161	spt4-rev	gtggcagcagcctaggttaaCTAATCCCATGA-TGTACCATTC	cloning
6163	Spt5-Nend-rev	taacgccaagTAATCCTTTCATAGGACC-TCTG	cloning
6164	CTR-fwd	gaaaggattaCTTGGCGTTATTAAGGATAC	cloning
6166	Spt5-T1E-rev	atttatatctccttcttaaagttaaacTTAAGGAGA-TGGCGGTAC	cloning

6.18 Antibodies used in this study

Table 6.4: Antibodies used in this study

Antibody	**Origin**	**Application**	**Source**
RNAPII (8WG16)	mouse	ChIP (15 μl per sample), WB (1:1,000)	Millipore
α-Ser2P(3E10)	rat	ChIP (15 μl per sample), WB (1:1,000)	Millipore
α-Ser5P(3E8)	rat	ChIP (15 μl per sample), WB (1:1,000)	Millipore
α-Ser7P(4E12)	rat	WB (1:1,000)	Active Motif
α-Tyr1P(3D12)	rat	ChIP (15 μl per sample), WB (1:1,000)	Active Motif
α-Thr4P(6D7)	rat IgG2b	ChIP (15 μl per sample), WB (1:1,000)	Active Motif
α-Flag-HRP	mouse	WB (1:5,000)	Sigma-Aldrich
α-His_6-HRP	mouse	WB (1:5,000)	Clontech
α-mouse	goat	WB (1:10,000)	Sigma-Aldrich
α-rat	goat	WB (1:10,000)	Sigma-Aldrich
α-rat IgG2b	goat	WB (1:10,000)	NA

Bibliography

Allen, B. L. and Taatjes, D. J. (2015). The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology*, 16(3):155–166.

Allepuz-Fuster, P., Martínez-Fernández, V., Garrido-Godino, A. I., Alonso-Aguado, S., Hanes, S. D., Navarro, F., and Calvo, O. (2014). Rpb4/7 facilitates RNA polymerase II CTD dephosphorylation. *Nucleic acids research*, 42(22):13674–88.

Arigo, J. T., Eyler, D. E., Carroll, K. L., and Corden, J. L. (2006). Termination of Cryptic Unstable Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. *Molecular Cell*, 23(6):841–851.

Austenaa, L., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B., de Pretis, S., et al. (2015). Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Molecular Cell*, 60(3):460–474.

Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K. C., Boltendahl, A., Rus, P., Esslinger, S., et al. (2017). Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Molecular Cell*, 66(1):38–49.e6.

Bähler, J., Wu, J.-Q., Longtine, M. S., Shah, N. G., Mckenzie III, A., Steever, A. B., Wach, A., Philippsen, P., and Pringle, J. R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting inSchizosaccharomyces pombe. *Yeast*, 14(10):943–951.

Barilla, D., Lee, B. A., and Proudfoot, N. J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 98(2):445–450.

Baskaran, R., Dahmus, M. E., and Wang, J. Y. (1993). Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America*, 90(23):11167–71.

Bataille, A. R., Jeronimo, C. L., Jacques, P.-É. T., Laramé, L., Fortin, M.-V., Forest, A., Bergeron, M., Hanes, S. D., and Robert, F. O. (2012). A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. *MOLCEL*, 45:158–170.

Becker, R., Loll, B., and Meinhart, A. (2008). Snapshots of the RNA Processing Factor SCAF8 Bound to Different Phosphorylated Forms of the Carboxyl-terminal Domain of RNA Polymerase II. *Journal of Biological Chemistry*, 283(33):22659–22669.

Bernecky, C., Plitzko, J. M., and Cramer, P. (2017). Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nature Structural & Molecular Biology*, 24(10):809–815.

Bird, L. E. (2011). High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. *Methods*, 55(1):29–37.

Bonhoure, N., Bounova, G., Bernasconi, D., Praz, V., Lammers, F., Canella, D., Willis, I. M., Herr, W., Hernandez, N., Delorenzi, M., et al. (2014). Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome research*, 24(7):1157–68.

Booth, G. T., Parua, P. K., Sansó, M., Fisher, R. P., and Lis, J. T. (2018). Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. *Nature Communications*, 9(1):543.

Booth, G. T., Wang, I. X., Cheung, V. G., and Lis, J. T. (2016). Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome research*, 26(6):799–811.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K. J., Carroll, T., Xie, S. Q., Stock, J. K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell stem cell*, 10(2):157–70.

Buratowski, S. (2003). The CTD code. *Nature Structural & Molecular Biology*, 10(9):679–680.

Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E., and Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science (New York, N.Y.)*, 320(5878):935–8.

Carpy, A., Krug, K., Graf, S., Koch, A., Popic, S., Hauf, S., and Macek, B. (2014). Absolute proteome and phosphoproteome dynamics during the cell cycle of fission yeast. *Molecular & cellular proteomics : MCP*, pages 1925–1936.

Casañal, A., Kumar, A., Hill, C. H., Easter, A. D., Emsley, P., Degliesposti, G., Gordiyenko, Y., Santhanam, B., Wolf, J., Wiederhold, K., et al. (2017). Architecture of eukaryotic mRNA 3'-end processing machinery. *Science (New York, N.Y.)*, 358(6366):1056–1059.

Chan, S., Choi, E.-A., and Shi, Y. (2011). Pre-mRNA 3'-end processing complex assembly and function. *Wiley interdisciplinary reviews. RNA*, 2(3):321–35.

Chen, Y., Yamaguchi, Y., Tsugeno, Y., Yamamoto, J., Yamada, T., Nakamura, M., Hisatake, K., and Handa, H. (2009). DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes & development*, 23(23):2765–77.

Cheng, B. and Price, D. H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *The Journal of biological chemistry*, 282(30):21901–12.

Cheng, B. and Price, D. H. (2008). Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Research*, 36(20):e135–e135.

Churchman, L. S. and Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution.

Churchman, L. S., Weissman, J. S., Churchman, L. S., and Weissman, J. S. (2012). Native Elongating Transcript Sequencing (NET-seq). In *Current Protocols in Molecular Biology*, pages 14.4.1–14.4.17. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Ciurciu, A., Duncalf, L., Jonchere, V., Lansdale, N., Vasieva, O., Glenday, P., Rudenko, A., Vissi, E., Cobbe, N., Alphey, L., et al. (2013). PNUTS/PP1 Regulates RNAPII-Mediated Gene Expression and Is Necessary for Developmental Growth. *PLoS Genetics*, 9(10):e1003885.

Cohen, P. T. W. (2002). Protein phosphatase 1–targeted in many directions. *Journal of cell science*, 115(Pt 2):241–56.

Corden, J. L. (2013). RNA polymerase II C-terminal domain: Tethering transcription to transcript and template.

Coudreuse, D., Van Bakel, H., Dewez, M., Soutourina, J., Parnell, T., Vandenhaute, J., Cairns, B., Werner, M., and Hermand, D. (2010). A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. *Current Biology*, 20(12):1053–1064.

Craft, J. W. and Legge, G. B. (2005). An AMBER/DYANA/MOLMOL Phosphorylated Amino Acid Library Set and Incorporation into NMR Structure Calculations. *Journal of Biomolecular NMR*, 33(1):15–24.

Dengl, S. and Cramer, P. (2009). Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro. *Journal of Biological Chemistry*, 284(32):21270–21279.

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M. A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife*, 3:e02105.

Doamekpor, S. K., Sanchez, A. M., Schwer, B., Shuman, S., and Lima, C. D. (2014). How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes & development*, 28(12):1323–36.

Doamekpor, S. K., Schwer, B., Sanchez, A. M., Shuman, S., and Lima, C. D. (2015). Fission yeast RNA triphosphatase reads an Spt5 CTD code. *RNA (New York, N.Y.)*, 21(1):113–23.

Eaton, J. D., Davidson, L., Bauer, D. L. V., Natsume, T., Kanemaki, M. T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity.

Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., and Sekine, S.-i. (2017). Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science*, 357(6354):921–924.

Eick, D. and Geyer, M. (2013). The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chemical Reviews*, 113(11):8456–8490.

Eser, P., Wachutka, L., Maier, K. C., Demel, C., Boroni, M., Iyer, S., Cramer, P., and Gagneur, J. (2016). Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. *Molecular systems biology*, 12(2):857.

Fahrner, K., Yarger, J., and Hereford, L. (1980). Yeast histone mRNA is polyadenylated. *Nucleic Acids Research*, 8(23):5725–5737.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M., Sheridan, R., Nguyen, T., Karp, S., and Bentley, D. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Molecular Cell*, 60(2):256–267.

Fuda, N. J., Buckley, M. S., Wei, W., Core, L. J., Waters, C. T., Reinberg, D., and Lis, J. T. (2012). Fcp1 dephosphorylation of the RNA polymerase II C-terminal domain is required for efficient transcription of heat shock genes. *Molecular and cellular biology*, 32(17):3428–37.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741.

Ghazy, M. A., Gordon, J. M. B., Lee, S. D., Singh, B. N., Bohm, A., Hampsey, M., and 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 research*, 40(3):1214–25.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345.

Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R. H., Michaelis, J., and Werner, F. (2011). The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Molecular cell*, 43(2):263–74.

Guo, J. and Price, D. H. (2013). RNA Polymerase II Transcription Elongation Control. *Chemical Reviews*, 113(11):8583–8603.

Haag, J. R. and Pikaard, C. S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nature Reviews Molecular Cell Biology*, 12(8):483–492.

Harlen, K., Trotta, K., Smith, E., Mosaheb, M., Fuchs, S., and Churchman, L. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Reports*, 15(10):2147–2158.

Hartzog, G. A. and Fu, J. (2013). The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochimica et biophysica acta*, 1829(1):105–15.

He, X. and Moore, C. (2005). Regulation of yeast mRNA 3' end processing by phosphorylation. *Molecular cell*, 19(5):619–29.

He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., and Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. *Nature*, 533(7603):359–365.

Heidemann, M., Hintermair, C., Voß, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(1):55–62.

Heo, D.-h., Yoo, I., Kong, J., Lidschreiber, M., Mayer, A., Choi, B.-Y., Hahn, Y., Cramer, P., Buratowski, S., and Kim, M. (2013). The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome. *The Journal of biological chemistry*, 288(51):36676–90.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., et al. (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *The EMBO journal*, 31(12):2784–97.

Hintermair, C., Voß, K., Forné, I., Heidemann, M., Flatley, A., Kremmer, E., Imhof, A., and Eick, D. (2016). Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression. *Scientific Reports 2016 6*, 6(1):27401.

Hollingworth, D., Noble, C. G., Taylor, I. A., and Ramos, A. (2006). RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. *RNA (New York, N.Y.)*, 12(4):555–60.

Hsin, J.-P., Sheth, A., and Manley, J. L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science (New York, N.Y.)*, 334(6056):683–6.

Hsu, P. L., Yang, F., Smith-Kinnaman, W., Yang, W., Song, J.-E., Mosley, A. L., and Varani, G. (2014). Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD. *Journal of molecular biology*, 426(16):2970–81.

Hu, B., Petela, N., Kurze, A., Chan, K.-L., Chapard, C., and Nasmyth, K. (2015). Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Research*, 43(20):gkv670.

Jasnovidova, O., Krejcikova, M., Kubicek, K., and Stefl, R. (2017). Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO reports*, 18(6):906–913.

Jeronimo, C., Collin, P., and Robert, F. (2016). The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. *Journal of Molecular Biology*.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J. H., Pollock, D. D., Megee, P. C., and Bentley, D. L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature structural & molecular biology*, 17(10):1279–86.

Kim, M., Ahn, S.-H., Krogan, N. J., Greenblatt, J. F., and Buratowski, S. (2004a). Transitions in RNA polymerase II elongation complexes at the 3 0 ends of genes. *The EMBO Journal*, 23:354–364.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. (2004b). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432(7016):517–522.

Kim, M., Vasiljeva, L., Rando, O. J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct Pathways for snoRNA and mRNA Termination. *Molecular Cell*, 24(5):723–734.

Kim, Y.-J., Björklund, S., Li, Y., Sayre, M. H., and Kornberg, R. D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, 77(4):599–608.

Kinoshita, N., Ohkura, H., and Yanagida, M. (1990). Distinct, essential roles of type 1 and 2A protein phosphatases in the control of the fission yeast cell division cycle. *Cell*, 63(2):405–15.

Kwak, H. and Lis, J. T. (2013). Control of Transcriptional Elongation. *Annual Review of Genetics*, 47(1):483–508.

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

Laitem, C., Zaborowska, J., Isa, N. F., Kufs, J., Dienstbier, M., and Murphy, S. (2015). CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II–transcribed genes. *Nature Structural & Molecular Biology*, 22(5):396–403.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359.

Lawson, M. R., Ma, W., Bellecourt, M. J., Landick, R., Schulten, K., Berger Correspondence, J. M., Artsimovitch, I., Martin, A., and Berger, J. M. (2018). Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. *Molecular Cell*, 71.

Lemay, J.-F., Marguerat, S., Larochelle, M., Liu, X., van Nues, R., Hunyadkürti, J., Hoque, M., Tian, B., Granneman, S., Bähler, J., et al. (2016). The Nrd1-like protein Seb1 coordinates cotranscriptional 3' end processing and polyadenylation site selection. *Genes & development*, 30(13):1558–72.

Lemieux, C., Marguerat, S., Lafontaine, J., Barbezier, N., Bähler, J., and Bachand, F. (2011). A Pre-mRNA Degradation Pathway that Selectively Targets Intron-Containing Genes Requires the Nuclear Poly(A)-Binding Protein. *Molecular Cell*, 44(1):108–119.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Licatalosi, D. D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B., and Bentley, D. L. (2002). Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Molecular cell*, 9(5):1101–11.

Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H., Ranish, J., Shokat, K. M., and Hahn, S. (2009). Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Molecular and cellular biology*, 29(17):4852–63.

Loya, T. J. and Reines, D. (2016). Recent advances in understanding transcription termination by RNA polymerase II. *F1000Research*, 5.

Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nature Structural & Molecular Biology*, 17(10):1195–1201.

Luo, J. and Hall, B. D. (2007). A Multistep Process Gave Rise to RNA Polymerase IV of Land Plants. *Journal of Molecular Evolution*, 64(1):101–112.

Luo, W., Johnson, A. W., and Bentley, D. L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes & development*, 20(8):954–65.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., and Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Science*, 15(1):182–189.

Mata, J. (2013). Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA biology*, 10(8):1407–14.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. *Science*, 336(6089):1723–1725.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nature Structural & Molecular Biology*, 17(10):1272–1278.

Meinhart, A. and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, 430(6996):223–226.

Milligan, L., Huynh-Thu, V. A., Delan-Forino, C., Tuck, A., Petfalski, E., Lombraña, R., Sanguinetti, G., Kudla, G., and Tollervey, D. (2016). Strand-specific, high-resolution mapping of modified RNA polymerase II. *Molecular systems biology*, 12(6):874.

Mischo, H. E. and Proudfoot, N. J. (2013). Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(1):174–185.

Missra, A. and Gilmour, D. S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25):11301–6.

Moorhead, G. B. G., Trinkle-Mulcahy, L., and Ulke-Lemée, A. (2007). Emerging roles of nuclear protein phosphatases. *Nature Reviews Molecular Cell Biology*, 8(3):234–244.

Nedea, E., Nalbant, D., Xia, D., Theoharis, N. T., Suter, B., Richardson, C. J., Tatchell, K., Kislinger, T., Greenblatt, J. F., and Nagy, P. L. (2008). The Glc7 Phosphatase Subunit of the Cleavage and Polyadenylation Factor Is Essential for Transcription Termination on snoRNA Genes. *Molecular Cell*, 29(5):577–587.

Nemec, C. M., Yang, F., Gilmore, J. M., Hintermair, C., Ho, Y.-H., Tseng, S. C., Heidemann, M., Zhang, Y., Florens, L., Gasch, A. P., et al. (2017). Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner. *Proceedings of the National Academy of Sciences of the United States of America*, 114(20):E3944–E3953.

Ni, Z., Xu, C., Guo, X., Hunter, G. O., Kuznetsova, O. V., Tempel, W., Marcon, E., Zhong, G., Guo, H., Kuo, W.-H. W., et al. (2014). RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nature Structural & Molecular Biology*, 21(8):686–695.

Noble, C. G., Hollingworth, D., Martin, S. R., Ennis-Adeniran, V., Smerdon, S. J., Kelly, G., Taylor, I. A., and Ramos, A. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nature Structural & Molecular Biology*, 12(2):144–151.

Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N. J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*, 161(3):526–540.

Ohkura, H., Kinoshita, N., Miyatani, S., Toda, T., and Yanagida, M. (1989). The fission yeast dis2+ gene required for chromosome disjoining encodes one of two putative type 1 protein phosphatases. *Cell*, 57(6):997–1007.

Orlando, D., Chen, M., Brown, V., Solanki, S., Choi, Y., Olson, E., Fritz, C., Bradner, J., and Guenther, M. (2014). Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome. *Cell Reports*, 9(3):1163–1170.

Osheim, Y. N., Sikes, M. L., and Beyer, A. L. (2002). EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3' end cleavage of nascent transcripts. *Chromosoma*, 111(1):1–12.

Parua, P. K., Booth, G. T., Sansó, M., Benjamin, B., Tanny, J. C., Lis, J. T., and Fisher, R. P. (2018). A Cdk9–PP1 switch regulates the elongation–termination transition of RNA polymerase II. *Nature*, 558(7710):460–464.

Pei, Y. and Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *The Journal of biological chemistry*, 277(22):19639–48.

Pei, Y. and Shuman, S. (2003). Characterization of the Schizosaccharomyces pombe Cdk9/Pch1 protein kinase: Spt5 phosphorylation, autophosphorylation, and mutational analysis. *The Journal of biological chemistry*, 278(44):43346–56.

Peters, J. M., Mooney, R. A., Grass, J. A., Jessen, E. D., Tran, F., and Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes & development*, 26(23):2621–33.

Pineda, G., Shen, Z., de Albuquerque, C. P., Reynoso, E., Chen, J., Tu, C.-C., Tang, W., Briggs, S., Zhou, H., and Wang, J. Y. J. (2015). Proteomics studies of the interactome of RNA polymerase II C-terminal repeated domain. *BMC Research Notes*, 8(1):616.

Porrua, O. and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nature Reviews Molecular Cell Biology*, 16(3):190–202.

Qiu, Y. and Gilmour, D. S. (2017). Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing. *The Journal of biological chemistry*, 292(13):5555–5570.

Reinstein, J., Vetter, I. R., Schlichting, I., Rösch, P., Wittinghofer, A., and Goody, R. S. (1990). Fluorescence and NMR investigations on the ligand binding properties of adenylate kinases. *Biochemistry*, 29(32):7440–50.

Rodríguez-Molina, J., Tseng, S., Simonett, S., Taunton, J., and Ansari, A. (2016). Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. *Molecular Cell*, 63(3):433–444.

Roguev, A., Shevchenko, A., Schaft, D., Thomas, H., Stewart, A. F., and Shevchenko, A. (2004). A comparative analysis of an orthologous proteomic environment in the yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Molecular & cellular proteomics : MCP*, 3(2):125–32.

Ryan, K. and Bauer, D. L. (2008). Finishing touches: Post-translational modification of protein factors involved in mammalian pre-mRNA 3' end formation. *The International Journal of Biochemistry & Cell Biology*, 40(11):2384–2396.

Sadowski, M., Dichtl, B., Hübner, W., and Keller, W. (2003). Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *The EMBO journal*, 22(9):2167–77.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16(3):129–143.

Sansó, M., Levin, R. S., Lipp, J. J., Wang, V. Y.-F., Greifenberg, A. K., Quezada, E. M., Ali, A., Ghosh, A., Larochelle, S., Rana, T. M., et al. (2016). P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes & development*, 30(1):117–31.

Schaughency, P., Merran, J., and Corden, J. L. (2014). Genome-Wide Mapping of Yeast RNA Polymerase II Termination. *PLoS Genetics*, 10(10):e1004632.

Schlackow, M., Marguerat, S., Proudfoot, N. J., Bähler, J., Erban, R., and Gullerova, M. (2013). Genome-wide analysis of poly(A) site selection in Schizosaccharomyces pombe. *RNA (New York, N.Y.)*, 19(12):1617–31.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N. J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Molecular cell*, 65(1):25–38.

Schneider, S., Pei, Y., Shuman, S., and Schwer, B. (2010). Separable functions of the fission yeast Spt5 carboxyl-terminal domain (CTD) in capping enzyme binding and transcription elongation overlap with those of the RNA polymerase II CTD. *Molecular and cellular biology*, 30(10):2353–64.

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A. R., Keller, W., Zavolan, M., and Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes & development*, 28(21):2381–93.

Schreieck, A., Easter, A. D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L. A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nature Structural & Molecular Biology*, 21(2):175–179.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Molecular Cell*, 61(2):305–314.

Schulz, S., Gietl, A., Smollett, K., Tinnefeld, P., Werner, F., and Grohmann, D. (2016). TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 113(13):E1816–25.

Schwer, B. and Shuman, S. (2011). Deciphering the RNA polymerase II CTD code in fission yeast. *Molecular cell*, 43(2):311–8.

Shah, N., Maqbool, M. A., Yahia, Y., El Aabidine, A. Z., Esnault, C., Forné, I., Decker, T.-M., Martin, D., Schüller, R., Krebs, S., et al. (2018). Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Molecular cell*, 69(1):48–61.e6.

Shah, S., Wittmann, S., Kilchert, C., and Vasiljeva, L. (2014). lncRNA recruits RNAi and the exosome to dynamically regulate pho1 expression in response to phosphate levels in fission yeast. *Genes & development*, 28(3):231–44.

Shetty, A., Kallgren, S. P., Demel, C., Maier, K. C., Spatt, D., Alver, B. H., Cramer, P., Park, P. J., and Winston, F. (2017). Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Molecular Cell*, 66(1):77–88.e5.

Shi, Y. (2009). Serine/Threonine Phosphatases: Mechanism through Structure. *Cell*, 139(3):468–484.

Shi, Y., Di Giammartino, D. C., Taylor, D., Sarkeshik, A., Rice, W. J., Yates, J. R., Frank, J., and Manley, J. L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. *Molecular cell*, 33(3):365–76.

Sikorski, T. W. and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Current opinion in cell biology*, 21(3):344–51.

Søgaard, T. M. M., Svejstrup, J. Q., and Svejstrup, J. Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *The Journal of biological chemistry*, 282(19):14113–20.

Stachora, A. A., Schäfer, R. E., Pohlmeier, M., Maier, G., and Ponstingl, H. (1997). Human Supt5h protein, a putative modulator of chromatin structure, is reversibly phosphorylated in mitosis. *FEBS Letters*, 409:74–78.

Steinmetz, E. J., Conrad, N. K., Brow, D. A., and Corden, J. L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*, 413(6853):327–331.

Suh, H., Ficarro, S., Kang, U.-B., Chun, Y., Marto, J., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Molecular Cell*, 61(2):297–304.

Sullivan, S. L. and Gottesman, M. E. (1992). Requirement for E. coli NusG Protein in Factor-Dependent Transcription Termination. Technical report.

Swanson, M. S., Malone, E. A., and Winston, F. (1991). SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Molecular and cellular biology*, 11(6):3009–19.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription Termination and Nuclear Degradation of Cryptic Unstable

Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. *Molecular Cell*, 23(6):853–864.

Tietjen, J. R., Zhang, D. W., Rodríguez-Molina, J. B., White, B. E., Akhtar, M. S., Heidemann, M., Li, X., Chapman, R. D., Shokat, K., Keles, S., et al. (2010). Chemical-genomic dissection of the CTD code. *Nature Structural & Molecular Biology*, 17(99):1154–1161.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., et al. (2014). Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Molecular Cell*, 55(3):467–481.

Vanoosthuyse, V., Legros, P., van der Sar, S. J. A., Yvert, G., Toda, K., Le Bihan, T., Watanabe, Y., Hardwick, K., and Bernard, P. (2014). CPF-associated phosphatase activity opposes condensin-mediated chromosome condensation. *PLoS genetics*, 10(6):e1004415.

Vasiljeva, L. and Buratowski, S. (2006). Nrd1 Interacts with the Nuclear Exosome for 3' Processing of RNA Polymerase II Transcripts. *Molecular Cell*, 21(2):239–248.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature Structural & Molecular Biology*, 15(8):795–804.

Viktorovskaya, O. V., Appling, F. D., and Schneider, D. A. (2011). Yeast transcription elongation factor Spt5 associates with RNA polymerase I and RNA polymerase II directly. *The Journal of biological chemistry*, 286(21):18825–33.

Volanakis, A., Kamieniarz-Gdula, K., Schlackow, M., and Proudfoot, N. J. (2017). WNK1 kinase and the termination factor PCF11 connect nuclear mRNA export with transcription. *Genes & development*, 31(21):2175–2185.

Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018a). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*, 560(7720):607–612.

Vos, S. M., Farnung, L., Urlaub, H., and Cramer, P. (2018b). Structure of paused transcription complex Pol II–DSIF–NELF. *Nature*, 560(7720):601–606.

Washington, K., Ammosova, T., Beullens, M., Jerebtsova, M., Kumar, A., Bollen, M., and Nekhai, S. (2002). Protein phosphatase-1 dephosphorylates the C-terminal domain of RNA polymerase-II. *The Journal of biological chemistry*, 277(43):40442–8.

Werner, F. (2012). A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *Journal of molecular biology*, 417(1-2):13–27.

Werner, F. and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology*, 9(2):85–98.

West, S., Gromak, N., and Proudfoot, N. J. (2004). Human 5' -¿ 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016):522–525.

West, S., Proudfoot, N. J., and Dye, M. J. (2008). Molecular Dissection of Mammalian RNA Polymerase II Transcriptional Termination. *Molecular Cell*, 29(5):600–610.

Wittmann, S., Renner, M., Watts, B. R., Adams, O., Huseyin, M., Baejen, C., El Omari, K., Kilchert, C., Heo, D.-H., Kecman, T., et al. (2017). The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. *Nature Communications*, 8:14861.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. *Molecular Cell*, 21(2):227–237.

Yurko, N., Liu, X., Yamazaki, T., Hoque, M., Tian, B., and Manley, J. L. (2017). MPK1/SLT2 Links Multiple Stress Responses with Gene Expression in Budding Yeast by Phosphorylating Tyr1 of the RNAP II CTD. *Molecular Cell*, 68(5):913–925.e3.

Zaborowska, J., Egloff, S., and Murphy, S. (2016). The pol II CTD: new twists in the tail. *Nature Structural & Molecular Biology*, 23(9):771–777.

Zhang, H., Rigo, F., and Martinson, H. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Molecular Cell*, 59(3):437–448.

Zhang, Z., Fu, J., and Gilmour, D. S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes & Development*, 19(13):1572–1580.

Zhang, Z. and 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. *Molecular Cell*, 21(1):65–74.

Zhou, K., Kuo, W. H. W., Fillingham, J., and Greenblatt, J. F. (2009). Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17):6956–61.