Structural and Functional Studies of the Transcriptional Regulator Seb1 in Fission Yeast

A dissertation submitted for the degree of
Doctor of Philosophy



Department of Biochemistry
University of Oxford

Sina Wittmann
Exeter College
Trinity Term 2016

ABSTRACT

Structural and Functional Studies of the Transcriptional Regulator Seb1 in Fission Yeast

Sina Wittmann, Exeter College, University of Oxford
DPhil, Department of Biochemistry
Trinity Term 2016

RNA polymerase II (Pol II) is responsible for the transcription of all protein-coding and some non-coding genes in eukaryotes. Its largest subunit, Rpb1, contains a unique C-terminal domain (CTD) which consists of repeats of the heptad YSPTSPS. It acts as a binding platform for proteins that control the different stages of transcription and their recruitment is regulated mainly by differential phosphorylation of residues contained within the CTD. Previous studies could unveil proteins containing a CTD-interacting domain (CID) as important players that specifically bind to certain phosphorylation types of the CTD. More precisely, they were shown to be important for the last step of transcription, termination. Despite extensive research over the past 30 years, the exact mechanism of how these proteins facilitate the dislodgement of Pol II from the DNA template, still remains unknown. The work presented here contains detailed studies of the CID-containing protein Seb1 from the fission yeast *Schizosaccharomyces pombe*, revealing an unexpectedly broad role of this protein in transcription termination.

In addition to a CID, Seb1 also contains an RNA recognition motif (RRM) which allows direct binding to RNA. Here, I present high-resolution crystal structures of both domains of Seb1. While the CID has a very conserved fold, the RNA binding regions contains an unusual arrangement of a canonical RRM intertwined with a second domain that are both important for RNA binding. Structure-based mutations were introduced and a combination of *in vitro* and genome-wide *in vivo* studies uncover Seb1 as an essential player in transcription termination. Importantly, both domains are required to promote the full function of Seb1. Despite its homology to the well-studied budding yeast protein Nrd1, the role of Seb1 in fission yeast is quite different. This thesis therefore provides important insight into the mechanisms that underlie eukaryotic transcription termination.

TABLE OF CONTENTS

I ACKNOWLEDGEMENTS....VII

II DECLARATION....VIII

III ABBREVIATIONS....X

IV LIST OF FIGURES....XV

V LIST OF TABLES....XVII

CHAPTER 1: INTRODUCTION....1
1.1 The transcription cycle of Pol II....2
1.1.1 Initiation....2
1.1.2 Elongation....4
1.1.3 Termination....5
1.2 The CTD code....6
1.3 3'end processing and transcription termination....11
1.3.1 Protein-coding vs. non-coding genes: general considerations....12
1.3.2 Mechanisms of 3'end processing and termination in *S. cerevisiae*....13
1.3.3 Mechanisms of 3'end processing and termination in *H. sapiens*....19
1.3.4 Known 3' end processing and termination factors in *S. pombe*....23
1.3.5 Transcription termination: the allosteric vs. torpedo model....27
1.4 The role of CID-containing proteins in transcription termination....29
1.5 Objectives of this work....32

CHAPTER 2: *IN VIVO* RECRUITMENT PROFILE AND NECESSITY OF SEB1 FOR CELL VIABILITY. 35
2.1 The structured regions of Seb1 are essential for cell viability....36
2.2 Seb1 binds to the 3'end of genes genome-wide....38
2.3 Seb1 co-purifies with the 3'end processing machinery....44
2.4 Discussion....47

CHAPTER 3: STRUCTURAL AND BIOCHEMICAL STUDIES OF THE CID....51
3.1 The Seb1-CID preferentially binds to S2P-CTD....52
3.2 Conserved amino acids in the Seb1-CID recognise the S2P moiety....54
3.3 Recruitment of Seb1 to S2P-Pol II is essential *in vivo*....61
3.4 Discussion....62

CHAPTER 4: STRUCTURAL AND BIOCHEMICAL STUDIES OF THE RRM....65
4.1 Construct design for expression of the Seb1 RNA-binding region....67
4.2 The Seb1-RRM has a unique second domain....68
4.3 The RNA sequence does not solely determine Seb1 binding *in vivo*....74
4.4 Point mutations within the Seb1-RRM can abolish RNA binding....78
4.5 Other CID-RRM proteins likely have a similar RNA-binding region....81
4.6 RNA-binding is essential for cell viability....83
4.7 Discussion....84

CHAPTER 5: CELLULAR CONSEQUENCES OF SEB1 POINT MUTATIONS....87
5.1 Seb1 mutations cause read-through transcription genome-wide....88
5.2 Seb1-dependent genes terminate more efficiently in WT....95
5.3 Genes that recruit Seb1 have termination defects in Seb1 mutants....96
5.4 Seb1 and Pcf11 are both required for transcription termination....99
5.5 Discussion....101

CHAPTER 6: GENERAL DISCUSSION AND CONCLUSION....103
6.1 Does Seb1 function in 3'end processing or termination?....104
6.2 What role do CID-RRM proteins play on different types of genes?....108
6.3 How are ncRNAs targeted to the exosome in *S. pombe*?....110
6.4 Is Seb1 required for the termination of cryptic transcription?....111

CHAPTER 7: MATERIALS AND METHODS....113
7.1 List of oligonucleotides used in this study....113
7.2 List of plasmids used in this study....117
7.3 List of *S. pombe* strains used in this study....118
7.4 Antibodies used in this study....120
7.5 *In vitro* protein biochemistry methods....120
7.5.1 Cloning of expression plasmids....120
7.5.2 Transformation of *E. coli* cells and plasmid extraction....121
7.5.3 Recombinant expression and purification of proteins....122
7.5.3.1 Full-length Seb1....122
7.5.3.2 Seb1-CID$_{1-152}$....123
7.5.3.3 Seb1-RRM$_{388-540}$....123
7.5.4 Crystallisation, data collection and refinement....124
7.5.4.1 Seb1-CID$_{1-152}$....124
7.5.4.2 Seb1-RRM$_{388-540}$....125
7.5.5 Small angle X-ray scattering (SAXS)....126
7.5.6 Peptide binding assay....127
7.5.7 Fluorescence anisotropy assay....127
7.5.7.1 Seb1-CID$_{1-152}$....127
7.5.7.2 Seb1-RRM$_{388-540}$....128
7.5.8 Electrophoretic mobility shift assay (EMSA)....129
7.5.9 SDS-PAGE....130
7.5.10 Western blotting....130
7.6 *S. pombe in vivo* methods....131
7.6.1 Media and growth conditions for *S. pombe*....131
7.6.2 Genetic manipulation and transformation....132
7.6.3 Protein extraction for Western blot (TCA method)....133
7.6.4 Spot test....133
7.6.5 RNA extraction....133
7.6.6 Northern blotting....134
7.6.7 Reverse transcription coupled with quantitative PCR (RT-qPCR)....135
7.6.8 Chromatin immunoprecipitation (ChIP) with detection by qPCR....135
7.6.9 Protein purification and mass spectrometry....136
7.7 Genome-wide and bioinformatics methods....138
7.7.1 ChIP-sequencing (ChIP-Seq)....138
7.7.2 Photoactivatable ribonucleotide-enhanced crosslinking and immuno-precipitation (PAR-CLIP)....139
7.7.3 RNA-Sequencing (RNA-Seq)....140
7.7.4 Correlation between different genome-wide datasets....141
7.7.4.1 Metagene profile plots....141
7.7.4.2 Determination of binding events in certain region....141
7.7.4.3 Boxplots of transcriptional read-through in dependence of binding....141
7.7.4.4 Visualisation of aligned reads....142
7.7.5 Protein bioinformatics....142
7.7.6 Data access....142

8 REFERENCES....143

APPENDIX I: LIST OF PROTEINS IDENTIFIED BY MS....161

APPENDIX II: LIST OF PUBLICATIONS....177

I ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor Lidia for giving me a valuable *DPhil* experience. I very much enjoyed working on this project and Lidia has given me the opportunity to explore a range of different areas, thereby enhancing my knowledge, increasing my set of skills and, most importantly, satisfying my curiosity. I was able to expel myself and I believe that this *DPhil* has provided me with a good foundation to continue along the bumpy road of science.

I am also grateful to all the people who have worked with me in the Vasilieva lab. I would not have been able to finish all those experiments without the help of two very talented students, Oliver and Miles. I also had the pleasure to closely work with Beth and together we were trying to understand what it is that Seb1 is doing. But also non-scientifically, as the only constant British lab member, Beth greatly increased my understanding of the British culture and, of course, the meaning of tea time. I also would like to thank Cornelia and Tea for always being up for discussions about any aspect of science and research. Furthermore, Dong-Hyuk and Tea were kind enough to provide me with their S2P data and I feel very happy to have been working in a lab with a very friendly atmosphere between all lab members. Further, I would like to thank former lab members Sneha and Niko for helping me in my early days and getting me started in the lab. Also, Monika and Doğukan for many nice chats during lab outings or on the way home, I always enjoyed talking to you. I would also like to mention my two previous students Jess and Lars, even though your contribution did not make it into my thesis, I really enjoyed having you around.

Many of the experiments would not have been possible without the help of people from outside the lab. This includes all members of StruBi who I've been in touch with and who made me feel welcome even though I am not a structural biologist. I am specifically grateful to Max, Jon, Karl, Tom, Guido, Cedric, Kamel, and Tao who each contributed individually in the process of obtaining the crystal structures from setting up trays to crystal manipulation, data collection, phasing, and refinement. Furthermore, I would like to thank Patrick Cramer for the PAR-CLIP data and Carlo for always promptly responding to my emails, patiently answering all my questions even on Christmas day!

Furthermore, I would like to thank mainly Ronja but also Harry, Angelika and Hamish for initial help with bioinformatics. In addition, the whole third floor, predominantly the Brockdorff and Klose labs in the early years of my *DPhil* and the Nasmyth and Akiyoshi labs in later years, provided good company and I am sure that each of you have helped me with something over the years.

The University of Oxford provided me with a very inspiring environment and thanks to joint meetings with other labs inside and outside the department as well as through many talks by visiting speakers, I was able to get a deeper insight into science and develop my own ideas which will help me throughout my career as a scientist.

To Max, I would like to give a special thanks. Not just for introducing me to the realm of structural biology but also for your support outside the lab during all those years we have been together. This is the third, and most likely last, thesis that we are writing together but hopefully not the last period of our lives together.

II DECLARATION

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

- The Seb1 PAR-CLIP experiment was performed and the initial bioinformatic analysis done by Dr Carlo Bäjen, Cramer lab, MPI Göttingen, Germany and the practical part of the Seb1 ChIP-Seq was done by Dr Beth Watts, Vasilieva lab (section 2.2).
- Trypsin digestion and mass spectrometric analysis of the Ysh1 purification was performed by the Central Proteomics Facility, Dunn School of Pathology, University of Oxford (section 2.3).
- Purification of Seb1 from *S. pombe* was done by Mr Miles Huseyin, Vasilieva lab, under my supervision and subsequent mass spectrometric analysis was done by the Advance Proteomics Facility, Department of Biochemistry, University of Oxford (section 2.4).
- The CID and RRM structures were generated in collaboration with Prof Jonathan Grimes, Division of Structural Biology, University of Oxford (sections 3.2 and 4.2). Cryo-freezing of crystals, data collection, and refinement was performed with the help of the following people from the division: Mr Max Renner, Dr Karl Harlos and Dr Guido Paesen. Phasing of the RRM data was done by Dr Kamel El Omari from Diamond Light Source, Didcot.
- Recombinant expression and purifications of all mutated RRM domains and some preparations of the WT RRM with subsequent fluorescence anisotropy assays were carried out by Mr Oliver Adams and Mr Miles Huseyin, Vasilieva lab, under my supervision (sections 4.2, 4.3, and 4.4).
- Depletion of rRNA, library preparation and sequencing of RNA-Seq samples was performed by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics, University of Oxford (section 5.1).

- All single-gene ChIP experiments were carried out by Dr Beth Watts and the *adh1* Northern blot was generated by Mr Oliver Adams (section 5.1).
- The practical part of the Pcf11 and S2P-Pol II ChIP-Seq experiments as well as subsequent alignment of the data to the genome were done by Mr Miles Huseyin, Dr Dong-Hyuk Heo, and Ms Tea Kecman from the Vasilieva lab (section 5.4).
- Some strains used in this thesis were made by Dr Cornalia Kilchert, Mr Oliver Adams, Dr Beth Watts, and Dr Sneha Shah or were obtained from outside the lab (section 7.3) and some plasmids were generated by Mr Oliver Adams and Mr Miles Huseyin (section 7.2).
- 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.

III ABBREVIATIONS

Å	Ångström (10^{-10} m)
aa	amino acid
adh1	alcohol dehydrogenase 1
air1/air2	*arginine methyltransferase-interacting RING finger proteins 1 and 2*
APT	associated with Pta1
ATP	adenosine triphosphate
bp	base pair(s)
brr5	*bad response to refrigeration 5*
bur1	*bypass UAS requirement 1*
CBC	cap-binding complex
CBP	Calmodulin binding protein
cDNA	copy DNA
Cdk or CDK	cyclin-dependent protein kinase
CF	cleavage factor
cft1/cft2	*cleavage factor two proteins 1 and 2*
cgt1	*ceramide glucosyltransferase*
ChIP	chromatin immunoprecipitation
CID	CTD-interacting domain
clp1	*cleavage/polyadenylation factor IA subunit*
c-Myc	*avian myelocytomatosis virus oncogene cellular homolog*
CPF	cleavage and polyadenylation factor
CPSF	cleavage and polyadenylation specificity factor
CstF	cleavage stimulation factor
CTD	C-terminal domain of Rpb1 in Pol II
CTDK1	CTD kinase 1
ctk1	*CTD kinase 1*
CUT	cryptic unstable transcript
dhp1	*DNase I homolog protein 1*
DIG	digoxigenin
dis2	*defective in sister chromatid disjoining*
DLS	Diamond Light Source
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DRB	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole
DSE	downstream element
DSIF	DRB-sensitivity factor
DUE	downstream U-rich element
E. coli	*Escherichia coli*
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid

EE	efficiency element
EMMG	Edinburgh minimal medium + glutamate
EMSA	electrophoretic mobility shift assay
ess1	*essential 1*
FAM	5'-fluorescein amidite
fcp1	*TFIIF-associating component of CTD phosphatase 1*
fip1	*factor interacting with pA polymerase 1*
fks2	*FK506 Sensitivity 2*
glc7	*glycogen 7*
h	hour(s)
H. sapiens	*Homo sapiens* (human)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	horse radish peroxidase
hrp1	*heterogenous nuclear ribonucleoprotein 1*
INT	integrator
INT11	integrator protein 11
IP	immunoprecipitation
iss1	*insertional suppressor of sme2-delta*
K	Kelvin
K_d	dissociation constant
kDa	kilo dalton
kin28	*kinase 28*
LB	Lysogeny broth
lsk1	*latrunculin sensitive kinase 1*
MAPK	mitogen-activated protein kinase
mcs6	*mitotic catastrophe suppressor 6*
min	minute(s)
mpe1	*mutant Pcf11 extragenic suppressor 1*
mpk1	*mitogen-activated protein kinase 1*
mRNA	messenger RNA
MS	mass spectrometry
msi2	*mRNA cleavage factor complex subunit Msi2*
mtl1	*mid-two like 1*
mtr4	*mRNA transport 4*
MTREC	Mtl1-Red1 core
nab3	*nuclear polyadenylated RNA-binding 3*
ncRNA	non-coding RNA
NDSB	non-detergent sulfobetain
NELF	negative elongation factor
NEXT	nuclear exosome targeting complex
NiNTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance

NNS	Nrd1-Nab3-Sen1
nrd1	*nuclear pre-mRNA down-regulation 1*
nt	nucleotide(s)
OD_{600}	optical density measured at 600 nm
ORF	open reading fram
pap1	*pA polymerase 1*
PDBID	protein data bank identifier
pfs2	*polyadenylation factor subunit 2*
pta1	*pre-tRNA accumulation 1*
P-TEFb	positive elongation factor b
pA	polyadenylations
pac1	*pat1 compensation*
PAGE	polyacrylamide gel electrophoresis
PAP	Peroxidase-α-Peroxidase or pA polymerase
pap1	*pA polymerase 1*
PAR-CLIP	Photoactivatable ribonucleotide-enhanced crosslinking and immunoprecipitation
PBS	phosphate-buffered saline
pcf11	*protein 1 of cleavage and polyadenylation factor I*
PCR	polymerase chain reaction
pct1	*polynucleotide 5'-triphosphatase 1*
PE	positioning element
PF	polyadenylation factor
pho1	*phosphatase 1*
pin1	*peptidyl-prolyl cis-trans isomerase NIMA-interacting 1*
pla1	*pA polymerase 1*
PMSF	phenylmethylsulfonyl fluoride
Pol II	RNA polymerase II
PP1	protein phosphatase 1
ppn1	*pombe pnuts*
PROMPT	promoter upstream transcript
prt	*pho1-repressive transcript*
pti1	*Pta1 interaction protein 1*
PVDF	polyvinylidene fluoride
qPCR	quantitative PCR
rai1	*Rat1p interacting protein 1*
rat1	*ribonucleic acid trafficking 1*
Rbbp6	Retinoblastoma-binding protein 6
reb1	*RNA polymerase I enhancer binding protein 1*
red1	*reductional division 1*
ref2	*RNA end formation 2*
rhn1	*Rtt103 homolog that localizes to the nucleus 1*
RMSD	root mean square deviation
RNA	ribonucleic acid

rna14/rna15	*poly(A) mRNA metabolism 14 and 15*
RNase	ribonuclease
RNP	ribonucleoprotein
rnt1	*RNase three 1*
RPAP2	RNA polymerase II-associated protein 2
rpm	rounds per minute
Rprd1A/Rprd1B /Rprd2	regulation of nuclear pre-mRNA domain-containing protein 1A, 1B and 2
rps2202	40S ribosomal protein S22-B
RRM	RNA recognition motif
rRNA	ribosomal RNA
RS	arginine/serine
RT	reverse transcription or room temperature
rtr1	*regulator of transcription 1*
rtt103	*regulator of Ty1 transposition 103*
S. cerevisiae	*Saccharomyces cerevisiae* (budding yeast)
S. pombe	*Schizosaccharomyces pombe* (fission yeast)
SAD	single wavelength anomalous dispersion
SAXS	small-angle X-ray scattering
Scaf4/Scaf8	SR-related and CTD-associated factor 8
SCP1	small CTD phosphatase 1
SDS	sodium dodecyl sulfate
seb1	*Rpb7-binding protein 1*
sec	second(s)
sen1	*splicing endonuclease 1*
Seq	sequencing (refers to next-generation sequencing methods)
SETX	Senataxin
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
spt4/spt5	*suppressor of Ty's 4 and 5*
srb10	*suppressor of RNA polymerase B*
SSC	Saline-sodium citrate buffer
ssDNA	single-stranded salmon sperm DNA
ssu72	*suppressor of sua7, gene 2*
SUMO	small ubiquitin-like modifier
SUT	stable unannotated transcript
swd2	*Set1c, WD40 repeat protein*
syc1	*similar to Ysh1 C-terminal*
TBE	Tris-buffered EDTA
TBST	Tris-buffered saline + Tween
TCA	trichloroacetic acid
TEV	tobacco edge virus
TFII	Pol II transcription factors
TRAMP	Trf4/5-Air1/2-Mtr4

trf4/trf5	*topoisomerase 1-related function 4 and 5*
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
ts	temperature-sensitive
TSS	transcription start site
TTS	transcription termination signal
U	units of enzyme activity
UTR	untranslated region
UUE	upstream U-rich element
Wdr33/Wdr82	WD repeat-containing proteins 33 and 82
WT	wild-type
xrn1/xrn2	*5'-3' exoribonucleases 1 and 2*
YES	yeast extract broth + supplements
ysh1	*yeast seventy-three kDa homolog 1*
yth1	*yeast thirty kDa homolog*
4-SU	4-thiouridine

IV LIST OF FIGURES

Figure 1.1: CTD phosphorylation profile across protein-coding genes....8
Figure 1.2: Architecture of the 3'end processing complex in *S. cerevisiae*....15
Figure 1.3: Cleavage, polyadenylation and termination of mRNAs in budding yeast....16
Figure 1.4: Processing and termination of short ncRNAs in budding yeast....17
Figure 1.5: Architecture of the 3'end processing complex in mammals....20
Figure 1.6: Cleavage, polyadenylation and termination of mRNAs in metazoans....22
Figure 1.7: Processing and termination of snRNAs in metazoans....23
Figure 1.8: Comparison of *cis* elements required for 3'end processing....24
Figure 1.9: CID-domain containing proteins in *S. cerevisiae* and their homologues in *S. pombe* and *H. sapiens*....29
Figure 2.1: Seb1 is largely disordered with two structured domains....35
Figure 2.2: Both domains of Seb1 are required for cell viability....37
Figure 2.3: Seb1 is recruited to the 3'end of genes in ChIP-Seq and PAR-CLIP....39
Figure 2.4: Seb1 recognises a motif at the 3'end of genes genome-wide....41
Figure 2.5: The majority of genes recruits Seb1 to their 3'ends during transcription....42
Figure 2.6: Binding at the TSS and pA site are independent of each other....43
Figure 2.7: Seb1 binds to non-coding genes and antisense to TSS and pA sites....44
Figure 2.8: Purification of Ysh1-HA-TAP....45
Figure 2.9: Full-length sequence alignment of Seb1 and Nrd1....49
Figure 3.1: Purification of recombinant Seb1-CID_{1-152}....52
Figure 3.2: Seb-CID_{1-152} has highest binding affinity to S2P-CTD....53
Figure 3.3: Co-transcriptional recruitment of Seb1 is better on long genes....54
Figure 3.4: Comparison between the Seb1-CID_{1-152} and published CID structures....56
Figure 3.5: Sequence alignment and binding specificity of different CID proteins....58
Figure 3.6: Point mutations in the CID affect CTD binding *in vitro*....60
Figure 3.7: Binding of Seb1 to S2P-Pol II is required for cell viability....62
Figure 4.1: RNA recognition by canonical RRM domains and Seb1....65
Figure 4.2: RRM solution structure of $Nrd1_{307-491}$....67
Figure 4.3: Only long RRM constructs are soluble *in vitro*....68
Figure 4.4: Purification of the Seb1-$RRM_{388-540}$ for structural studies....69
Figure 4.5: Crystal structure of the Seb1-$RRM_{388-540}$ domain....71
Figure 4.6: The Seb1-$RRM_{388-540}$ crystal structure represents its solution behaviour....72
Figure 4.7: Example dimensionless Kratky plots for a range of different proteins....74
Figure 4.8: The SUMO-tag does not interfere with RNA binding of the Seb1-$RRM_{388-540}$....75
Figure 4.9: The Seb1-$RRM_{388-540}$ has a broad RNA sequence specificity *in vitro*....77
Figure 4.10: Interactions between the two domains are required for protein solubility....78
Figure 4.11: Some $RRM_{388-540}$ point mutations disrupt RNA binding....79
Figure 4.12: Neither α-helix 2 nor the β-sheet on domain 2 bind to RNA....80
Figure 4.13: The canonical RRM and amino acids on domain 2 interact with RNA....81

Figure 4.14: Other CID-RRM proteins have sequences similar to the Seb1-RRM$_{388-540}$.....82
Figure 4.15: Seb1-RNA binding is necessary for cell viability.....84
Figure 4.16: Hypothetical model for RNA binding of the Seb1-RRM$_{388-540}$.....86
Figure 5.1: RNA-Seq is highly reproducible between replicates.....88
Figure 5.2: Decreased Seb1 recruitment causes read-through on *adh1* and *pho1*.....90
Figure 5.3: *adh1* mRNA levels are not changed in Seb1 mutants.....91
Figure 5.4: Read-through transcripts in Seb1 mutants can be detected genome-wide.....92
Figure 5.5: All classes of Pol II transcripts show read-through transcription.....93
Figure 5.6: Read-through transcription is highly correlated between the F445A and Y64K mutant.....94
Figure 5.7: Read-through transcription interferes with nearby gene expression.....95
Figure 5.8: Genes with low read-through transcription levels depend more on Seb1.....96
Figure 5.9: Seb1 binding prevents read-through transcription.....97
Figure 5.10: Binding of Seb1 at the TSS does not influence transcription termination.....98
Figure 5.11: Seb1 and Pcf11 are simultaneously recruited to the same genes.....99
Figure 5.12: Seb1 and Pcf11 are both recruited to sn/snoRNAs.....101
Figure 6.1: Hypothetical model showing the function of Seb1 at the 3'end of genes.....106
Figure 6.2: Transcription termination pathways differ in budding and fission yeast.....109

V LIST OF TABLES

Table 1: CTD kinases and phosphatases in *S. pombe* with homologues in *S. cerevisiae* and *H. sapiens*.....11
Table 2: Components of the 3' end processing machinery in human, budding and fission yeast.....26
Table 3: Seb1 and known 3'end processing proteins identified by MS.....46
Table 4: Data collection and refinement statistics for the Seb1-CID_{1-152}.....55
Table 5: Data collection and refinement statistics for the Seb1-$RRM_{388-540}$.....70
Table 6: DNA oligonucleotides used in this study.....113
Table 7: Plasmids used in this study.....117
Table 8: *S. pombe* strains used in this study.....118
Table 9: Antibodies used in this study.....120
Table 10: List of Mascot hits after mass spectrometric analysis of purified proteins.....161

CHAPTER 1: INTRODUCTION

Transcription is essential to all life and describes the process of RNA synthesis using a DNA template which stores the genetic information of a given organism. This reaction is performed by RNA polymerases. While prokaryotes only need one RNA polymerase, eukaryotes have evolved a division of labour which utilises three of these enzymatic complexes. Every one of these polymerases specifically transcribes a different subset of RNAs. RNA polymerase I is necessary for the production of most ribosomal RNAs (rRNAs) and RNA polymerase III makes the 5S rRNA, transfer (t)RNAs, and the U6 small nuclear (sn)RNA. Both of them therefore produce non-coding (nc)RNAs that are needed only for the translation step of the protein synthesis process.

RNA polymerase II (Pol II), on the other hand, plays quite a different role in protein expression. It produces messenger (m)RNAs which encode the protein sequences, as well as a number of ncRNAs including small nucleolar (sno) and snRNAs and other ncRNAs which are mostly of regulatory nature. While the other polymerases can therefore influence gene expression only on a global scale, Pol II regulates gene expression on a single gene level. This regulation is a complex process which not only includes adjusting the amount of mRNA produced but also ensuring that the resulting transcripts are functional. For this, the pre-mRNA needs to undergo extensive processing during transcription. Correct regulation of transcription is crucial because not producing the correct amount of each individual protein can be detrimental to the cell. Also, it gives cells the necessary flexibility to adapt to changing environmental conditions and it is critical for development and cell identity. Failure in regulating gene expression is a hallmark of cancer and other diseases, further underpinning the importance of this process. While a lot of research was conducted to understand transcription initiation, the mechanisms involved in termination are less well studied. Therefore, my thesis focuses specifically on the last step of Pol II transcription and how it influences gene expression using the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) as a model organism.

1.1 The transcription cycle of Pol II

The transcription cycle of Pol II consists of three distinct stages: initiation, elongation and termination. RNA processing occurs co-transcriptionally during these different stages by a multitude of proteins and every stage is subject to regulation. Important aspects of every stage will be described briefly below.

1.1.1 Initiation

During initiation, a DNA sequence called promoter is recognised and a complex is formed consisting of general transcription factors (TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) as well as all 12 subunits of Pol II. After formation of this so-called preinitiation complex, the DNA is melted to form a transcription bubble which allows synthesis of the first RNA nucleotides complementary to the template at the transcription start site (TSS) (reviewed in Sainsbury et al. 2015; Sikorski and Buratowski 2009). For efficient transcription initiation, transcriptional activators as well as enhancers are frequently required. In addition, negative regulation can occur through transcriptional repressors. The effects of all these different factors are usually passed to the transcription machinery via a conserved multiprotein complex called mediator (reviewed in Allen and Taatjes 2015).

The best studied example of how cells adjust their protein pool to environmental changes via transcription initiation is probably the heat-shock response. As a result of high temperatures, expression of proteins such as chaperones that are required for cell growth under those conditions is induced by the binding of a transcriptional activator to a specific DNA element that all heat-shock genes have in common (reviewed in Verghese et al. 2012).

Another feature of promoters I would like to discuss here, is directionality (reviewed in Bagchi and Iyer 2016). It has been proposed that some, if not all, promoters are inherently bidirectional and transcription is initiated in both directions as a result of the open chromatin state formed at the promoter. Some of the resulting antisense transcripts

are stable. However, most transcripts are immediately degraded by the nuclear degradation machinery (the exosome complex) and can only be visualised upon knock-down of exosome components. Examples of this include promoter upstream transcripts (PROMPTs) in mammalian cells (Preker et al. 2011), and cryptic unstable transcripts (CUTs) in budding and fission yeast (Xu et al. 2009; Zhou et al. 2015). Antisense transcription of these unstable RNAs is thought to be terminated soon after initiation due to the presence of general termination sequences which are absent in the sense direction (Schulz et al. 2013; Ntini et al. 2013). Recent research even suggests that termination might be the default state, working in sense as well as in antisense direction, and can only be overcome if Pol II can transition into the elongation state (Guo and Price 2013; Milligan et al. 2016). This is further supported by reports that many short, unstable transcripts are produced not only in the antisense but also in the sense direction (Brannan et al. 2012; Tuck and Tollervey 2013).

Another way in which initiation can be influenced is by transcription of an upstream gene that continues through the promoter of the downstream gene resulting in its repression. This phenomenon was termed transcription interference and was first described by Prescott and Proudfoot (2002). It also works in the antisense direction in which case transcription antisense to a promoter occurs. One particular example of this form of gene expression regulation in *S. pombe* was discovered in the Vasilieva lab (Shah et al. 2014). Here, it is the *pho1* gene which is regulated. It encodes an acid phosphatase that is only needed under low phosphate conditions. If phosphate levels are high, however, expression of the mRNA is repressed by transcription of a lncRNA termed *prt* (pho1-repressive transcript) that is initiated from a promoter upstream of the mRNA TSS.

1.1.2 Elongation

During elongation, Pol II is tightly and stably associated with the DNA template and quickly transcribes through the body of the gene. However, the transition from initiation to elongation is rather complex. In metazoans, a large number of genes have stalled Pol II just downstream of the TSS (Zeitlinger et al. 2007; Muse et al. 2007; Guenther et al. 2007). This promoter-proximal pausing is caused by the presence of the DRB-sensitivity inducing factor (DSIF) and the negative elongation factor (NELF) (reviewed in Guo and Price 2013). Upon recruitment of the positive elongation factor b (P-TEFb), both DSIF (via its subunit Spt5) and NELF are phosphorylated which leads to the release of Pol II from the paused state and productive elongation can occur. The purpose of this pausing is thought to provide a “window of opportunity” for Pol II to recruit all necessary RNA processing and elongation factors, and was therefore termed the “early-elongation checkpoint”. However, this mechanisms can also be used for regulatory purposes. On certain genes, Pol II can either be stalled until a stimulus allows gene expression which is controlled by an activator (e.g. in the case of the *c-Myc* gene) (reviewed in Kwak and Lis 2013) or to synchronise the synthesis of a subset of genes during development (reviewed in Levine 2011).

For a long time it has been thought that early transcriptional pausing is specific to higher eukaryotes and to not occur in yeast, partly because NELF is not conserved (Guo and Price 2013). Recent studies, however, have challenged this view. Strand-specific, genome-wide analyses of transcribing Pol II have detected an accumulation of polymerase shortly after the TSS in budding yeast (Churchman and Weissman 2011; Milligan et al. 2016) however, promoter-proximal relief does not seem to be controlled by P-TEFb but by the unrelated kinase Kin28 (Rodríguez-Molina et al. 2016). In addition, termination factors are frequently recruited to the Pol II pause site and failure to exchange initiation with elongation factors has been suggested to result in early termination (Tuck and Tollervey 2013). In contrast, fission yeast seems to utilise a mechanism more similar to humans as Spt5 was shown to be phosphorylated by P-TEFb early in transcription. In

addition, P-TEFb can also directly interact with Pct1, one of the two enzymes required for cap formation in *S. pombe*. This ensures mRNA elongation can proceed only upon successful capping and therefore prevents unnecessary rounds of transcription if early RNA processing fails (Pei et al. 2003). Furthermore, a recent genome-wide study could demonstrate that deletion of *spt4*, the binding partner of Spt5, prevents early transcriptional pausing (Booth et al. 2016). This strongly indicates that an early elongation checkpoint exists in yeasts as well. However, more research is needed to fully understand the mechanisms involved.

Furthermore, extensive RNA processing takes place during elongation. Shortly after initiation, the nascent RNA is 5' capped (Shatkin 1976) and introns are co-transcriptionally spliced (reviewed in Saldi et al. 2016). At the end of elongation, the mRNA is cleaved and polyadenylated which is collectively called 3'end processing. Co-transcriptional splicing and 3'end processing are both also accompanied by Pol II pausing (Alexander et al. 2010; Carrillo Oesterreich et al. 2010; Nojima et al. 2015; Yonaha and Proudfoot 1999) illustrating how tight the connection is between transcription and RNA processing. Furthermore, Pol II elongation rates were suggested to influence assembly of the spliceosome (Saldi et al. 2016) and polyadenylation (pA) site selection (Pinto et al. 2011; Fusby et al. 2016) constituting another way to regulate gene expression. Both splicing and pA site selection can affect the coding potential of the transcript as well as its stability (Volanakis et al. 2013; Shi and Manley 2015). In addition, checkpoints have been described for both processes (Chathoth et al. 2014; Laitem et al. 2015), further supporting the notion that a plethora of mechanisms are in place to ensure correct processing of RNA transcripts before nuclear export and translation.

1.1.3 Termination

After elongation is complete, Pol II has to be removed from the template in order to be recycled for the next round of transcription and to prevent transcription interference with downstream genes (Greger et al. 2000). Furthermore, various examples have been

found in which premature termination is used to control the expression of specific genes. The budding yeast protein Nrd1, for example, promotes premature termination of its own transcript if Nrd1 protein levels are high, thereby regulating the amount of Nrd1 in the cell utilising a negative feedback loop (Arigo et al. 2006). The Fks2 protein which encodes a 1,3-β-D-glucan synthase important for cell wall synthesis uses a different mechanism. In this case, the mitogen-activated protein kinase (MAPK) Mpk1 prevents premature termination of Fks2 specifically under cell wall stress thereby ensuring a rapid response to changing conditions (Kim and Levin 2011). However, termination plays an even broader role in gene regulation. At bidirectional promoters, termination of transcription in the antisense direction is essential to protect the transcriptome from uncontrolled transcriptional interference (reviewed in Porrua and Libri 2015).

Because 3'end processing and transcription termination are a major focus of this thesis, both processes will be discussed in more detail in section 1.3.

1.2 The CTD code

The previous section only scratched the surface of the intricate mechanisms that are in place to ensure proper gene expression. But how can this complex network of factors be recruited in a timely manner to perform and regulate the different stages of transcription and to guarantee correct processing of the newly synthesised RNA? At the heart of this lies a domain that is unique to Pol II but conserved in eukaryotes from fungi to humans (Stiller and Hall 2002): the C-terminal domain (CTD) of its largest subunit, Rpb1. It consists of repeats of the heptad $Y_1S_2P_3T_4S_5P_6S_7$ and is subject to extensive post-translational modifications. First discovered in 1985 (Allison et al. 1985; Corden et al. 1985), a lot of research has been performed since and the CTD field is still fast moving, which is reflected by the large number of reviews that have been published in recent years (Egloff et al. 2012; Hsin and Manley 2012; Jeronimo et al. 2013; Corden 2013; Eick and Geyer 2013; Heidemann et al. 2013; Jeronimo et al. 2016; Zaborowska et al. 2016).

The number of repeats varies in every organisms, however, it is approximately correlated with genomic complexity. For example, the CTD in *Saccharomyces cerevisiae* (*S. cerevisiae*) consists of 27, in *S. pombe* of 29, in *Drosophila melanogaster* of 45, and in humans of 52 repeats (Chapman et al. 2008). All three serines as well as Y1 and T4 in the $Y_1S_2P_3T_4S_5P_6S_7$ consensus can be phosphorylated, S2P and S5P being the best studied modifications. In addition, both prolines can be present as either the *cis* or the *trans* isoform. The phosphorylation pattern of Pol II varies between different stages of the transcription cycle (Figure 1.1) leading to a readable code – the CTD code – which provides a binding platform for a variety of proteins that can interact only with certain modification types of the CTD. This ensures timely association and dissociation of transcription and processing factors (Jeronimo et al. 2013; Corden 2013; Eick and Geyer 2013). For example, the S5P modification which is placed shortly after initiation is recognised by capping enzymes in many organisms and leads to capping of the nascent transcript early during transcription, preventing its immediate degradation (Pei et al. 2003). How fundamental this process is has been demonstrated in *S. pombe*. Mutation of the S5 position to the non-phosphorylateable amino acid alanine in all repeats of the CTD is lethal, however, if the capping enzyme is fused to the CTD the S5A mutation is tolerated (Schwer and Shuman 2011). This demonstrates that the ability to phosphorylate S5 is essential in order to cap and therefore stabilise all cellular mRNAs. Similarly, the S2P mark increases towards the 3'end of a gene and several 3'end processing and termination factors were shown to specifically bind upon modification of this residue (Meinhart and Cramer 2004; Noble et al. 2005; Lunde et al. 2010). Here, the situation is more complex than in the case of capping though, as S2A mutants are viable in budding and fission yeast (Schwer and Shuman 2011; Cassart et al. 2012).

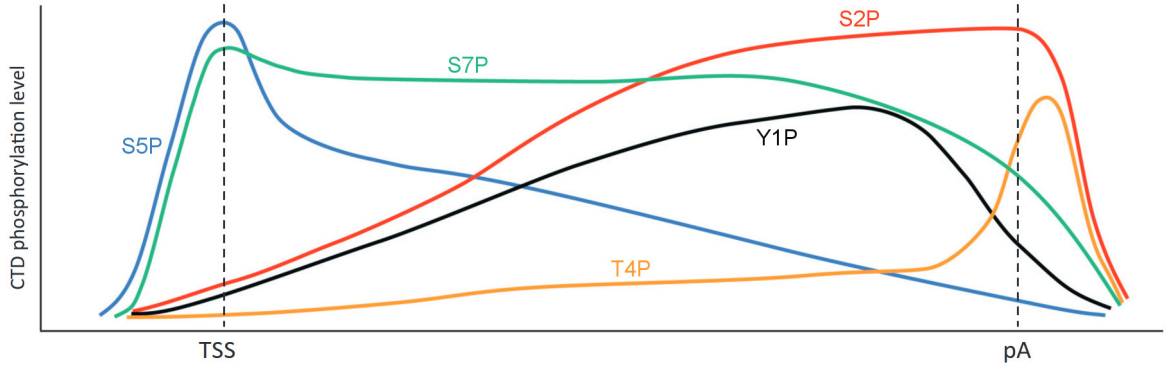


Figure 1.1: CTD phosphorylation profile across protein-coding genes
Schematic representation of the phosphorylation levels of the different CTD residues along a typical protein-coding gene in yeast. S2P (red) levels increase whereas S5P (blue) decreases throughout transcription of a gene. S7P (green) is high throughout the gene body, similar to Y1P (black) but with maximum levels on opposite ends of the gene. T4P (yellow) is the only modification that peaks after the pA site has been transcribed. In mammals, the profile only differs for S7P which peaks towards the 3'end and Y1P which shows a peak over both the TSS and pA sites. TSS: transcription start site; pA: polyadenylation; Reprinted with permission from Hajheidari et al. (2013).

In recent years it has become clear that dephosphorylation of Y1 as well as phosphorylation of T4 is also part of the intricate regulatory network that controls transcription termination (Mayer et al. 2012; Harlen et al. 2016). Similar dependencies of factor recruitment on CTD modifications have been demonstrated for Pol II initiation, elongation, pausing and splicing (Jeronimo et al. 2013) and a variety of domains were described that can specifically recognise and bind certain CTD forms (reviewed in Jasnovidova and Stefl 2013).

But how is it possible that this relatively short repeat can coordinate the recruitment of so many different proteins? Until earlier this year, the argument was often used that in theory, 2^7 = 128 different isoforms of every heptad repeat exists (that is every amino acid can exist in two different states which is either phosphorylated or not for Tyr, Ser, and Thr or in *cis* or *trans* in the case of Pro). And this is ignoring other modifications which would be beyond the scope of this thesis (Eick and Geyer 2013). However, mass spectrometric analyses of the CTD have now shown that the majority of repeats is unphosphorylated and most phosphorylated repeats contain only one modification with S2P and S5P being the most prevalent (Suh et al. 2016; Schüller et al. 2016).

Furthermore, only 70% of all proline bonds were observed to be of the *trans* isoform (Noble et al. 2005). This markedly reduces the number of different CTD states that are actually being used. However, proteins do not merely recognise whether a phosphate group is present; the situation is actually more complex. The CTD itself is disordered which gives it a great flexibility to adopt different secondary structures (Corden 2013). The CTD structure was described in solution and molecular dynamics simulations tried to predict its behaviour *in silico*. Nuclear magnetic resonance (NMR) and circular dichroism studies suggest that the CTD fluctuates between an unordered state and β-turn as well as polyproline II helix conformations resulting in a small subpopulation of peptides in folded states (Cagas and Corden 1995; Bienkiewicz et al. 2000). Phosphorylation has been suggested to restrict the amount of possible conformations and to influence the equilibrium between different states (Bienkiewicz et al. 2000; Yonezawa 2014). Furthermore, most proteins are expected to specifically bind to only one of the two proline isoforms. *In vivo,* a cis-trans isomerase (Ess1 in *S. cerevisiae* and Pin1 in *S. pombe* and humans) can catalyse the *cis/trans* conversion reaction. Indeed, the enzyme has been linked to transcription regulation (reviewed in Hanes 2014). In addition, interactions between the tyrosine residues of different repeats have been observed, resulting in a more ordered state for long stretches of unbound repeats (Bienkiewicz et al. 2000). Making matters even more complicated, crystal and NMR structures of different proteins in complex with CTD peptides show that not all factors recognise the same length of the CTD (Jasnovidova and Stefl 2013). The most extreme cases observed are only four residues in the case of SCP1 (Zhang et al. 2006), a serine phosphatase, and three repeats in the case of Cgt1, a guanylyltransferase (Fabrega et al. 2003). The different conformational states present in solution, which are influenced by posttranslational modifications and *cis/trans* isomerisation as well as the amount of (un)bound repeats, combined with an induced fit mechanisms of binding (Noble et al. 2005) would therefore regulate recruitment and binding of different factors throughout the transcription cycle.

Because CTD kinases and phosphatases are themselves CTD binding proteins,

their recruitment underlies the same rules as is the case for other CTD interactors. A neat system has evolved of sequential phosphorylation and dephosphorylation that is in turn regulated by the marks which are placed by these proteins themselves (Jeronimo et al. 2013; Corden 2013). Not all “writers” and “erasers” of the CTD code have been identified today and contradicting results have led to confusion in the field. However, many important players have been characterised (Table 1). In summary, Pol II is recruited to promoters in an unphosphorylated state and, in yeast, the S5P and S7P marks are placed by the general transcription factor TFIIH before transcription even begins. S5P then recruits the capping enzyme. The S5P mark is subsequently removed within promoter-proximal regions, in yeast by Rtr1, and to a lesser extend by Ssu72. Next, in humans as well as *S. pombe*, P-TEFb promotes the transition from initiation to elongation, not only by phosphorylating Spt5 but also S2 directly. However, CTDK1 is responsible for the majority of S2 phosphorylation in yeast and therefore establishes the modification during elongation. S5P and S2P are both required in order to regulate elongation and splicing. As part of the 3’end processing complex, Ssu72 removes the S5P and S7P modifications at the end of the transcription cycle which allows termination to occur. The activity of this phosphatase is restricted to the beginning and end of genes probably as a result of *cis/trans* isomerisation. Following transcription termination, S2P is completely removed by Fcp1 as erasing all modifications is necessary for the reinitiation of Pol II (Hsin and Manley 2012; Jeronimo et al. 2013; Eick and Geyer 2013; Jeronimo et al. 2016).

Phosphorylation of Y1 and T4 is less well understood. In budding yeast, Y1P levels increase throughout transcription and the mark was suggested to prevent premature termination (Mayer et al. 2012). While the kinase depositing the mark is currently unknown, the phosphate group was shown to be erased by Glc7 which is part of the 3’end processing machinery (Schreieck et al. 2014). T4P also plays a role in transcription termination, however, it has an opposite effect. T4P was demonstrated to help termination by recruiting the transcription termination factor Rtt103. Furthermore, together with S5P it seems to be important for co-transcriptional splicing (Harlen et al. 2016).

Table 1: CTD kinases and phosphatases in *S. pombe* with homologues in *S. cerevisiae* and *H. sapiens*
grey: only inferred by homology

S. pombe	***S. cerevisiae***	***H. sapiens***	**Complex**	**CTD Specificity**
CTD kinases				
Mcs6	Kin28	CDK7	TFIIH	S5P, S7P
Lsk1	Ctk1	CDK12, CDK13	CTDK1	S2P
Cdk9	Bur1	CDK9	P-TEFb	S2P, S5P, S7P
Cdk8	Srb10	CDK8	Mediator	S5P
CTD phosphatases				
Ssu72	Ssu72	Ssu72	CPF	S5P, S7P
Fcp1	Fcp1	Fcp1		S2P, S5P, T4P?
Rtr1	Rtr1	RPAP2		S5P
Dis2	Glc7	PP1	CPF	Y1P

1.3 3'end processing and transcription termination

3'end processing and transcription termination are two distinct processes which both take place at the end of transcription. They are so closely linked, however, that it is difficult to study one, without influencing the other. Extensive research over the past decades have helped to elucidate how these processes work. However, there are still a lot of open questions remaining. In the following sections, I aim to give an overview of the research conducted and unanswered questions. The protein machinery and RNA sequences involved in 3'end processing have been extensively studied in budding yeast and humans. However, this is not the case for fission yeast. Because *S. pombe* is evolutionary closer to *S. cerevisiae* than to humans, it is expected to be more similar to the former (Yanagida 2002). However, a number of examples have been described in which *S. pombe* resemble *H. sapiens* more than *S. cerevisiae* (Sipiczki 1995) which makes it difficult to know, without direct testing, which organism fission yeast is more similar to in regards of 3'end processing and transcription termination. I will therefore first introduce what is known in humans and budding yeast and then discuss where fission yeast stands in relation to them.

1.3.1 Protein-coding vs. non-coding genes: general considerations

In all three organisms, protein-coding genes rely on defined pA signals for their processing which are recognised by the cleavage and polyadenylation machinery, a conserved multi-protein complex (reviewed in Mischo and Proudfoot 2013; Shi and Manley 2015; Porrua and Libri 2015; Loya and Reines 2016). The purpose of this is to correctly position an endonuclease (Ysh1/Brr5 in *S. cerevisiae*, CPSF-73 in humans, Ysh1 in *S. pombe*) at the pA site which cleaves the nascent transcript, thereby defining its 3'end. Following cleavage, the RNA is polyadenylated and released from the chromatin. This, however, is not sufficient to terminate Pol II. Instead, it continues transcription until it is dislodged from the template DNA. Therefore, the 3'end of the transcript and the Pol II transcription termination signal (TTS) are not identical. Two different theories about how transcription termination is achieved have been proposed and will be discussed in section 1.3.5.

Many non-coding RNAs do not rely on pA signals for their processing. Even though some of the same factors are used that are also part of the cleavage and polyadenylation machinery, they are organised into different distinct complexes. Additional factors are utilised as well. For pA-independent processing, there are large differences between budding yeast and humans, indicating that the machinery involved is not conserved. In *S. cerevisiae*, a distinct mechanism is responsible for the termination of short transcripts such as sn/snoRNAs and unstable RNAs such as CUTs. In *H. sapiens*, on the other hand, snoRNAs are usually encoded within introns and therefore are not terminated separately. snRNAs are processed and terminated by pA-independent mechanisms, which does not seem to be the case for cryptic transcripts. However, the existence of short unstable transcripts was discovered much later in humans than in budding yeast which explains why the latter has been studied in much more detail. In *S. pombe*, the mechanism of pA-independent transcription termination is unknown (reviewed in Arndt and Reines 2015; Tudek et al. 2015).

1.3.2 Mechanisms of 3'end processing and termination in *S. cerevisiae*

Protein-coding genes rely on *cis* elements for correct positioning of the 3'end processing machinery and, in the case of *S. cerevisiae*, four of these have been described (Figure 1.2, reviewed in Mandel et al. 2008; Tian and Graber 2012). The efficiency element (EE) is an AU-rich element (UAUAUA being the optimal sequence) and is located ~ 25-40 nt upstream of the cleavage site. It increases the efficiency of 3'end processing, however, many base pair mutations can be tolerated without affecting cleavage. The positioning element (PE) is generally A-rich (with AAUAAA being the most effective sequence) and is located ~ 10-30 nt upstream of the cleavage site. When mutated, the position of the pA site changes. However, the efficiency of cleavage is usually not affected (Guo and Sherman 1996). The cleavage or pA site itself is defined by a pyrimidine which is followed by multiple adenosines after which cleavage takes place (Heidmann et al. 1992). Finally, U-rich elements are located directly up- and downstream of the cleavage site (UUE and DUE), respectively. Mutating them does not abolish cleavage, however, in combination with cleavage site mutations, the efficiency of the reaction is dramatically reduced indicating a synergistic role of the two elements (Graber et al. 1999; Dichtl and Keller 2001). Interestingly, the very sequence of the pA signals affect transcript stability. For example, mRNAs with short half-lives often contain guanines in their EE motifs (Graber 2003) and poly-U stretches in the UUE increase mRNA stability. In addition, the terminal nucleotide also has a large influence on the half-life of the transcript (Geisberg et al. 2014), demonstrating the importance of cleaving at the correct position.

The polyadenylation and cleavage machinery consists of 20 proteins in budding yeast that are organised into several sub-complexes (Figure 1.2) which are called cleavage factor (CF) IA and IB and cleavage and polyadenylation factor (CPF) with its sub-complexes CFII and polyadenylation factor I (PFI, reviewed in Mandel et al. 2008; Shi et al. 2009). *In vitro* studies have shown that the cleavage reaction requires CFIA, CFIB, and CFII while polyadenylation is performed by CFIA, CFIB, and CPF, suggesting recruitment of additional factors or their rearrangement after cleavage (Chen and Moore

1992). The cleavage and polyadenylation machinery recognises pA signals on the nascent RNA and bind them as soon as they emerge from the polymerase while simultaneously interacting with Pol II. Specifically, the CFIA proteins Rna14 and Rna15 form a very tight heterodimeric complex which can directly interact with the PE while Pcf11 (which is also part of CFIA) binds to S2P-CTD (Minvielle-Sebastia et al. 1994; Gross and Moore 2001; Barillà et al. 2001; Licatalosi et al. 2002; Noble et al. 2004). At the same time, the EE is bound by Hrp1, the sole component of CFIB, which in turn interacts with a Rna14-Rna15 tetramer in CFIA (Pérez-Cañadillas 2006; Barnwal et al. 2012). The last component of CFIA, Clp1, forms a bridge to the CPF which contains the catalytic activity for transcript cleavage, thereby positioning it close to the pA site (Holbein et al. 2011).

However, this model has been challenged by a recent report demonstrating that Rna15 does not bind upstream but downstream of the pA site *in vivo* (Baejen et al. 2014). This is consistent with *in vitro* studies indicating that simultaneous deletion of the PE and EE does not abolish cleavage and that CFIB is dispensable for the reaction. Rather it was suggested that CFIB binding to RNA prevents assembly of the 3'end processing complex and pre-mature cleavage at the EE (Minvielle-Sebastia et al. 1998; Dichtl and Keller 2001). More research is necessary to fully understand the role of the CF in *S. cerevisiae*.

Several subunits of the CPF also directly bind to RNA. The CFII component Cft1/Yhh1 binds near the cleavage site and is thought to position the endonuclease Brr5/Ysh1 correctly, which on its own has little sequence specificity (Dichtl et al. 2002; Mandel et al. 2006; Baejen et al. 2014). Another subunit of the CFII complex, Pta1, probably serves as a scaffold (Mandel et al. 2008) and seems to be regulated by the phosphatase Glc7 (which was shown to dephosphorylate Y1P-CTD, Table 1.1). Deletion or inhibition of Glc7 leads to accumulation of phosphorylated Pta1 and shortened pA tails (He and Moore 2005).

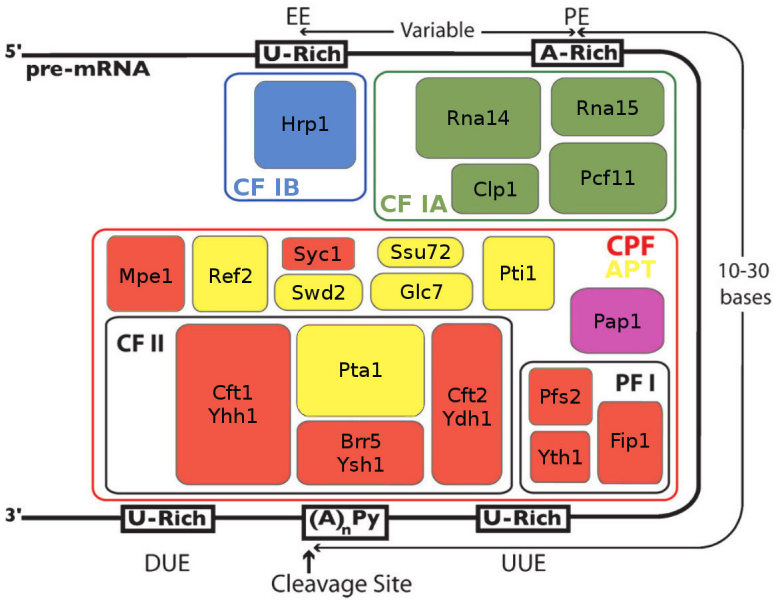


Figure 1.2: Architecture of the 3'end processing complex in *S. cerevisiae*
Schematics representation of the 20 proteins forming the cleavage factor (CF) and cleavage and polyadenylation factor (CPF) sub-complexes of the 3'end processing machinery in budding yeast. The *cis* elements on RNA which are recognised by the complex are also indicated. See text for details. Reprinted slightly modified with permission from Mandel et al. (2008).

After cleavage of the nascent RNA, a pA tail of about 70 nt length is added to the 5' cleavage product by Pap1 (Bienroth et al. 1993). The enzyme does not require any additional proteins for the polyadenylation reaction *in vitro.* However, other factors are necessary to regulate pA tail length (Hector et al. 2002). The PFI proteins Yth1 and Fip1 are thought to recognise the UUE and DUE and to bring Pap1 close to the pA site (Preker et al. 1995; Barabino et al. 2000; Kaufmann et al. 2004; Ezeokonkwo et al. 2011).

A large network of protein-protein interactions exists within and between the CF and CPF complexes and some proteins seem to solely function as a scaffold or for correct positioning of enzymatic subunits (Mandel et al. 2008). However, some proteins have additional and quite diverse functions. Syc1, for example, was indicated in negatively regulating 3'end processing (Zhelkovsky et al. 2006) and Ssu72 dephosphorylates S5P and S7P on the Pol II-CTD, preparing it for a new round of transcription (Krishnamurthy et al. 2004; Zhang et al. 2012). Swd2 seems to be dispensable for 3'end processing but might play a larger role in transcription termination (Cheng et al. 2004).

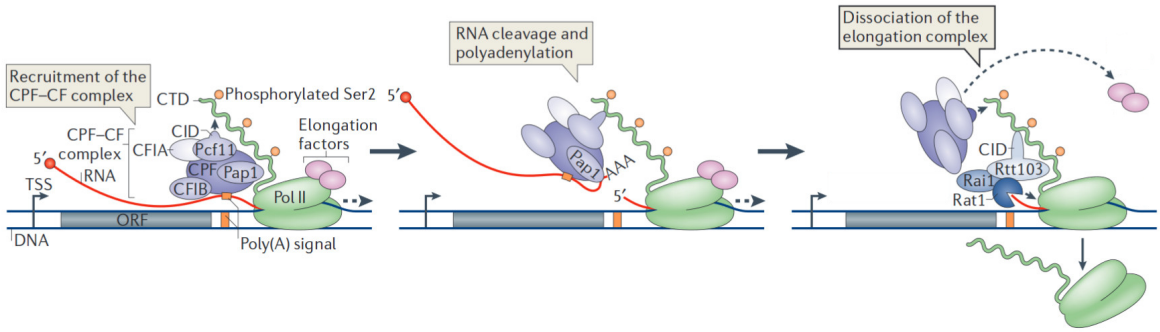


Figure 1.3: Cleavage, polyadenylation and termination of mRNAs in budding yeast As soon as the pA signals have been transcribed, components of the cleavage factor (CF) and cleavage and polyadenylation factor (CPF) form a stable complex with the transcription machinery via RNA and S2P-CTD interactions. This takes place at the end of elongation. After cleavage of the nascent RNA, the 5' cleavage product is polyadenylated. Next, elongation factors dissociate from Pol II and the 3' cleavage product is degraded by the Rat1-Rai1-Rtt103 complex. Pol II is subsequently removed from the DNA template. TSS: transcription start site; ORF: open reading frame; CID: CTD-interacting domain; Reprinted with permission from Porrua and Libri (2015).

Binding of CPF and CF proteins which promote cleavage and polyadenylation is necessary to produce functional mRNA but does not terminate Pol II (Figure 1.3). After release of the cleaved RNA, Pol II continues to transcribe the DNA for another 150-200 nt (Mischo and Proudfoot 2013; Porrua and Libri 2015). The 3' cleavage product which is still being synthesised by Pol II is, in contrast to the mRNA, not capped and creates an entry point for the nuclear 5'-3' exonuclease Rat1, which degrades the nascent transcript until it reaches Pol II. In addition, Rat1 is bound by two other proteins, Rai1 and the S2P-CTD-interacting protein Rtt103 (Kim et al. 2004b; Lunde et al. 2010). However, how exactly Pol II is removed from the DNA template is still not fully understood and will be discussed in more detail in section 1.3.5.

In *S. cerevisiae*, ncRNAs can be divided into two general classes: housekeeping ncRNAs that function in modifying or processing of other RNAs and unstable transcripts that are produced by pervasive transcription. The former class constitutes snRNAs which are important for mRNA splicing, and snoRNAs which associate with other proteins to form ribonucleoprotein particles (RNP) that can modify other RNA molecules such as rRNAs (Tudek et al. 2015). In contrast, pervasively transcribed RNAs have no function

beyond the process of their transcription. The two most important classes are the unstable CUTs (Wyers et al. 2005) and stable unannotated transcript (SUTs) (David et al. 2006). SUTs have a similar length as mRNAs with defined 3'ends and are thought to be terminated mainly be the CF-CPF pathway described above. However, they are later degraded by the cytoplasmic 5'-3' exonuclease Xrn1 or the exosome, a 3'-5' degradation machinery (Marquardt et al. 2011; Gudipati et al. 2012).

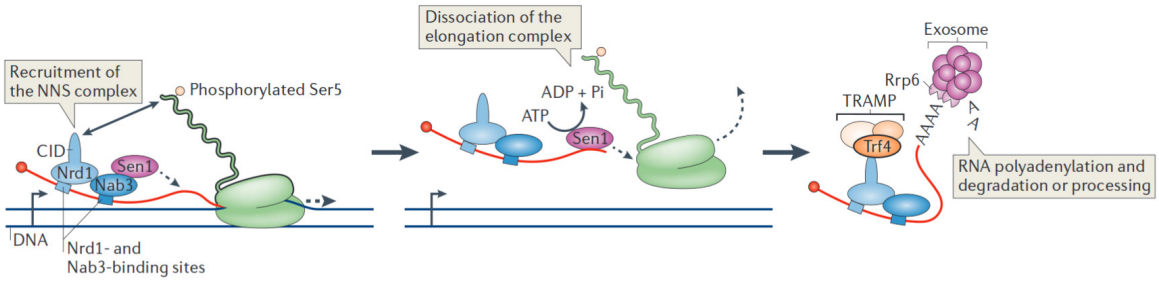


Figure 1.4: Processing and termination of short ncRNAs in budding yeast
The Nrd1-Nab3-Sen1 (NNS) protein complex is recruited to sites of active transcription via RNA and S5P-CTD interaction early after initiation. The helicase Sen1 then terminates Pol II in an ATP-dependent manner. Nrd1 and Nab3 lead to recruitment of the Trf4-Air1/2-Mtr4 (TRAMP) complex which in turn recruits the exosome for processing or degradation of the released transcript. Reprinted with permission from (Porrua and Libri 2015).

CUTs and sn/snoRNAs, on the other hand, mainly rely on a different mechanism for processing and termination (Figure 1.4) which utilises the Nrd1-Nab3-Sen1 (NNS) complex (Steinmetz et al. 2001; Arigo et al. 2006; Thiebaut et al. 2006). Both Nrd1 and Nab3 are RNA-binding proteins and they recognise the motifs UGUAG and UCUUG, respectively (Carroll et al. 2004, 2007; Creamer et al. 2011; Jamonnak et al. 2011). In addition, Nrd1 can bind to S5P-CTD which occurs early in transcription (Figure 1.1) and explains why the NNS pathway only efficiently terminates up to ~ 1 kb after the TSS (Vasiljeva et al. 2008a; Gudipati et al. 2008). In the current model, the Nrd1-Nab3 heterodimer is bound to Pol II near the RNA exit channel and can interact with the nascent transcript as soon as their recognition sequences emerge from the polymerase (Heo et al. 2013; Tudek et al. 2014). They subsequently recruit the DNA and RNA helicase Sen1 which can terminate Pol II likely by moving along and thereby unwinding the DNA:RNA hybrid until it reaches the polymerase (Porrua and Libri 2013). Direct contacts between

Sen1 and the CTD have also been reported which might be additionally necessary to remove Pol II from the template (Ursic et al. 2004). In contrast to protein-coding genes, there is no transcript cleavage step involved (Porrua and Libri 2015).

After transcription termination, Nrd1 and Nab3 are thought to stay bound to the RNA and to recruit the Trf4-Air1/2-Mtr4 (TRAMP) complex via direct interaction between Nrd1 and Trf4 (Vasiljeva and Buratowski 2006; Tudek et al. 2014). TRAMP then adds a short (~ 40 nt) pA tail to the RNA which, in contrast to adenylation of pre-mRNAs by Pap1, destabilises the transcript through the recruitment of the exosome complex (Wyers et al. 2005; LaCava et al. 2005). In the case of CUTs, the transcripts are then completely degraded while sn/snoRNAs are only trimmed which is part of their maturation pathway (Vasiljeva and Buratowski 2006). In the latter case, it was suggested that co-transcriptional binding of RNP proteins stabilises the RNA and restricts degradation to a certain position (Morlando et al. 2004; Ballarino et al. 2005; Coy et al. 2013).

However, termination of CUTs and sn/snoRNAs is not completely independent of CF-CPF proteins (Kim et al. 2006). CFIA and a sub-complex of the CPF termed APT (associated with Pta1) which consists of Glc7, Swd2, Pti1, Pta1, Ref2 and Ssu72 are required for proper processing and termination of sn/snoRNAs (Morlando et al. 2002; Nedea et al. 2003; Kim et al. 2006). Neither Brr5/Ysh1 nor Pap1 are part of this complex indicating that the function of the CF-CPF components here is not cleavage and polyadenylation. On the contrary, polyadenylation would lead to RNA export and needs to be suppressed to ensure nuclear retention of sn/snoRNAs (Morlando et al. 2002). Why the CF-CPF components are required at sn/snoRNAs though, is not fully understood. It was proposed to constitute a “fail-safe” termination pathway as CF-CPF termination signals have been found after NNS termination signal on some snoRNAs and CUTs (Steinmetz and Brow 2003; Grzechnik and Kufel 2008; Castelnuovo et al. 2014). Furthermore, the expression of some mRNAs was found to be regulated by the NNS complex (Webb et al. 2014), demonstrating that the two pathways are more closely linked than previously thought.

A number of alternative termination pathways exist in addition to the two described above. Another fail-safe mechanism on protein-coding genes is Rnt1-dependent termination. In this case, cleavage would be performed by the RNase III homologue Rnt1, rather than Ysh1, triggering the same Rat1-dependent termination mechanism that acts on protein-coding genes (Ghazal et al. 2009; Rondón et al. 2009). Rnt1 was also implicated in normal processing of the U2 snRNA (Elela et al. 1996). In the case of road-block termination, Reb1 is stably bound to DNA and acts as a barrier for the transcription machinery which leads to ubiquitination and probably degradation of Pol II (Tudek et al. 2015). Removal of the polymerase complex by degradation is also used as a last resort mechanism after DNA damage (Wilson et al. 2013). In summary, the fact that so many fail-safe termination mechanisms exist demonstrates how important it is to restrict Pol II transcription to defined regions of the genome in order to prevent interference with the expression of downstream or antisense genes.

1.3.3 Mechanisms of 3'end processing and termination in *H. sapiens*

In general, the mechanism and the machinery involved that dictate 3'end processing and termination of protein-coding genes are quite conserved between yeast and mammals. However, there are considerable differences in the *cis* elements which are recognised (Figure 1.5). In human, three primary and two auxiliary sequence elements exist which define the pA site and regulate the cleavage and polyadenylation reactions, respectively (Mandel et al. 2006). The first element identified is a conserved hexamer with the consensus AAUAAA (Proudfoot and Brownlee 1976) which shows very little sequence variation between different genes (Tian et al. 2005). In addition, not only the primary sequence but also its distance to the pA site is highly conserved and deleting the region between the two leads to cleavage further downstream (Fitzgerald and Shenk 1981). It is therefore homologous to the PE in yeast, however, mutations are only rarely tolerated. The AAUAAA hexamer is followed by the cleavage or pA site with CA being the optimal sequence (Sheets et al. 1990). On most genes, the G/U-rich downstream element (DSE)

is located at ~ 30 nt after the pA site (Takagaki and Manley 1997), however, the exact sequence is not very conserved and point mutations often have little influence on cleavage efficiency (McDevitt et al. 1986; Hutchins et al. 2008). An auxiliary upstream element is not present on all genes and is generally U-rich (Hu et al. 2005). The presence of this element enhances cleavage efficiency and, interestingly, can often be found on intronless genes where it increases expression levels (Moreira et al. 1995; Huang et al. 1999). Additional, G-rich auxiliary downstream elements (DSE) can be found on some genes, however, neither their position nor sequence are very conserved (Bagga et al. 1995). Instead of being important for 3'end processing, they have been implicated in Pol II pausing and efficient termination of transcription (Gromak et al. 2006).

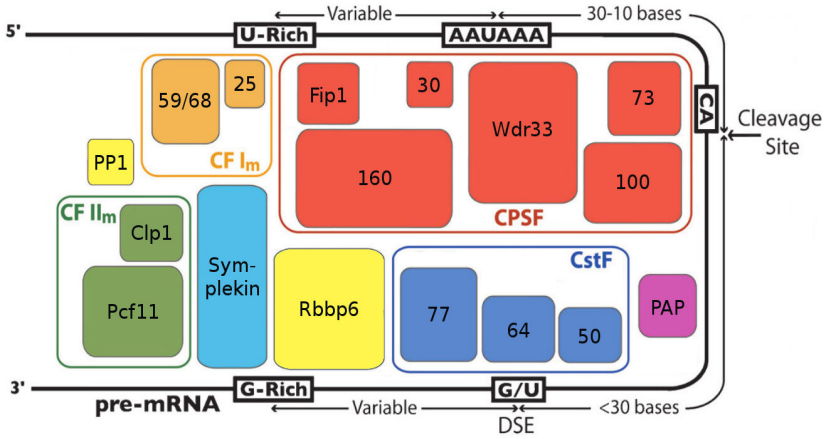


Figure 1.5: Architecture of the 3'end processing complex in mammals
Schematics representation the proteins forming the mammalian cleavage factors (CFs), the cleavage and polyadenylation specificity factor (CPSF), and the cleavage stimulation factor (CstF) sub-complexes of the 3'end processing machinery. The *cis* elements on RNA which are recognised by the complex are also indicated. See text for details. Reprinted slightly modified with permission from (Mandel et al. 2008).

Overall, homologues of the majority of mRNA 3'end processing proteins from yeast are also implicated in 3'end formation in mammalian cells (see also Table 2), apart from a few exceptions such as Ssu72 which does not associate with the 3'end processing complex in humans (St-Pierre et al. 2005). However, the organisation into sub-complexes is different from yeast (Figure 1.5) and consists of the CFI and CFII, the cleavage and polyadenylation specificity factor (CPSF), and the cleavage stimulation factor (CstF). The

AAUAAA hexamer is recognised by CPSF (Bienroth et al. 1991; Murthy and Manley 1992; Shi et al. 2009) via its subunits Wdr33 and CPSF-30 (Schönemann et al. 2014; Chan et al. 2014). CPSF additionally binds U-rich sequences upstream of the hexamer via Fip1 (Kaufmann et al. 2004; Martin et al. 2012; Lackford et al. 2014). Together with CPSF-160, Fip1 recruits the pA polymerase (PAP) to the 3'end processing complex (Murthy and Manley 1992; Kaufmann et al. 2004). Also part of the CPSF sub-complex is the protein that cleaves the transcript, however, for a long time its identity remained unclear as both CPSF-73 or CPSF-100 contain metallo-β-lactamase domains. Finally, cleavage activity could be demonstrated for CPSF-73, and CPSF-100 lacks amino acids that are necessary for the reaction (Callebaut et al. 2002; Mandel et al. 2006).

Another factor was suggested to bind U-rich sequences upstream of the AAUAAA hexamer: CFI, which consists of CFI-25 and either CFI-59 or CFI-68. All three proteins have been implicated in RNA interaction, recognising the consensus motif UGUA (Rüegsegger et al. 1996; Martin et al. 2012). This was suggested to regulate pA site selection (Yang et al. 2010, 2011). Similar to yeast, CFI functions early in 3'end processing and helps to recruit CPSF proteins (Rüegsegger et al. 1996, 1998). Interestingly, both CFI-59 and CFI-68 contain arginine-serine (RS) repeats and associate with the spliceosome which is thought to stimulate 3'end processing (Rüegsegger et al. 1998; Rappsilber et al. 2002; Zhou et al. 2002).

The CstF sub-complex was found to bind the DSE downstream of the AAUAAA hexamer via its subunit CstF-64 (MacDonald et al. 1994). Because the DSE elements are highly variable, only a subset of genes that have a GU-rich region in the first 20 nt downstream of the pA site followed by a U-rich sequence are bound by CstF. Whether the complex can be recruited by protein-protein interactions only is currently unknown (Yao et al. 2012).

Two additional proteins were only recently identified as *bona fide* members of the 3'end processing machinery - Rbbp6 and PP1 (Shi et al. 2009). Rbbp6 most likely associates with CstF and engages in direct RNA interaction, however, without any

apparent sequence specificity (Di Giammartino et al. 2011). Furthermore, inhibition of PP1 seems to decrease polyadenylation of pre-mRNA (Shi et al. 2009).

The mechanism of 3'end processing and transcription termination of mRNAs was suggested to be analogous to yeast (Figure 1.6). After cleavage by the CPSF complex, the Rat1 homologue Xrn2 is thought to degrade the nascent 3' cleavage product until it reaches Pol II which is subsequently removed from the template (West et al. 2004). Interestingly, the human Sen1-homologue Senataxin (SETX) was proposed to be involved in mRNA termination by unwinding the RNA:DNA hybrid formed at downstream G-rich regions after cleavage. In contrast to Nrd1-dependent termination of sn/snoRNAs in yeast, however, SETX would still require Xrn2 for termination (Skourti-Stathaki et al. 2011).

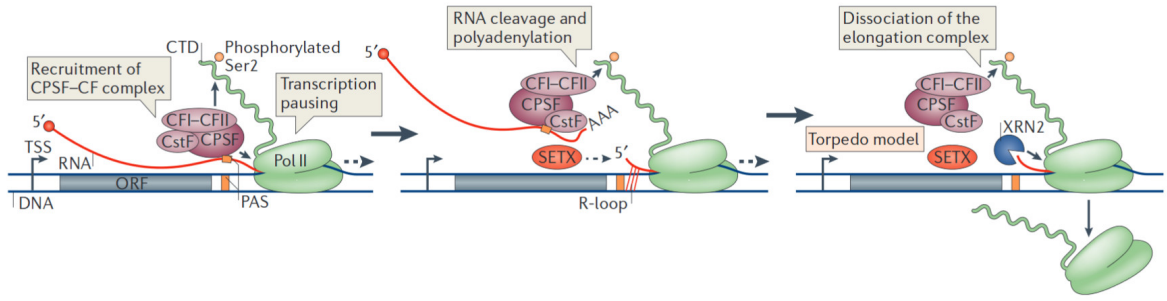


Figure 1.6: Cleavage, polyadenylation and termination of mRNAs in metazoans
Components of the 3'end processing machinery are recruited to the end of transcription units via interaction with S2P-Pol II and RNA. After the nascent transcript has been cleaved, the 5' cleavage product is polyadenylated and the 3' product is degraded by Xrn2. At G-rich regions, the RNA:DNA hybrid is unwound by Senataxin (SETX). Following this, Pol II is removed from the DNA template. TSS: transcription start site; ORF: open reading frame; CPSF: cleavage and polyadenylation specificity factor; CF: cleavage factor; CstF: cleavage stimulation factor; Reprinted with permission from (Porrua and Libri 2015).

Mammalian cells likely use a different mechanism than yeast cells to process and terminate their non-coding genes as no analogues of Nrd1 or Nab3 have been identified and knock-down of the Sen1 homologue SETX has no effect on snRNA gene termination (Porrua and Libri 2013; O'Reilly et al. 2014). Their processing, however, also differs from mRNAs as these transcripts are not polyadenylated. Instead of pA signals, a so-called 3'box is located at the end of snRNA genes which dictates processing and termination (Figure 1.7). This RNA sequence is recognised by a protein complex called Integrator that contains homologues of some CPSF proteins, such as INT11, which can cleave the

nascent transcript (like CPSF-73) (Baillat et al. 2005; Ezzeddine et al. 2011). In addition, this subunit also directly binds to S7P-Pol II (Egloff et al. 2008). Furthermore, Xrn2 only seems to play a minor role in terminating these genes, instead a nucleosome-free region spanning the whole transcription unit combined with the action of NELF is thought to terminate transcription (O'Reilly et al. 2014; Yamamoto et al. 2014; Fong et al. 2015). Furthermore, the CFII subunits Clp1 and Pcf11 as well as Ssu72 were suggested to be involved (Hallais et al. 2013; O'Reilly et al. 2014). Analogous to yeast genes, processing occurs early in transcription probably by linking the cap to elongating Pol II (Porrua and Libri 2015). In conclusion, processing of human snRNAs appears to have more in common with mRNAs than with yeast sn/snoRNAs.

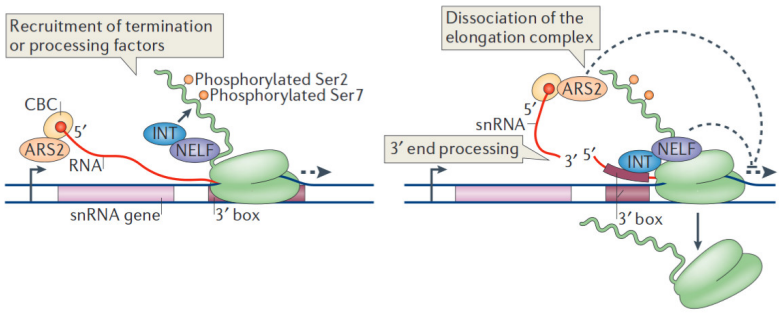


Figure 1.7: Processing and termination of snRNAs in metazoans
In metazoans, the Integrator (INT) complex is recruited to the 3'end of snRNA genes which leads to 3'end processing after the 3'box has been transcribed. The negative elongation factor (NELF) as well as other factors such as the cap-binding complex (CBC) then leads to termination of Pol II ~ 1 kb further downstream. Reprinted with permission from (Porrua and Libri 2015).

1.3.4 Known 3' end processing and termination factors in *S. pombe*

RNA 3'end formation in *S. pombe* is less well understood compared to humans and *S. cerevisiae*. Early studies that aimed to characterise the *cis* elements necessary for 3'end processing of mRNAs in fission yeast tried to swap the human and fission yeast signals as well as the fission and budding yeast signals. In the latter case, both yeasts could terminate mRNAs efficiently using each others pA signals, however, this was not the case for humans. Neither were human cells able to properly process transcripts containing fission yeast pA signals nor vice versa. This indicates that, overall, fission yeast pA signals

resemble *S. cerevisiae* signals more than the ones found in mammalian cells (Humphrey et al. 1991). Two recent genome-wide studies aimed to identify the *cis* elements that are used by *S. pombe*. Juan Mata experimentally mapped all pA sites by using a method that sequences polyadenylated RNAs from the 3'end (Mata 2013). Instead of a precise cleavage site, he found considerable heterogeneity between individual transcripts indicating that most genes rather have a region of cleavage. This is thought to result from an imprecise nature of the 3'end processing machinery and has been observed in other organisms as well (Pauws et al. 2001; Tian et al. 2005). He also found a general enrichment for adenines and uridines around the cleavage site and, similar to budding yeast and mammalian cells, the sequence AAUAAA was the most abundant hexamer upstream of the pA site (with a peak at 22 nt) (Mata 2013). Schlackow et al. took this analysis even further. They could identify the same hexamer, however, in addition they found a UGUA motif which, again, has the same consensus as the CFI recognition site in human cells and resembles the U-rich regions found in *S. cerevisiae*. Overall and consistent with previous studies, they found the pA signals to be more similar to budding yeast than to mammalian cells (Schlackow et al. 2013 and Figure 1.8).

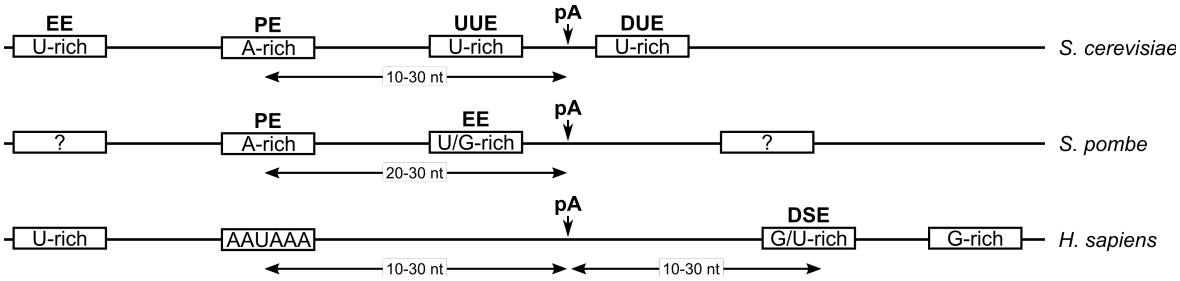


Figure 1.8: Comparison of *cis* elements required for 3'end processing
The RNA *cis* elements which are recognised by the 3'end processing complex and are required for proper cleavage and polyadenlyation are shown for *S. cerevisiae* (top), *S. pombe* (middle) and *H. sapiens* (bottom). *S. pombe* has been studied in less detail than the other two organisms and the question marks indicate that more elements could potentially be necessary which have not yet been discovered. EE: efficiency element; PE: positioning element; UUE: upstream U-rich element; DUE: downstream U-rich element; DSE: downstream element

The protein components that are required for mRNA 3'end processing have been identified in *S. pombe* by purification of homologues of known factors in other organism (Ysh1 and Swd2.2) as well as by homology (Roguev et al. 2004; Wood et al. 2011). Overall, all proteins that are conserved in both humans and budding yeast can be found in *S. pombe* as well (Table 2). However, no evidence for their involvement in 3'end processing is available at the experimental level for most of them. The mammalian CFI complex which is not conserved in budding yeast can also not be found in fission yeast. Concerning proteins that are present in budding yeast but not humans, some are found in fission yeast and some are not. Specifically, the *S. cerevisiae* CFIA complex (but not CFIB) is conserved in *S. pombe*, which is in contrast to human CFI. However, not all CPF components can be found in fission yeast. The number of proteins involved in 3'end processing is therefore closer to mammalians. This therefore puts *S. pombe* in-between the other two organisms. How the factors are organised into different sub-complexes has not been determined in fission yeast. This, however, would help to better understand where *S. pombe* stands from a evolutionary point of view and how the individual proteins promote 3'end processing in this organism.

Considering that the mechanism of mRNA 3'end processing and termination is somewhat conserved from *S. cerevisiae* to human, *S. pombe* is expected to function in the same way. Further supporting this, a Rat1/Xrn2-homologue was identified in fission yeast and this protein, which was named Dhp1, can rescue a *rat1* temperature sensitive (ts) mutant of *S. cerevisiae* (Sugano et al. 1994).

Table 2: Components of the 3' end processing machinery in human, budding and fission yeast
Subcomplex are indicated as follows: CF: cleavage factor, CPF: cleavage and polyadenylation factor, PF: polyadenylation factor, CPSF: cleavage and polyadenylation specificity factor, CstF: cleavage stimulation factor. Rows show homologous proteins and proteins in grey have not been implicated in 3'end processing. The table is based on Roguev et al. (2004), Shi et al. (2009), Wood et al. (2011), and Chan et al. (2011).

S. cerevisiae	***H. sapiens***	***S. pombe***
-	CFI-68kDa	-
-	CFI-59kDa	-
-	CFI-25kDa	-
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
Cft1/Yhh1 (CFII)	CPSF-160	Cft1
Cft2/Ydh1 (CFII)	CPSF-100	Cft2
Brr5/Ysh1 (CFII)	CPSF-73	Ysh1
Pta1 (CFII)	Symplekin	Pta1
Pfs2 (PFI)	Wdr33	Pfs2
Yth1 (PFI)	CPSF-30	Yth1
Fip1 (PFI)	Fip1 (CPSF)	Iss1
Mpe1 (CPF)	Rbbp6	Mpe1
Ref2 (CPF)	-	-
Syc1 (CPF)	-	-
Swd2 (CPF)	Wdr82	Swd2.2
Ssu72 (CPF)	Ssu72	Ssu72
Glc7 (CPF)	PP1	Dis2
-	-	Ppn1
Pti1 (CPF)	-	-
Pap1	PAP	Pla1

Processing of sn/snoRNAs in *S. pombe* seems to be different from the main pathways in *S. cerevisiae* and human. The RNase III-like nuclease Pac1 which is homologous to the *S. cerevisiae* protein Rnt1 is the only sn/snoRNA 3'end processing enzyme described to date. This is in contrast to *S. cerevisiae*, in which Rnt1 seems to play only a minor role in processing of both mRNAs and snRNAs (see above). It was demonstrate that Pac1 cleaves at least some sn/snoRNAs after a conserved hairpin which provides an entry point for both the 3'-5' degrading exosome complex and the 5'-3' exonuclease Dhp1. While the former is necessary for transcript trimming and maturation, the latter leads to termination of Pol II (Potashkin and Frendewey 1990; Zhou et al. 1999; Rotondo and Frendewey 2001; Nabavi and Nazar 2008, 2010b, 2010a).

An interesting fail-safe termination mechanism has been proposed in *S. pombe*, which acts on backtracked Pol II if re-initiation of transcription fails. In that case, it was hypothesised that the exposed 3'end of the transcript provides an entry point for degradation by the exosome which, in turn, leads to termination of Pol II (Lemay et al. 2014; Shah et al. 2014). However, whether the exosome is directly involved in transcriptional termination remains to be determined.

1.3.5 Transcription termination: the allosteric vs. torpedo model

In the previous sections, I described proposed mechanisms that lead to Pol II termination. However, how exactly the polymerase complex is removed from the DNA template is still not entirely understood. Almost 30 years ago, two different models were proposed: the allosteric and the torpedo model. The former suggests that Pol II dissociates as a result of conformational changes upon transcription of the pA site which is caused by the dissociation and/or association of factors (Logan et al. 1987). The torpedo model proposes that cleavage of the nascent transcript provides an uncapped 5'end which is degraded by a 5'-3' exonuclease until it catches up with the elongating Pol II and leads to its dislodgement (Connelly and Manley 1988; Proudfoot 1989). The two mechanisms are not mutually exclusive and lots of research has been conducted in favour of either model.

The allosteric model is supported by the observation that elongation factors are removed from Pol II before termination occurs which could destabilise the elongation complex (Ahn et al. 2004; Kim et al. 2004a). In a different scenario, the binding of proteins to Pol II could lead to the terminating conformational change. This is supported by experiments showing that the CTD is required for optimal termination (McNeil et al. 1998) and the CTD-binding protein Pcf11 seems to be able to dislodge Pol II *in vitro* (Zhang et al. 2005).

In support of the torpedo model, mutants of the 5'-3' exonuclease show read-through transcription *in vivo* (Kim et al. 2004b; West et al. 2004). Also, this model predicts a kinetic competition between the exonuclease and Pol II to allow the exonuclease to catch up with the elongating polymerase after transcript cleavage. In support of this, mutants that slow down transcription were found to terminate closer to the pA site that faster mutants (Fong et al. 2015). However, several studies indicate that the Rat1-dependent mechanism is not sufficient to dislodge Pol II (Kim et al. 2004b, 2006; Dengl and Cramer 2009). Also, in this model cleavage would be a prerequisite for termination. However, several mutations have been observed that disrupt cleavage without affecting termination (Dichtl et al. 2002; He et al. 2003). Furthermore, it was proposed that the removal of RNA itself is what triggers termination, however, degrading the RNA *in vivo* by a different exonuclease does not dismantle Pol II (Luo et al. 2006).

Instead, a combined allosteric-torpedo model has been proposed (Luo et al. 2006; Zhang et al. 2015). Zhang et al. could demonstrate that termination is completely inhibited by α-amanitin which suggests that it blocks an essential conformational change. Furthermore, they suggest that this conformational change can be brought about by either Pcf11 which requires pA signals (but not necessarily cleavage) or by a 5'-3' exonuclease which requires cleavage of the transcript (Zhang et al. 2015).

As described above, a third mechanism acts on non-coding genes in *S. cerevisiae*. In that case, Pol II is also in kinetic competition with a termination factor, however, in that case it is not with an exonuclease but with the helicase Sen1 (Hazelbaker et al. 2013). Similar to bacterial rho-dependent termination, Sen1 "chases" down Pol II thereby unwinding the DNA:RNA hybrid in a ATP-dependent reaction. Furthermore, Sen1 was shown to be able to dislodge Pol II *in vitro* (Porrua and Libri 2013; Martin-Tumasz and Brow 2015). In contrast, the human homologue of Sen1 still requires the exonuclease Xrn2 for termination (see above) and whether the *S. pombe* homologue of Sen1 has any function in transcription has not been tested. This, therefore, might constitute a mechanism that is unique to lower eukaryotes.

1.4 The role of CID-containing proteins in transcription termination

One specific protein domain seems to be particularly important for transcription termination: the CTD-interacting domain (CID) which can directly bind to certain phosphorylation/isomerisation types of the CTD. In *S. cerevisiae*, the three known CID-containing proteins are Nrd1, Pcf11, and Rtt103 (www.uniprot.org) which all participate in transcription termination of Pol II (which was described in section 1.3.2). In *S. pombe* and humans, one or more homologues can be found for each protein (Figure 1.9). It should be mentioned, however, that additional CID-containing proteins are present in these organisms (www.uniprot.org) which do not seem to play any role in transcription termination.

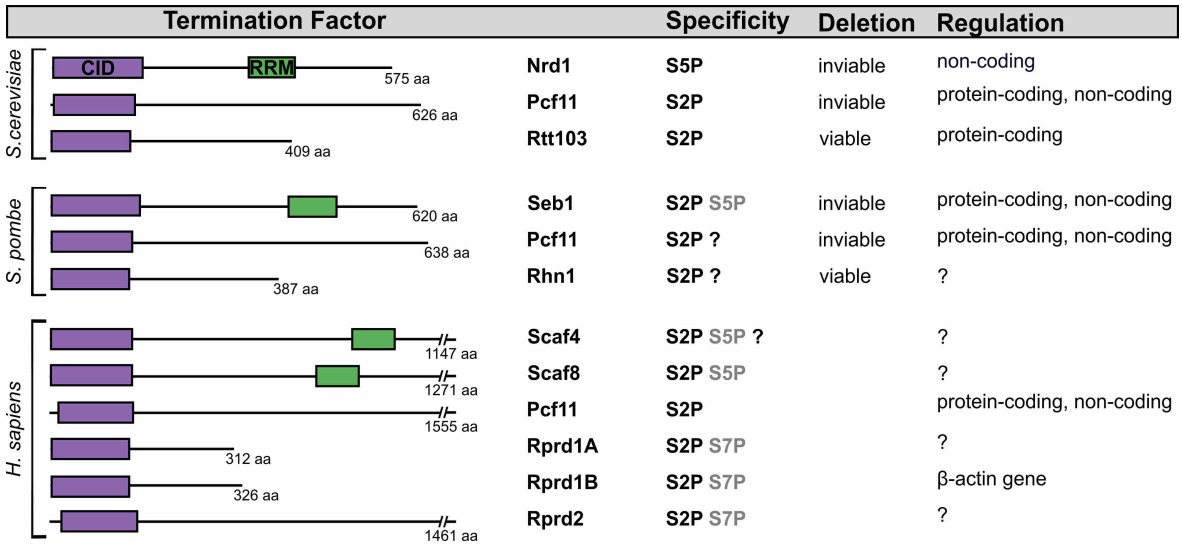


Figure 1.9: CID-domain containing proteins in *S. cerevisiae* and their homologues in *S. pombe* and *H. sapiens*
Schematics showing known CID-containing proteins from *S. cerevisiae* and all homologues in *S. pombe* and *H. sapiens.* The domain organisations are approximately to scale and CTD binding preferences are based on this study (for Seb1, see later), published data (Vasiljeva et al. 2008a; Noble et al. 2005; Lunde et al. 2010; Becker et al. 2008; Mei et al. 2014; Ni et al. 2014) or inferred from sequence alignments (indicated by a question mark). The affinity to grey coloured modifications is lower than to black coloured modifications. The class of transcripts that are terminated by the corresponding protein is shown, a question mark indicates that no studies are published that show any involvement in transcription termination for this protein. For budding and fission yeast, it is also indicated whether the protein is essential for cell viability. CID: CTD-interacting domain; RRM: RNA recognition motif

S. cerevisiae Pcf11 and Rtt103 bind to S2P-CTD (Meinhart and Cramer 2004; Lunde et al. 2010) which reaches its peak late in transcription while Nrd1 recognises S5P-CTD (Vasiljeva et al. 2008a) which occurs early in transcription (Figure 1.1). In addition, Y1P inhibits binding of CID proteins during elongation thereby preventing premature recruitment of termination factors (Mayer et al. 2012; Schreieck et al. 2014). Furthermore, the isomerisation state of the CTD also plays a role in CID factor binding: while Pcf11 and Rtt103 can only bind to the all-*trans* isomer (Noble et al. 2005; Lunde et al. 2010), Nrd1 seems to prefer the S5-P6 bond to be in the *cis* conformation (Singh et al. 2009; Kubicek et al. 2012). An emerging view is therefore, that the termination pathway is dictated by the phosphorylation marks (and proline states) of Pol II and therefore the position within the gene. This is also consistent with the observation that the NNS pathway predominantly terminates short RNAs (like CUTs and sn/snoRNAs) while the Pcf11-dependent CF-CPF pathway acts on the relatively longer mRNA genes (see section 1.3.2).

When it comes to Rtt103, it's main function seems to help recruit Rat1 to S2P-Pol II after transcript cleavage and Y1 dephosphorylation by the CF-CPF (Schreieck et al. 2014). However, *rtt103* deletion does not result in a very strong termination defect (Kim et al. 2004b; Luo et al. 2006). Consistent with this, Pcf11 was shown to also be able to recruit Rat1 (Luo et al. 2006), which indicates that Rtt103 is dispensable for termination. The human homologues, Rprd1A, Rprd1B, and Rprd2 were shown to play a role in dephosphorylation of S5P and only one study has been published so far that indicates an involvement of Rprd1B in transcription termination (Morales et al. 2014; Ni et al. 2014). It is therefore not clear if Rprd1A, Rprd1B and Rprd2 have any function in this process.

While Nrd1 itself is an essential protein, its CID domain is not (Vasiljeva et al. 2008a). Recruitment of Nrd1 to the CTD rather increases the efficiency of termination than being necessary for it (Tudek et al. 2014). And Pcf11 has been shown to be important even for the NNS termination pathway (Morlando et al. 2002; Kim et al. 2006; Grzechnik et al. 2015). This argues that Pcf11 can dislodges Pol II from the template independent of the gene class and that this function is also conserved in humans (see section 1.3.3). In

contrast to humans, however, the NNS complex can act as a fail-safe termination pathway. It was demonstrated that the CF-CPF and NNS complexes have partially overlapping sequence requirements and can therefore be recruited to each others binding site. Which complex is then utilised for termination only depends on the distance from the TSS (Gudipati et al. 2008; Porrua et al. 2012). This is consistent with reports that Nrd1 and Nab3 regulate several hundreds of mRNAs via interaction with their 3'UTR, arguing for a kinetic competition between the two complexes (Webb et al. 2014).

Taken together, this would lead to the following model for termination of NNS-dependent genes: Nrd1 is recruited shortly after transcription initiation via S5P-Pol II and RNA interactions. In fact, simultaneous binding to both was shown to be required for Nrd1 function (Heo et al. 2013; Tudek et al. 2014). For this interaction to occur, the S5-P6 bond of the CTD would need to be in the *cis* isoform which also favours the recruitment of the phosphatase Ssu72 (Xiang et al. 2010; Werner-Allen et al. 2011). The combined action of Ssu72 in removing S5P and Ess1 in increasing the amount of *trans* isoforms would then lead to the dissociation of Nrd1 and association of Pcf11 (Singh et al. 2009). Because this happens so early in transcription, Pol II is thought to not have transitioned into an elongation complex yet. This would then lead to binding of Sen1 to the CTD and Pol II release via ATP-dependent unwinding of the RNA:DNA hybrid (Grzechnik et al. 2015).

After the Pcf11-CID was found to interact with RNA directly, a model was proposed upon which Pcf11 would lead to Pol II dissociation via forces that are transmitted onto the polymerase through simultaneous interaction with both the CTD and RNA (Zhang et al. 2005). This, however, is inconsistent with later reports demonstrating that the CTD and RNA interactions are mutually exclusive (Hollingworth et al. 2006). Instead, in addition to protein-protein interactions, RNA binding could help to recruit Pcf11 to the 3'end of genes before the Y1P marks have been removed from the CTD. This is consistent with reports that mutations in the CID only affect termination but not 3'end processing of mRNAs. In fact, the 3'end processing and termination function of Pcf11 seem to be independent of one another (Sadowski et al. 2003).

1.5 Objectives of this work

If the function of Rtt103 is not absolutely required, termination of short RNAs is not achieved by Nrd1 in *S. pombe* and higher eukaryotes, and Pcf11 is able to terminate both mRNAs and ncRNAs *in vivo*, then why are Nrd1-like proteins which contain a CID and an RNA recognition motif (RRM) so highly conserved? Not only is the Nrd1 homologue Seb1 essential, most metazoans have evolved two copies of the protein. In this thesis, I aim to characterise Seb1 in the model organisms fission yeast and to test whether it plays any role in transcription termination.

The only study about Seb1 to date describes its involvement in the maintenance of constitutive heterochromatin at centromeric regions (Marina et al. 2013). This is reminiscent of Nrd1 which also functions in maintaining the heterochromatic state of permanently silenced regions through immediate termination of any transcription initiated by Pol II (Vasiljeva et al. 2008b). Furthermore, in her *DPhil* thesis which she conducted in the Vasiljeva lab, Dr Beth Watts described a mechanisms by which Seb1 helps to regulate gene expression via local, constitutive heterochromatin that is established in response to environmental changes (Watts 2015). Using a different mechanism, Nrd1 was also implicated in the response to changing nutrient availability (Darby et al. 2012; Webb et al. 2014). Seb1 therefore seems to have similar functions to Nrd1.

Initial experiments performed by Dr Beth Watts showed that Seb1 co-purifies with several factors of the canonical 3'end processing machinery. Furthermore, neither proteins of the exosome complex nor Nab3 or Sen1 homologues could be identified as binding partners. This suggests that Seb1 functions differently in transcription than Nrd1. In addition, Seb1 was found to localise to the 3'end of at least three protein-coding genes (*adh1*, *rps2202*, and *pho1*) as determined by chromatin immunoprecipitation (ChIP) (Watts 2015). Taken together, this suggests that, in contrast to Nrd1, Seb1 is involved in 3'end processing and/or transcription termination of long, protein-coding genes.

In this thesis, I undertake *in vivo* studies in order to understand whether Seb1 indeed plays a role in 3'end processing and/or transcription termination and which transcripts are targeted by the protein. This is accompanied by extensive *in vitro* studies which include biochemical assays and structural characterisation of the two domains of Seb1, the CID and the RRM. Lastly, the results obtained *in vitro* are then utilised to elucidate in detail how Seb1 functions in the cells with the aim to determine whether Seb1 really constitutes a general transcription termination factor in fission yeast.

CHAPTER 2: *IN VIVO* RECRUITMENT PROFILE AND NECESSITY OF SEB1 FOR CELL VIABILITY

As discussed above (section 1.3.2), the *S. cerevisiae* protein Nrd1 functions mainly in terminating short ncRNAs. This is in contrast to preliminary data from our group which found the *S. pombe* homologue Seb1 to be associated with the 3'end of protein-coding genes and to interact with components of the CF-CPF complex (Watts 2015). This strongly suggests that Seb1 has a different role in *S. pombe* than Nrd1 in *S. cerevisiae.*

Nrd1 elicits its function with the help of two domains, a Pol II-interacting CID and an RNA-binding RRM domain which are both conserved in Seb1. Interestingly, based on domain predictions and in homology to Nrd1, these domains constitute only ~ 36% of Seb1. A disorder prediction indicates that most regions outside the two domains are disordered (Figure 2.1). There is one exception, however, which is the region around the predicted RRM especially C-terminally. In the case of Nrd1, it was demonstrated that, in addition to the canonical RRM, the region directly downstream of it also participates in RNA interaction (Bacikova et al. 2014). It is therefore likely that the ordered region following the RRM constitutes a second RNA-binding region in Seb1, too.

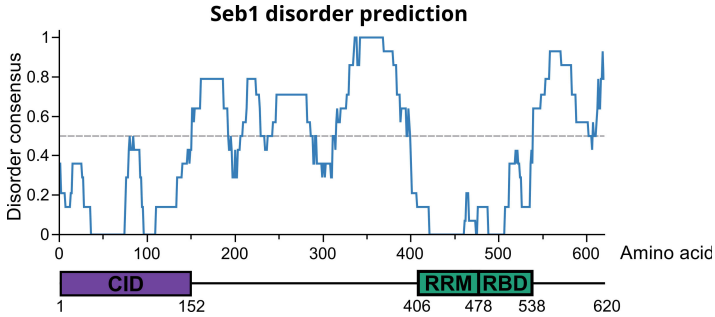


Figure 2.1: Seb1 is largely disordered with two structured domains
A disorder prediction of the full-length Seb1 protein was performed using the Dismeta server (Huang et al. 2014). According to www.uniprot.org, Seb1 contains two domains and an additional ordered region after the second domain which are schematically shown below the plot, with first and last amino acid indicated. Generally, a disorder consensus above 0.5 indicates disorder (shown by grey dotted line).
CID: CTD-interacting domain; RRM: RNA-recognition motif; RBD: putative RNA-binding domain

While Nrd1 itself is essential, deletion of the CID domain has no effect on cell growth (Vasiljeva et al. 2008a). In contrast, the RRM domain is required for cell viability (Steinmetz and Brow 1998). This indicates that binding of Nrd1 to RNA is more important than to the CTD and that the latter interaction is not absolutely necessary to promote the function of Nrd1. Similarly, Seb1 was also shown to be essential for cell viability, however, more detailed studies looking at the individual domains have not been conducted (Mitsuzawa et al. 2003). Therefore, I wanted to test which regions of Seb1 are necessary for cell viability as this could help us understand which interactions are required for the function of Seb1 *in vivo*.

2.1 The structured regions of Seb1 are essential for cell viability

To understand which parts of Seb1 are functionally important, I designed seven deletion mutants (Figure 2.2A). Since deletion of only parts of Seb1 could be lethal, a system was used in which FLAG-tagged wild-type (WT) *seb1* was expressed in addition to an TAP-tagged mutated version (Figure 2.2A). WT-*seb1* was, however, under control of the repressible *nmt1* promoter which shuts down gene expression in the presence of thiamine. In contrast, mutated *seb1* contained the endogenous *seb1* promoter and was located at an ectopic position in the genome. Therefore, cells are always viable in medium lacking thiamine and upon shift to thiamine-containing medium, the effects of the deletions can be observed.

The seven different deletion mutants were lacking either a whole domain, the ordered region after the RRM domain, parts of the disordered region between the domains, or the very C-terminus of the protein (as indicated in Figure 2.2A). Expression of all mutated as well as WT-Seb1 was confirmed by Western blotting (Figure 2.2B). As expected, all proteins can be detected in the absence of thiamine, however, WT-Seb1 levels are considerably lower upon shift to thiamine-containing medium for 24 h. Expression of the deletion mutants is not affected in these conditions. This confirms that

the expression system works as expected and can be used for further analyses.

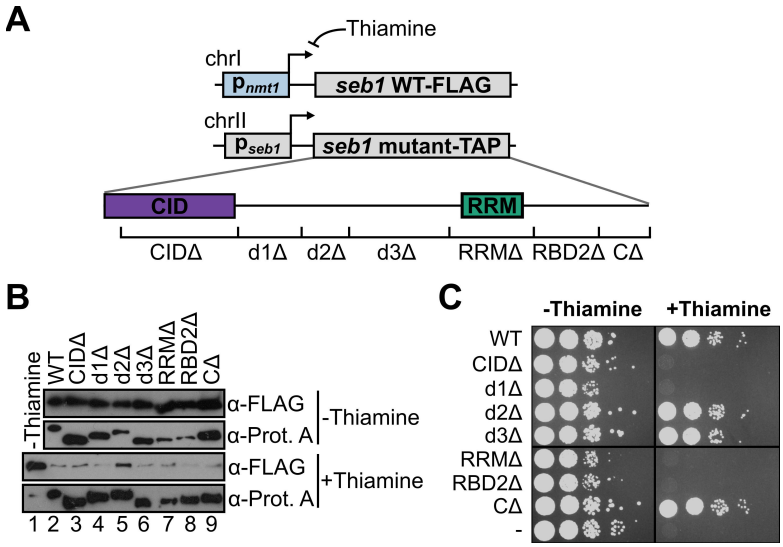


Figure 2.2: Both domains of Seb1 are required for cell viability
(A) Schematic representation of the expression system used in order to assess effects of Seb1 deletions on cell viability. WT-*seb1* was expressed from a thiamine-repressible *nmt1* promoter while the indicated deletions were expressed from the endogenous promoter at an ectopic location. The deletion schematics is approximately to scale.
(B) Expression of the WT and mutated proteins in media containing and lacking thiamine was tested by Western blotting. Due to the different tags used, the WT copy can be specifically detected using an α-FLAG antibody, while the mutated copy reacts with α-Protein A (α-Prot. A) antibodies.
(C) Spot test assessing the effects of the mutation on plates containing and lacking thiamine is shown. For this, serial dilutions of the indicated strains were pipetted onto agar plates and incubated at 30°C for several days. A strain containing two wild-type copies (WT) and a strain lacking a second, non-WT copy (-) were used as controls.
CID: CTD-interacting domain; d1-3: disordered regions 1-3; RRM: RNA recognition motif; RBD2: RNA-binding domain 2; C: C-terminus

Growth of all strains was determined on agar plates ± thiamine (Figure 2.2C). As expected from the protein expression analysis (Figure 2.2B), all strains grew normally on plates lacking thiamine, in which case the WT copy is expressed. A control strain containing only one copy of the *seb1* gene (WT) grows at the same rate as all other strains, showing that overexpression of Seb1 does not influence growth rates. On plates containing thiamine this strain is not viable, confirming that Seb1 is an essential protein. In addition, several deletion mutants do not support growth either, including both domain deletions. This indicates that interactions between Seb1 and Pol II as well as RNA are both necessary for cell viability which is in contrast to Nrd1 where only RNA binding is required. Furthermore, the regions directly following both domains of Seb1 are also

essential. In the case of the disordered region 1 which is positioned after the CID, this could be caused by effects on the CID domain (e.g. it cannot properly fold any more). However, it is also possible that it constitutes an essential protein-binding site as is the case in Nrd1. There, deletion of the region following the CID results in slow growth and lethality at high temperatures which is thought to be caused by a loss of Nab3 binding (Vasiljeva et al. 2008a). In the case of the region following the RRM domain, however, lethality could be caused by decreased or abolished RNA binding as this part could be involved in RNA interactions, similarly to the region around the RRM domain in Nrd1.

2.2 Seb1 binds to the 3'end of genes genome-wide

After it has been established that both the CID and the RRM domain are required for cell viability, I wanted to determine the recruitment profile of Seb1. Nrd1 only efficiently terminates transcripts up to ~ 1 kb after the TSS (Gudipati et al. 2008), however, single-gene ChIP found Seb1 to localise to the 3'end of two genes that are longer than this (*adh1* and *pho1,* Watts 2015). To assess whether Seb1 plays a general role in transcription and which type of transcripts it binds, genome-wide studies were used which provide a powerful tool to understand recruitment profiles to the whole transcriptome.

Therefore, ChIP combined with next-generation sequencing (ChIP-Seq) was conducted which is used to pull-down a protein of interest and to determine its association with chromatin through sequencing of the co-purified DNA (Robertson et al. 2007). Because Seb1 has no DNA binding domain, the association with chromatin is likely to be mediated by other proteins such as Pol II. In this case, the Pol II-CTD can either be bound by Seb1 directly or indirectly via interactions with CF-CPF proteins. ChIP-Seq was therefore used to assess where Seb1 plays a role during active transcription. The experiment was performed by Dr Beth Watts (Vasilieva lab), while I did data analysis. However, it should be noted that ChIP-Seq is not strand-specific.

Even though Seb1 probably does not interact with DNA, it is expected to bind RNA via its RRM domain. Therefore, in order to identify which transcripts are targeted by Seb1

and to elucidate its interaction sites on RNA, another genome-wide method was utilised. Photo-activatable ribonucleotide-enhanced crosslinking and immunoprecipitation (PAR-CLIP) allows the purification and subsequent sequencing of any RNA that is directly bound by Seb1. Before harvesting, the cells were treated with 4-thiouridine (4-SU) that is incorporated into nascent RNA. During the sequencing reaction, the crosslinked 4-SU base is recognised by the reverse transcriptase as a guanine, resulting in thymidine to cytosine mutations (Hafner et al. 2010). PAR-CLIP therefore allows to determine the exact binding position of Seb1 on all cellular RNAs, however, it does not distinguish between co- and post-transcriptional events. Combining both methods should therefore give as complete a picture of the *in vivo* localisation of Seb1 as possible. Seb1 PAR-CLIP was performed and analysed by Dr Carlo Bäjen (Cramer lab, MPI Göttingen, Germany), bioinformatic comparisons with the ChIP-Seq data was done by me. All analyses were conducted using the most recent annotations published for *S. pombe* (Eser et al. 2016).

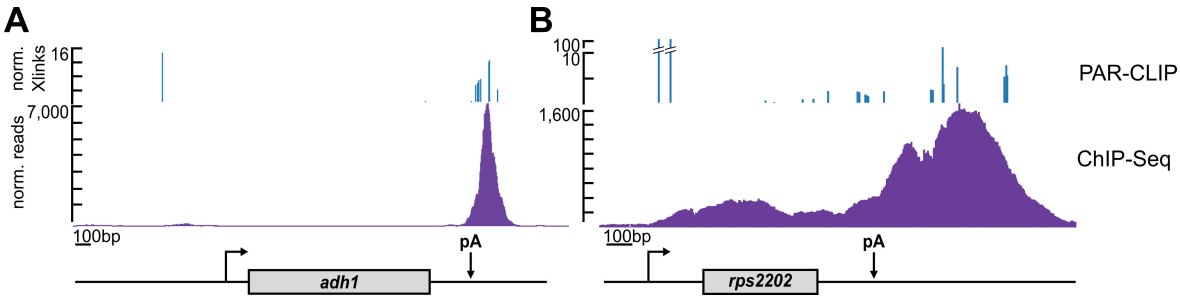


Figure 2.3: Seb1 is recruited to the 3'end of genes in ChIP-Seq and PAR-CLIP
Crosslinking sites of Seb1 were determined by PAR-CLIP and are shown in blue. The height of the peak represents the amount of crosslinks found at any given site, normalised to the expression level of the RNA. Recruitment of Seb1 to the chromatin was assessed by ChIP-Seq and is shown in purple. The profile was normalised to 0.01% of the input. The position of the crosslinks and ChIP peaks in relation to the two protein-coding genes *adh1* (A) and *rps2202* (B) is indicated by the schematics below the profiles.

To first test whether the result of the previously performed single-gene ChIP (Watts 2015) could be reproduced in the genome-wide ChIP-Seq, binding profiles of Seb1 on the two house-keeping genes encoding the alcohol dehydrogenase 1 (Adh1) and the 40S ribosomal protein S22-B (Rps2202) were assessed (Figure 2.3). Indeed, in both cases recruitment of Seb1 to the 3'end could be detected. Furthermore, many PAR-CLIP

crosslinks are observed in the regions of the ChIP-Seq peaks, indicating that Seb1 directly interacts with RNA during transcription. In addition, PAR-CLIP crosslinks around the TSS can be detected on both genes which do not show any corresponding signal in ChIP. This could be due to either post-transcriptional RNA binding of Seb1 or very weak interactions with the transcription machinery.

Because of the T-to-C mutations introduced at crosslinking sites during library preparation, the exact location of Seb1 can be determined for all transcripts. This also allows to search for a specific sequence that might be recognised by the protein. In a window of ± 25 nt around mutation sites, a clear binding motif could be found (Figure 2.4A).

Next, the average binding profile of Seb1 across all genes was determined (Figure 2.4B). In both ChIP-Seq and PAR-CLIP, clear recruitment of Seb1 can be observed just after the pA site. In the case of PAR-CLIP, the highest amount of crosslinking sites are located ~ 80 nt after the pA site; in the case of ChIP-Seq this occurs further downstream, ~ 160 nt after the pA site. This indicates that Seb1 associates with the transcription machinery only after the binding motif has been transcribed. Whether it associates with RNA and Pol II simultaneously cannot be determined from these data. Interestingly, Seb1 also engages with RNA near the TSS (Figure 2.4A left). This is not accompanied by a corresponding binding event detectable by ChIP-Seq and is therefore consistent with results obtained for *adh1* and *rps2202* (Figure 2.3). This promoter-proximal binding could therefore either occur post-transcriptionally or the interaction with Pol II may be too transient for detection by ChIP.

In order to test where in the genome the UGUA occurs, the regions 500 nt around TSS and pA sites were analysed for the presence of this sequence. As expected, a strong enrichment of this motif can be observed exactly where Seb1 crosslinks in PAR-CLIP (Figure 2.4B). Strikingly however, the motif, though not completely absent, is strongly depleted from the gene body and TSS region. This indicates that the motif, while probably not being the sole determinant, at least helps to recruit Seb1. However, even though

protein-RNA interactions occur close to the TSS, no enrichment of the motif could be detected in that region.

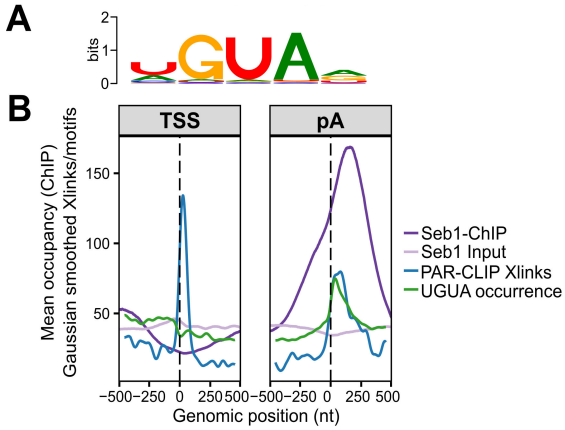


Figure 2.4: Seb1 recognises a motif at the 3'end of genes genome-wide
(A) The binding motif of Seb1 was determined in a window of 25 nt around T-to-C mutation sites detected by PAR-CLIP using the XXmotif server (Luehr et al. 2012). The motif occurs in 42.74% of cases (E-value: $5.34 \cdot 10^{-55}$).
(B) Average occupancy profile of Seb1 at all genes with a minimum length and distance to the neighbouring gene of 500 nt is shown around the TSS and pA site, as indicated. The ChIP input (normalised to library size) is plotted as a background control for the recruitment profile (light and dark purple, respectively). PAR-CLIP crosslinking sites were smoothed using the Gaussian function (blue). A search for the Seb1 RNA binding motif UGUA was performed in the indicated window and plotted after Gaussian smoothing (green). The PAR-CLIP and motif plots were multiplied by 2000 in order to match the scale of the ChIP-Seq profile.

Because genome-wide occupancy profiles only show where a protein binds in relation to the 5' and 3' ends of genes but not how many genes are bound, ChIP peak occurrences and PAR-CLIP crosslinks around the TSS and pA site were compared with the presence of the UGUA motif on individual transcripts (Figure 2.5). Overall, 63% of the 4,228 protein-coding and non-coding genes included in the analysis show recruitment of Seb1 to the region 250 nt ± pA by ChIP-Seq and ~ 36% by PAR-CLIP. These numbers increase to 72% and 45%, respectively, if the quartile with lowest gene expression (as determined by RNA-Seq, see Chapter 5) is omitted from the analysis (n = 2,943). The lower number of binding events detected by PAR-CLIP could be caused by the low abundance of RNA downstream of the pA site since the 3' cleavage product is highly unstable as it is immediately degraded by the exonuclease Dhp1 (*S. cerevisiae* Rat1, human Xrn2). Strikingly though, ~ 74% of the genes that bind Seb1 in PAR-CLIP, also

have detectable peaks in ChIP-Seq, indicating that Seb1 does indeed bind to RNA co-transcriptionally. Also, ~ 81% of the genes that show more than the average number of UGUA motifs in their pA region recruit Seb1 which is detectable by either ChIP-Seq or PAR-CLIP, further emphasising the importance of the binding motif in Seb1 recruitment.

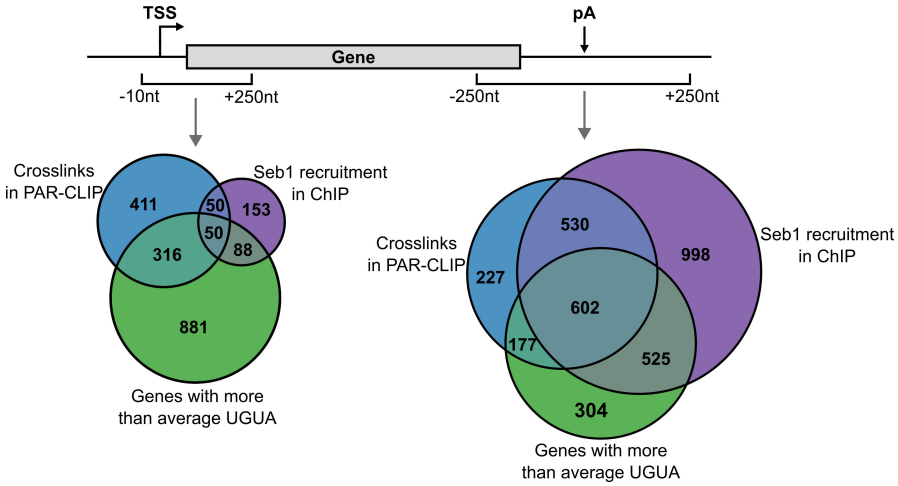


Figure 2.5: The majority of genes recruits Seb1 to their 3'ends during transcription
(Top) Schematic representation of the regions in which Seb1 binding by PAR-CLIP and ChIP-Seq as well as UGUA occurrence was analysed. Protein-coding and non-coding genes of at leas 500 nt length and a minimum distance of 250 nt to their downstream gene were included. No genes with less than 10 nt distance to any neighbouring gene are annotated in *S. pombe*.
(Bottom) The overlap between the amount of PAR-CLIP crosslinking events (blue), ChIP-Seq peak summits (purple), and more than the average number of UGUA motifs (green) in the indicated regions are shown as Venn diagrams.

In the case of the TSS region (from 10 nt before to 250 nt after the TSS), the number of genes showing ChIP-Seq peaks is smaller than those showing PAR-CLIP crosslinks (Figure 2.5). In addition, there is little overlap between Seb1 binding events and UGUA motif occurrence in this region, further supporting the notion that the sequence plays only a minor role in recruiting Seb1 to the TSS. All of this is consistent with the overall recruitment profile shown in Figure 2.4A.

In order to compare the number of binding events at the TSS and pA site, the window that was used to look for binding events around the pA site was reduced to the same size as was used for the TSS (260 nt). Overall, less Seb1 binding events could be detected at the TSS (-10 nt before to 250 nt after) than at the pA site (-50 nt before to

210 nt after, Figure 2.6). In both ChIP-Seq and PAR-CLIP, less than half of the TSS binding events are accompanied by simultaneous binding at the pA site. This indicates that neither pA site binding is caused by TSS binding nor *vice versa* and points towards the possibility that Seb1 has different functions at the 5' and 3' ends.

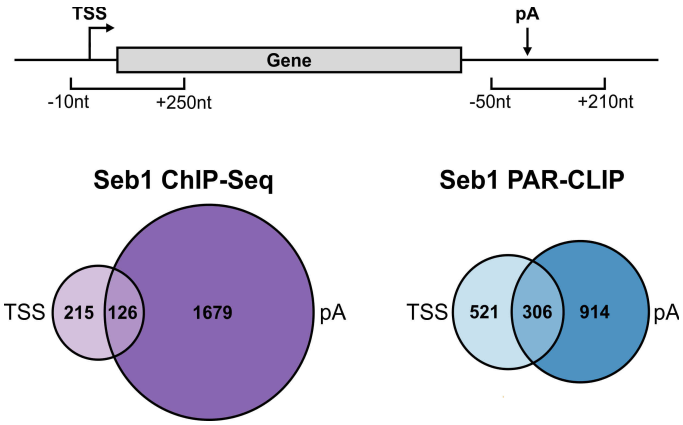


Figure 2.6: Binding at the TSS and pA site are independent of each other
Venn diagram showing the overlap between TSS and pA binding events detected by ChIP-Seq (purple) and PAR-CLIP (blue). The regions used are indicated in the schematics above. The same subset of genes was used as in Figure 2.5.

After establishing that Seb1 is recruited to most genes in the genome, I wanted to check if it is specifically recruited to non-coding genes as well. Therefore, ncRNAs were extracted from the genes used in Figure 2.4 and plotted separately (Figure 2.7A). The same TSS peak exclusively seen by PAR-CLIP can be observed as before. Recruitment of Seb1 to the 3'end of non-coding transcripts is also detectable by both PAR-CLIP and ChIP-Seq. However, the peaks are much broader than on protein-coding genes. Since the 3'end of non-coding genes are not formed by pA site cleavage, this could be due to larger variations in the distance between the annotated 3'ends of the transcripts and the Pol II termination sites. An enrichment of UGUA motifs co-occuring with PAR-CLIP crosslinks can be seen shortly after the 3'end (Figure 2.7A), similar to what was observed for all genes of which the majority is protein-coding (Figure 2.4). This argues that Seb1 is indeed involved in transcription at the 3'end of both, pA-dependent and -independent RNAs. Because 3'end formation is very different between protein-coding and non-coding genes, Seb1 is more likely to be involved in termination rather than 3'end processing.

Finally, the question arises whether Seb1 can also bind antisense to genes. Therefore, the same regions were analysed as before, however, the strand opposite to the transcription unit was used to determine binding profiles (Figure 2.7B). Because ChIP-Seq is not strand-specific, only PAR-CLIP crosslinking events can be used for the analysis. Interestingly, here, too, a peak is visible very close to the TSS and pA sites. This is also accompanied by a higher frequency of UGUA motifs compared to the region antisense to gene bodies, however, no clear peaks are visible in the UGUA profiles.

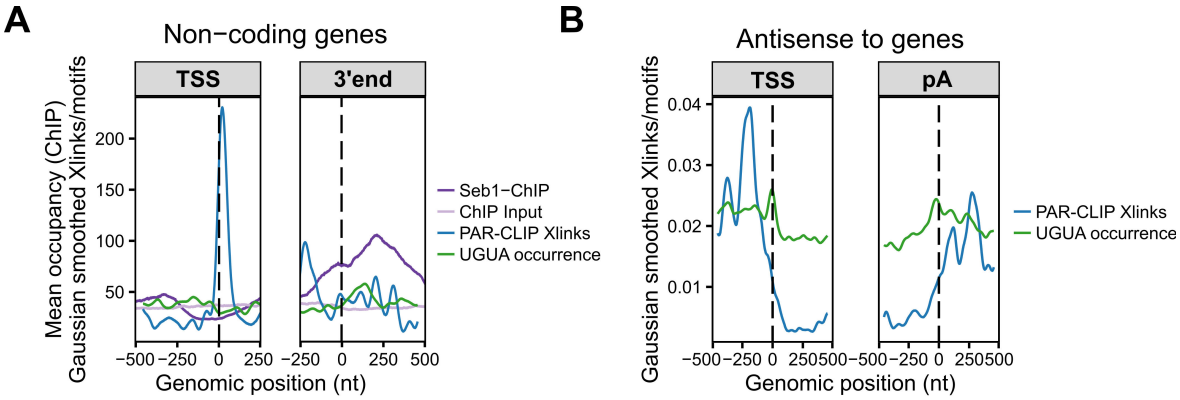


Figure 2.7: Seb1 binds to non-coding genes and antisense to TSS and pA sites
(A) The occupancy profile of Seb1 recruitment as detected by ChIP-Seq and PAR-CLIP as well as UGUA motif occurrence was determined as in Figure 2.4. Because of the fact that ncRNAs are generally shorter than protein-coding genes, only 250 nt into the gene body is shown.
(B) The same subset of genes was used as in Figure 2.4, however, PAR-CLIP crosslinks and presence of the UGUA motif was determined antisense to the indicated regions. Because ChIP-Seq is not strand-specific, the profile was omitted from this plot.

2.3 Seb1 co-purifies with the 3'end processing machinery

In the previous section, Seb1 was clearly shown to be recruited to the 3'end of genes. The Seb1 purifications that were performed by Dr Beth Watts also indicate a direct interaction with the 3'end processing machinery. In this experiment, the highest enrichment was observed for the CFIA protein Rna14 followed by components of CFII (Watts 2015). In order to confirm this interaction, I wanted to test whether Seb1 can also be recovered by purifying the 3'end processing complex. For this, I used a strain that contained a C-terminal HA-TAP-tag on Ysh1 (Figure 2.2A) which is part of the CFII subcomplex of the CPF and performs the cleavage reaction (as discussed in section 1.3).

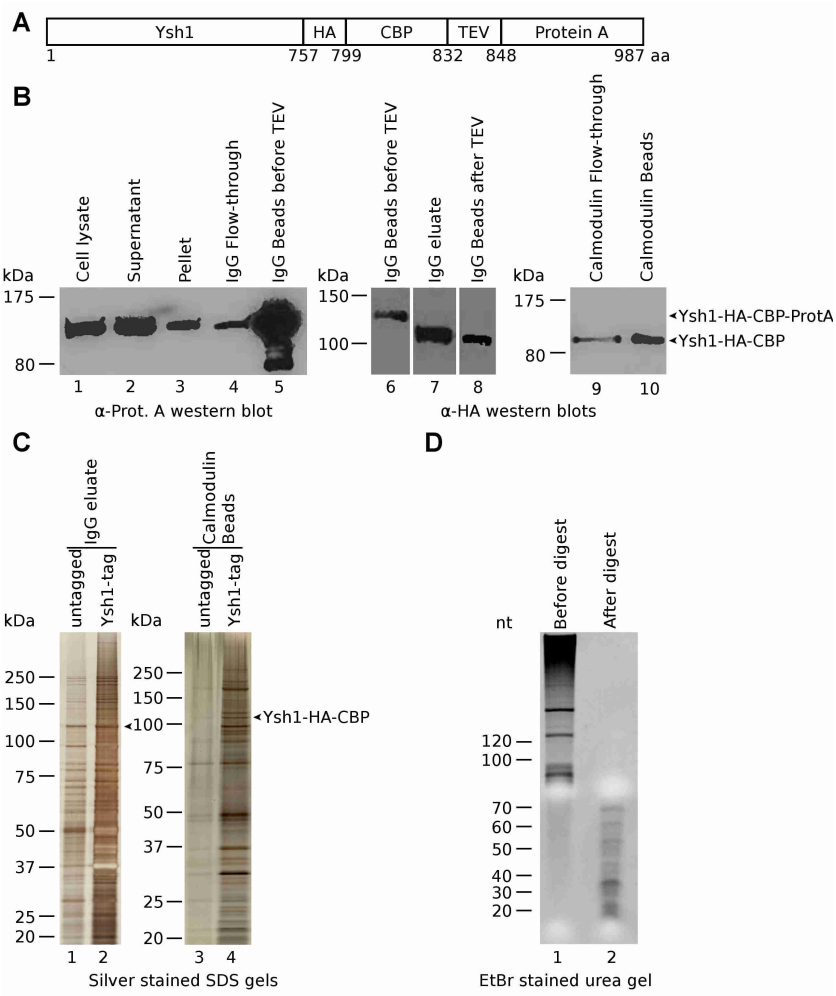


Figure 2.8: Purification of Ysh1-HA-TAP
(A) Schematics showing the Ysh1-HA-TAP construct that was used for the purification, amino acid (aa) positions of the different elements are indicated below the figure. First, Ysh1 in complex with nascent proteins was purified utilising the Protein A moiety and IgG beads. After washing, the protein was subsequently eluted by cleavage with a tobacco edge virus (TEV) protease and, in a second purification step, the now exposed Calmodulin binding protein (CBP) could be bound to Calmodulin beads. Elution for mass spectrometry was performed by boiling the beads in Lämmli buffer, SDS gel electrophoresis and excision of the whole sample from the gel.
(B) Western blot showing the presence of tagged Ysh1 after each purification step. Note that after cleavage of the protein A moiety, a shift in protein size from 110 kDa to 103 kDa occurs. Before TEV cleavage, the protein can be detected by Peroxidase-α-Peroxidase (PAP) antibodies which recognize Protein A; after TEV cleavage α-HA antibodies were used.
(C) Silver stained SDS gels showing proteins that are purified from an untagged and Ysh1-HA-TAP-tagged strain at different purification steps. Arrows indicate the band corresponding to Ysh1.
(D) Ethidium bromide stained urea gel showing 1 µg of DNA and RNA extracted from the cell lysate before and after treatment with various nucleases (RNase A, RNase T1, DNase RQ1, and Benzonase).

Purification of Ysh1 was performed in duplicate from 16 l of initial culture volume and a non-tagged control was included. The tagged protein was followed throughout the

purification on Western blots and the protein composition of the resulting sample was analysed by silver-stained SDS gels (Figure 2.2B and C). In addition, the extract was treated with nucleases before purification of Ysh1 in order to ensure that only proteins which directly bind to the CF-CPF complex can be recovered (Figure 2.2D). The identity of the proteins that were present in the purified complex was determined by mass spectrometry. The total number of proteins detected was 370 and 423 for the two replicates, respectively. In contrast, only 74 proteins could be identified in the untagged control (appendix I).

As expected, many known 3' end processing components could be detected (Table 3), none of which were present in the untagged control. However, most of the recovered proteins are homologous to proteins that are part of the budding yeast CPF but not the CF complex. This was to be expected since Ysh1 is also a member of the CPF in *S. cerevisiae*, however, it also indicates that the different subcomplexes are not very tightly bound to one another and dissociate under the purification conditions used.

Table 3: Seb1 and known 3'end processing proteins identified by MS
CFIA: cleavage factor IA; CFII: cleavage factor II; CPF: cleavage and polyadenylation factor; PFI: polyadenylation factor I

Protein name	***S. cerevisiae* homologue subcomplex**	**Replicate 1**	**Replicate 2**
Rna14	CFIA		x
Ysh1	CPF (CFII)	x	x
Cft1	CPF (CFII)	x	x
Pta1	CP (CFII)	x	x
Cft2	CPF (CFII)	x	x
Pfs2	CPF (PFI)	x	x
Fip1	CPF (PFI)	x	x
Seb1			**x**

Not only known components of the 3' end processing machinery but also other proteins were co-purified. The 3' end processing machinery is known to cross-talk with several cellular pathways including other steps of transcription (He et al. 2003), splicing (Albulescu et al. 2012), and mRNA export (Brodsky and Silver 2000). Factors associated with all of these processes could be recovered (see appendix I). Most interestingly though,

Seb1 was present in one repeat (Table 3), the same repeat in which the CFI component homologue Rna14 was detected. This indicates that Seb1 might be interacting with CFI rather than the CPF and could explain why it was recovered in one repeat only. This is also supported by the fact that the Seb1 purification yielded Rna14 as the most enriched component of the 3'end processing complex and strongly suggests that a direct interaction really exists between Seb1 and the fully assembled 3'end processing machinery which is located at actively transcribed protein-coding genes.

2.4 Discussion

The genome-wide data presented in section 2.2 strongly indicate an active role for Seb1 in 3'end processing or transcription termination. This is supported by Seb1 and Ysh1 purification data, suggesting a direct interaction between Seb1 and the 3'end processing machinery, or at least that they are in close proximity to each other. Furthermore, PAR-CLIP analysis of Seb1 binding sites detected a sequence motif that is recognised by the protein *in vivo* (Figure 2.4B). Interestingly, this motif is almost identical to the Nrd1 recognition site (Carroll et al. 2004, 2007; Jamonnak et al. 2011) which indicates high conservation of the RNA binding region between the two proteins and further supports a role for the region following the predicted RRM domain in RNA interaction. In addition, I found the UGUA motif to be localised just after the pA site of most genes genome-wide. This is in agreement with a previous analysis predicting sequences that are important for 3'end formation of protein-coding genes in *S. pombe* (Schlackow et al. 2013). Strikingly, the exact same sequence was found to be enriched around pA sites, indicating that Seb1 is of general importance for 3'end processing and/or transcription termination. Not only *S. cerevisiae* and *S. pombe* use this motif, it was also found in humans as part of the U-rich sequence upstream of the cleavage site in which case the CFI complex recognises the same consensus (Rüegsegger et al. 1996; Martin et al. 2012). The Ysh1 and Seb1 protein purification data point towards a stronger interaction between Seb1 and the CFI

complex rather than the CPF (Table 3 and (Watts 2015)). CFI might therefore be recruited to the UGUA motif via interactions with Seb1 in fission yeast, while the human CFI complex binds to UGUA directly. In this scenario, CFI would bind just after the pA site in *S. pombe*, similar to what has been reported for *S. cerevisiae* (Baejen et al. 2014) but in contrast to humans. While the UGUA sequence motif only plays a role in non-coding transcription in budding yeast, it can be found on protein-coding genes in human. In contrast, fission yeast contains the motif on both gene classes. This demonstrates that the UGUA motif is highly conserved from fungi to humans as a signal for 3'end processing/transcription termination further supporting a role for Seb1 in termination of protein-coding as well as non-coding genes on a genome-wide scale.

To understand how Seb1 promotes its function at the 3'end of genes, in depth analysis of the interaction with RNA and Pol II is required. The deletion mutations give a first indication that both interactions are important as the CID and RRM are both essential (section 2.1). Except for the region directly following the CID, all other disordered regions are dispensable for the cell. While it is very likely that the CID and RRM interact with the Pol II-CTD and RNA, respectively, the function of the disordered regions is not very well understood. Generally, those regions have been implicated in the formation of higher-ordered structures such as hydrogels and amyloid-like fibers *in vitro* which are necessary for the formation of membrane-less organelles such as stress granules or P-bodies *in vivo* (reviewed in Weber and Brangwynne 2012; Hyman et al. 2014). Furthermore, RNA was found in these organelles and many disordered regions were shown to directly interact with RNA (Järvelin et al. 2016).

The Seb1 homologue Nrd1 was found to contain a Q/P rich stretch at its very C-terminus that was indicated in the formation of filamentous structures *in vitro* (O'Rourke et al. 2015). However, this sequence is not conserved in Seb1 (Figure 2.9). Another motif that is known to promote the formation of higher order structures constitutes repeats of the consensus [R/S]Y[R/S] (Kato et al. 2012; Han et al. 2012). However, this sequence is also not present in Seb1. Therefore, it is not clear if the ability of Nrd1 to form filamentous

structures *in vitro* is conserved in Seb1 and if the protein can localise to granules *in vivo*.

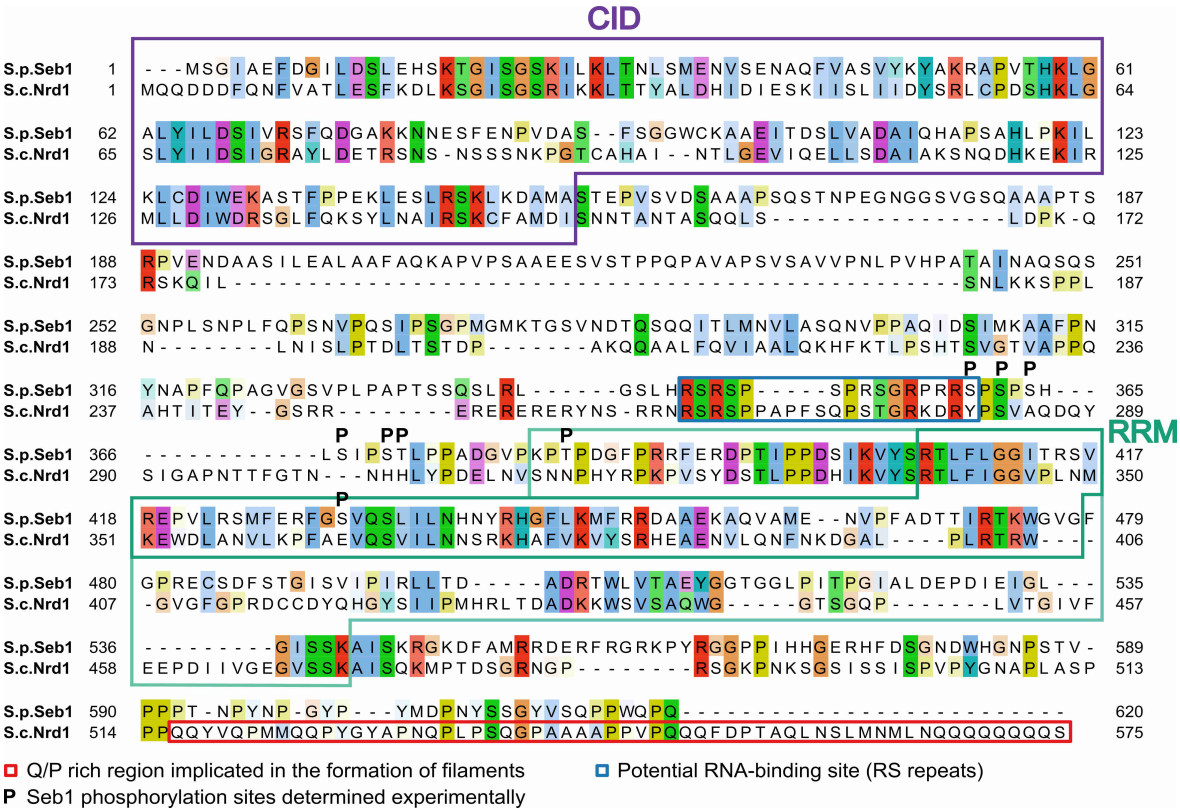


Figure 2.9: Full-length sequence alignment of Seb1 and Nrd1
The full-length sequences of Seb1 and Nrd1 were aligned using Promals3d (Pei et al. 2008). The CID is indicated by a purple and the canonical RRM domain by a turquoise box, respectively. Ordered regions around the RRM which might participate in RNA binding are marked in light turquoise. A potential RNA binding site in the disordered region and the filament-forming region of Nrd1 are indicated in blue and red, respectively. Seb1 phosphorylation sites are marked above the residue by the letter P. The sites were determined by purification of FLAG-tagged Seb1 from *S. pombe* and followed by mass spectrometry (MS). The experiment was performed by the *DPhil* rotation student Miles Huseyin under my supervision, MS was done by the Advance Proteomics Facility, Department of Biochemistry, Oxford.

In contrast, a region containing several RS repeats is conserved between Nrd1 and Seb1 (Figure 2.9 blue box) and several examples have been described in the literature in which these repeats engage direct RNA interactions (Järvelin et al. 2016; Castello et al. 2016). Furthermore, RS repeats were implicated in increasing the binding affinity of RRM domains (Jamros et al. 2015). It therefore seems likely that this region participates in RNA binding in both Nrd1 and Seb1. In addition, phosphorylation of RS repeats in the human protein SRSF1 was shown to decrease their flexibility and to thus

influence RNA binding (Xiang et al. 2010). Intriguingly, Seb1 is heavily phosphorylated just downstream of the RS repeat region, as determined by mass spectrometry (Figure 2.9). However, if the disordered region of Seb1 can indeed interact with RNA and if it is able to form higher order structures such as filaments needs to be tested experimentally.

Because both the CID and the RRM domain of Seb1 are essential, it seems likely that they are important for the function of the protein. Therefore, to further elucidate how transcription termination in *S. pombe* is promoted by Seb1 a more extensive analysis of both domains was performed which will be discussed in the following chapters.

CHAPTER 3: STRUCTURAL AND BIOCHEMICAL STUDIES OF THE CID

The CID is a conserved domain that is specifically prevalent in proteins that participate in 3'end processing and/or termination (Figure 1.9). *In vitro* studies of several different CIDs have shown a binding preference to specifically phosphorylated CTD peptides. Most of them, including Pcf11, Rtt103, Scaf8, Prpd1A, Prprd1B, and Rprd2, have the highest affinity to S2P peptides (Noble et al. 2005; Becker et al. 2008; Lunde et al. 2010; Mei et al. 2014; Ni et al. 2014). The Nrd1-CID is the only example characterised to date which prefers the S5P-CTD (Vasiljeva et al. 2008a). Furthermore, phosphorylation of Y1 prevents CID binding (Mayer et al. 2012; Schreieck et al. 2014).

Crystal and NMR structures of several different CIDs have been solved and helped to elucidate the molecular basis of these interactions. They all show a very similar overall fold containing eight α-helices that are arranged into a right-handed superhelical structure. Co-structures with CTD peptides revealed that about 1.5 consensus repeats are generally recognised. Furthermore, the amino acids involved in contacting the CTD could be determined and, to some extent, the structures could explain why domains from different proteins show differing binding specificities (Meinhart and Cramer 2004; Noble et al. 2005; Becker et al. 2008; Lunde et al. 2010; Kubicek et al. 2012; Ni et al. 2014).

The behaviour of Seb1 *in vivo* can, to some degree, be predicted using published information about other CIDs. However, the binding specificity of the CID should be confirmed experimentally and mutational analysis is necessary to understanding, in detail, how the interaction between Seb1 and Pol II is promoted. This information can then be used to disrupt CTD binding in living cells and studying the effects will help to elucidate the cellular function of the Seb1 protein in transcription termination and how CTD binding contributes to it. Therefore, the Seb1-CID was closely examined *in vitro* using structural and biochemical approaches, which will be described in this chapter.

3.1 The Seb1-CID preferentially binds to S2P-CTD

First, I wanted to test which phosphorylation type of the CTD is preferred by the Seb1-CID. Therefore, I recombinantly expressed and purified the CID_{1-152} from bacteria with a His_8-tag at its C-terminus (Figure 3.1). The resulting protein was used for fluorescence anisotropy measurements in order to determine the binding affinity of the CID_{1-152} to differently phosphorylated CTD peptides. The peptides used constituted two repeats of the $Y_1S_2P_3T_4S_5P_6S_7$ consensus which were either unmodified or phosphorylated at S2, S5, or S7 on both repeats. Also, they contained an N-terminal 6-carboxyfluorescein (FAM) label which is necessary to measure anisotropy.

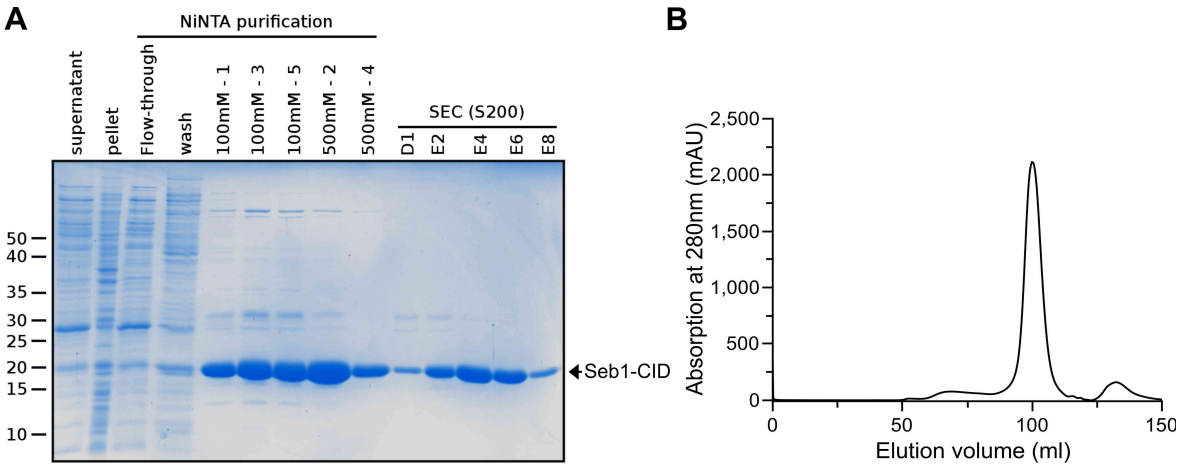
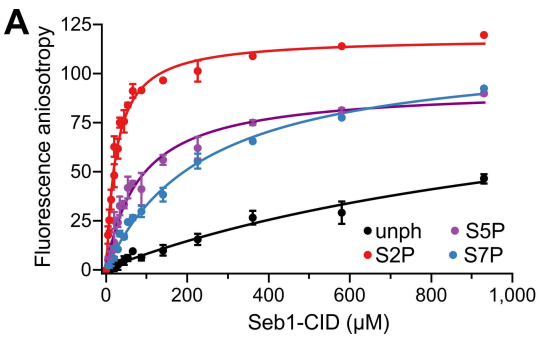


Figure 3.1: Purification of recombinant Seb1-CID_{1-152}
(A) An exemplar Coomassie-stained SDS gel following the whole purification procedure of the Seb1-CID_{1-152} is shown. After lysis of *E. coli* cells using a French Press, the lysate was cleared by centrifugation (supernatant and pellet) and applied to a Nickel nitrilotriacetic acid (NiNTA) column (Flow-through). After washing of the resin (wash), the His_8-tagged Seb1-CID_{1-152} was eluted using increasing amounts of imidazole (100 mM-500 mM). The protein was further purified using size-exclusion chromatography (SEC). The purity of the peak elution fractions from the chromatogram depicted in B is shown (SEC (S200)).
(B) Elution profile of the Seb1-CID_{1-152} from a HiLoad Superdex S200 16/60 PrepGrade column (GE Healthcare). The elution of the protein was measured by absorption at 280 nm and is plotted against the elution volume.

Strikingly, the highest affinity of the Seb-CID_{1-152} could be detected to CTD peptides phosphorylated at S2 (Figure 3.2). This is in contrast to Nrd1 but in agreement with the genome-wide binding profile of Seb1 which shows highest recruitment to the 3'end of genes, where S2P levels are highest (Figure 2.4). Intriguingly, the human homologue of

Seb1, Scaf8, also prefers CID peptides that are phosphorylated at S2 (Becker et al. 2008). The Seb1-CID$_{1-152}$ also shows modest affinity to S5P-CTD and very little to S7P-CTD. Furthermore, no detectable binding is observed to unphosphorylated peptides. Similar results have been obtained from *in vivo* experiments. Both Western blot detection of specific phosphorylation forms of Pol II that co-purified with Seb1 as well as peptide binding assays with Seb1 purified in complex with interacting partners from yeast showed interactions with S2P and S5P but not with S7P or unphosphorylated CTD (Watts 2015). Taken together, this indicates that S2P-Pol II is responsible for the recruitment of Seb1 to the 3'end of genes and that the presence of other phospho-serines does not influence binding negatively. This does not exclude a role for RNA binding in Seb1 recruitment.



B

Peptide	K_d
unph CTD	> 1000 μM
S2P-CTD	23.4 ± 2.7 μM
S5P-CTD	83.7 ± 12.6 μM
S7P-CTD	241.8 ± 28.7 μM

Figure 3.2: Seb-CID$_{1-152}$ has highest binding affinity to S2P-CTD
(A) A fluorescence anisotropy assay was performed to measure the affinity between Seb1-CID$_{1-152}$ and double-repeat peptides that are either unphosphorylated (unph) or phosphorylated on both repeats at S2, S5, or S7, as indicated. A plot of the anisotropy measured for each Seb-CID$_{1-152}$ concentration used is shown. Error bars indicated the standard deviation of at least three repeats.
(B) The equilibrium dissociation constant (K_d) for all peptides was determined graphically from the plot shown in A.

Because Seb1 has a preference for S2P-Pol II and levels of this mark are known to be lower on short genes (Bataille et al. 2012), I wanted to test whether Seb1 binding is influenced by gene length. Therefore, I split all genes into two groups, those that show binding of Seb1 by ChIP-Seq and those that do not have any detectable peaks. Indeed, the gene lengths of the two groups differ significantly (Figure 3.3) which suggests that the S2P mark is important for Seb1 recruitment *in vivo*. In contrast, PAR-CLIP does not show any such preference, indicating that Seb1-Pol II interactions are not required for RNA binding *per se*.

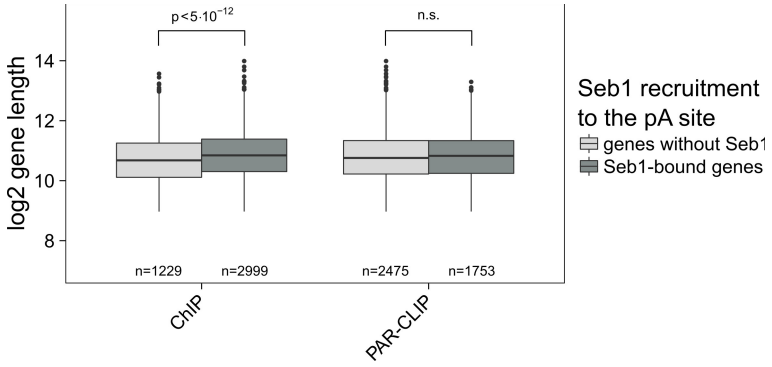


Figure 3.3: Co-transcriptional recruitment of Seb1 is better on long genes
The box plot depicts two different groups of genes, those that are bound by Seb1 at the pA site ± 250 nt (dark grey) and those that do not have any detectable Seb1 peaks (light grey) as determined by either ChIP-Seq (left) or PAR-CLIP (right). Log2 scaling was used for gene lengths. The significance of the differences in gene lengths was calculated using the Wilcoxon-Mann-Whitney test and is indicated above the boxes. The amount of genes forming each group is shown below the boxes. The same subset of genes was used as in Figure 2.5.

3.2 Conserved amino acids in the Seb1-CID recognise the S2P moiety

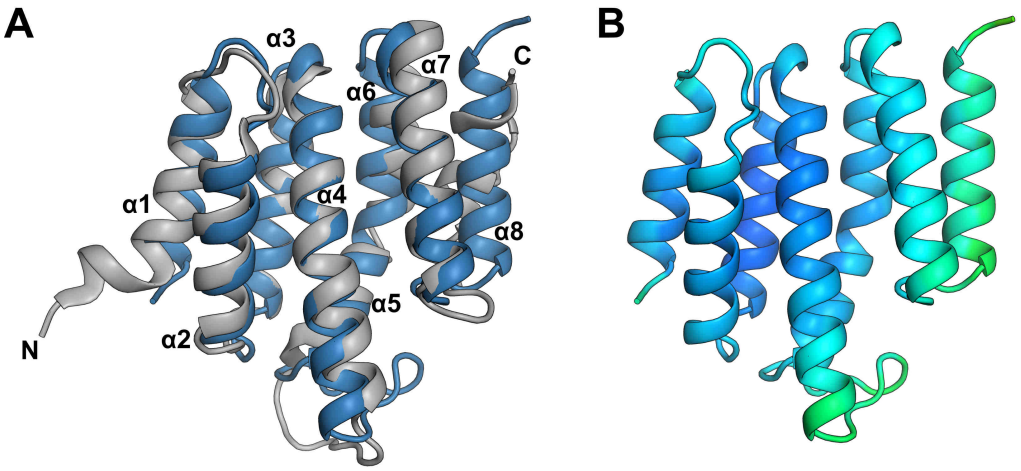
To be able to study what functional consequences loss of interaction between Seb1 and different phosphorylation types of Pol II has, it is necessary to understand which amino acids are responsible for the recognition. It was therefore attempted to selectively disrupt binding to either S2P or S5P-CTD without affecting interactions with the other peptide, respectively. Therefore, in order to determine whether important amino acids are topologically conserved in comparison to published CIDs, crystallisation of the Seb1-$CID_{1\text{-}152}$ was performed and a 1.6 Å resolution structure of the domain could be obtained (Table 4). This was done with the help of Dr Karl Harlos (Division of Structural Biology, University of Oxford) who was involved in cryo-freezing and Mr Max Renner from Prof Jonathan Grimes' lab (Division of Structural Biology, University of Oxford) who performed data collection, phasing by molecular replacement using the Nrd1-CID structure (PDBID: 3CLJ), model building, and refinement.

Table 4: Data collection and refinement statistics for the Seb1-CID$_{1-152}$
Values in parenthesis are for the highest resolution shell

Data collection	
Space group	P 31 2 1
Cell dimensions	
a, b, c (Å)	55.6, 55.6, 131.2
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	48.1 - 1.6
R_{merge}	9.8 (215.1)
I/σ	14.2 (1.1)
$CC_{1/2}$	1.0 (0.5)
Completeness (%)	99.3 (98.5)
Redundancy	8.5 (4.8)
Refinement	
Resolution (Å)	48.1 - 1.6
No. reflections	30426 (2968)
R_{work}/R_{free}	18.9/20.8
No. atoms	
overall	1323
protein	1162
B factors	
overall	52.1
protein	50.3
solvent	62.54
r.m.s. deviations	
bonds (Å)	0.014
angles (°)	1.53
Ramachandran	
favoured (%)	98.7
allowed (%)	1.3
outliers (%)	0

The Seb-CID$_{1-152}$ folds into an eight helix bundle which adopts a right-handed superhelical structure in line with what has been published for other known CIDs (Figure 3.4A). The B-factors, also called atomic displacement parameters, quantify the attenuation of the X-rays due to thermal motion and other factors therefore giving an indication of the static disorder within the structure. In other words, regions with high B-factors are thought to be more flexible compared to low B-factor regions. The Seb1-CID$_{1-152}$ forms a very rigid core with some flexibility present only at the α4-α5 loop and in C-terminal α-helix 8 (Figure 3.4B). However, none of those regions have been reported to directly contact the CTD peptide upon binding (Meinhart and Cramer 2004; Becker et al. 2008; Lunde et al. 2010; Kubicek et al. 2012; Ni et al. 2014), indicating that the binding site is pre-formed and rigid,

forming a stable interaction surface. Furthermore, the highest sequence identity of the Seb1-CID$_{1-152}$ is to its *S. cerevisiae* homologue Nrd1 (Figure 3.4C) to which it is also most similar structurally together with Pcf11. The higher root mean square deviation (RMSD) to the Nrd1 NMR structure as compared to the crystal structure can be explained by the fact that NMR is performed in solution which allows the protein to move and show conformational flexibility.



C

CID	Bound peptide	Organism	Method	PDBID	Sequence identity	RMSD	Citation
Nrd1	-	*S. cerevisiae*	Xtal	3CLJ	29.45%	0.981Å	Vasiljeva et al. 2008a
Nrd1	S5P	*S. cerevisiae*	NMR	2LO6	28.94%	1.242Å	Kubicek et al. 2012
Pcf11	S2P	*S. cerevisiae*	Xtal	1SZA	22.91%	0.826Å	Meinhart and Cramer 2004
Scaf8	-	Human	Xtal	3D9I	21.73%	1.111Å	Becker et al. 2008
Rprd1A	S7P	Human	Xtal	4JXT	15.38%	3.679Å	Ni et al. 2014
Rtt103	S2P	*S. cerevisiae*	NMR	2L0I	13.97%	4.869Å	Lunde et al. 2010

Figure 3.4: Comparison between the Seb1-CID$_{1-152}$ and published CID structures
(A) The crystal structure of the Seb1-CID$_{1-152}$ is shown in blue, aligned to the Nrd1-CID structure in grey (PDBID 2LO6).
(B) The Seb1-CID$_{1-152}$ was coloured by B-factor from blue (lowest value: 25 Å²) to green (highest value: 105 Å²). Generally, a high B-factor indicates high average displacement of an atom around its mean position.
(C) A structural alignment between the Seb1-CID$_{1-152}$ and the indicated published structures was performed using PyMOL (Schrödinger 2015). To calculate the sequence identity, amino acid compositions of the corresponding structures were used. RMSD: root mean square deviation; Xtal: X-ray crystallography; NMR: nuclear magnetic resonance

Because soaking experiments and crystallisation trials of the protein in the presence of CTD peptides has been unsuccessful, amino acids involved in CTD binding can only be inferred from similarities with other CID structures. The interaction between CIDs and S2P-CTD has been studied in detail for Rtt103, Pcf11, and Scaf8. The binding is

usually mediated by direct contacts between the phosphate moiety and one or two basic amino acids (R or K) which are conserved in most CIDs (Figure 3.5A). In Rtt103, the amino acid R108 (Figure 3.5C) contributes more to the interaction while K105 plays a lesser role in binding as it only directly contacts S2P in some states of the NMR ensemble. Both of the amino acids are topologically conserved in Seb1, corresponding to K124 and K121, respectively (Figure 3.5B). This is in agreement with the fluorescence anisotropy assay showing highest affinity of Seb1 to CTD peptides that are phosphorylated at S2 (Figure 3.2). Furthermore, Scaf4 as well as Scaf8 both contain basic amino acids at the these positions (Figure 3.5A), however, the Scaf8 crystal structures in complex with S2P-CTD (or double phosphorylated peptides) only show a direct interaction between the phosphate moiety and the stronger binding amino acid, the one that corresponds to K124 in Seb1 or R108 in Rtt103 (Becker et al. 2008). This strongly indicates that only one of the basic amino acids determines whether a given CID has highest affinity to S2P or not. This is also in agreement with Nrd1 binding studies showing that the protein does not bind to S2P-CTD very strongly even though one of the basic amino acids is conserved, however, the other one corresponding to K121 in Seb1 is not (Figure 3.5A). There seems to be one exception to the rule: Pcf11. Like Nrd1, only one of the basic amino acids is conserved, the one that contributes less to S2P binding (corresponding to K121 in Seb1, Figure 3.5A) but Pcf11 is recruited to the 3'end of genes in budding and fission yeast ((Kim et al. 2004a) and see later). Consistent with the model of S2P recognition, no direct contacts to the phosphate moiety in the *S. cerevisiae* Pcf11 crystal structure can be observed (Meinhart and Cramer 2004) and the affinity of the Pcf11-CID to S2P-CTD is considerably lower than is the case for other S2P binders (Lunde et al. 2010). This suggests that affinity to a specifically phosphorylated CTD is not the only way in which binding of a protein at a certain stage of transcription is achieved, rather it only helps to recruit the protein.

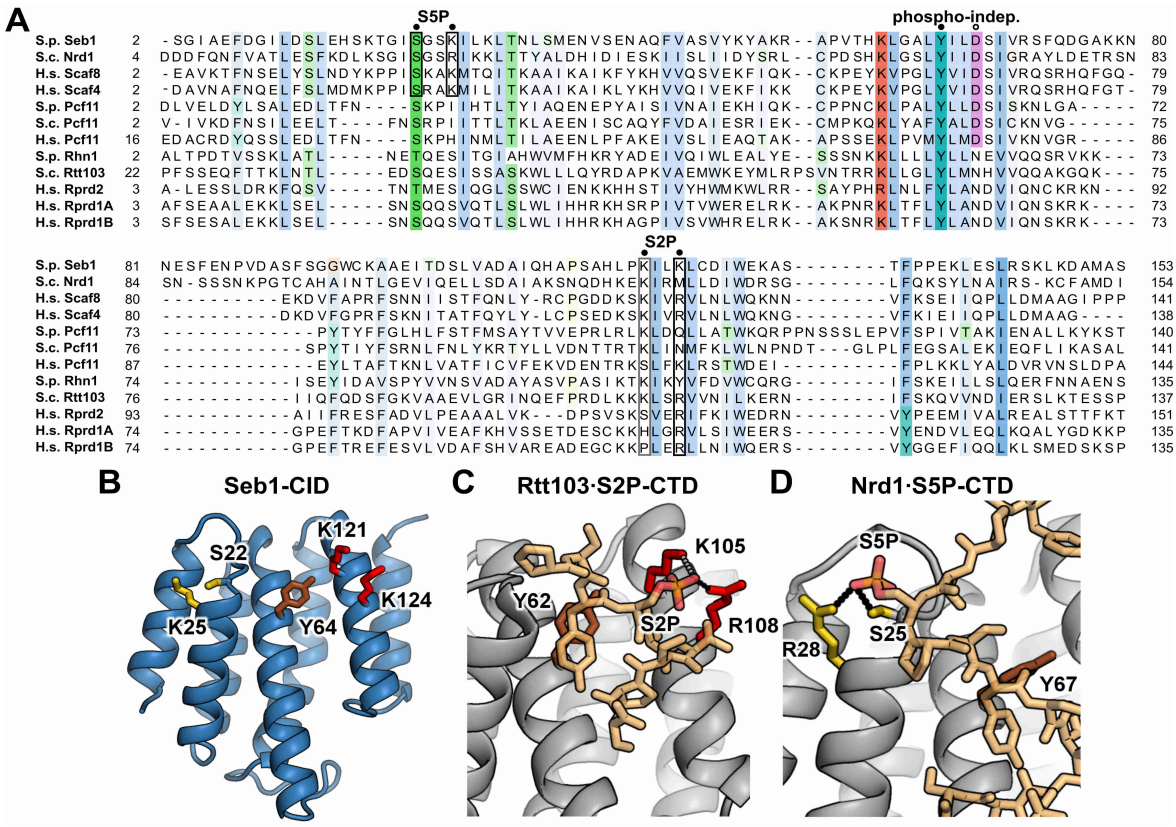


Figure 3.5: Sequence alignment and binding specificity of different CID proteins
(A) A multiple sequence alignment of the CIDs from all proteins shown in Figure 1.9 is depicted. The colour indicates amino acid conservation. Residues previously shown to be involved in S2P or S5P recognition are marked by boxes. A conserved region that interacts with the CTD in a phospho-independent (indep.) way is also indicated. Closed circles highlight amino acids that were mutated in Seb1 *in vitro* and *in vivo*, open circles indicate *in vitro* analysis only. The alignment was done using Promals3D and included information about all 3D structures of CIDs that are known.
(B) The crystal structure of the Seb1-CID$_{1\text{-}152}$ is shown in the same orientation as in Figure 3.4. Amino acids that were mutated *in vivo* are depicted as sticks. Yellow coloured amino acids are involved in S5P binding, red coloured amino acids are important for S2P recognition and Y64 which is shown in brown interacts with the CTD in an phosphorylation-independent manner.
(C) The NMR structure of the *S. cerevisiae* Rtt103 is shown as a cartoon representation and the S2P-CTD is depicted as sticks (PDBID 2L0I). Residues which are involved in S2P-CTD recognition are shown in red and the tyrosine corresponding to the highlighted tyrosine in B is depicted in brown. Specifically, R108 contacts the phosphate moiety on S2 in most states of the NMR ensemble (dotted black line) while K105 binds only in some states (dotted grey line).
(D) The NMR structure of *S. cerevisiae* Nrd1 is shown in complex with an S5P-CTD (PDBID 2LO6). The representation of the structures and the amino acids colour coding is as in B and C. Residues that directly contact the phosphate on S5 are conserved in Seb1 and are highlighted in yellow.

In the case of Nrd1, the highest affinity can be observed to S5P-CTD (Vasiljeva et al. 2008a). The crystal structure reveals two amino acids that form direct contacts with the

phosphate moiety, a serine and an arginine (Figure 3.5D). A serine or threonine is conserved at the former position in all CID proteins, however, only Seb1, Scaf4, and Scaf8 contain a basic amino acid at the second position (Figure 3.5A). The affinity of at least Seb1 and Scaf8 to S2P-CTD is still higher than to S5P-CTD, however, Scaf8 binds stronger to S2P-S5P-CTD as compared to either singly phosphorylated peptide (Becker et al. 2008). Whether that is the case for Seb1 as well was not tested because recent research suggests that double phosphorylated CTD repeats are very uncommon *in vivo* and the interaction might therefore not have any function in the cell (Schüller et al. 2016; Suh et al. 2016). Of course, most of the interactions between CIDs and the CTD are independent of any phospho-groups. Most notably, a tyrosine (corresponding to Y64 in Seb1) is conserved in all CIDs (Figure 3.5A) and forms an aromatic edge-to-face interaction with the Y1 of the CTD in all known co-structures (Figure 3.5C and D). It is located within a patch of conserved amino acids and mutation of a nearby aspartate in Nrd1 (corresponding to D67 in Seb1) was found to result in cell death (Kubicek et al. 2012). It is conserved in all Seb1 and Pcf11 homologues (Figure 3.5A).

In order to test the contribution of the individual Seb1-CID residues on CTD binding, all the amino acids mentioned above were mutated. Because this experiment was performed temporally before the fluorescence anisotropy assay had been established (Figure 3.2), the technically simpler peptide binding assay was utilised in that case. For this, C-terminally His_8-tagged full-length protein carrying one or more mutations was recombinantly expressed and affinity purified using NiNTA as before. Binding of the different mutants was subsequently tested using biotinylated peptides that consisted of four repeats of the $Y_1S_2P_3T_4S_5P_6S_7$ consensus and were either unphosphorylated or phosphorylated at position S2, S5, or S7 in all repeats, respectively (Figure 3.6). As expected from the fluorescence anisotropy assay, binding of WT-Seb1 to all types of phosphorylated CTD could be observed but not to unphosphorylated peptides (Figure 3.6, lane 1). Affinity to S2P-CTD was lost when mutating either K121 or K124 to the positively charged amino acid glutamate, as predicted (Figure 3.6, lanes 6 and 8). However,

whether this is due to charge repulsion between the mutated amino acid and the phosphate moiety is not clear because an interaction with the corresponding neighbouring basic amino acid (e.g. between K121 and K124E) could simply make it inaccessible for phosphate binding. Therefore, both positions were also mutated to the neutral amino acid alanine which had no effect on S2P-CTD binding (Figure 3.6, lanes 7 and 9). This indicates that both basic amino acids are important for the recognition of S2P in Seb1, however, the presence of either one of them is sufficient for binding *in vitro*.

Interestingly, whenever the interaction with S2P-CTD is lost, binding to S7P-CTD is also affected. The two phosphate moieties are in close proximity and the same amino acids might therefore contribute to binding of both peptides, with K121 probably having a larger role (Figure 3.6, compare lane 6 and 8). However, since the overall affinity to S7P-CTD is weak (Figure 3.2B), this interaction probably has no function *in vivo*.

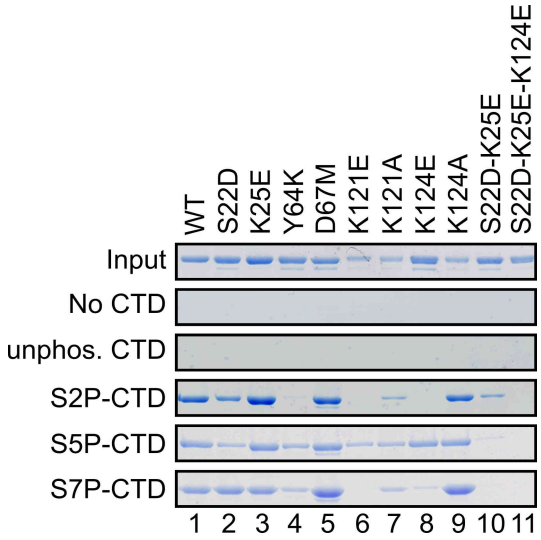


Figure 3.6: Point mutations in the CID affect CTD binding *in vitro*
The Coomassie-stained SDS gel depicts a peptide binding assay performed between recombinantly expressed and purified full-length Seb1 harbouring the indicated mutations and biotinylated peptides of four CTD consensus repeats that were either unphosphorylated (unphos.) or phosphorylate at S2, S5, or S7 on all repeats. For this, the purified protein was incubated with Streptavidin beads coupled to the indicated peptide and the beads were subjected to SDS-PAGE after extensive washing. Wild-type (WT) protein as well as beads not containing any peptides (No CTD) were included as controls.

The S5P-CTD interaction was predicted to be mediated via S22 and K25. Mutation of neither amino acid on its own has much effect on CTD binding. However, in combination the interaction is lost (Figure 3.6, lanes 2,3, and 10). Y64 and D67 are thought to mediate CTD binding in a phosphorylation-independent manner. However, mutating D67 to the larger methionine has no effect on the interaction (Figure 3.6, lane 5). In contrast, in the Y64K mutant binding to S2P-CTD was completely abolished. Surprisingly though, interactions with other types of CTD could still be detected, indicating that Y64 plays a larger role for S2P-CTD recognition (Figure 3.6, lane 4). This can be explained by the spatial proximity between Y64 and S2 (Figure 3.5B and C) and the Y1-Y64 interaction could lead to a stabilisation of the peptide in the S2 binding region. In addition, the S22D-K25E double mutant was combined with a K124E mutation which, as expected, leads to loss of binding to any type of CTD peptide (Figure 3.6, lane 11).

3.3 Recruitment of Seb1 to S2P-Pol II is essential *in vivo*

Next, I wanted to assess the effects of these mutations on a cellular level. Therefore, mutations influencing binding to different types of CTD were introduced into *S. pombe* cells and their growth was tested by spot assays (Figure 3.7). The same system using a thiamine repressible WT copy of *seb1* in addition to a version carrying the mutation was used as before. Protein expression from both *seb1* loci was tested by Western blotting and showed the expected result (Figure 3.7A). Interestingly, neither mutation affecting the interaction between Seb1 and S5P-Pol II had any effect on cell growth (Figure 3.7B, S22D and S22D-K25E). The two mutations introduced to affect S2P binding showed very different outcomes, however. While K121E had only a minor effect on cell growth, cells carrying the Y64K mutation were not able to grow at all (Figure 3.7B, Y64K and K121E). This indicates that Y64 is more important for S2P-CTD binding than K121 which is consistent with published data suggesting that K121 contributes only little to S2P recognition. As expected, the S22D-K25E-K124E mutant was also inviable. CTD

binding cannot occur in this mutant which should lead to the same phenotype as was observed for CID deletion (Figure 2.2). In summary, the *in vivo* mutation analysis strongly indicate that the ability of Seb1 to interact with S2P-Pol II is necessary for cell viability and that co-transcriptional recruitment of Seb1 to the 3'end of genes is essential.

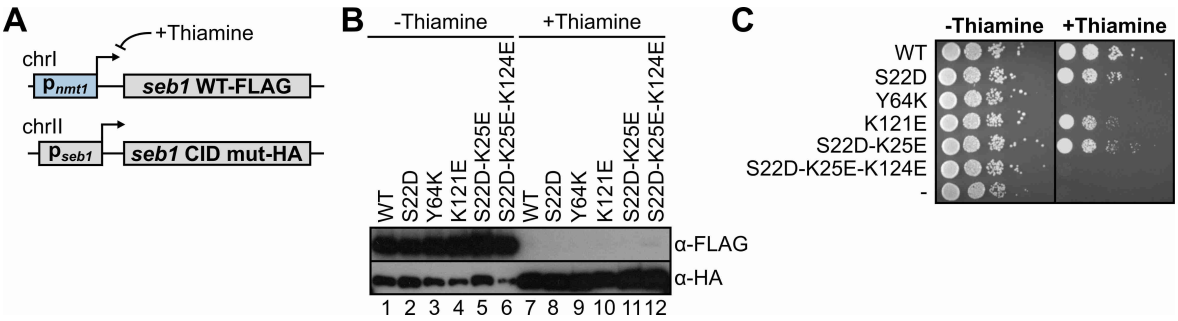


Figure 3.7: Binding of Seb1 to S2P-Pol II is required for cell viability
(A) A schematic representation of the expression system used to assess effects of Seb1-CID point mutations on cell viability is shown. See Figure 2.2A for details.
(B) Expression of the WT and mutated proteins in media containing and lacking thiamine was tested by Western blotting. Due to the different tags used, the WT and mutated copies can be specifically detected by α-FLAG and α-HA antibodies, respectively.
(C) The effects of the CID mutations were tested by spot assays as described in Figure 2.2C and the same controls were used as previously.

3.4 Discussion

In this chapter, I was able to demonstrate that the Seb1-CID has highest affinity to S2P-CTD (Figure 3.2), which is consistent with its genome-wide distribution *in vivo* (Figure 2.3). Overall, the Seb1-CID$_{1\text{-}152}$ structure has a similar fold to other known CIDs and many amino acids involved in the recognition of the phosphate moieties on the CTD are conserved and show the expected effect on CTD recognition (Figure 3.5 and 3.6). In contrast, it is not well understood how preference for proline isoforms is achieved and it remains unclear if Seb1 prefers one of the proline bonds to be present as the *cis* isomer like Nrd1 (Kubicek et al. 2012) or if it has a higher affinity to the more prevalent all-*trans* isoform as seems to be the case for all other CIDs (Meinhart and Cramer 2004; Noble et al. 2005; Becker et al. 2008; Ni et al. 2014). A prediction about the proline bond preference by homology is also difficult because even though the Seb1-CID has highest homology to the Nrd1-CID (31% identical, 48% similar) its binding specificity more closely

matches that of the Scaf8-CID (22% identical, 35% similar). This might also indicates that the cellular function of Seb1 may be more similar to homologues from higher eukaryotes than to Nrd1. However, more research is needed to understand what those proteins are doing in the cell.

Generally, affinities between CIDs and CTDs are quite low with all measured K_d values being in the micromolar range (Noble et al. 2005; Becker et al. 2008; Vasiljeva et al. 2008a; Lunde et al. 2010; Kubicek et al. 2012; Ni et al. 2014). Seb1 is no exception, its affinity to differently phosphorylated peptides is similar to published data from other CIDs (Figure 3.2B). However, recruitment of CID-containing proteins *in vivo* is probably aided by the repetitive nature of the CTD which leads to a large number of binding sites being available, thereby increasing the binding affinity *in vivo*. In support of this, it was shown that CIDs can rapidly dissociate and reassociate with neighbouring CTD repeats, which also suggests that they might be able to move along the CTD (Noble et al. 2005). Furthermore, cooperative binding of two CIDs to four repeats could be demonstrated *in vitro*, and recruitment of CID-containing proteins was shown to be increased by the presence of other CIDs even if they were part of a different protein (Lunde et al. 2010; Ni et al. 2014). All of this supports the notion that the *in vivo* binding strength of CID-containing proteins is generally higher than suggested by *in vitro* affinity measurements.

Furthermore, the fact that Pcf11 localises to the 3'end of genes even though it does not show strong binding to S2P-Pol II (Lunde et al. 2010) indicates that other factors influence the recruitment as well. For example, Pcf11 also interacts with the Rna14-Rna15 tetramer as part of the 3'end processing complex which makes it possible that recruitment of Pcf11 to the polymerase occurs indirectly via Rna14-Rna15 which can bind to nascent RNA (Minvielle-Sebastia et al. 1994; Gross and Moore 2001). Binding to the CTD could then be achieved by close proximity of the Pcf11-CID and CTD, independent of the phosphorylation state. This is also supported by the fact that Pcf11 seems to play a role in terminating short ncRNAs on which Pol II-S2P levels are not very high (Morlando et al. 2002; Kim et al. 2006; Grzechnik et al. 2015). Therefore, it is possible that the S2P mark

on the CTD is not the only determinant that recruits Seb1 to the 3'end of actively-transcribed genes and that other protein-protein interactions play a role as well. In addition, Seb1 is expected to be able to bind to RNA directly via its RRM domain which can also contribute to its recruitment and will be examined in the next chapter.

Nonetheless, a large number of different proteins with different domains are in direct contact with the CTD co-transcriptionally and binding competition is likely to exist between all of these proteins. This makes it very difficult to recreate the situation during active transcription in *in vitro* experiments. Consequently, *in vivo* studies are indispensable. Introducing point mutations into the Seb1-CID, I was able to show that loss of S2P binding *in vivo* does not support cell viability which demonstrates that this interaction is indeed crucial for the cells (Figure 3.7). However, whether this is caused by effects on transcription termination cannot be concluded merely from a spot assay and will be assessed in Chapter 5.

CHAPTER 4: STRUCTURAL AND BIOCHEMICAL STUDIES OF THE RRM

The RRM is the most abundant and best studied RNA-binding domain in eukaryotic cells (reviewed in Maris et al. (2005) and Cléry et al. (2008)). Each domain is approximately 90 amino acids in length and adopts a typical β_1-α_1-β_2-β_3-α_2-β_4 topology whereby the four β-strands fold into an antiparallel β-sheet that is packed against the two α-helices. With the help of two ribonucleoprotein (RNP) motifs located on β_1 and β_3 it can directly interact with single-stranded RNA (Figure 4.1). The binding is usually mediated by stacking interactions of two conserved aromatic amino acids with RNA nucleotides and hydrophobic contacts between a third aromatic residue and a sugar ring in the RNA backbone. In addition, a positively charged side chain can form a salt bridge to a phosphate moiety from the RNA. However, not all of these contacts can be found in all RRM structures, and based on the amino acid sequence, only two stacking interactions are conserved in Seb1 (Figure 4.1). In addition, other parts of the RRM are often involved in RNA binding, such as β_2 and β_4, loops, α-helices, or C- and N-terminal regions and some RRMs were even shown to bind other proteins rather than RNA. This shows that RRM domains are very versatile and each case has to be assessed individually.

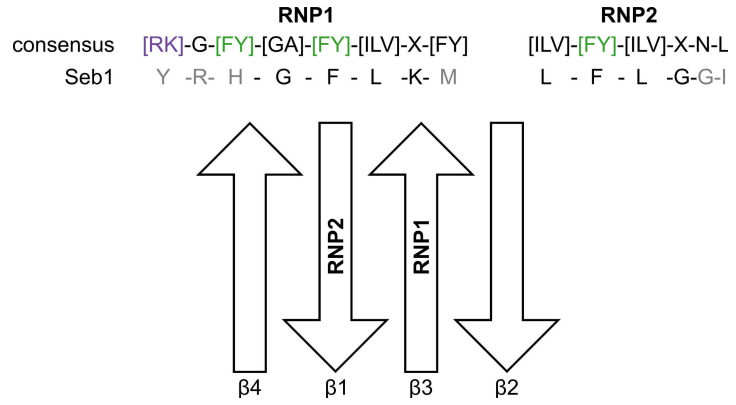


Figure 4.1: RNA recognition by canonical RRM domains and Seb1
The consensus sequence of the binding motifs ribonucleoprotein (RNP) 1 and 2 that are conserved in RNA recognition motifs (RRM) are shown at the top (X can be any amino acid). A basic amino acid (purple) binds RNA via a salt bridge and three aromatic amino acids (green) are involved in nucleotide base stacking and hydrophobic interactions with an RNA backbone sugar. Below are RNP1 and RNP2 from Seb1 with consensus amino acids in black and non-consensus residues in grey. Schematically depicted is the position of the motif on the RRM β-sheet. Figure based on (Maris et al. 2005).

Furthermore, 65% of RRM proteins contain at least one additional RRM domain or another type of RNA binding domain such as a zinc finger. This helps to modulate RNA-binding affinities and specificities (Maris et al. 2005). A single, canonical RRM can only bind four nucleotides while a combination of two RRMs can expand the recognition site to 10 nucleotides and the linker region between the domains is often involved in RNA interaction as well. For example, the interdomain region of both the pA binding protein and nucleolin forms salt bridges with the RNA, thereby mediating cooperative binding of the flanking RRM domains (Deo et al. 1999; Allain et al. 2000). In the case of Nrd1, it is thought that the second RNA binding domain is provided by its interaction partner Nab3 and cooperative binding drastically increases the RNA affinity of both proteins (Carroll et al. 2007; Hobor et al. 2011). Whether Seb1 also has an RNA-binding interaction partner is not known.

Of all RRM domains with known structures, the one from Nrd1 has the highest homology to Seb1. The solution structure of the Nrd1-RRM$_{307-491}$ including flanking regions on both sides could recently be solved by NMR (Bacikova et al. 2014). Interestingly, in addition to a canonical RRM fold, a flexible helix-loop bundle could be observed which is formed by regions located N- and C-terminally of the RRM (Figure 4.1A). Comparing the 20 lowest energy structures of the NMR ensemble shows that the orientation of this additional region is rather flexible in relation to the canonical RRM but also within itself (Figure 4.1B). It was reported that this region is important for the recognition of guanine-containing sequences and that the binding specificity of the isolated RNA-binding region is rather broad.

The regions homologous to the helix-loop bundle that flanks the Nrd1-RRM are also predicted to be ordered in Seb1 (Figure 2.1) and they therefore are likely to engage in direct RNA interactions, too. This indicates that RNA-binding of Seb1 is more complex than a classical RRM-RNA interaction and requires more in depth examination. Therefore, in order to understand how RNA recognition is achieved, the RNA-binding region was assayed similarly to the CID using a range of *in vitro* experiments in combination with

structural methods. The results were then used to disrupt RNA binding *in vivo* which will be discussed at the end of this chapter and in the next one.

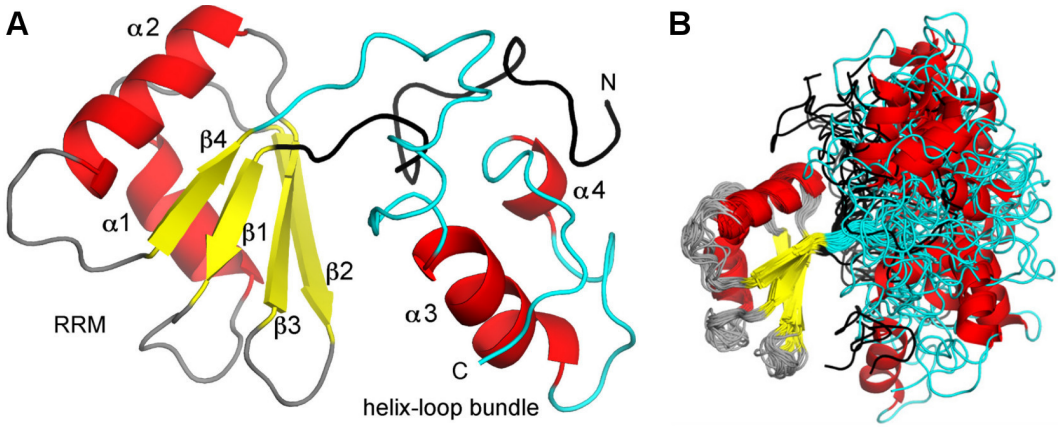


Figure 4.2: RRM solution structure of Nrd1$_{307-491}$
(A) The NMR structure of the Nrd1 RNA binding region (amino acids 307-491) shows a canonical RRM fold in addition to an adjacent helix-loop bundle which is formed by regions N- (black) and C-terminally (turquoise) of the RRM domain (grey).
(B) The 20 lowest energy structures of the Nrd1$_{307-491}$ NMR ensemble were aligned at the RRM domain and their overlay is shown. The orientation is the same as in B.
Reprinted under the creative commons license from Bacikova et al. (2014).

4.1 Construct design for expression of the Seb1 RNA-binding region

It was reported that the Nrd1-RRM only results in soluble protein if regions N- and C-terminally of the canonical domain are added to the canonical RRM domain (Bacikova et al. 2014). Because of the high homology between Seb1 and Nrd1 and the disorder prediction which indicates that the RNA-binding region of Seb1 is not restricted to the predicted RRM domain, several different constructs extending the canonical RRM region were designed, recombinantly expressed in *E. coli* and tested for solubility (Figure 4.3). Because crystallisation studies were attempted, the aim was to find the shortest construct possible. As observed in Nrd1, if too few amino acids were added to the predicted RRM domain both N- and C-terminally, protein expression levels were low and the construct was insoluble. This was the case even in the presence of a small ubiquitin-like modifier (SUMO) tag which is known to increase protein solubility (Marblestone et al. 2006). Eventually, a construct could be found containing amino acids 388-540 of the Seb1 protein that was soluble and showed decent expression levels. The construct was

designed such that the Seb1 RNA-binding region is expressed with an N-terminal His_6-SUMO-tag which allows affinity purification in addition to solubilisation and the tag is cleavable due to the presence of a 3C protease recognition site (Figure 4.3, bottom construct).

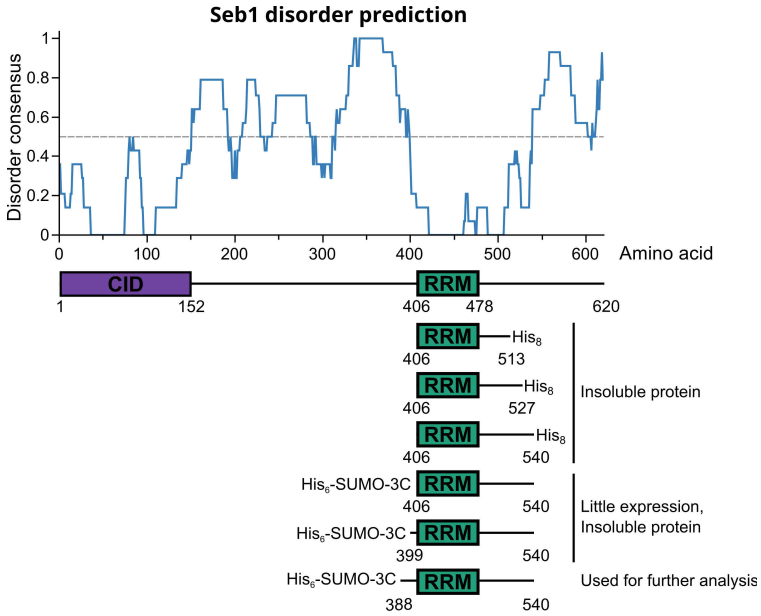


Figure 4.3: Only long RRM constructs are soluble *in vitro*
The Seb1 disorder prediction with the position of the domains below is shown as in Figure 2.1. Six different constructs with either a C-terminal His_8-tag or an N-terminal His_6-small ubiquitin-like modifier (SUMO) tag containing a 3C cleavage site were test expressed as indicated on the right. The first and last amino acid of each construct is shown and the schematics is approximately to scale.

4.2 The Seb1-RRM has a unique second domain

After successful expression of the soluble construct (Figure 4.3), the His_6-SUMO-tagged Seb1-$RRM_{388-540}$ protein was purified as shown in Figure 4.4. After cleavage of the solubilising His_6-SUMO-tag which is necessary to perform crystallisation studies, the $RRM_{388-540}$ cleavage product had to be kept in buffer containing 500 mM non-detergent sulfobetain (NDSB) 201 as it was not soluble otherwise. The protein purification was performed with the help of Mr Oliver Adams (Part II student, Vasilieva lab) and the resulting $RRM_{388-540}$ was used for crystallisation trials which gave rise to rod-shaped crystals in various condition after one day (Figure 4.4E). Cryo-freezing of several crystals from different conditions was performed by Dr Guido Paesen (post-doctoral researcher)

and Mr Max Renner (*DPhil* student) who are both part of Prof Jonathan Grimes' lab. Data collection was performed with the help of Mr Max Renner and yielded diffraction up to 1.0 Å resolution.

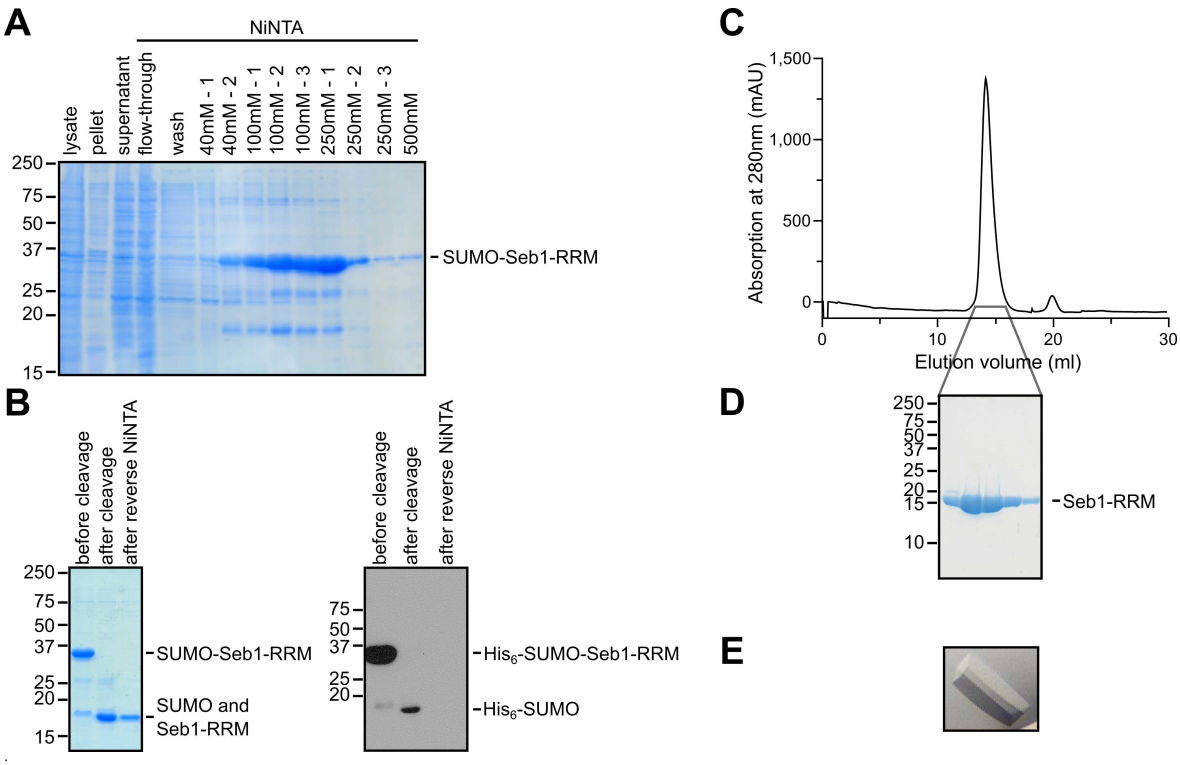


Figure 4.4: Purification of the Seb1-RRM$_{388-540}$ for structural studies
(A) A Coomassie-stained SDS gel of the Seb1-RRM$_{388-540}$ purification is shown. After cell lysis (lysate), the lysate was cleared by centrifugation (pellet and supernatant) and the supernatant was subjected to affinity purification using NiNTA. The resin was washed (wash) and His-tagged protein was eluted using 40-500 mM imidazole as indicated.
(B) A buffer exchange of the eluted protein was performed into 50 mM Tris pH 7.5, 300 mM NaCl and 500 mM non-detergent sulfobetain (NDSB) 201 and the protein solution (before cleavage) was incubated with 3C protease overnight which removed the SUMO tag from the RRM$_{388-540}$ (after cleavage). The His-tagged SUMO cleavage product was separated from the RRM$_{388-540}$ using NiNTA beads (after reverse NiNTA). Note that both cleavage products show the same mobility on the SDS gel shown (left) but only the uncleaved protein and His$_6$-SUMO cleavage product can be detected by Western blotting using α-His antibodies (right).
(C) The cleaved RRM$_{388-540}$ was then further purified by size-exclusion chromatography using a Superdex 75 10/100 GL column (GE Healthcare) and a plot of the protein eluting from the column as determined by absorption at 280 nm against the elution volume is shown.
(D) The indicated elution fractions from C were analysed for the presence of Seb1-RRM$_{388-540}$ by SDS-PAGE.
(E) The protein-containing fractions determined in D were pooled, concentrated and used in crystallisation trials and a typical crystal of the RRM$_{388-540}$ is depicted.

Solving of the phase problem by molecular replacement was attempted using the Nrd1-RRM structure as a search model, however, no meaningful solution could be found. This indicates that the Seb1 and Nrd1 RRM domains are structurally quite different. Eventually, Dr Kamel El-Omari (Diamond Light Source, Didcot, UK) was able to successfully phase the diffraction data using sulphur single wavelength anomalous dispersion (S-SAD, Table 5). Data refinement was subsequently performed by Mr Max Renner.

Table 5: Data collection and refinement statistics for the Seb1-RRM$_{388-540}$
Values in parenthesis are for the highest resolution shell; SAD: sulphur single wavelength anomalous dispersion

	SAD	**native**
Data collection		
Space group	C 1 2 1	C 1 2 1
Cell dimensions		
a, b, c (Å)	111.3, 47.2, 32.3	111.1, 47.2, 32.4
α, β, γ (°)	90.0, 98.9, 90.0	90.0, 98.9, 90.0
Resolution (Å)	55.0 - 2.0	43.3 - 1.0
R_{merge}	14.0 (25.9)	4.5 (74.8)
I/σ	67.1 (8.5)	15.5 (1.3)
$CC_{1/2}$	1.0 (1.0)	1.0 (0.5)
Completeness (%)	93.3 (42.5)	94.6 (57.5)
Redundancy	167.5 (15.4)	6.0 (3.2)
Refinement		
Resolution (Å)		43.3 - 1.0
No. reflections		79452 (5110)
R_{work}/R_{free}		13.5/15.2
No. atoms		
overall		1491
protein		1244
B factors		
overall		16.20
protein		13.79
solvent		28.57
r.m.s. deviations		
bonds (Å)		0.010
angles (°)		1.48
Ramachandran		
favoured (%)		97.5
allowed (%)		2.5
outliers (%)		0

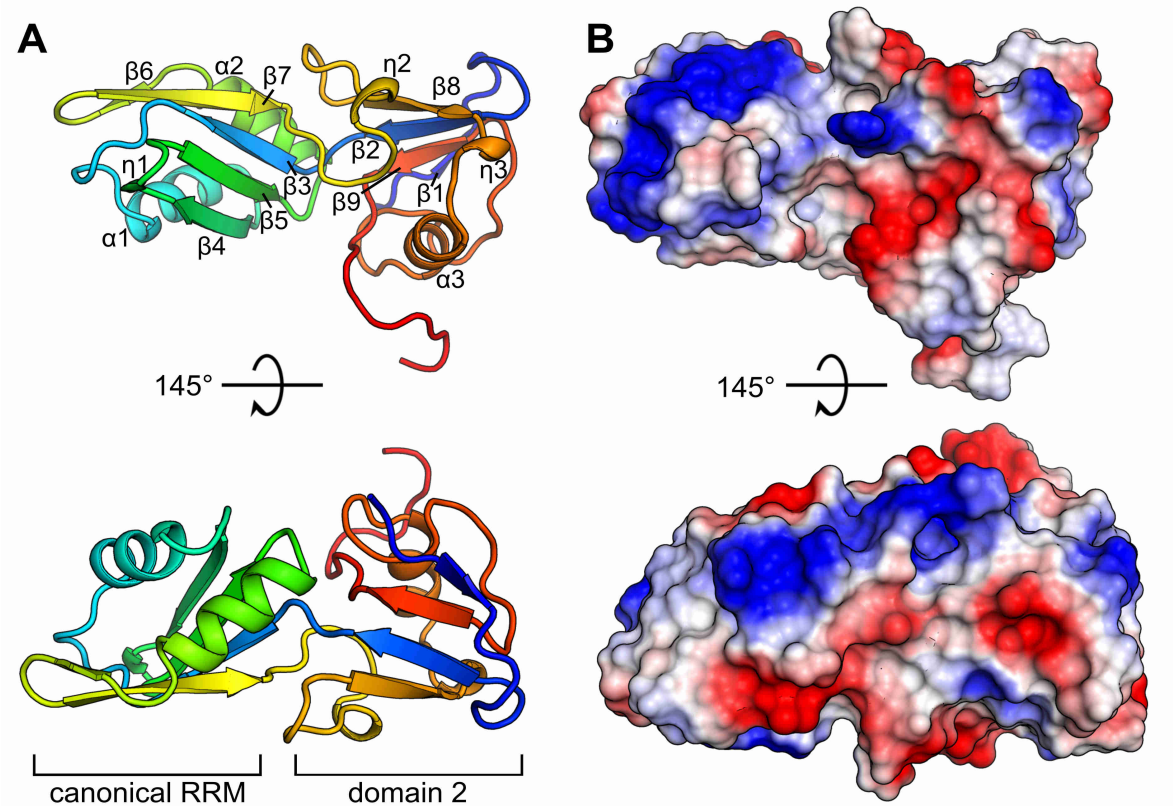


Figure 4.5: Crystal structure of the Seb1-RRM$_{388-540}$ domain
(A) The Seb1-RRM$_{388-540}$ crystal structure is depicted as a cartoon representation, coloured in blue to red from the N- to the C-terminus, respectively. The position of the canonical RRM and the additional domain (domain 2) are indicated below the structure. Note how the two domains are interwoven and cross over between β2 and β3.
(B) The electrostatic surface of the Seb1-RRM$_{388-540}$ is shown in the same orientations as in A. Positively charged areas are indicated by blue colour and negatively charged areas are shown in red. In canonical RRM domains, RNA binding is usually achieved by interactions of the negatively charged phosphate backbone with positively charged, basic amino acids located on the β-sheet. Clear patches of negative charges are visible for the canonical Seb1-RRM domain as well (top). They are surrounded by uncharged regions which can form hydrophobic and stacking (in the case of aromatic residues) interactions with the RNA bases. Both features cannot be found on the β-sheet of domain 2 (bottom).

In addition to a canonical RRM, a second domain is present in the Seb1$_{388-540}$ protein that is formed by folding together of the regions N- and C-terminally of the RRM domain (Figure 4.5). Surprisingly, however, this structure is very different from the reported Nrd1-RRM structure. While in Nrd1, the additional domain is very flexible and contains a lot of loops (Figure 4.2), the second domain in Seb1 has a defined structure with only a small number of loops. Overall, it somewhat resembles an RRM domain as it has four β-strands that form a β-sheet, however, instead of antiparallel, the sheet is of mixed topology. Interestingly, two of the β-strands are provided by the region N-terminal

and the other two by the region C-terminal of the canonical RRM domain. This also explains why the protein could not be expressed without enough amino acids N-terminal to the canonical RRM domain (Figure 4.3). In contrast to canonical RRM domains, the sheet of the second domain is packed against only one α-helix and two η-helices are additionally present. Using the Dali server (Holm and Rosenström 2010), I was unable to find any similar structure on the protein data bank (PDB) consisting of an RRM domain with the adjacent regions interwoven to form a second domain in this way or of two adjacent RRM domains in similar orientations.

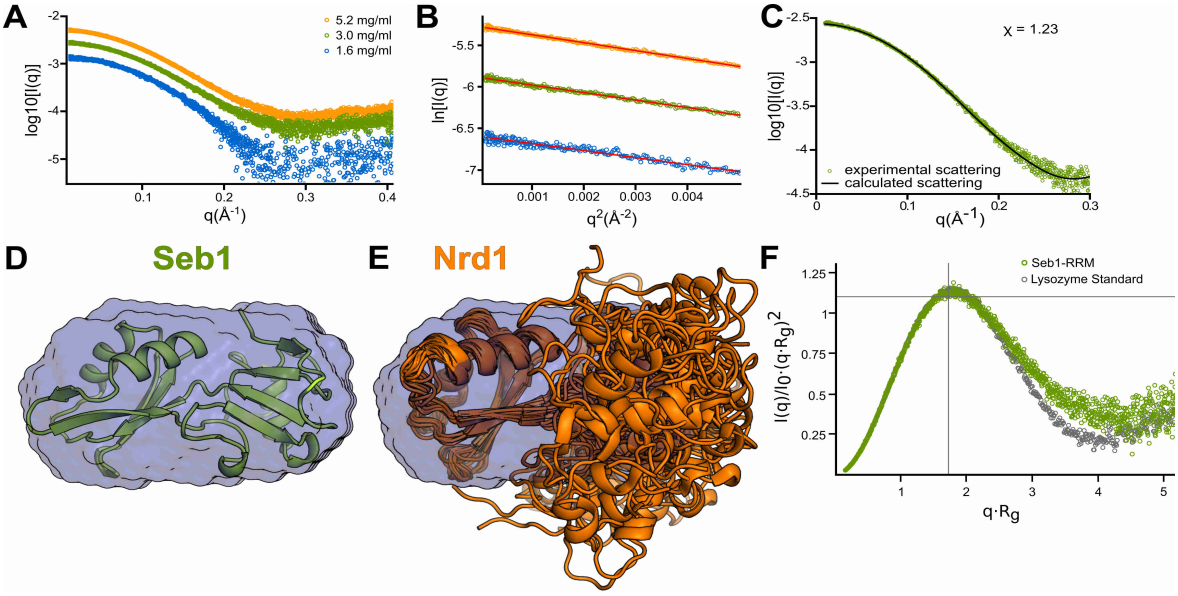


Figure 4.6: The Seb1-RRM$_{388-540}$ crystal structure represents its solution behaviour
(A) Plot showing the small angle X-ray scattering (SAXS) curves of the Seb1-RRM$_{388-540}$ at three different concentrations as indicated. The logarithmic signal intensity is plotted against the momentum transfer q which depends on the scattering angle.
(B) Guinier plots of the data shown in A with the same colours used for the corresponding concentrations and fitting (red lines) confirm linear behaviour of the SAXS data in the low q region for all measured concentrations.
(C) Fit of a theoretically calculated SAXS curve from the X-ray structure (black) to the solution SAXS curve of the Seb1-RRM$_{388-540}$ at 3.0 mg/ml is shown. The quality of the fit is expressed in χ and is indicated in the plot.
(D) An *ab initio* bead model was calculated from the SAXS data into which the Seb1-RRM$_{388-540}$ crystal structure was fitted.
(E) The Nrd1-RRM$_{307-491}$ NMR solution structure (PDBID 2M88) was aligned to the Seb1-RRM$_{388-540}$ structure and fit into the same bead model as shown in D. Amino acids 25-105 of all 20 Nrd1 structures reported in the ensemble are shown.
(F) Dimensionless Kratky plotting was used to compare the flexibility of the Seb1-RRM$_{388-540}$ (green) and a lysozyme standard (grey, BioisisID: LYSOZP). The intersection of the grey lines shows the position of the Guinier-Kratky point ($\sqrt{3}$, 1.104) which indicates the peak position of an ideal globular and rigid protein. Rigid proteins can be recognised by a characteristic parabolic shape with a peak at the Guinier-Kratky point, while curves of unfolded proteins plateau with increasing q-values.

Because of the difference between the Nrd1 and the Seb1 structures, the behaviour of the Seb1-RRM$_{388-540}$ domain was tested in solution in order to determine whether the domain is similarly flexible as the Nrd1 RNA-binding region. Therefore, I repeated the Seb1-RRM$_{388-540}$ purification as before and the resulting protein was used for small angle X-ray scattering (SAXS, Figure 4.6A and B). Data collection was performed with the help of Mr Max Renner and data analysis was done exclusively by him.

Strikingly, a comparison between the experimentally obtained SAXS data with theoretical scattering data calculated from the crystal structure shows very high correlation (Figure 4.6C). Furthermore, an *ab initio* bead model was calculated from the SAXS data and fitting of the crystal structure into the resulting model shows very similar shapes (Figure 4.6D), confirming that the protein conformation in solution is, indeed, represented by the crystal structure. In contrast, fitting the Nrd1-RRM structure into the bead model calculated from the Seb1-RRM scattering data shows that the range of conformations sampled in the Nrd1 NMR ensemble is not compatible with the *ab initio* envelope (Figure 4.6E). In addition, the Nrd1-RRM structure shows a substantial amount of disorder and flexibility (Figures 4.2, 4.6E and Bacikova et al. (2014)) which is in contrast to the Seb1-RRM$_{388-540}$. Therefore, the flexibility of the domain in solution was determined from the SAXS data (Figure 4.6F). The dimensionless Kratky analysis performed indicates that the Seb1-RRM$_{388-540}$ is remarkably inflexible, similar to the lysozyme standard used. For a protein such as the Nrd1-RRM that contains ordered regions which are connected by a flexible linker, a different profile would have been expected (Figure 4.7). In summary, the Seb1-RRM$_{388-540}$ and the Nrd1-RRM$_{307-491}$ domains show very different behaviours in solution with the Seb1 RNA-binding region being a lot less disordered and flexible than its *S. cerevisiae* homologue.

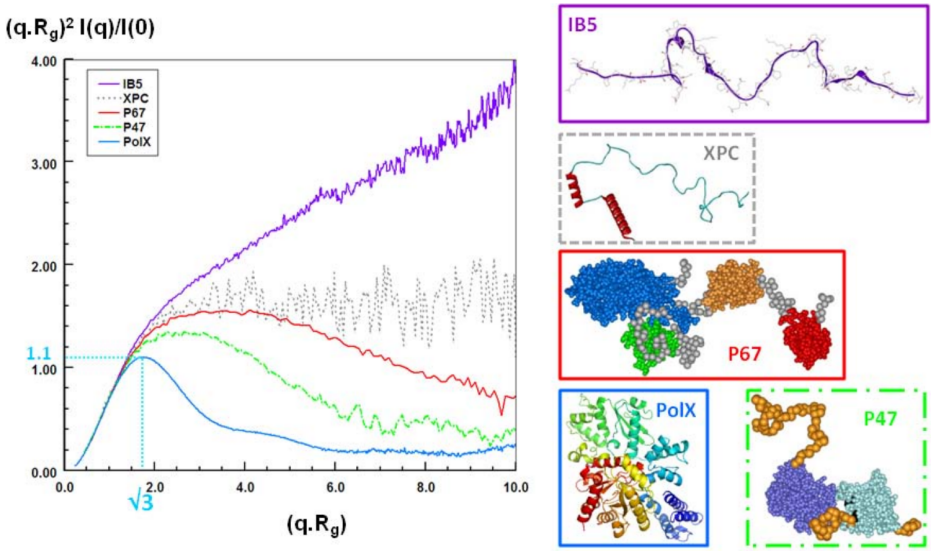


Figure 4.7: Example dimensionless Kratky plots for a range of different proteins
Analysis of SAXS data using the dimensionless Krakty plot results in characteristically shaped curves which depend only on the overall shape and flexibility of the protein but not its size. For each example, the dimensionless Kratky curve is shown on the left and the protein structure on the right using the same colour as for the curve. A perfectly globular protein such as PolX produces a bell-shaped curves (blue) with a maximum at the Guinier-Kratky point ($\sqrt{3}$, 1.104). The maximum increases and the shape of the curve broadens with increasing randomness of the structure. This can be seen increasingly so for a protein consisting of several domains which are connected by flexible linkers (P47, green), a protein with extended conformation (P67, red), and a very disordered protein with only a small amount of secondary structure elements (XPC, grey). A fully disordered and extended protein with short polyproline repeats does not show any bell shape at all but rather the $(q \cdot R_g)^2 \cdot I(q)/I_0$ values keep increasing with increasing $q \cdot R_g$ (IB5, purple). Reprinted from Receveur-Brechot and Durand (2012).

4.3 The RNA sequence does not solely determine Seb1 binding *in vivo*

Next, I wanted to determine whether the purified Seb1-RRM domain can indeed bind to RNA. Therefore, fluorescence anisotropy measurements were performed by Mr Oliver Adams under my supervision using 5'-FAM-labelled 10 nt RNA (sequence: AUUAGUAAAA) containing the Seb1 recognition motif inferred from PAR-CLIP (Figure 2.4A). In addition, it was tested whether presence of the His_6-SUMO-tag would interfere with binding. Therefore, the protein was affinity purified by Mr Oliver Adams as before (Figure 4.4A), however, instead of overnight cleavage, the protein solution was directly subjected to size-exclusion chromatography (Figure 4.8A and B). The K_d values obtained for both proteins His_6-SUMO-tagged $RRM_{388-540}$ and untagged $RRM_{388-540}$ were remarkably similar, however, the untagged protein showed slightly less affinity which is probably

caused by the presence of 75 mM NDSB-201 necessary to prevent the protein from precipitating (Figure 4.8C). This confirms that the purified protein is functional and can indeed engage with RNA of a sequence similar to its cellular target RNAs. Furthermore, the presence of the N-terminal His_6-SUMO tag does not interfere with RNA interaction and tagged protein was therefore used for all further *in vitro* assays.

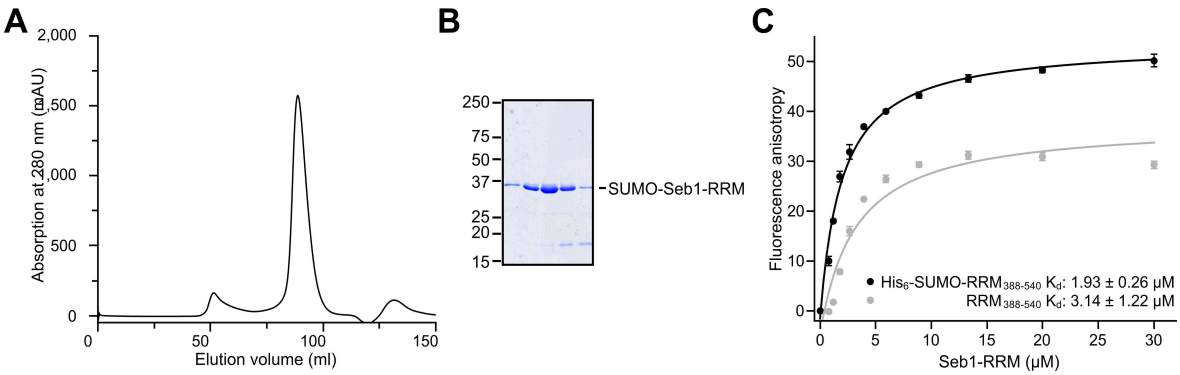


Figure 4.8: The SUMO-tag does not interfere with RNA binding of the Seb1-RRM$_{388-540}$
(A) The Seb1-RRM$_{388-540}$ was affinity purified as shown in Figure 4.4A and the NiNTA eluate was directly used for size-exclusion chromatography. The elution profile as determined by absorption at 280 nm from a HiLoad Superdex 200 16/60 column (GE Healthcare) plotted against the elution volume is shown.
(B) SDS-PAGE was performed to determine the purity of the peak elution fractions from the chromatogram shown in A.
(C) Fluorescence anisotropy assay of the Seb1-RRM$_{388-540}$ with and without a SUMO-tag was performed to measure binding to a 10 nt RNA of the sequence AUUAGUAAAA. The equilibrium dissociation constants (K_d) were determined by curve fitting and are indicated in the plot. Error bars indicate the standard deviation of two independent measurements.

Next, the RNA binding specificity of the His_6-SUMO-RRM$_{388-540}$ was assessed. For this, RNAs were designed that consisted of sequences from the *prt* ncRNA region upstream of the *pho1* gene which shows strong binding of Seb1 *in vivo* (Figure 4.9A). In addition to PAR-CLIP and ChIP-Seq, Seb1 interaction with this region was observed by single-gene ChIP experiments (Watts 2015). Therefore, this region was chosen to assess the sequence specificity of the Seb1 RNA interaction *in vitro* using electrophoretic mobility shift assays (EMSA). Eight 55 nt long RNAs were used that overlapped by 5 nt with the neighbouring RNA sequence on both ends (sequences can be found in section 7.5.8). Although all regions are bound by Seb1 *in vivo*, PAR-CLIP crosslinking levels vary notably

and only region 1 shows strong interaction (Figure 4.9A) even though the only region containing the UGUA motif is region 6. Furthermore, the GUA nucleotides which are more enriched in the Seb1 recognition motif (Figure 2.4B) can additionally be found in region 3. Therefore, if the Seb1-RRM domain does indeed specifically recognise this motif, the highest affinities should be observed to RNAs 3 and 6. However, if the behaviour of Seb1 *in vivo* and *in vitro* is identical, strongest binding should be seen to RNA 1. Indeed, the UGUA motif is specifically recognised as the K_d values of only RNAs 3 and 6 are below 10 µM (Figure 4.9B-D). However, binding to RNAs 1 and 2 occurs with only marginally less affinity (the K_d values are about 2-fold higher) while all other RNAs show weaker binding. In summary, this confirms the *in vivo* specificity of Seb1. However, RNAs of other sequence are also bound, supporting the notion that the Seb1-$RRM_{388-540}$ domain has a relatively broad specificity. Also, the fact that *in vitro* and *in vivo* binding is not identical further indicates that the RNA sequence is not the only determinant of Seb1 recruitment in the cell. What role the structure of the RNA plays is unclear and *in vitro* binding could have been positively or negatively influenced by stable secondary structures that quickly reformed after boiling and cooling on ice. In addition, the stable unfolding of any secondary structure elements which may be recognised by Seb1 could have also influenced the binding reaction.

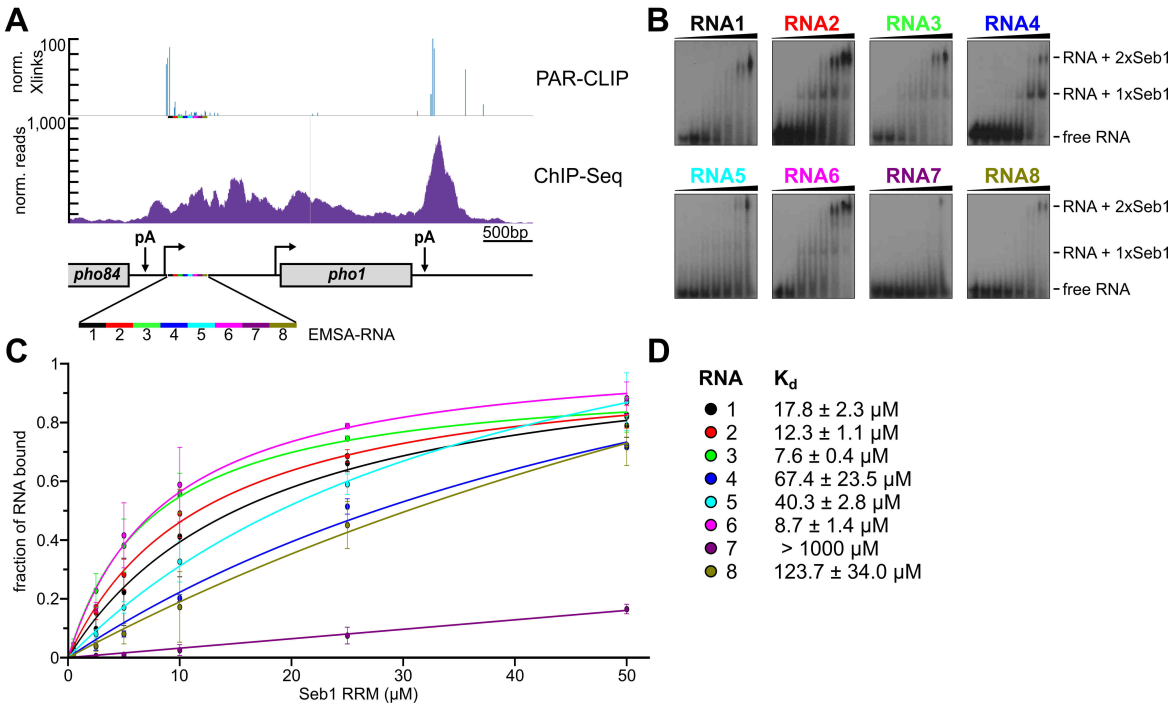


Figure 4.9: The Seb1-RRM$_{388-540}$ has a broad RNA sequence specificity *in vitro*
(A) The Seb1 binding profile at the *pho1* locus as determined by PAR-CLIP (blue) and ChIP-Seq (purple) is shown. Below is a schematic representation of the genomic region including the position of the ncRNA and mRNA promoters (bent arrows) in front of the *pho1* gene, open-reading frames (grey boxes), and pA sites (straight arrow). The sequence of the RNAs used for EMSAs were derived from the region unique to the ncRNA and is indicated by coloured boxes which is repeated directly below the PAR-CLIP track for better visibility of the crosslinking positions.
(B) Binding of the His$_6$-SUMO-Seb1-RRM$_{388-540}$ was assessed by electrophoretic mobility shift assays (EMSAs). All RNAs used were derived from the regions shown in A and are indicated above the gels. A constant amount of ^{32}P-labelled RNA was incubated with increasing protein concentrations (from left to right) and RNA binding was observed by a shift in RNA mobility on native gels which was detected by radiography.
(C) Quantification of the radioactive signals in B was performed and the fraction of RNA that is protein-bound was plotted against the RRM$_{388-540}$ concentration. The colours of the curves correspond to the colours of the regions shown in A. Error bars represent the standard deviation of duplicates.
(D) Dissociation constants (K_d) for each RNA was determined from the plot shown in C.

4.4 Point mutations within the Seb1-RRM can abolish RNA binding

In an attempt to map the RNA binding site on the $RRM_{388-540}$ domain, different point mutations of predominantly basic or aromatic amino acids were designed, which are typical for RNA interactions. The recombinantly expressed and purified proteins were then used for fluorescence anisotropy assays using the same RNA as before. All experiments described in this section were performed under my supervision by two students of the Vasilieva lab, Mr Oliver Adams (Part II) and Mr Miles Huseyin (*DPhil* rotation student).

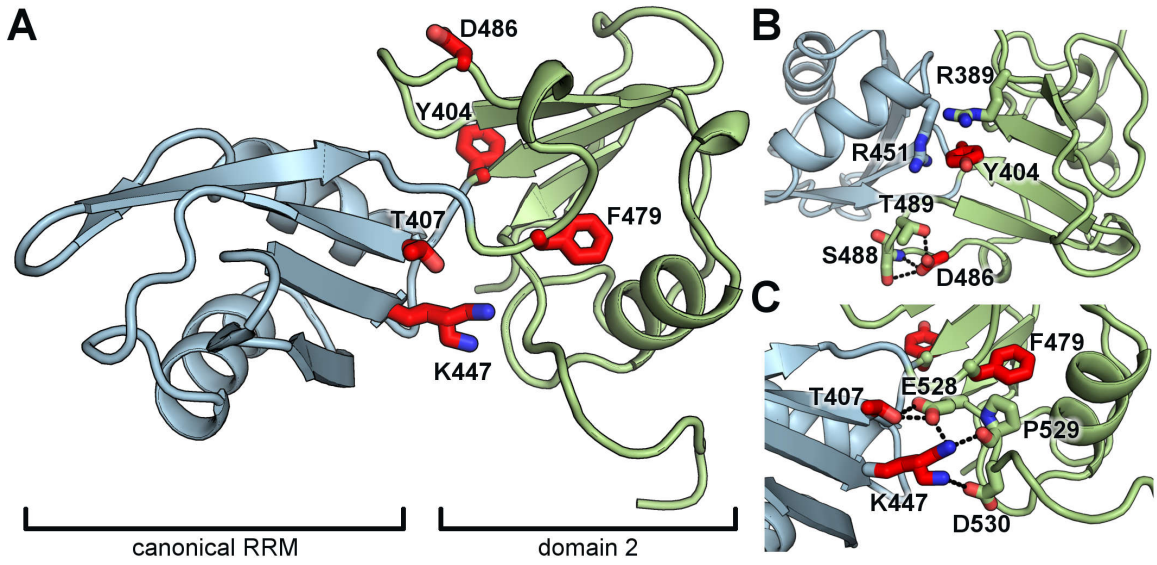


Figure 4.10: Interactions between the two domains are required for protein solubility
(A) Overview of the Seb1-$RRM_{388-540}$ structure with amino acids highlighted in red that, when mutated, lead to insolubility of the domain. Note that two alternative conformations of K447 were clearly distinguishable in the electron density and the residue was modelled accordingly.
(B) Close-up view of the molecular interactions between D486 and Y404 (red) with other amino acids in the structure (shown in blue if part of the canonical RRM or in green if part of domain 2). Polar contacts are indicate by dotted lines.
(C) The molecular interactions between K447, T407, and F479 (red) and other amino acids are depicted and colour coded as in B.

Some of the amino acid mutations tested yielded no expression or unstable protein. Interestingly, most of them are positioned at the interface between the canonical RRM and the additional domain (Figure 4.10A). K447 and T407 which are located on the canonical RRM form polar contacts with residues on domain 2, helping to correctly orient the two domains to each other (Figure 4.10C). Y404 on domain 2 is similarly involved in π

stacking interactions with two arginines of the canonical RRM (Figure 4.10B). This indicates that the rigid interfaces between the domains are critical for the structure as a whole. Furthermore, D486 and F479 seem to stabilise domain 2 overall (Figure 4.10B and C). Taken together, this shows that any perturbation of this domain or the orientation of the two domains towards each other dramatically impedes the solubility of this arrangement.

Mutations of all other amino acids tested yielded highly pure protein which could be used for fluorescence anisotropy measurements (Figure 4.11A). Most of the point mutations tested had little effect on RNA binding, however, some drastically reduced or even completely disrupted RNA interactions (Figure 4.11B and C). The mutations F445A, R472E, and F487A had a particularly severe effect while S492A only mildly decrease the protein's affinity to RNA.

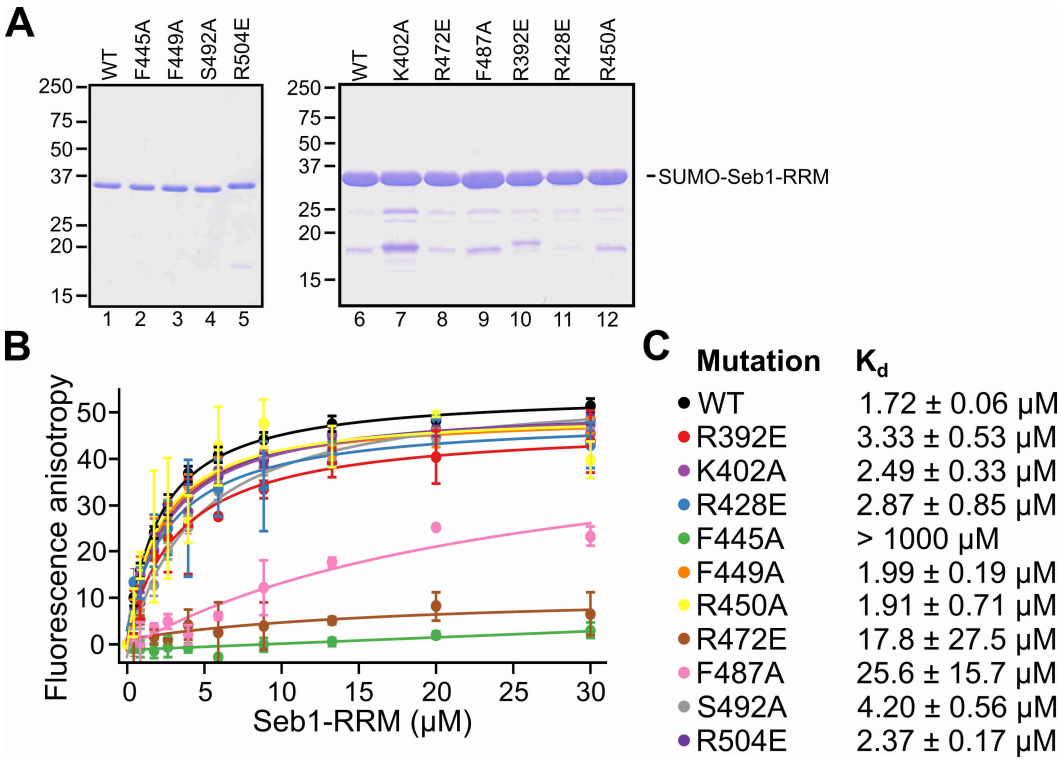


Figure 4.11: Some $RRM_{388-540}$ point mutations disrupt RNA binding
(A) SDS-PAGE of all His_6-SUMO-$RRM_{388-540}$ point mutants used for the fluorescence anisotropy assay. The mutations are indicated above the lanes.
(B) Fluorescence anisotropy of the proteins shown in A was measured as described in Figure 4.8C.
(C) K_d values of the protein-RNA interaction measured in B was determined as before. The colours of the mutants are as in B.

Seb1-RRM$_{388-540}$ possesses two basic patches which are flanked by hydrophobic residues (Figure 4.5B) and could form essential RNA contacts. Therefore, most of the amino acid mutations tested were positioned on or near one of those positively charged areas. However, looking at the location of mutations that do not affect RNA binding (Figure 4.12), the basic region formed by α-helix 2 of the canonical RRM domain which continues into domain 2 (Figure 4.5B, bottom) does not seem to be involved in any RNA interactions. Furthermore, basic amino acids on or near the β-sheet which is formed by domain 2 do not directly interact with the RNA either. This excludes the possibility that the additional domain engages with RNA in a similar way as canonical RRM domains, which is also consistent with the fact that neither the RNP1 nor the RNP2 motif can be found.

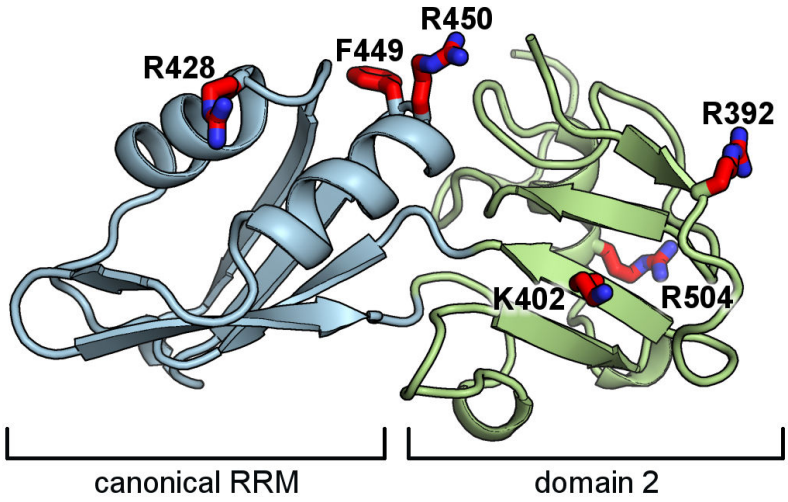


Figure 4.12: Neither α-helix 2 nor the β-sheet on domain 2 bind to RNA
Overview of the Seb1-RRM$_{388-540}$ structure with amino acids highlighted in red that were mutated but do not affect RNA binding as determined by fluorescence anisotropy (Figure 4.11). The orientation of the structure is similar to Figure 4.5 bottom.

In contrast, the other basic patch as well as neighbouring aromatic amino acids which are located on the β-sheet of the canonical RRM do seem to be important for RNA binding (Figures 4.11 and 4.13). Particularly severe is the effect of the F445A mutation, which is part of the conserved RNP1 motif on the 3rd β-strand of the canonical RRM (Figure 4.1). This indicates that the Seb1-RRM binds RNA similarly to other known RRM domains. This is further supported by mutation analysis of R472 located on the 4th RRM-β-strand which also affects RNA interaction. In addition, a second phenylalanine (F487)

which is part of domain 2 seems to be necessary for RNA binding and mutating S492 to alanine on the same domain mildly affects RNA affinity. This indicates that the additional domain observed in the Seb1-RRM$_{388-540}$ structure directly interacts with RNA.

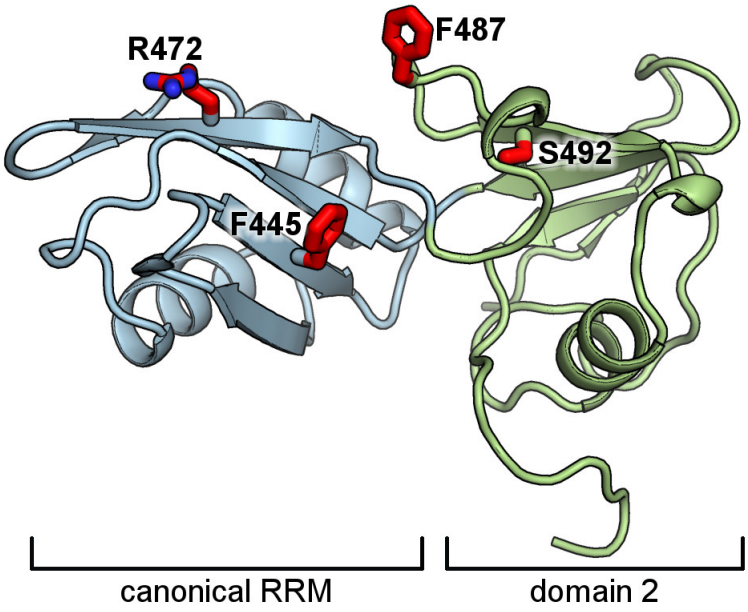


Figure 4.13: The canonical RRM and amino acids on domain 2 interact with RNA
Overview of the Seb1-RRM$_{388-540}$ structure with amino acids highlighted in red whose mutation led to a decrease in affinity to RNA as determined by fluorescence anisotropy (Figure 4.11). The orientation of the structure is identical to Figure 4.5 top.

4.5 Other CID-RRM proteins likely have a similar RNA-binding region

Because of the unusual fold of this RNA-binding region, the RRM$_{388-540}$ sequence was used to search for proteins containing similar sequences. For this, two different servers were utilised and the results compared. Both HMMER (Finn et al. 2011) and PSI-BLAST (Altschul et al. 1997; Schäffer et al. 2001) could not find any significant hits in prokaryotes and plants. However, most other eukaryotes seem to contain proteins with domains similar to the Seb1-RRM$_{388-540}$ (Figure 4.14A). This indicates that the protein evolved in the opisthokonts (which include fungi and metazoans) only after branching off of the plant kingdom. Due to the fact that a canonical RRM region is included in the search sequence, all proteins containing RRM domains show some level of similarity (Figure 4.14B, compare top architecture with the other two). However, protein hits with

higher significance had larger regions of sequence similarity indicating that the domain 2 might be present in addition to a canonical RRM. Manually sorting the results by visually inspecting the length of the aligned region enriched the results for CID-RRM proteins like Seb1 on both servers. Even though some proteins did not have either the CID or RRM domain annotated, upon closer inspection all of them turned out to contain both domains (e.g. Seb1 is part of the second architecture type in Figure 4.14B). This result strongly suggests that this novel $RRM_{388-540}$ fold is unique to CID-RRM proteins.

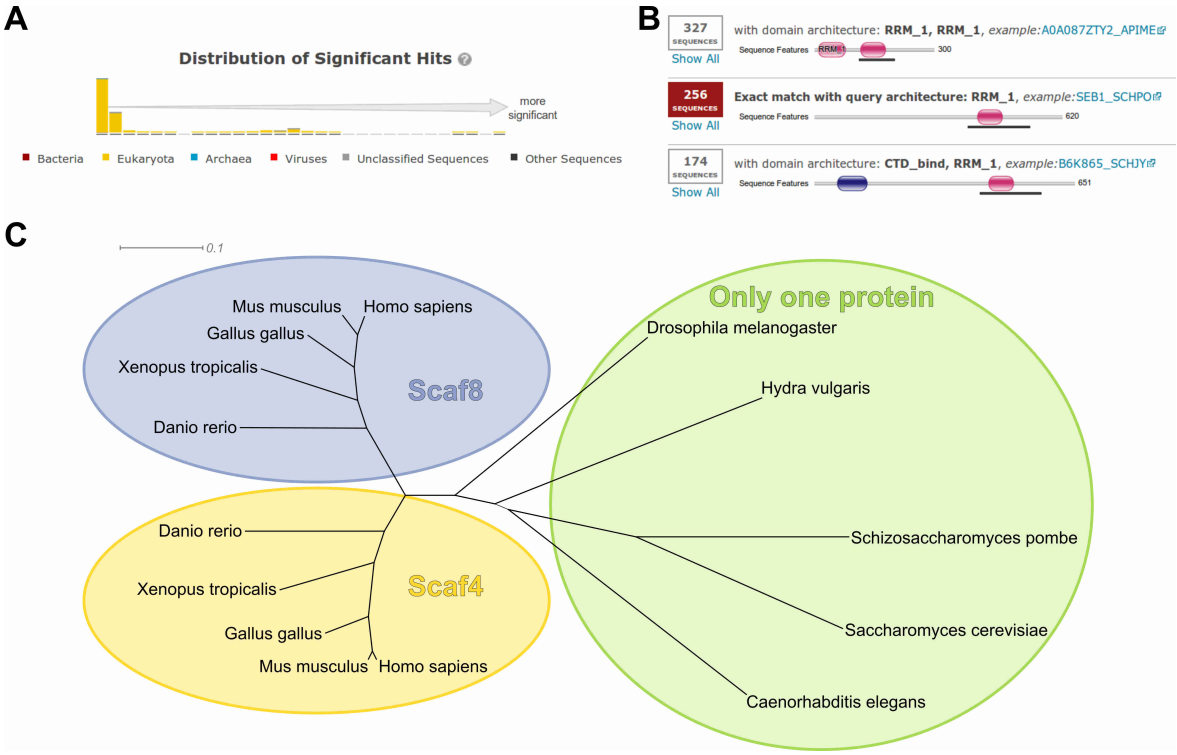


Figure 4.14: Other CID-RRM proteins have sequences similar to the Seb1-$RRM_{388-540}$
(A) HMMER output showing the organisms in which proteins with sequences similar to the Seb1-$RRM_{388-540}$ region could be found, sorted by the significance. Note that only eukaryotes could be identified.
(B) HMMER output showing three examples of protein domain architectures of proteins that were found to contain sequences similar to the Seb1-$RRM_{388-540}$. The region of alignment is indicated by a black bar below the schematics. Note how the first example aligns to a much shorter region than the other two. Also note that Seb1 was found as an example of the second type because the CID is not annotated in the database used by the server.
(C) Phylogenetic comparison between regions homologous to the Seb1-$RRM_{388-540}$ from CID-RRM proteins in different organisms. Lower eukaryotes only contain one CID-RRM protein (green), while higher eukaryotes have evolved two such proteins (Scaf4 in yellow and Scaf8 in blue). *Schizosaccharomyces pombe*: fission yeast, *Saccharomyces cerevisiae*: budding yeast, *Caenorhabditis elegans*: a nematode worm, *Hydra vulgaris*: fresh-water polyp, *Drosophila melanogaster*: fruit fly, *Danio rerio*: zebrafish, *Xenopus tropicalis*: tropical clawed frog, *Gallus gallus*: red junglefowl, *Mus musculus*: house mouse, *Homo sapiens*: human

To further understand the evolution of this fold, a phylogenetic comparison was performed between some of the regions homologous to the $RRM_{388-540}$. Exemplar CID-RRM proteins from fungi, lower metazoans such as worms and higher metazoans such as mammals were chosen, aligned, and the evolutionary distance calculated (Figure 4.14C). Interestingly, around the time vertebrates occurred, the CID-RRM gene must have split and two different CID-RRM proteins that evolved separately can be observed (called Scaf4 and Scaf8 in most organisms). Both proteins are approximately equally far away from the Seb1-$RRM_{388-540}$. The fact that the Seb1-$RRM_{388-540}$ structure seems to be conserved until humans might indicate that the function is conserved, too. However, why higher eukaryotes have developed two CID-RRM proteins and what cellular role they play, is currently not understood.

4.6 RNA-binding is essential for cell viability

In order to test whether RNA-binding is important for the function of Seb1 *in vivo*, two amino acids in the Seb1-$RRM_{388-540}$ domain that lead to a decrease in RNA affinity *in vitro* were introduced into *S. pombe* cells. This experiment was performed by Mr Oliver Adams under my supervision. The mutation F445A was chosen as it had the strongest effect on binding and is part of the conserved RNP1 motif, and S492A which had only little effect but is part of domain 2. Both mutations were introduced using the same *nmt1*-repressible system for WT-Seb1 as previously (Figures 2.2, 3.7 and 4.15A). Similarly to before, protein expression of WT and mutated Seb1 in thiamine-lacking and -containing media was as expected (Figure 4.15B) and spot assays could therefore be performed (Figure 4.15C). The S492A mutation has no effect on cell growth, indicating that the decrease in RNA affinity is not sufficient to influence the function of Seb1 *in vivo*. In contrast, the F445A mutation does not support cell viability, suggesting that an interaction between Seb1 and RNA is, indeed, required for the cells. This therefore demonstrates that Seb1 needs to bind both RNA and the Pol II-CTD in order to function properly *in vivo*.

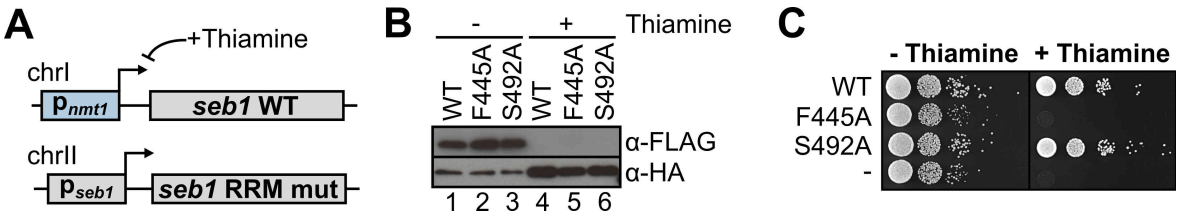


Figure 4.15: Seb1-RNA binding is necessary for cell viability
(A) A schematic representation of the expression system used to assess effects of Seb1-RRM point mutations on cell viability is shown. See Figure 2.2A for details.
(B) Expression of the WT and mutated proteins in media containing (+) and lacking (-) thiamine was tested by Western blotting. Due to the different tags used, the WT and mutated copies can be specifically detected by α-FLAG and α-HA antibodies, respectively.
(C) The effects of the RRM mutations were tested by spot assays as described in Figure 2.2C and the same controls were used as previously.

4.7 Discussion

Crystallisation and SAXS studies of the Seb1 RNA-binding region ($RRM_{388-540}$) illuminated an unusual structure consisting of a canonical RRM domain and a second domain 2 which is formed by folding together of regions that lie N- and C-terminally of the RRM (Figure 4.5 and 4.6). A structural homology search with the Dali server was performed using only the second domain and did not result in any confident hits, indicating that no structure with a similar fold is known. In addition to it being essential for protein solubility and most likely proper functioning of the canonical RRM domain (Figure 4.3), *in vitro* binding assays confirm that this region is, indeed, important for RNA binding (Figure 4.8 and 4.9). Furthermore, mutating amino acids that are necessary for this interaction results in lethality (Figure 4.15).

Comparing the Seb1-$RRM_{388-540}$ structure with the Nrd1-RRM structure shows that Seb1 is more ordered and less flexible in solution (Figures 4.7 and 4.6). The RNA affinity of both domains are comparable, with Seb1 having a K_d value between 1.72 and 7.6 μM depending on the method used (Figures 4.8C and 4.9D). Nrd1 has a K_d value of 10.1 μM to a 4 nt RNA of the sequence GUAA by fluorescence anisotropy and, interestingly, binding to longer RNAs containing the same motif was even stronger (up to K_d = 1.5 μM, Bacikova et al. 2014). Furthermore, the RNA-binding regions of both proteins are able to interact with RNAs of different sequences and still show affinities comparable to their

respective recognition motif (Figure 4.9 and Bacikova et al. (2014)). This indicates that despite structural differences, both domains behave in a similar way. It is therefore possible that the Nrd1-RNA binding region adopts a similar fold as the Seb1-RMM$_{388-540}$ when bound to RNA and flexibility differences are only apparent in an RNA-free state. This is also supported by the unusual arrangement that both domains show in which regions C- and N-terminally of the canonical RRM interact with each other. However, no experiments have been performed to date to address this.

Mutational analysis of the Seb1-RRM$_{388-540}$ helped to elucidate the RNA binding region on the protein (Figures 4.11 and 4.13). As expected, conserved amino acids on the β-sheet that is formed by the canonical RRM domain seem to be important for RNA interactions, similar to other RRMs. Surprisingly, however, amino acids that are part of the additional domain are also important for RNA binding. Even though this additional domain is somewhat reminiscent of RRM domains, it is not the β-sheet that binds RNA. Instead, a phenylalanine on a loop which is in close proximity to the canonical RRM domain might directly contact the RNA (Figure 4.13). In Nrd1, the homologous amino acid is R413 (Figure 2.9) which could equally be able to form stacking interactions with RNA bases. Unfortunately, in Nrd1-RRM mutation studies, this amino acid was not included.

The Dali server was again used to find structures that are similar to the Seb1-RRM$_{388-540}$ but have RNA bound. The most significant hits were the human Hu antigen R (HuR, PDBID 4ED5) and the Sex-lethal protein from *Drosophila melanogaster* (PDBID 4QQB). Both of them engage with RNA in a very similar way, and aligning these structures to the canonical RRM domain of Seb1 shows high structural conservation (RMSD of the HuR and Seb1 RRMs: 1.1 Å). Furthermore, many of the amino acids important for RNA recognition seem to be conserved and RNA binding of the Seb1-RRM domain is therefore likely to be equivalent to the HuR or Sex-lethal RRM domains. I therefore propose a model of binding whereby the RNA is recognised by amino acids on the β-sheet of the canonical RRM (Figure 4.16 F445, F409, and R472). In addition R415 and R442 which are located on loops in close proximity to the β-sheet could also engage in direct RNA

contacts. In the case of HuR, RNA binding is performed by two neighbouring RRM domains and the RNA is sandwiched between them (Wang et al. 2013). In a similar way, amino acids on the additional domain could also interact with the RNA, specifically R482 which would be in close proximity to the RNA. Furthermore, F487 was shown to be important for RNA binding and movements of the loop that contains the amino acid could bring it closer to the RNA upon binding. More experiments are necessary to test this model, however, it would render the Seb1-RNA binding region and those of homologous proteins unique, as no similar mode of binding has been observed to my knowledge.

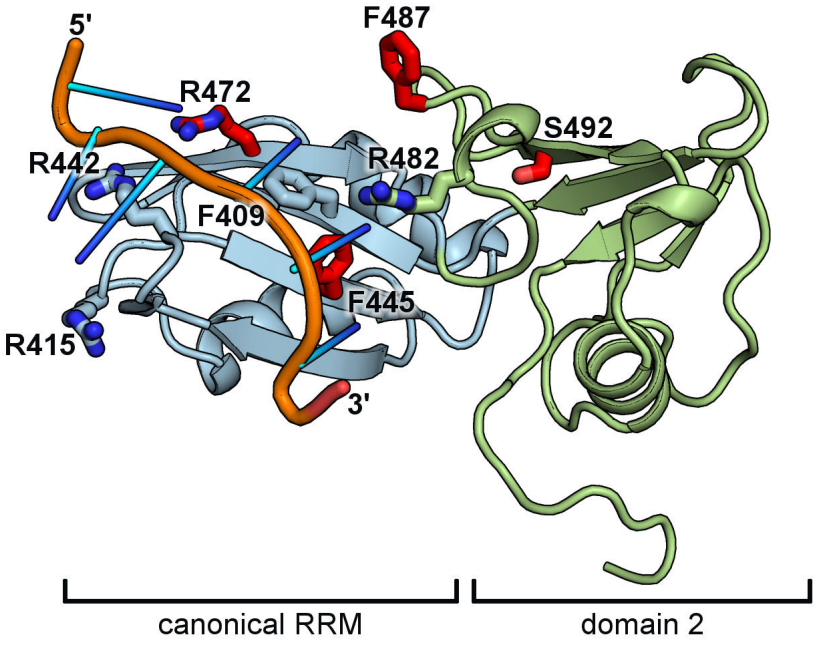


Figure 4.16: Hypothetical model for RNA binding of the Seb1-RRM$_{388-540}$
The canonical RRM of Seb1 (blue) was aligned to the RRM domain of the human Hu antigen R (HuR) bound to RNA (PDBID 4ED5). Six nucleotides of the RNA are shown with the direction indicated. Amino acids which when mutated in Seb1 affected RNA binding are highlighted in red, other amino acids that could be important for RNA recognition but were not mutated are shown in either blue (if part of the canonical RRM) or green (if part of domain 2). The orientation of the structure is similar to Figures 4.5 top and 4.13.

CHAPTER 5: CELLULAR CONSEQUENCES OF SEB1 POINT MUTATIONS

In Chapters 3 and 4, I could demonstrate that introducing point mutations into either the CID or RRM domain and thereby preventing Pol II-CTD or RNA interaction, respectively, results in lethality (Figures 3.7 and 4.15). However, the reason for this strong effect on cell viability is unknown. In Chapter 2, it was shown that Seb1 binds to the 3'end of protein-coding and non-coding genes and that it interacts with the canonical 3'end processing machinery (Figure 2.5 and Table 3). The assumption is therefore that Seb1 is important for 3'end processing and/or transcription termination and that preventing Seb1 from being recruited to the 3'end of genes by either abolishing RNA or Pol II binding, impedes 3'end formation and/or termination which causes the lethal phenotype.

To test whether this is actually the case, cells carrying CID and RRM point mutations were examined in more detail utilising RNA sequencing (RNA-Seq) and other *in vivo* methods. Concerning CID mutations, the S22D-K25E double mutant was used which prevents recruitment to S5P-CTD, Y64K which cannot bind to S2P-CTD, and the S22D-K25E-K124E triple mutant which cannot bind to any type of CTD independent of the phosphorylation mark. The only viable CID mutation chosen is S22D-K25E which might not have a very strong effect on transcription. However, comparing the Y64K with the triple mutant should additionally help us to understand whether the S5P-CTD interaction plays any role or if only binding to S2P-CTD is important for Seb1 functioning. Furthermore, RNA binding can be completely prevented in the F445A mutant, while the S492A mutation only has a mild effect on RNA affinity. Both mutants were used to examine the importance of Seb1-RNA interactions *in vivo*. Also, Seb1 will be compared to Pcf11 at the end of this chapter in order to better understand the relationship between these two CID-containing proteins during 3'end formation and transcription termination. All experiments performed in this chapter that use Seb1 mutants were done in the strains shown in Figures 3.7 and 4.15 after depletion of WT-Seb1 in thiamine-containing medium for 24 h.

5.1 Seb1 mutations cause read-through transcription genome-wide

RNA-Seq was performed in all the strains mentioned above that carry mutations in either the Seb1-CID or -RRM domain and steady-state levels of all cellular RNAs except rRNAs were determined. Ribodepletion, library preparation, and sequencing was done by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (Oxford). All strains carrying lethal mutations (Y64K, S22D-K25E-K124E, and F445A) as well as the WT were sequenced in triplicates, S22D-K25E and S492A in duplicates. The resulting reads were aligned to the *S. pombe* genome and replicates of the same strain show very high reproducibility (Figure 5.1). As expected, the two strains containing non-lethal mutations (S22D-K25E and S492A) are more similar to WT than any of the other strains. Interestingly, the similarity between the two CID mutants Y64K and S22D-K25E-K124E is as high as between replicates showing that disruption of the S2P-CTD interaction has the same effect as completely preventing CTD binding. Any phenotype that may result from abolishing S5P-CTD binding in the triple mutant is therefore masked by disrupting the interaction with S2P. Furthermore, the lethal RRM mutation F445A also causes a similar phenotype as the lethal CID mutations, however, they are not identical.

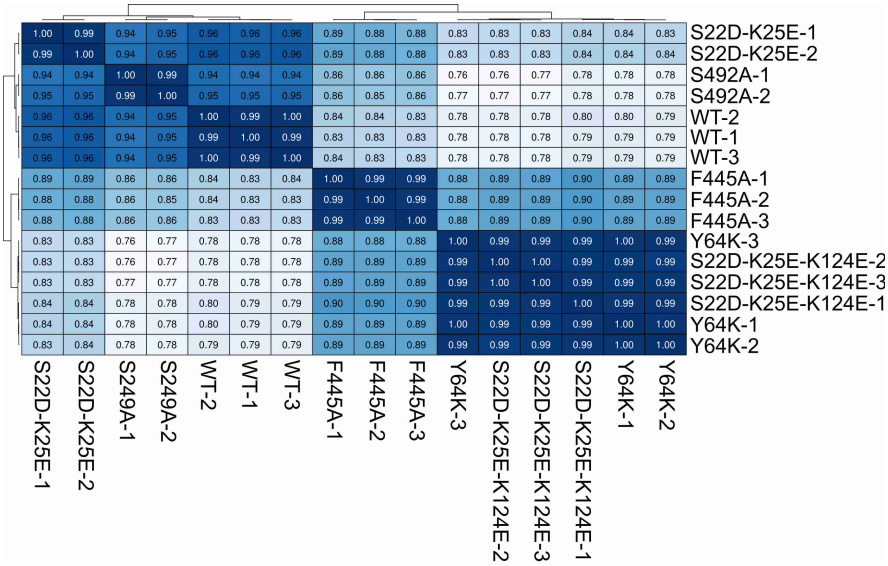


Figure 5.1: RNA-Seq is highly reproducible between replicates
Spearman correlation coefficients were calculated between all repeats and all strains used for RNA-Seq. For this, the genome was split into 100 nt binds and reads were counted in the bam alignment files for each bin and compared using deepTools (Ramírez et al. 2014). Colours indicate low (white) to high (blue) correlation.

To understand the effects that Seb1 mutations have on transcription, two highly expressed protein-coding genes, *adh1* and *pho1*, were investigated more closely. On both genes, higher levels of RNA past the pA site can clearly be observed in the F445A RRM mutant and additionally in all CID mutants on the *pho1* locus (Figure 5.2A and B, indicated by arrows). In order to confirm that this phenotype is caused by decreased Seb1 recruitment, single-gene ChIP-qPCR was performed. Indeed, all mutations show reduced recruitment to either gene as compared to WT-Seb1 which is especially pronounced for mutations causing lethality (Figure 5.2C and D). This strongly suggests that the absence of Seb1 binding to Pol II or RNA is what actually causes read-through transcription.

The higher levels of RNA past the pA site observed at the *adh1* and *pho1* genes, was confirmed by reverse transcription (RT) and cDNA detection by qPCR. Strikingly, both methods detect increased transcript levels past the pA site in the same strains (Figure 5.2E and F). This confirms the result observed by RNA-Seq and suggests that read-through transcription is indeed what causes the lethal phenotype in many of the mutants.

Next, ChIP-qPCR was utilised again to detect Pol II levels across both genes. At the *adh1* locus, all mutants except F445A show very similar amounts of Pol II until the pA site as compared to WT (Figure 5.2G). However, even though F445A has already decreased levels of polymerase before the pA site in comparison to WT, the levels are higher past the pA site which is consistent with the high level of read-through observed in this mutant. At the *pho1* locus, all mutants show the same amount of polymerase until just after the pA site (Figure 5.2H). Strikingly, however, all CID mutants have higher levels of Pol II past the pA site, which is consistent with the higher RNA levels observed in this region. Taken together, this indicates that the higher levels of transcripts after the pA site on both loci is caused by transcription defects rather than an increase in transcript stability. Furthermore, the decrease in Seb1 levels at 3'ends in these mutants is not caused by lower polymerase levels but rather by a lack of recruitment due to disruption of Pol II or RNA binding.

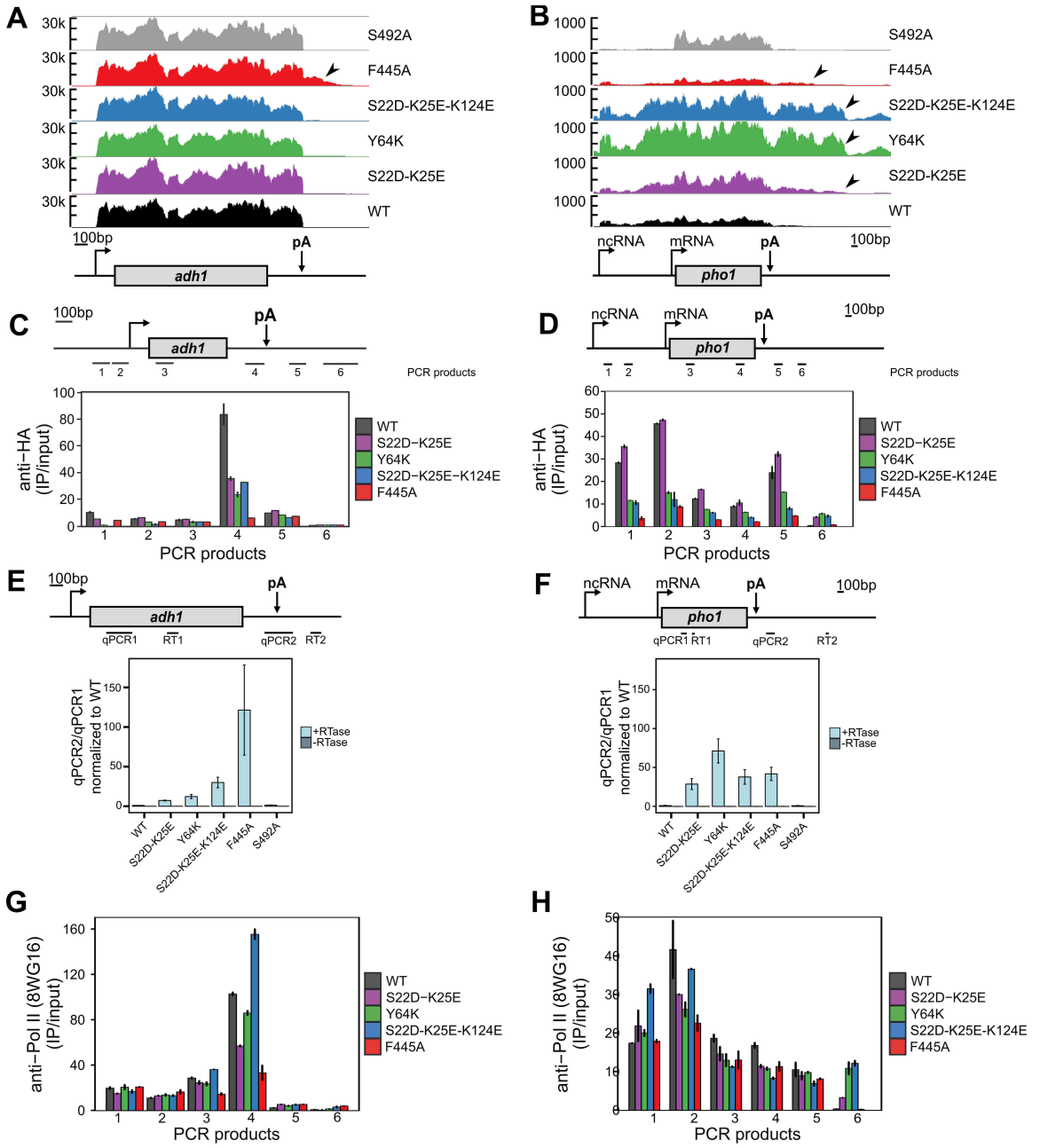


Figure 5.2: Decreased Seb1 recruitment causes read-through on *adh1* and *pho1*
(A, B) Reads resulting from RNA-Seq in the indicated mutants were aligned to the *S. pombe* genome and one replicate of each strain is shown for the *adh1* (A) and *pho1* (B) loci. Arrows indicate read-through transcription.
(C, D) Recruitment of Seb1-HA was determined by ChIP-quantitative (q)PCR in the indicated strains across the *adh1* (C) and *pho1* (D) genes. The positions of the qPCR products relative to the genes is shown schematically above the plots. Error bars indicate standard errors of two replicates.
(E, F) Reverse-transcription (RT) in combination with qPCR was used to detect read-through transcripts on *adh1* (E) and *pho1* (F) in the indicated mutants. The position of the RT primers and qPCR products are depicted schematically above the plots. The read-through signal (qPCR2) was normalised to the gene expression level (qPCR1) and the mutants were normalised to the WT (= 1). A negative control lacking reverse transcriptase (-RTase) was used. Error bars indicate standard deviations of three replicates.
(G, H) ChIP-qPCR using the same primers as in C and D was performed to detect Pol II across the *adh1* (G) and *pho1* (H) genes. The phosphorylation-independent antibody 8WG16 was utilised. Error bars indicate standard errors of two replicates.

In order to better understand the phenotype observed at the *adh1* gene, Northern blotting was performed by Mr Oliver Adams in all CID and RRM mutants. Interestingly, the level of *adh1* mRNA produced does not change a lot in any of the strains, indicating that the mRNA levels are not affected despite increased transcription past the pA site (Figure 5.3). Because defective 3'end cleavage is known to cause degradation of the pre-mRNA (Minvielle-Sebastia et al. 1991; Torchet et al. 2002; Libri et al. 2002; Milligan et al. 2005), this suggests that 3'end processing is still functional and that the read-through transcripts are caused by defective termination.

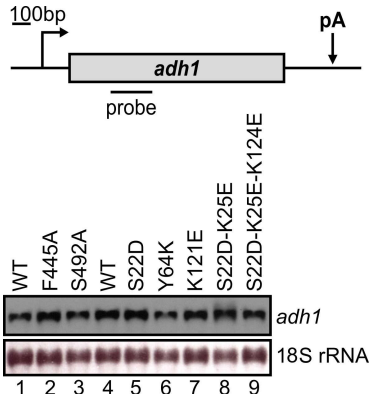


Figure 5.3: *adh1* mRNA levels are not changed in Seb1 mutants
Strand-specific Northern blotting was performed in all Seb1 mutants constructed in this study to detect *adh1* mRNA levels using a probe against the indicated region. Methylene blue staining of the 18S rRNA was used as a loading control.

Due to the observed read-through transcription, a greater proportion of the genome is transcribed in the mutants as compared to WT which makes data normalisation by library size inappropriate. However, since expression of *adh1* does not change, the signal observed on this gene was used to normalise RNA-Seq data shown in this thesis.

To understand the degree of read-through transcription caused by mutating Seb1, the analysis was extended to the whole genome. For this, reads past pA sites were normalised to gene expression levels and compared between WT and the different mutants (Figure 5.4). Strikingly, all mutants except S492A show significantly more read-through than WT suggesting that Seb1 is indeed necessary for proper transcription termination on a genome-wide scale. Surprisingly, S492A has the opposite phenotype.

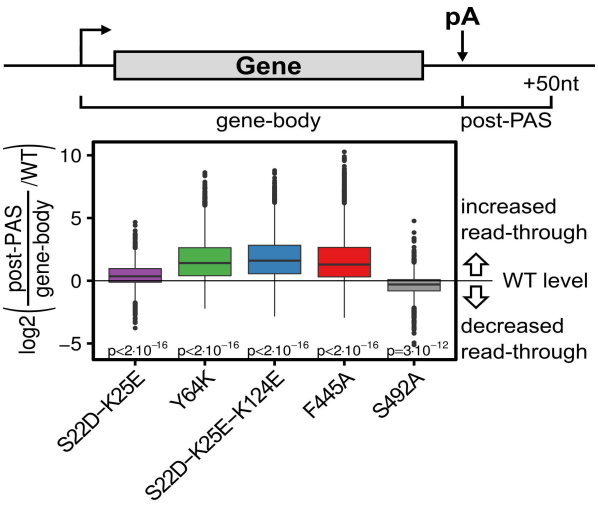


Figure 5.4: Read-through transcripts in Seb1 mutants can be detected genome-wide Read-through transcript levels were quantified in different point mutants for each of the 5119 non-multicistronic, annotated genes by counting the reads in a window from the pA site to 50 nt downstream of it and dividing each count by the total amount of reads coming from the body of the gene (as indicated in the schematics). The change in read-through compared to WT was plotted for the different mutants using a log2 scale. The significance of the overall difference between WT and each mutant was determined using the Wilcoxon-Mann-Whitney test and is indicated below each box.

Furthermore, the level of transcriptional read-through was assessed for every gene in the genome. The S22D-K25E and S492A mutants do not contain any one gene that shows significantly more read-through than WT (i.e. $p < 0.5$). This is consistent with the mild phenotype observed in these strains. In contrast, a large number of genes in strains carrying lethal Seb1 mutations show a significant increase in read-through transcription. More specifically, 10.7% of all non-coding and 62.3% of all protein-coding genes have significantly increased reads past their annotated 3'end in all three mutants. Strikingly, for both protein-coding and non-coding genes, it is largely the same transcripts that are affected in the different mutants (Figure 5.5A). Furthermore, less than 5% of protein-coding genes that have significantly more read-through in the F445A RRM mutant do not show the same phenotype in at least one of the CID mutants. This suggests that binding of Seb1 to both RNA and Pol II is necessary for proper termination. Also, genes with more read-through transcription show higher expression levels and longer half lives (Figure 5.4B) and read-through transcripts are inherently unstable which might indicate that the actual number of affected genes is even higher but escape detection by RNA-Seq.

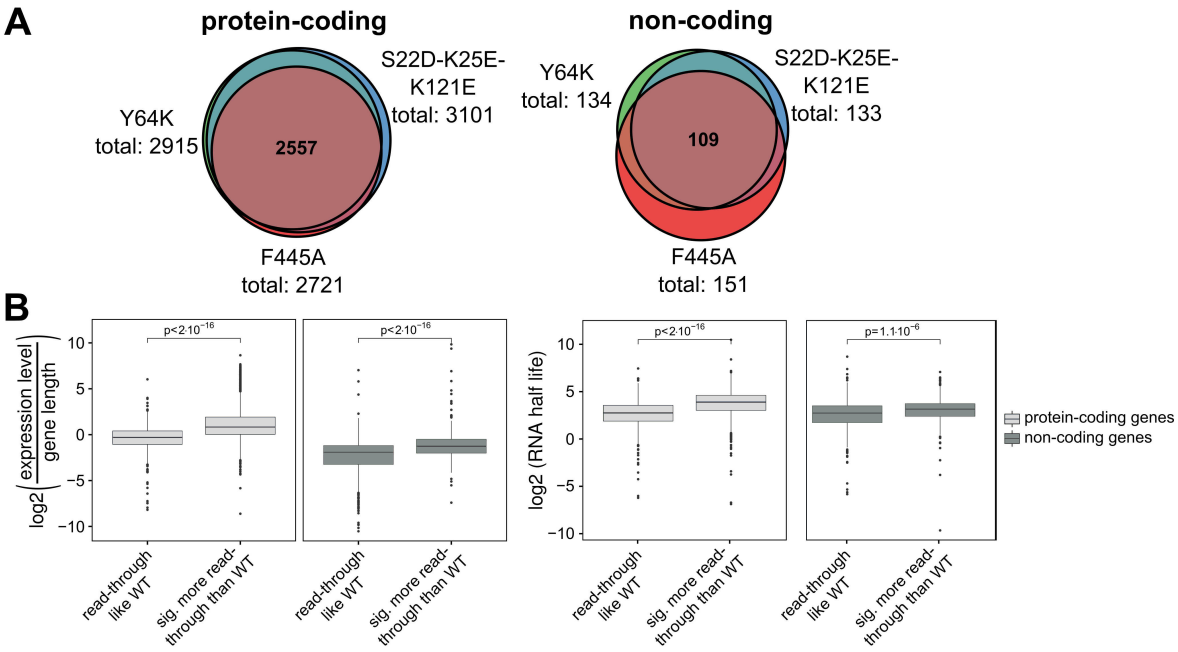


Figure 5.5: All classes of Pol II transcripts show read-through transcription
(A) The significance of the difference in read-through between WT and the different mutants calculated in Figure 5.4 was assessed for every gene individually using the Kruskal-Wallis test. In the case of the S22D-K25E and S492A mutants, no genes showing significantly more read-through than the WT could be detected. For all other strains, the overlaps between genes with significantly ($p < 0.05$) more read-through than WT are depicted as Venn diagrams for protein-coding and non-coding genes separately.
(B) All genes were split into two groups, those that show significantly more read-through in all three mutants (overlap between all three circles in A), and those that do not. The length-normalised gene expression level in WT (left) and the RNA half life from (Eser et al. 2016, right) were plotted as log2 for protein-coding (light grey) and non-coding (dark grey) genes separately. The significance of the differences between the two groups was calculated using the Wilcoxon-Mann-Whitney test and is indicated above the boxes. Number of genes in each group is as follows: read-through like WT: protein-coding: 1548, non-coding: 904; more read through than WT: protein-coding: 2557, non-coding: 109

To take this analysis even further, it was investigated whether some genes depended more on one binding event, or if both RNA and Pol II binding are necessary for proper termination. For this, the extent to which read-through transcription changes was compared between RRM and CID mutants. Strikingly, there is a significant and very high correlation between the F445A RRM mutant and both CID mutants (e.g. Y64K in Figure 5.6) which additionally indicates that binding to RNA and CTD is equally important for transcription termination on all genes. As expected, the correlation between the two CID mutants is even higher ($\rho = 0.98$, $p < 2 \cdot 10^{-16}$, data not shown), further supporting the notion that binding to S2P but not S5P is required for termination.

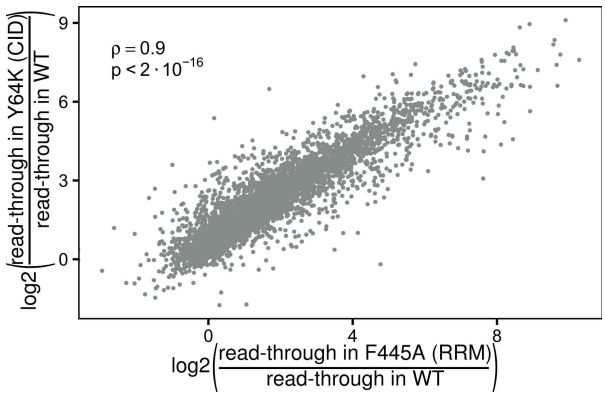


Figure 5.6: Read-through transcription is highly correlated between the F445A and Y64K mutant
Scatter plot showing the log2-fold change in read-through levels for every gene calculated as in Figure 5.4 of the CID mutant Y64K and the RRM mutant F445A. The Spearman correlation coefficient ρ and significance p are indicated in the plot.

The consequences that read-through transcription has for the expression of neighbouring genes will be illustrated using a representative example region on chromosome III. Transcriptional read-through (indicated by arrows in Figure 5.7) can interfere with downstream genes (e.g. *SPCC297.06c* reading into *set7* in Figure 5.7) as well as antisense genes (e.g. *SPNCRNA.1239* and *SPCC1223.14* in Figure 5.7) which can cause down-regulation of gene expression. However, extended transcripts can lead to a seemingly higher amount of reads in genes that are not expressed in WT and are likely to be neither expressed in the mutants (e.g. SPNCRNA.1240 in Figure 5.7). The situation is even more complicated when multiple genes are in close proximity and termination defects seem to merge the resulting transcripts (indicated by red bar in Figure 5.7). However, whether any transcripts that long are actually present *in vivo* cannot be determined by RNA-Seq. Also, how these transcripts are processed and which parts are translated into protein, is not clear. All this makes it impossible to assess if expression levels of individual transcripts change in the mutants. Nonetheless, this massive deregulation of transcription termination seen in the Seb1 mutants does have drastic effects on gene expression overall, influencing protein-coding as well as non-coding RNAs and is most likely the reason for the observed lethality.

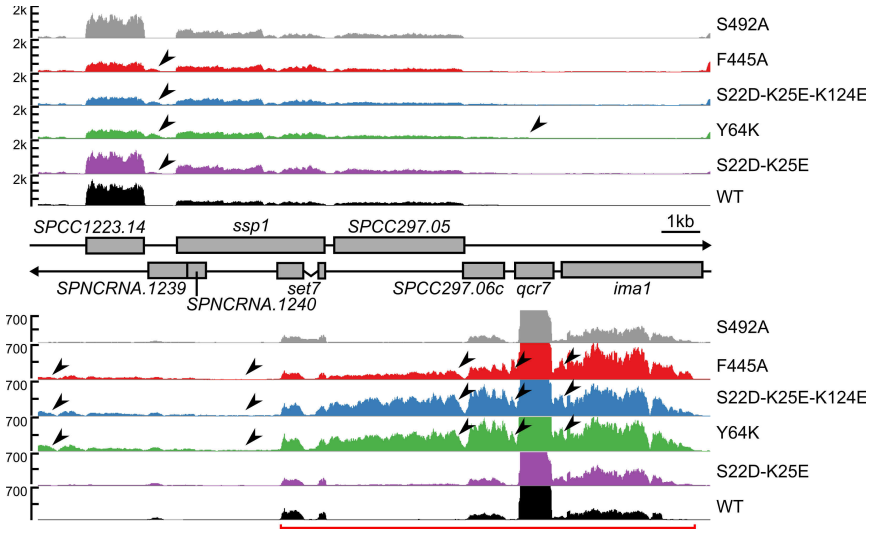


Figure 5.7: Read-through transcription interferes with nearby gene expression
An alignment of reads to a region on chromosome III of the *S. pombe* genome as determined by RNA-Seq in the indicated mutants is shown. Arrows indicate regions of transcriptional read-through, the red bar marks a region with uninterrupted transcription.

5.2 Seb1-dependent genes terminate more efficiently in WT

Changes of transcriptional read-through observed in the Seb1 mutants were correlated with a number of different features. As expected, no correlation (Spearman correlation coefficient $\rho < \pm 0.3$ or $p > 0.5$) with the distance to up- or downstream genes, the number of introns, the length of 5' and 3' UTRs, or the length of the gene itself could be found (data not shown). Surprisingly, there is also no correlation with the amount of UGUA motifs in the pA region indicating that multiple motifs are not necessary for proper functioning of Seb1. In contrast, changes in read-through transcription did negatively correlate with read-through levels detected in WT cells in all three mutants (e.g. Y64K in Figure 5.8; S22D-K25E-K124E: $\rho = -0.82$, $p < 2 \cdot 10^{-16}$; F445A: $\rho = -0.72$, $p < 2 \cdot 10^{-16}$). This strongly suggests that Seb1 is an important termination factor and the observed effects are not caused by already inefficiently terminated genes (e.g. due to weaker pA signals). On the contrary, Seb1 seems to be part of a very effective termination pathway because genes that are dependent on Seb1 show less read-through transcription than genes that do not.

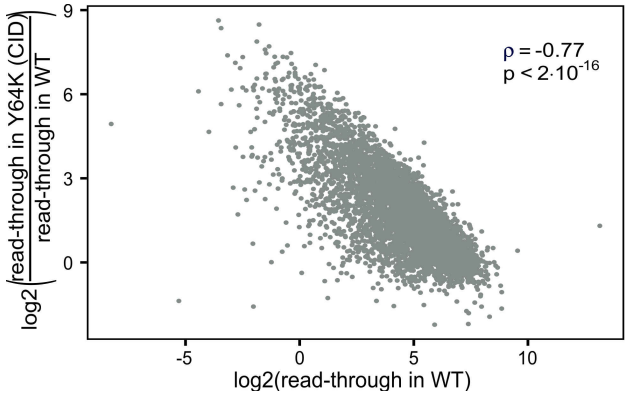


Figure 5.8: Genes with low read-through transcription levels depend more on Seb1
The log2-fold change in read-through levels for every gene in the Y64K mutant calculated as in Figure 5.4 is plotted against basal read-through levels in WT (corrected to gene length). The Spearman correlation coefficient ρ and significance p are indicated in the plot.

5.3 Genes that recruit Seb1 have termination defects in Seb1 mutants

To finally test if the defects in transcription termination seen in the mutants can indeed be caused by a reduction in Seb1 recruitment, it was assessed if those genes do actually recruit Seb1. Therefore, the RNA-Seq data was compared with the Seb1 ChIP-Seq and PAR-CLIP data sets described in Chapter 2. For this, all genes were split into two groups, those that bind Seb1 in the regions shown in Figure 2.5 and those that do not. As expected, genes that are bound by Seb1 around the pA site in ChIP-Seq or PAR-CLIP show significantly more read-through transcription in all mutants except S492A, which has the opposite effect (Figure 5.9A and B). Furthermore, 80.6% of genes that have a significantly higher amount of reads past their pA site as compared to WT in all three lethal mutants detectably recruit Seb1 by either PAR-CLIP or ChIP-Seq (Figure 5.9C). This further supports the notion that Seb1 is indeed important for proper transcription termination in *S. pombe*.

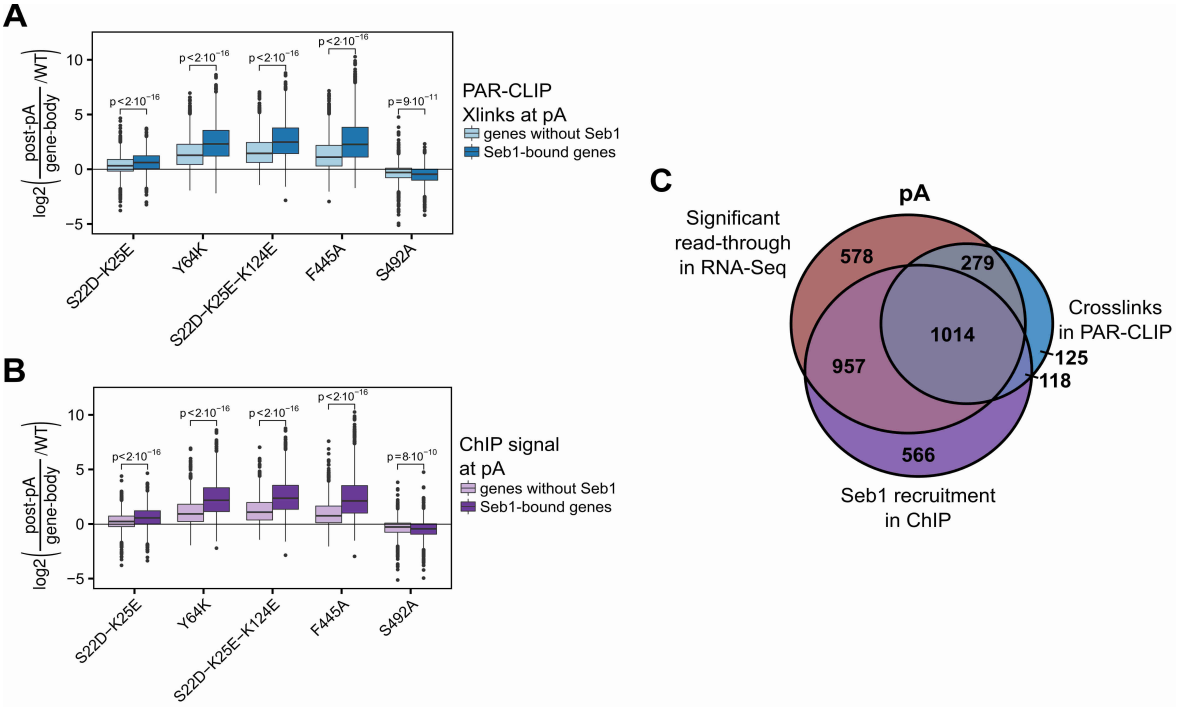


Figure 5.9: Seb1 binding prevents read-through transcription
(A, B) All genes were split into two groups, those that bind Seb1 in the region 250 nt ± pA by PAR-CLIP (A) or ChIP-Seq (B), and genes that do not have any detectable peaks. Changes in read-through transcription calculated as in Figure 5.4 were plotted for all mutants and the significance of any difference between the two groups was determined by the Wilcoxon-Mann-Whitney test and is indicated above the boxes. Genes that are shorter than 500 nt or closer than 250 nt to their downstream gene were excluded from the analysis. Number of genes in each group: genes bound by Seb1 in PAR-CLIP: 1536 and ChIP-Seq: 2655; without Seb1 by PAR-CLIP: 2692 and ChIP-Seq: 1573
(C) Venn diagram depicting the overlap between genes that show significant read-through by RNA-Seq in all three mutants (Y64K, S22D-K25E-K124E, and F445A, brown), or have detectable peaks in the same region as in A and B by PAR-CLIP (blue) or ChIP-Seq (purple).

The TSS region is slightly more ambiguous. Genes that bind Seb1 by PAR-CLIP at the TSS show more read-through than genes that do not (Figure 5.10A). However, genes that recruit Seb1 near the TSS are significantly more likely to bind Seb1 at the pA site as well (Wilcoxon-Mann-Whitney test, $p = 8.4 \cdot 10^{-11}$). This effect could in part be caused by the fact that binding of Seb1 to lowly expressed and unstable transcript cannot be detected by PAR-CLIP and all genes falling into this category would therefore automatically be part of the group of non-bound transcripts. Therefore, a second analysis was performed using only genes that do not have simultaneous crosslinks at the pA site (Figure 5.10B). In this case, the result is as expected. Transcripts that recruit Seb1 to the

TSS site only have less read-through transcription than the control group. While this might indicate that genes which bind Seb1 at the TSS are more likely to recruit Seb1 to the pA site as well, it also shows that recruitment of Seb1 to the TSS region does not affect transcription termination at the 3'end. This is further supported by the fact that no significant differences ($p > 0.05$) in transcriptional read-through can be detected between genes that recruit Seb1 to the TSS region and genes that do not, as determined by ChIP-Seq (Figure 5.10C).

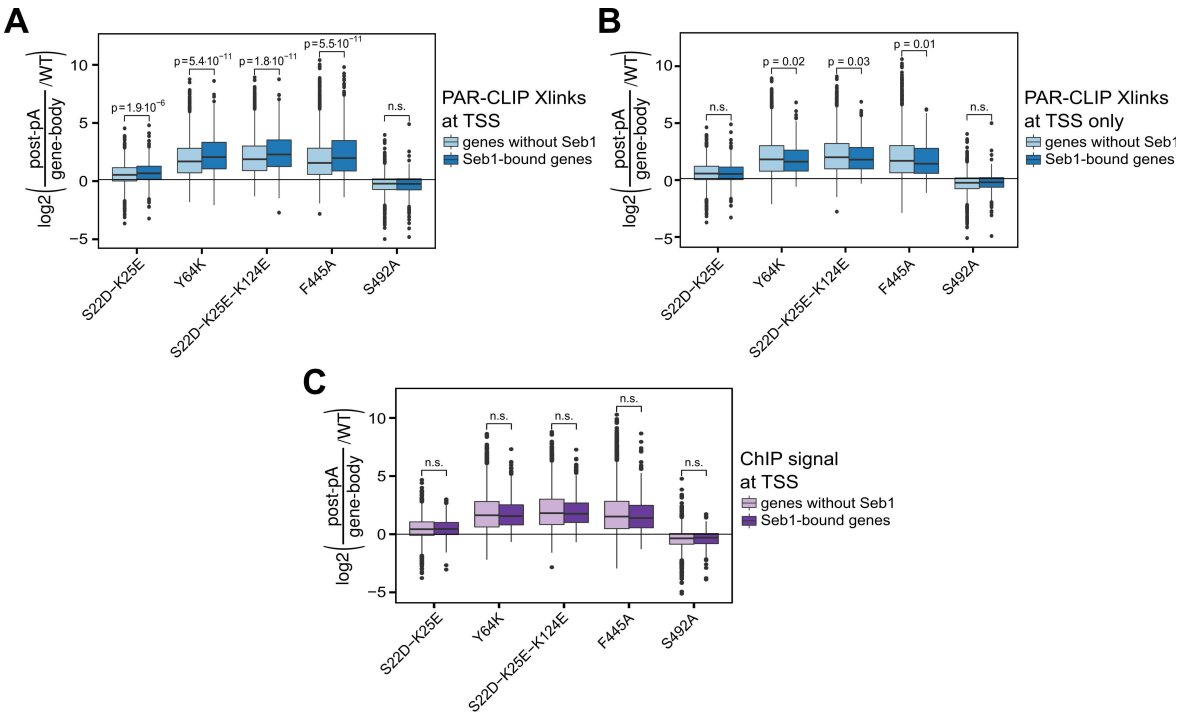


Figure 5.10: Binding of Seb1 at the TSS does not influence transcription termination
(A) All genes were split into two groups, those that have detectable peaks of Seb1 binding by PAR-CLIP in the region of 10 nt upstream of the TSS to 250 nt downstream, and genes that do not recruit Seb1. Changes in read-through transcription calculated in Figure 5.4 were plotted for all mutants and the significance of any differences between the two groups was determined by the Wilcoxon-Mann-Whitney test and is indicated above the boxes. Genes that are shorter than 500 nt or closer than 250 nt to their downstream gene were excluded from the analysis. Number of genes in each group: genes bound by Seb1: 827; not bound by Seb1: 3401
(B) Same as A with the only difference being that genes included in the group of Seb1 binding do not simultaneously have PAR-CLIP crosslinks at their pA sites. Number of genes in each group: genes bound by Seb1: 392; not bound by Seb1: 3836
(C) Same as A but ChIP-Seq rather than PAR-CLIP data was used to determine Seb1 recruitment. Number of genes in each group: genes bound by Seb1: 341; not bound by Seb1: 3887

5.4 Seb1 and Pcf11 are both required for transcription termination

Seb1 is not the only CID-containing protein that was shown to be involved in 3'end processing and transcription termination. Pcf11 is also required for both processes. However, in contrast to Seb1 it cannot bind to RNA directly but requires Rna14/Rna15 for the recognition of RNA motifs. Therefore, in order to determine if Seb1 and Pcf11 act together on the same genes or if some genes depend on one protein and some on the other, the recruitment profile of Pcf11 was determined by ChIP-Seq and compared to Seb1. Furthermore, both proteins can bind to S2P-Pol II and their binding profiles were therefore correlated with S2P levels. Pcf11 ChIP-Seq and alignment of the reads to the genome was performed by Mr Miles Huseyin, S2P-Pol II ChIP-Seq was done by Dr Dong-Hyuk Heo (post-doctoral researcher, Vasilieva lab) and the data was aligned to the *S. pombe* genome by Ms Tea Kecman (*DPhil* student, Vasilieva lab). All further data analysis was performed by myself as detailed below.

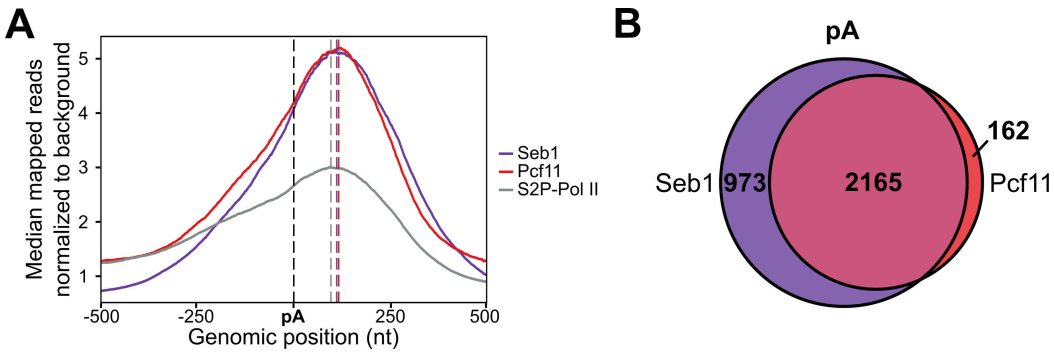


Figure 5.11: Seb1 and Pcf11 are simultaneously recruited to the same genes
(A) The occupancy profiles of Seb1 and Pcf11 recruitment and the S2P levels on Pol II as detected by ChIP-Seq was determined as in Figure 2.4. Background used for normalization were untagged control samples in the case of Seb1 and Pcf11, and a non-immunoprecipitated input in the case of S2P-Pol II. Dotted lines indicate the position of the pA site (black) and maximum values for the profiles of Seb1 (purple), Pcf11 (red), and S2P-Pol II (grey).
(B) The overlap between protein-coding genes that binding Seb1 (purple) and Pcf11 (red) is shown as a Venn diagram and was determined for the region 250 nt ± pA as described in Figure 2.5. 3531 genes were included in the analysis.

Interestingly, the binding profiles of Seb1 and Pcf11 are almost identical (Figure 5.11A) which suggests that they are recruited at the same time. This is also supported by the fact that they both interact with the same type of phosphorylated Pol II and that Seb1

co-purifies with the 3'end processing complex (Lunde et al. 2010; Watts 2015). As expected, S2P-Pol II is most enriched just before Pcf11 and Seb1 binding peaks, indicating that S2P levels do indeed play a role in recruiting these factors. However, whether both proteins actually bind to the same transcripts was assessed by comparing the protein-coding genes that recruit Seb1 or Pcf11 to their pA region (Figure 5.11B). Strikingly, 93% of protein-coding genes that bind Pcf11 also recruit Seb1 by ChIP-Seq. Furthermore, 61% of all protein-coding genes included in the analysis recruit both Seb1 and Pcf11 and even more strikingly 93% recruit either of them (or both). Taken together, this indicates that the majority, if not all, protein-coding genes are dependent on Seb1 and Pcf11 for transcription termination and that this requires simultaneous binding of both proteins to S2P-Pol II. If they actually work together or act independently of one another is not clear and further research is necessary to address this question.

In contrast to Seb1, the *S. cerevisiae* homologue Nrd1 is not terminating protein-coding genes but sn/snoRNAs and CUTs. In addition, Pcf11 was shown to be important on this class of genes as well. Therefore, it was determined whether Seb1 also acts on sn/snoRNAs and if Pcf11 from fission yeast is also recruited to those genes (Figure 5.12). Indeed, both proteins seem to bind sn/snoRNAs, however, their recruitment coincides suggesting that they are recruited simultaneously which indicates that they act at the same time. This is in contrast to budding yeast where Nrd1 and Pcf11 were shown to function consecutively (Grzechnik et al. 2015). Also in contrast to *S. cerevisiae* in which recruitment of Nrd1 is achieved by S5P, *S. pombe* has high levels of the S2P mark on sn/snoRNAs. Furthermore, Seb1 CID and RRM mutations cause transcriptional read-through on those genes (indicated by arrows), further underpinning the importance of Seb1 for transcription termination of this class of transcripts.

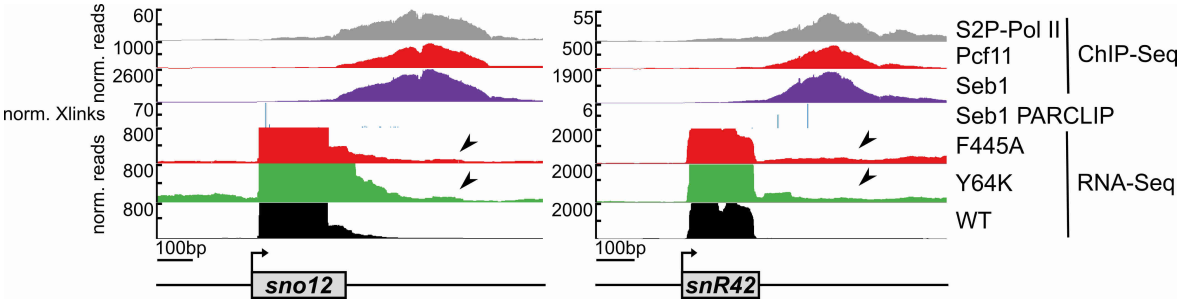


Figure 5.12: Seb1 and Pcf11 are both recruited to sn/snoRNAs
Alignment of reads to two exemplar sn/snoRNA genes as indicated in the schematics below. For this, RNA-Seq data from WT (black), Y64K (green), and F445A (light red) cells, as well as recruitment of Seb1 as determined by PAR-CLIP (blue) and ChIP-Seq (purple) and binding of Pcf11 (dark red) as well as S2P-Pol II levels (grey) determined by ChIP-Seq is depicted. Arrows indicate regions of transcriptional read-through.

5.5 Discussion

In this chapter, I was able to demonstrate that Seb1 is not only recruited to protein-coding and non-coding genes, but that it is required for proper 3'end processing and transcription termination (Figures 5.2, 5.4, 5.5, 5.9, and 5.12). Furthermore, Pcf11 is recruited together with Seb1 to protein-coding transcripts (Figure 5.11) which indicates that Seb1 acts together with the 3'end processing machinery to properly process the nascent RNA and to efficiently terminate transcription. The extent to which these other factors are necessary for the function of Seb1 or whether they merely serve to recruit the protein is unknown. However, considering that Seb1 can interact with Pol II and RNA suggests that other factors might not be additionally required for its recruitment and that binding to 3'end processing factors has a more direct role.

Mutations in the Seb1-CID as well as RRM domain cause a similar phenotype (Figures 5.1 and 5.6) which shows that interactions with both the Pol II-CTD and the nascent RNA are equally important for its function. Furthermore, even though a decrease in Seb1 recruitment upon mutation of either domain was only assessed on two genes, *adh1* and *pho1* (Figure 5.2C and D), it is assumed that binding to other genes is affected in the same way which would indicate that Seb1 might need to interact with RNA and the

CTD simultaneously. Thus, forming a bridge between the nascent transcript and the elongating polymerase might be necessary for Seb1 to terminate transcription. This is similar to Nrd1 which also needs to interact with Pol II as well as RNA (Heo et al. 2013; Tudek et al. 2014), however, considering that only the Seb1-CID is essential but not the Nrd1-CID (Vasiljeva et al. 2008a) clearly demonstrates that the bridging function is more important in fission yeast. Another feature both organisms share is the involvement of Pcf11 for terminating sn/snoRNAs. Surprisingly though in *S. pombe*, Pcf11 binds simultaneously with Seb1 and is probably recruited by high Pol II-S2P levels (Figure 5.12), which is both in contrast to *S. cerevisiae.*

Taken together, this indicates that Seb1 is required for the termination of protein-coding genes as well as non-coding RNAs such as sn/snoRNAs and acts together with Pcf11 in all cases. This strongly suggests that *S. pombe* did not evolve two separate termination pathways for different Pol II transcript classes but utilises the same termination mechanism in all cases. This is in contrast to *S. cerevisiae* in which distinct mechanism are required for different types of genes. This puts *S. pombe* closer to human and demonstrates the importance of Seb1 for transcription. Even though Nrd1 also plays a role on some protein-coding genes, it seems to act as a fail-safe termination factor rather than being absolutely required for termination. In contrast, the human Seb1 homologues Scaf4 and Scaf8 are almost entirely uncharacterised and whether they play any role in terminating protein-coding or sn/snoRNAs is unknown. To my knowledge, there is only one report looking at a possible involvement of Scaf8 in transcription termination (O'Reilly et al. 2014). In this study, knockdown of the protein had no effect on Pol II levels on neither the protein-coding gene β-actin nor the U1 or U2 snRNA genes. However, Scaf4 was not simultaneously targeted nor were changes in the levels of this protein checked. It therefore remains a possibility that either Scaf4 is important for transcription termination and Scaf8 is not, or that Scaf4 and Scaf8 have overlapping functions and Scaf4 was able to rescue any phenotype caused by the Scaf8 knockdown. Therefore, further research is necessary to understand the function of both proteins in human.

CHAPTER 6: GENERAL DISCUSSION AND CONCLUSION

In this thesis, I was able to demonstrate that Seb1 is recruited to the 3'end of protein-coding and non-coding genes co-transcriptionally (Chapter 2). This is achieved by direct interactions between the Seb1-CID and S2P-Pol II (Chapter 3) on the one hand and the Seb1-RRM and RNA on the other hand (Chapter 2 and 4). Seb1 was found to process and/or terminate the transcripts it is recruited to, which requires binding of both domains (Chapter 5). Seb1 therefore constitutes a novel 3'end processing/termination factor in *S. pombe*. Furthermore, Seb1 was found to co-purify with the 3'end processing machinery (Chapter 2) and direct binding between them is therefore likely. Whether this is also necessary for the recruitment and function of Seb1 is unclear.

Some of the data presented here is directly supported by a recent publication from the Bachand lab (Lemay et al. 2016). They, too, could demonstrate that Seb1 interacts with proteins of the 3'end processing complex and does not co-purify with components homologous the the *S. cerevisiae* NNS complex or the exosome. Furthermore, they observe recruitment of Seb1 to the 3'end of actively transcribed genes and genome-wide RNA-binding studies could identify the same recognition motif as reported in this thesis (Figure 2.4). Upon depletion of Seb1, they were able to detect Pol II further downstream than in the presence of Seb1 genome-wide, indicative of transcription termination defects. This is consistent with single-gene ChIP experiments performed in this thesis using Seb1 mutants (Figure 5.2). While they did not study the interactions with Pol II and RNA in any detail, nor did they do biochemical or structural analyses, their focus is more on 3'end processing. Lemay et al. could show that Seb1 depletion leads to decreased recruitment of some 3'end processing factors (e.g. Rna14 and Cft2) but not others (e.g. Clp1 and Ysh1) and that more distal pA sites are being used. However, cleavage does not seem to be affected *per se* as mRNAs levels do not change overall which is consistent with the *adh1* data presented here (Figure 5.3).

6.1 Does Seb1 function in 3'end processing or termination?

Several experiments clearly show that Seb1 has an effect at the 3'end of genes. However, the question of whether it is involved in 3'end processing or transcription termination is difficult to answer, as transcript cleavage always affects termination as well (Kuehner et al. 2011; Mischo and Proudfoot 2013; Porrua and Libri 2015). Furthermore, the torpedo model requires pA cleavage as a prerequisite for exonucleolytic degradation of the 3' product by Dhp1/Rat1/Xrn2 which, in turn, leads to termination of Pol II. If this model is correct, uncoupling the two steps to study them separately is very difficult and can only be properly done *in vitro*. However, constituting the cleavage reaction using *S. pombe* proteins has never been successfully performed to date and recreating the disassembly of the polymerase complex is even more challenging.

One possibility could be that Seb1 is directly involved in pA site choice by recruiting the 3'end processing machinery to the cleavage site. However, this is difficult to imagine given that Seb1 is not an integral part of this machinery in *S. pombe* (Roguev et al. 2004). Furthermore, even though some protein-coding genes may use Nrd1 for termination in *S. cerevisiae* (Webb et al. 2014), the CF-CPF generally does not need Nrd1 for 3'end processing (Lykke-Andersen and Jensen 2007; Mandel et al. 2008) and the complex is quite conserved between the two yeasts (Table 2). In addition, the cleavage reaction *per se* appears to be unaffected by Seb1 depletion, instead different, more distal cleavage sites are being used. This suggests that Seb1 acts upstream of 3'end processing. Lemay et al. (2016) reason that a decrease in CF-CPF proteins at the proximal cleavage site in Seb1-depleted cells indicates that the protein is directly involved in positioning of the CF-CPF. However, several lines of evidence indicate that Seb1 is not directly recruiting this complex. Firstly, a study in humans has shown that changes in pA site usage lead to a decrease in 3'end processing factors at the original pA site (Fusby et al. 2016), and the observed reduction is therefore expected independently of what causes the observed change in pA site usage. Secondly, Seb1 is not only required at pA-

dependent genes but absence of the protein also leads to read-through transcription at sn/snoRNAs which are assumed to be pA-independent in *S. pombe* (Figure 5.12, Lemay et al. 2016). This further supports the notion that it is not 3'end processing by the CF-CPF that requires Seb1, but a step that is common to both types of genes.

Lemay et al. (2016) have performed an additional experiment in which they utilised 6-azauracil which is known to decrease transcription elongation rates (Mason and Struhl 2005). Strikingly, those conditions led to an increase in the amount of proximally cleaved transcripts after Seb1 depletion, partially reverting the phenotype back to WT (Lemay et al. 2016). This suggests the intriguing possibility that the main function of Seb1 is to slow down transcription, thereby affecting 3'end processing indirectly. It was proposed previously that Pol II kinetics influence the choice of pA site by changing the "window of opportunity" for pA factors to process transcripts (Bentley 2014). In this model, slower transcribing Pol II gives the CF-CPF complex more time to assemble at proximal sites before more distal sites emerge and start to compete for the binding of factors. In fact, cleavage of several genes in human and flies were shown to respond in the predicted way when Pol II transcription rates are changed (Enriquez-Harris et al. 1991; Peterson et al. 2002; Pinto et al. 2011). Deceleration of transcription alone could therefore explain why more distal pA sites are being used upon depletion of Seb1. Furthermore, transcription termination is also influenced by Pol II elongation rates (Glover-Cutter et al. 2008; Hazelbaker et al. 2013; Fong et al. 2015) which would explain the large amount of unstable read-through transcripts which can be observed far downstream of the pA site in Seb1 mutants (Figure 5.7). In this scenario, Seb1 would facilitate both 3'end processing as well as transcription termination, and I therefore would like to propose a model based on this idea (Figure 6.1). However, to determine if Seb1 really influences 3'end processing, transcription termination or both and whether this is indeed caused by changes in transcription rates, more experiments need to be conducted.

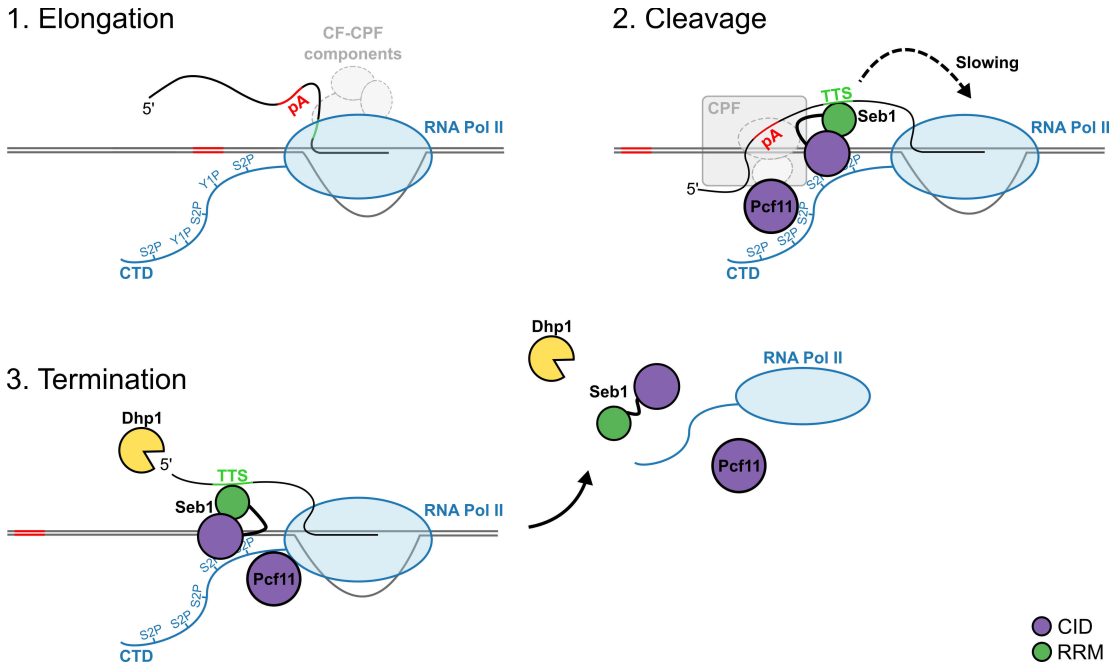


Figure 6.1: Hypothetical model showing the function of Seb1 at the 3'end of genes
During elongation, some of the CF-CPF components interact with the core polymerase. Interactions of CID proteins with the CTD are inhibited by the presence of Y1P. Transcription of the pA site causes Y1P to be removed, which allows binding of Seb1 and Pcf11. Seb1 is additionally recruited by binding to its recognition site on RNA which therefore acts as a transcription termination signal (TTS). Interactions with both the CTD and RNA are necessary for Seb1 to assert its full function which leads to a decrease in transcription rates. This allows the full assembly of the CF-CPF complex which result in cleavage at the pA site and degradation of the 3' cleavage product by the exonuclease Dhp1. The combined action of Seb1, Dhp1, and Pcf11 eventually leads to the dismantling of Pol II and thus transcription termination.

In human cells, it was shown that CPSF binding to Pol II during elongation occurs via interactions with the polymerase core rather than the CTD. After the pA site has been transcribed, the full 3'end processing complex assembles and translocates onto the CTD. However, this only occurs upon Pol II pausing (Nag et al. 2007). Although it was suggested that the pA site is responsible for pausing, another study has shown that the presence of pA signals is not sufficient (Fusby et al. 2016). In yeast, some 3'end processing factors are also known to bind after initiation and to travel with Pol II during elongation (Licatalosi et al. 2002; Kim et al. 2004a; Calvo and Manley 2005). Therefore, a model in which Seb1 is responsible for decreased Pol II elongation rates which subsequently leads to the recruitment of additional factors would be consistent with the observation that Seb1 depletion affects the levels of only some CF-CPF proteins.

In addition, the recognition of the pA site probably leads to the removal of Y1P marks on the CTD, which was suggested to be a prerequisite for binding of CID proteins in yeast such as Pcf11 and Seb1 (Mayer et al. 2012). Furthermore, at least in humans, pausing at the pA site is required to increase S2P levels, which should additionally strengthen Seb1 binding (Davidson et al. 2014; Fusby et al. 2016). However, Seb1 cannot function efficiently when it interacts with only Pol II or RNA (Figure 5.4) and mutating either domain negatively affects Seb1 recruitment during transcription (Figure 5.2). This suggests that the protein needs to be engaged with both domains. The mutation studies presented in this thesis, however, do not clarify which interaction comes first and it cannot be ruled out that binding to only the RNA is insufficient to detect co-transcriptional recruitment of Seb1 by ChIP. Nonetheless, both the RNA recognition site and the S2P-CTD only become available for Seb1 after the pA site has been transcribed. Therefore, both Seb1 and the CF-CPF may need to be bound to their respective RNA motifs to trigger full Pol II pausing. Via direct interactions with the Rna14/15 complex which recognises upstream pA signals, Seb1 could “know” of the presence of a pA site. This mechanism could prevent Seb1-induced pausing at UGUA motifs within the gene body. Simultaneous interaction of Seb1 with the CTD and the RNA might then transfer forces onto the polymerase that could lead to decreased transcription rates which are a prerequisite for both 3’end processing and termination.

This model of transcription termination is consistent with both the allosteric and the torpedo model (described in section 1.3.5). The fact that Seb1 depletion is lethal (Figure 2.2) and Seb1 mutations lead to massive transcription termination defects genome-wide (Figure 5.4) demonstrates that Seb1 is an absolute requirement for efficient termination in *S. pombe*. A combined allosteric-torpedo model has been proposed before (Luo et al. 2006; Zhang et al. 2015) and the data presented here also support this notion. Together, Pcf11 and Seb1 might be necessary to prime Pol II for termination by decreasing transcription rates and potentially inducing conformational changes which help Dhp1 to dislodge the complex. This is also supported by the fact that both proteins are recruited to

pA-independent transcription units in *S. pombe* (Figure 5.12). Therefore, the consorted action of all three proteins might be necessary to efficiently remove the incredibly stable transcription complex at the 3'end of genes.

However, arguing against this model is the fact that, upon Seb1 depletion, polymerase levels are increased in ChIP-Seq further downstream of the pA site than in the presence of Seb1 (Lemay et al. 2016). This indicates that pausing still occurs albeit at a more distal site than in WT cells. It is therefore possible that 3'end processing itself leads to transcriptional pausing, and in the absence of Seb1, this would occur further downstream than normally. In this scenario, Seb1 would only influence 3'end processing while termination would solely be dependent on the position of transcript cleavage. However, it can not be excluded that additional factors also play a role as transcription rates are known to be influenced by the presence of nucleosomes, the histone acetylation state or RNA secondary structure (Churchman and Weissman 2011; Bintu et al. 2012; Zamft et al. 2012). Therefore, pausing could be induced as Pol II leaves the active chromatin state and encounters more densely packed nucleosomes at the end of transcription units which could then promote transcription termination. Indeed, this mechanism was shown to play a role at human snRNA genes (O'Reilly et al. 2014).

6.2 What role do CID-RRM proteins play on different types of genes?

In this thesis, I was able to demonstrate that the CID-RRM protein Seb1 is required for the termination of protein-coding genes which depend on pA signals, as well as ncRNAs that do not have pA signals and therefore utilise a different 3'end processing pathway. This is in contrast to *S. cerevisiae* in which Nrd1 is not generally required at protein-coding genes but primarily acts at pA-independent genes as part of the NNS complex (Figure 6.2). In this case, Sen1, another protein of the NNS, unwinds the RNA:DNA hybrid until it eventually catches up with Pol II and dislodges the complex from the template (Porrua and Libri 2013, 2015). This helicase-dependent mechanisms of

termination is very different from the cleavage-dependent pathway that acts on protein-coding genes. In both cases, Pcf11 is required as well, demonstrating that CID-containing proteins are essential for transcription termination of all types of Pol II-transcribed genes (see Chapter 1).

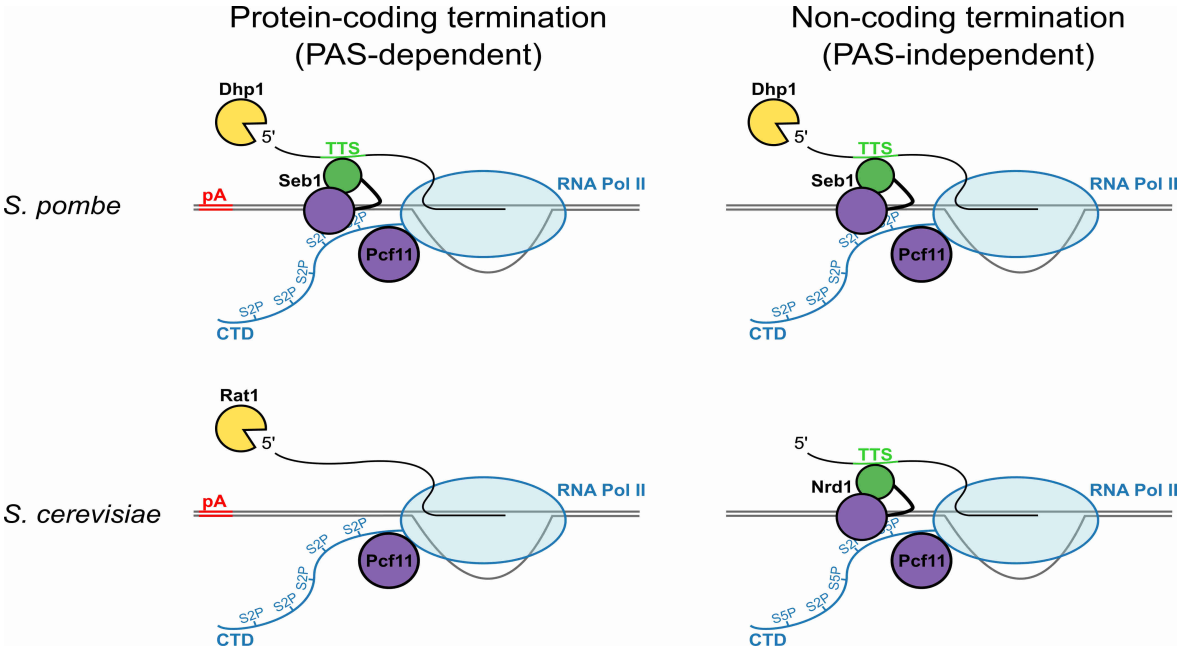


Figure 6.2: Transcription termination pathways differ in budding and fission yeast
Schematic depiction of CID proteins that participate in the termination of Pol II. Seb1 is required for both pA-dependent protein-coding genes and pA-independent ncRNAs in *S. pombe* (top). This is in contrast to *S. cerevisiae* that only requires Nrd1 at pA-independent genes and has no equivalent CID-RRM protein acting at protein-coding RNAs (bottom). The CID-protein Pcf11 seems to be required in all cases.
TTS: transcription termination signal

In contrast to budding yeast, termination of Pol II by a helicase does not seem constitute a major termination pathway in fission yeast and both pA-dependent and -independent genes seem to require transcript cleavage, followed by the action of the exonuclease Dhp1 (see Chapter 1 and Zhou et al. 2015; Lemay et al. 2016). Therefore, even though the different types of genes depend on different endonucleases for 3'end cleavage, downstream events that lead to Pol II termination could be the same. This might explain why Seb1 is required in both cases. It is surprising that, in contrast to fission yeast, the cleavage-dependent termination pathway in budding yeast does not rely on the

action of any CID-RRM protein and it is the helicase-dependent pathway that needs Nrd1. This aspect of cleavage-dependent termination is therefore not conserved between the two yeasts.

In human cells, too, pre-snRNAs have to be cleaved before transcription can terminate and, as is the case in fission yeast, a helicase-dependent mechanism is probably not employed. However, neither protein-coding nor non-coding genes are known to use CID-RRM proteins for this purpose. On the other hand, the two candidates, Scaf4 and Scaf8, have not been studied in enough detail to rule out an involvement in transcription termination. Future research is therefore necessary to understand what role these proteins play and it will be interesting to see whether Scaf4 and Scaf8 act in the same way as Seb1 does in *S. pombe*, or if they evolved to perform a different task in higher eukaryotes.

6.3 How are ncRNAs targeted to the exosome in *S. pombe*?

Nrd1 in *S. cerevisiae* directly recruits proteins that lead to the degradation of the bound RNA (Tudek et al. 2014). In *S. pombe*, the situation is clearly different. Not only does Seb1 not co-purify with the exosome, it also cannot functionally complement Nrd1 *in vivo* (Watts 2015; Lemay et al. 2016). However, recruitment of the exosome is necessary for the maturation of sn/snoRNAs and the degradation of aberrant transcripts. So how is this achieved in fission yeast?

In humans, promoter-proximal transcripts and snRNA precursors are bound by the nuclear exosome targeting complex (NEXT) which can recruit the exosome for degradation or processing. The NEXT complex can directly engage with U-rich RNA, however, it has little sequence specificity and interactions with the cap-binding complex are additionally required. The bound RNA is subsequently degraded if a 3'end is available in close proximity (Lubas et al. 2011, 2015; Hrossova et al. 2015). *S. pombe* was shown to target its sn/snoRNAs and CUTs in a very similar way. In this case, the exosome is

recruited by the MTREC (Mtl1-Red1 core) complex also via direct binding to RNA and the 5'cap (Lee et al. 2013; Zhou et al. 2015). Transcription termination of ncRNAs and their processing/degradation are therefore uncoupled in *S. pombe* and humans, unlike in *S. cerevisiae*. This thus constitutes another example in which the resemblance of fission yeast is higher to human cells than to budding yeast. However, Seb1, in contrast to Nrd1, does not specifically act on transcripts that need the exosome further downstream, instead, it seems to be rather broadly required on all types of Pol II-transcribed genes. Therefore, recruitment of the exosome is, in most cases, not desired. Utilising only one termination pathway as observed in *S. pombe* and humans might give cells a greater flexibility to regulate their transcriptomes than using different pathways for different types of genes, as is the case in budding yeast.

6.4 Is Seb1 required for the termination of cryptic transcription?

In fission yeast, three different types of CUTs have been described: PROMPT-like transcripts that initiate close to a TSS but in antisense direction, transcripts antisense to gene bodies, and transcripts in sense direction that start at the TTS. All of those were found to be targets of MTREC and are hyperpolyadenylated probably by the canonical pA machinery (Zhou et al. 2015). In addition, pA signals are enriched upstream of TSSs, further indicating that the 3'end processing complex might be involved (Schlackow et al. 2013). The question remains whether Seb1 is required for their termination as well.

Considering that the data presented in this thesis suggest that Pol II might require Seb1 for efficient termination at any type of gene, it seems likely that Seb1 is involved in terminating CUTs as well. This is supported by PAR-CLIP crosslinks before the TSS and after the pA site in antisense direction (Figure 2.7) which are accompanied by a simultaneous enrichment of the UGUA motif. In addition, Seb1 binding can be observed just after the TSS in sense direction (Figure 2.4). Transcription termination shortly after initiation is most likely induced as a result of the early elongation checkpoint, similar to

what has been observed in and mammals (described in section 1.1.2). While the NNS pathway is responsible for termination in *S. cerevisiae*, the canonical 3'end processing machinery is thought to be involved in mammals (Zhang et al. 2007; Natarajan et al. 2013; Tuck and Tollervey 2013). Therefore, it seems likely that Seb1 is recruited to facilitate the dislodgement of Pol II that failed to correctly transitioning into the elongation state.

The PAR-CLIP crosslinks near the TSS are not accompanied by a simultaneous enrichment of Seb1 by ChIP-Seq (Figure 2.4), which indicates that Seb1 might not interact with Pol II in those regions and is bound to only the RNA. More transient binding that escapes detection by ChIP cannot be excluded and would be consistent with the fact that Seb1 has a notably lower affinity to S5P-CTD than S2P-CTD (Figure 3.2), which is supposedly the predominant mark on the CTD near the TSS (Figure 1.1). Whether binding to Pol II antisense of the pA site occurs cannot be determined due to the nature of ChIP, which lacks strand-specificity. However, localisation of Pcf11 to the TSS could not be detected either (data not shown) which contrasts *S. cerevisiae* where Pcf11 is required for the termination of CUTs and can directly be recruited by Nrd1 (Grzechnik et al. 2015).

Another possibility is that binding of Seb1 to just the RNA without simultaneous interactions with Pol II could itself transfer information and lead to hyperpolyadenylation and degradation of the transcript, rather than normal 3'end processing. However, whether that is actually the case is purely speculative at this point and further experiments are necessary to test whether Seb1 plays any role at CUTs at all.

CHAPTER 7: MATERIALS AND METHODS

7.1 List of oligonucleotides used in this study

Table 6: DNA oligonucleotides used in this study

Number	Name	Sequence	Purpose
2469	adh1-1 fwd	CGGAAGCTGGTGAGAAGAAC	qPCR
2470	adh1-1 rev	CGTTGGAATGCGGAGTAGAG	qPCR
2471	adh1-2 fwd	CAACCTCCCATTTCCTCCTT	qPCR
2472	adh1-2 rev	GTGGACACATTTCGGGAATC	qPCR
2473	adh1-3 fwd	TCTCTCGCTTTCCTCATTCG	qPCR
2474	adh1-3 rev	GCCAACTGCTTGTCAGGAAT	qPCR
2475	adh1-4 fwd	GGTCCCGAGAACGTCAAGT	qPCR, Northern probe
2476	adh1-4 rev	ACTTGACACCAACACGGTCA	qPCR, Northern probe
2478	adh1 gene body	AATGGCAACAACACGCATAG	RT
2536	pho1 gene body	CAAACATACCATATCCATACCAAAGAG	RT
2547	pho1-5 fwd	AGCAGAGGATAGTTTATGTAGGAGATAATG	qPCR
2548	pho1-5 rev	TTTATATGGTGAGAGTATTGTCAAAGAAAC	qPCR
2857	pho1-6 rev	AAACTAAGTCTTGACAACTATAACGAAACC	qPCR
2857	pho1 read-through	AAACTAAGTCTTGACAACTATAACGAAACC	RT
3032	fbp1 rev	CAAGTGACGGCATAGGAACC	qPCR
3048	pho1-1 fwd	ACAATTATATCTTGGTCTGGGGAAC	qPCR
3049	pho1-1 rev	ATCATTAAATTGTGAATATCGCAAGAC	qPCR
3050	pho1-2 fwd	ATGTTTGAGATTTACGGGAAGTG	qPCR
3051	pho1-2 rev	TTTGTCCTAATTTTCCAAACAGC	qPCR
3052	pho1-3 fwd	TTTGTACCAACTTGGACTCCTG	qPCR
3053	pho1-3 rev	GCGTCCCATGTCAAATAACTC	qPCR
3054	pho1-4 fwd	CTTCGCCTTTACTCATGATGC	qPCR
3055	pho1-4 rev	TTGGTAGGAAGTAGGCAATGG	qPCR
3072	fbp1 fwd	ACTGCGATGAAGTCGAACG	qPCR
3089	adh1-5 fwd	GTACGACGATCCCTAATCCAAC	qPCR
3090	adh1-5 rev	ACGCAAATCTTGAAAAAGATCC	qPCR
3129	pho1-6 fwd	AAAATTCTATGTTTCTATACATGCCTCTG	qPCR
3422	Seb1 L3	AAGATTTTGCTATGCGTCGT	cloning

3423	Seb1 L4	TAATTAACCCGGGGATCCGTCGACCTTG GGGTTGCCAAGGAGGTT	cloning
3424	Seb1 L5	AAACGAGCTCGAATTCATCGATGATATGT GTTAAAA	cloning
3425	Seb1 L6	TCGCGATTTGATCTTTTTG	cloning
3620	adh1-6 fwd	CGAAAACGAAGCGCTTTACTC	qPCR
3621	adh1-6 rev	TCACTTTGCCATTCATCTGTCT	qPCR
3622	adh1-7 fwd	ATGCAACGTTGTGCAGTGAT	qPCR
3623	adh1-7 rev	CAGTCCATTTGTGCGTACGT	qPCR
3778	rps2202-1 fwd	CCGTATATGCCCTTCAGGTT	qPCR
3828	Seb1-NdeI-fwd	TAATTACATAATGTCGGGAATCGCTGAATT C	cloning
3829	Seb1-NotI-rev	TTATATGCGGCCGCTTGGGGTTGCCAAG	cloning
3831	Seb1 rev	TTAATTGGGGTTGCCAAGGAG	cloning
3878	K25E fwd	GGATCAGAAATTTTGAAATTGACTAACTT GTCG	site-directed mutagenesis
3879	K25E rev	AATTTCAAAATTTCTGATCCCGAGATTCCT G	site-directed mutagenesis
3881	K22D fwd dmut	CAAGACAGGAATCGACGGATCAGAAATT TT	site-directed mutagenesis
3882	K22D rev	TGATCCGTCGATTCCTGTCTTGGAATGTT	site-directed mutagenesis
3883	Y64K fwd	GGAGCTTTGAAATTCTAGACTCAATCGTT CGTAG	site-directed mutagenesis
3884	Y64K rev	GTCTAGAATTTTCAAAGCTCCTAATTTATG GGTGAC	site-directed mutagenesis
3885	D66M fwd	GTATATTCTAATGTCAATCGTTCGTAGCTT TCAGG	site-directed mutagenesis
3886	D66M rev	CGATTGACATTAGAATATACAAAGCTCCTA ATTTATG	site-directed mutagenesis
3887	K124E fwd	GCCTAAAATATTAGAACTTTGTGATATTTG GGAGAAG	site-directed mutagenesis
3888	K124E rev	CAAATATCACAAAGTTCTAATATTTTAGGC AAATGAGC	site-directed mutagenesis
3895	rps2202-1 rev	TGTATCTACAGGAGCAGTCACA	qPCR
3898	rps2202-3 fwd	CTGAACGGCCGTATCAACAA	qPCR
3899	rps2202-3 rev	ACAATCACACCAACTTGACGA	qPCR
3900	rps2202-4 fwd	TGTCCCATAATGAGGCTCGT	qPCR
3901	rps2202-4 rev	GCCAGCCTTTTCACCGTGA	qPCR
3902	rps2202-5 fwd	ACTTAGTCTCTGGTTTCGAGCA	qPCR
3903	rps2202-5 rev	TCAACGCCTCTCTCACTTCT	qPCR
3932	rps2202-6 fwd	ACTCTGGCACTGTCTGAAGA	qPCR

3933	rps2202-6 rev	TACTCTTCTACGGCGGCATT	qPCR
3934	rps2202-7 fwd	CGCTGATATGACTTGTGTACAGT	qPCR
3935	rps2202-7 rev	ACCGATTCCCATTTTGTGCT	qPCR
4027	rps2202-2 fwd	ACAAGATGTGAGCGGAAGTC	qPCR
4028	rps2202-2 rev	CGAGAAGCGCGTTAGTTTC	qPCR
4088	after CID delete rev 180	ACCAACATAACCACCATTCCC	cloning
4089	after CID delete fwd 278	GGATCGGTCAATGATACCCAGAG	cloning
4090	before RRM delete fwd 392	TGTCTTCATCCCCATGGGACC	cloning
4091	before RRM delete rev 280	CGTGACCCAACCATTCCACC	cloning
4092	RRM delete rev 392	CTCAAATCGGCGAGGAAATCC	cloning
4093	RRM delete short fwd 488	ACAGGAATCAGCGTTATCCCAATC	cloning
4094	RRM delete long fwd 562	TATAGGGGAGGTCCACCCATTC	cloning
4095	C delete rev 563	TGGTTTACGACCCCGGAATCG	cloning
4096	C delete TAP fwd 620	GGTCGACGGATCCCCGG	cloning
4105	K124A fwd	GCCTAAAATATTAGCACTTTGTGATATTTGGGAGAAG	site-directed mutagenesis
4106	K124A rev	CAAATATCACAAAGTGCTAATATTTTAGGCAAATGAGC	site-directed mutagenesis
4107	K121A fwd	TGCTCATTTGCCTGCAATATTAAAGCTTTG	site-directed mutagenesis
4108	K121A rev	CTTTAATATTGCAGGCAAATGAGCAGATGG	site-directed mutagenesis
4109	K121E fwd	TGCTCATTTGCCTGAAATATTAAAGCTTTG	site-directed mutagenesis
4110	K121E rev	CTTTAATATTTCAGGCAAATGAGCAGATGG	site-directed mutagenesis
4118	Seb1-AdcI-fwd	ATAAGGCGCGCCGAACCAAATGCACGAGTA	cloning
4252	Seb1-XhoI-152 rev	TATATACTCGAGTGCCATTGCATCTTTCAGC	cloning
4359	CID delete fwd 154	AGTACGGAACCGGTTAGTGTAGATTC	cloning
4360	after CID delete1 rev 152	TGCCATTGCATCTTTCAGCTTT	cloning

4361	after CID delete1 fwd 224	CCTGCCGTCGCACCATCC	cloning
4362	after CID delete2 rev 223	CTGCGGTGGAGTGCTAACTG	cloning
4363	RRM delete second rev 488	GGAAAAGTCAGAGCACTCTCGAG	cloning
4444	Seb1 L4-FLAG-pFa	CCGGGGATCCGTCGACCCCTACTTGTCATCGTCATCCTTGTAGTCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCTCCACCCCCGCCTCCCCCTTGGGGTTGCCAAGG	cloning
4524	Seb1 F445A fwd	GACACGGGGCCTTGAAAATGTTTC	site-directed mutagenesis
4525	Seb1 F445A rev	CATTTTCAAGGCCCCGTGTCTGTA	site-directed mutagenesis
4532	Seb1 S492A fwd	ACAGGAATCGCCGTTATCCCAATCC	site-directed mutagenesis
4533	Seb1 S492A rev	GGGATAACGGCGATTCCTGTGG	site-directed mutagenesis
4591	Seb1 KpnI R388 fwd	TAATTAGGTACCCGCCGATTTGAGCGTGAC	cloning
4592	Seb1 HindIII K540 rev	TTATATAAGCTTACTTAGAACTTATTCCTAATCCAATTTC	cloning
4676	ProtA_del fwd	TGAGGCGCGCCACTTCTAA	cloning
4677	ProtA_del rev	AGCGTAATCTGGAACGTCATATG	cloning
5074	HA-AscI rev	TATAGGCGCGCCTCAAGCGTAATCTGGAAC	cloning
5240	adh1 read-through	ACTTTGACGCTATAAGACATGCA	RT

7.2 List of plasmids used in this study

Table 7: Plasmids used in this study

Number	**Description**	**Purpose**	Constructed by
3932	pET41a(+)-Seb1(full-length)-His_8	recombinant expression	this study
3934	pET41a(+)-Seb1(S22D)-His_8	recombinant expression	this study
3935	pET41a(+)-Seb1(K25E)-His_8	recombinant expression	this study
3936	pET41a(+)-Seb1(Y64K)-His_8	recombinant expression	this study
3937	pET41a(+)-Seb1(D67M)-His_8	recombinant expression	this study
3938	pET41a(+)-Seb1(K124E)-His_8	recombinant expression	this study
3939	pET41a(+)-Seb1(S22D-K25E)-His_8	recombinant expression	this study
3940	pET41a(+)-Seb1(S22D-K25E-K124E)-His_8	recombinant expression	this study
3943	pET41a(+)-Seb1(K124A)-His_8	recombinant expression	this study
3944	pET41a(+)-Seb1(K121E)-His_8	recombinant expression	this study
3945	pET41a(+)-Seb1(K121A)-His_8	recombinant expression	this study
3967	pOPINS3C-SUMO-His_6-Seb1(388-540)	recombinant expression	this study
3981	pDUAL-p(seb1)-seb1(Δ2-152)-TAP	expression in *S. pombe*	this study
3982	pDUAL-p(seb1)-seb1(Δ153-223)-TAP	expression in *S. pombe*	this study
3983	pDUAL-p(seb1)-seb1(Δ224-278)-TAP	expression in *S. pombe*	this study
3984	pDUAL-p(seb1)-seb1(Δ279-391)-TAP	expression in *S. pombe*	this study
3985	pDUAL-p(seb1)-seb1(Δ392-488)-TAP	expression in *S. pombe*	this study
3986	pDUAL-p(seb1)-seb1(Δ489-562)-TAP	expression in *S. pombe*	this study
3987	pDUAL-p(seb1)-seb1(Δ563-620)-TAP	expression in *S. pombe*	this study
4028	pDUAL-p(seb1)-seb1-HA	expression in *S. pombe*	this study
4029	pDUAL-p(seb1)-seb1(S22D)-TAP	expression in *S. pombe*	this study
4030	pDUAL-p(seb1)-seb1(Y64K)-TAP	expression in *S. pombe*	this study
4031	pDUAL-p(seb1)-seb1(K121E)-TAP	expression in *S. pombe*	this study
4032	pDUAL-p(seb1)-seb1(S22D-K25E)-TAP	expression in *S. pombe*	this study
4033	pDUAL-p(seb1)-seb1(S22D-K25E-K124E)-TAP	expression in *S. pombe*	this study
4041	pOPINS3C-SUMO-His_6-Seb1(388-540)-T407A	recombinant expression	this study

4042	pOPINS3C-SUMO-His_6-Seb1(388-540)-F445A	recombinant expression	this study
4043	pOPINS3C-SUMO-His_6-Seb1(388-540)-K447D	recombinant expression	this study
4044	pOPINS3C-SUMO-His_6-Seb1(388-540)-F449A	recombinant expression	this study
4045	pOPINS3C-SUMO-His_6-Seb1(388-540)-D486K	recombinant expression	this study
4046	pOPINS3C-SUMO-His_6-Seb1(388-540)-S492A	recombinant expression	this study
4047	pOPINS3C-SUMO-His_6-Seb1(388-540)-R504E	recombinant expression	this study
4055	pOPINS3C-SUMO-His_6-Seb1(388-540)-K447A	recombinant expression	Oliver Adams (Vasilieva lab)
4056	pOPINS3C-SUMO-His_6-Seb1(388-540)-D486A	recombinant expression	Oliver Adams (Vasilieva lab)
4062	pDUAL-p(seb1)-seb1(F445A)-TAP	expression in *S. pombe*	Oliver Adams (Vasilieva lab)
4063	pDUAL-p(seb1)-seb1(S492A)-TAP	expression in *S. pombe*	Oliver Adams (Vasilieva lab)
4065	pOPINS3C-SUMO-His_6-Seb1(388-540)-K402A	recombinant expression	Miles Huseyin (Vasilieva lab)
4066	pOPINS3C-SUMO-His_6-Seb1(388-540)-Y404A	recombinant expression	Miles Huseyin (Vasilieva lab)
4068	pOPINS3C-SUMO-His_6-Seb1(388-540)-R472E	recombinant expression	Miles Huseyin (Vasilieva lab)
4071	pOPINS3C-SUMO-His_6-Seb1(388-540)-F479A	recombinant expression	Miles Huseyin (Vasilieva lab)
4072	pOPINS3C-SUMO-His_6-Seb1(388-540)-F487A	recombinant expression	Miles Huseyin (Vasilieva lab)
4083	pET41a(+)-Seb1(1-152)-His_8	recombinant expression	this study

7.3 List of *S. pombe* strains used in this study

Table 8: *S. pombe* strains used in this study

Name	**Genotype**	**Reference/ Constructed by**
PAR-CLIP	h-, seb1::seb1-TAP::KanMX	Damien Hermand
YP21	h+, ura4-294, leu1-32, ade6-M210	Bioneer Inc.
YP51	h+, ura4-Δ18, leu1-32, ade6-M216, imr1R(NcoI)::ura4+	(Bühler et al. 2008)
YP144	h+, ura4-Δ18, leu1-32, ade6-M216, his3-Δ1	(Lemieux et al. 2011)

YP122	h+, ura4-Δ18, leu1-32, ade6-M216, imr1R(NocI)::ura4+, ysh1::ysh1-HA-TAP::KanMX	Sneha Shah (Vasilieva lab)
YP294	h+, ura4-Δ18, leu1-32, ade6-M210, his3-Δ1, pcf11::pcf11-TAP::KanMX	(Lemieux and Bachand 2009)
YP475	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(F445A)-HA	Oliver Adams (Vasilieva lab)
YP476	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(S492A)-HA	Oliver Adams (Vasilieva lab)
YP291	h+, ura4-Δ18, leu1-32, ade6-M216, imr1R(NcoI)::ura4+, seb1::seb1-HA-TAP::KanMX	Beth Watts (Vasilieva lab)
YP510	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX	this study
YP512	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1-HA-TAP	this study
YP531	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(Δ2-152)-HA-TAP	this study
YP532	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(Δ153-223)-HA-TAP	this study
YP533	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(Δ224-278)-HA-TAP	this study
YP534	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(Δ279-391)-HA-TAP	this study
YP535	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(Δ392-488)-HA-TAP	this study
YP536	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(Δ489-562)-HA-TAP	this study
YP537	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(Δ563-620)-HA-TAP	this study
YP539	h+, ura4-294, leu1-32, ade6-M210, seb1::seb1-FLAG::NatMX	this study
YP587	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1-HA	Cornelia Kilchert (Vasilieva lab)
YP588	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(S22D)-HA	Cornelia Kilchert (Vasilieva lab)
YP589	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(Y64K)-HA	Cornelia Kilchert (Vasilieva lab)
YP590	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(K121E)-HA	Cornelia Kilchert (Vasilieva lab)
YP591	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(S22D-K25E)-HA	Cornelia Kilchert (Vasilieva lab)
YP592	h+, ura4-294, leu1-32, ade6-M210, seb1::p(nmt1)-seb1-FLAG::NatMX, leu1::p(seb1)-seb1(S22D-K25E-K124E)-HA	Cornelia Kilchert (Vasilieva lab)

7.4 Antibodies used in this study

Table 9: Antibodies used in this study

Antibody	Origin	Application and dilution	Manufacturer / Reference
8WG16	mouse	ChIP (5 µl per sample) and Western blot (1:2,000)	Millipore
α-S2P (ab5131)	rabbit polyclonal	Western blot (1:1,000)	abcam
α-S2P (3E10)	rat	ChIP (5 µl per sample)	Millipore
α-S5P (3E8)	rat	Western blot (1:2,000)	Millipore
α-S7P (4E12)	rat	Western blot (1:2,000)	Millipore
Peroxidase-α-Peroxidase (PAP)	rabbit	Western blot (1:2,000)	Sigma-Aldrich
α-HA (12CA5)	mouse	Western blot (1:1,000)	Steve Buratowski
α-His_6-HRP	mouse	Western blot (1:10,000)	Clontech
α-FLAG M2-HRP	mouse	Western blot (1:1,000)	Sigma-Aldrich
α-mouse	goat	Western blot (1:10,000)	Sigma-Aldrich
α-rabbit	goat	Western blot (1:10,000)	Sigma-Aldrich
α-rat	goat	Western blot (1:10,000)	Calbiochem

7.5 *In vitro* protein biochemistry methods

7.5.1 Cloning of expression plasmids

RNA extracted from *S. pombe* cells was used to reverse transcribe the Seb1 mRNA using a gene-specific primer (see Table 6 for all oligonucleotides used in this section). Via polymerase chain reaction (PCR), restriction sites were added to the resulting DNA which was subsequently cloned into pET41a(+). The resulting plasmid allows expression of full-length Seb1 with a C-terminal His_8-tag in *E. coli* (plasmid 3932 in Table 7).

CID point mutations were introduced into this plasmid using site-directed mutagenesis with primers that contain the necessary nucleotide changes. Double and triple mutants were constructed by several, subsequent site-directed mutagenesis steps (resulting in plasmids 3934-3949 in Table 7).

To be able to express just the CID_{1-152} domain, plasmid 3932 (Table 7) was used to amplify everything except the region from the nucleotide corresponding to aa 153 to the

end of the gene. The parental plasmid was DpnI digested and the PCR product religated. The resulting plasmid (4083 in Table 7) allows expression of CID_{1-152} with a C-terminal His_8-tag in *E. coli*.

The RRM domain was amplified from a plasmid 3932 (Table 7) with primers that contained the restriction site AscI on both ends. The pOPINS3C expression vector (Bird 2011) was modified such that a AscI restriction site is in frame with the encoded tag. The PCR product was cloned into the modified pOPINS3C vector resulting in plasmid 3967 (Table 7) which allows expression of $RRM_{388-540}$ with an N-terminal His_6-SUMO tag in *E. coli*. Point mutations were introduced into this plasmid by Mr Oliver Adams (Part II student, Vasilieva lab) and Mr Miles Huseyin (rotation student, Vasilieva lab) using site-directed mutagenesis resulting in vectors 4055-4072 (Table 7).

Successful construction of all plasmids used in this study was verified by sequencing.

7.5.2 Transformation of *E. coli* cells and plasmid extraction

All transformations were performed using chemically competent cells of *E. coli* strains TOP10 (Invitrogen, for cloning) or BL21(DE3) containing either the pLysS or Rosetta plasmid (for protein expression). Cells were removed from -80°C and thawed on ice before addition of 10 µl ligation reaction (for cloning) or ~ 50 ng of plasmid (for protein expression). The cells were subsequently incubated on ice for 5 min, followed by a heat shock step at 42°C for 30 sec and another incubation on ice for 5 min. 250 µl of SOC medium (20 g/l tryptone, 5 g/l yeast extract, 0.59 g/l NaCl, 0.186 g/l KCl, 2.03 g/l $MgCl_2$ x $6H_2O$, 2.46 g/l $MgSO_4$ x 7 H_2O) were added and the cells were allowed to recover for 1 h at 37°C and 200 rpm before being plated onto Lysogeny broth (LB) plates (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl, 20 g/l agar) containing the appropriate antibiotics (50 µg/ml Ampicillin, 25 µg/ml Kanamycin and/or 34 µg/ml Chloramphenicol) and incubated at 37°C overnight.

For cloning, single colonies were picked and used to inoculate 5 ml liquid LB containing the appropriate antibiotics and incubated at 37°C and 200 rpm overnight. The whole culture was then subjected to plasmid extraction using the QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's instructions.

7.5.3 Recombinant expression and purification of proteins

7.5.3.1 Full-length Seb1

WT full-length Seb1 as well as CID point mutants (plasmids 3932-3945 in Table 7) were expressed in Rosetta *E. coli* strain. For this, cells were grown at 37°C and 200 rpm in Terrific Broth (12 g/l tryptone, 24 g/l yeast extract, 3.5% (v/v) glycerol, 2.3 g/l KH_2PO_4, 12.3 g/l K_2HPO_4, 1% (w/v) glucose) supplemented with 50 µg/ml Kanamycin and 25 µg/ml Chloramphenicol until OD_{600} ~ 0.8. Protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside and the cells were further incubated at 18°C for 18-24 h. Harvesting was done by centrifugation at 4°C and 6,200 *g* for 15 min. Pellets were resuspended in lysis buffer (50 mM Tris(hydroxymethyl)aminomethane (Tris)-HCl, 5 mM imidazole, pH 8.0) and subsequently lysed using a French Press. NaCl to a final concentration of 600 mM was added, in addition to 0.1% Tween-20 and 1 mM phenylmethylsulfonyl fluoride (PMSF). The lysate was clarified by centrifugation at 4°C and 100,000 *g* for 1 h, and the supernatant was passed through a 45 µm filter. The resulting protein solution was subjected to affinity purification using NiNTA (nickel-nitrilotriacetic acid) agarose resin (Qiagen). The beads were washed with 50 mM Tris-HCl pH 8.0, 600 mM NaCl, 20 mM imidazole and 0.1% Tween-20 and eluted using the same buffer but with increasing amounts of imidazole (40 mM, 100 mM, and 500 mM) and decreasing amounts of NaCl (600 mM, 400 mM, and 300 mM). All protein concentrations were determined using a NanoDrop Spectrophotometer ND-1000 (PeqLab). A final concentration of 33% (v/v) glycerol was added to the elution fractions before freezing in liquid nitrogen and storage at -80°C. The two highest concentrated fractions (usually 2-4 column volumes of 100 mM imidazole) were used for peptide binding assays.

7.5.3.2 Seb1-CID$_{1-152}$

The Seb1-CID$_{1-152}$ was expressed, harvested and lysed as described in section 7.5.3.1. The lysate was supplemented with 1 M NaCl, 0.1% Tween-20 and 1 mM PMSF before filtering. The solution was then added onto a column containing equilibrated NiNTA beads. Beads were washed with wash buffer (50 mM Tris-HCl pH 8.0, 1 M NaCl, 20 mM imidazole) and bound protein was eluted with the the same buffer containing increasing concentrations of imidazole (40 mM, 100 mM, and 500 mM). Protein-containing fractions were pooled together and further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) using an ÄKTA purifier system (GE Healthcare) with buffer containing 25 mM Tris-HCl pH 7.5, 300 mM NaCl, and 5 mM β-mercapto ethanol. The resulting protein-containing fractions were combined and concentrated to ~ 10 mg/ml and subjected to crystallisation trials.

In preparations used for fluorescence anisotropy assays, Tween-20 was substituted with IGEPAL CA-630. Binding to NiNTA beads and elution was performed as described above, the final size-exclusion chromatography step was performed as above but using 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.3, 100 mM NaCl and 1 mM EDTA as a buffer. Protein-containing fractions were pooled and concentrated up to ~ 40 mg/ml.

7.5.3.3 Seb1-RRM$_{388-540}$

The protein was expressed from plasmids 3967 and 4055-4072 (Table 7) under the conditions described in section 7.5.3.1 but using pLysS *E. coli* cells and 30 µg/ml Ampicillin and 25 µg/ml Chloramphenicol as antibiotics in the medium. Harvesting, lysis and purification via NiNTA was performed as detailed in section 7.5.3.2. For fluorescence anisotropy assays and EMSAs, size-exclusion chromatography was performed using a HiLoad 16/60 Superdex 200 prep grade column in 50 mM Tris pH 8.0 and 200 mM NaCl and protein-containing fractions were concentrated to ~ 5 mg/ml. Some purifications were performed by Mr Oliver Adams and Mr Miles Huseyin, as detailed in the text.

For protein that was used for crystallisation trials and SAXS, the His_6-SUMO-tag was cleaved off. To achieve this, protein-containing NiNTA fractions were pooled and the buffer was exchanged into 50 mM Tris pH 7.5, 300 mM NaCl, and 500 mM non-detergent sulfobetain (NDSB)-201 using PD-10 desalting columns (GE Healthcare). The resulting protein solution was incubated with ~ 15 U 3C protease (Merck-Millipore) per mg of purified protein on a rotator at room temperature (RT) overnight. Precipitated protein was pelleted out by centrifugation at 25,000 *g* at 4°C and uncleaved protein as well as the His_6-SUMO cleavage product were removed by incubation with NiNTA. The $RRM_{388-540}$ containing flow-through was further purified by size-exclusion chromatography using a Superdex 75 10/300 GL column (GE Healthcare) in 20 mM HEPES pH 7.3, 150 mM NaCl, and 500 mM NDSB-201. Protein-containing fractions were pooled and concentrated to ~ 8.7 mg/ml for crystallisation or ~ 5 mg/ml for SAXS. For the latter, unconcentrated fractions were used as well. The purification for crystallisation trials was performed with the help of Mr Oliver Adams (Part II student, Vasilieva lab).

7.5.4 Crystallisation, data collection and refinement

Crystallisation was done in collaboration with Mr Max Renner (*DPhil* student) at the laboratory of Prof Jonathan Grimes at the Division of Structural Biology (Oxford University). The sitting-drop vapour diffusion method was used and trials were performed in CrystalQuick 96-well plates (Greiner) at 20.5°C using a Cartesian Technologies robotic pipetting system (Walter et al. 2005). 100 ng of protein solution were mixed with 100 ng reservoir (using a range of different screens) and equilibrated against 95 µl reservoir solution. Because of the high amount of NDSB-201 in the $RRM_{388-540}$ preparation, 10 µl 5 M NaCl was added to the reservoir prior to plate sealing and imaging.

7.5.4.1 Seb1-CID_{1-152}

For Seb1-CID_{1-152}, we obtained diamond-shaped crystals after one day in 100 mM Tris pH 8.0 and 4 M NaCl. The crystals were cryo-protected with 25% (v/v) glycerol and

flash-frozen in liquid nitrogen by Dr Karl Harlos (Division of Structural Biology, Oxford). Diffraction data was collected at Diamond Light Source (DLS), Didcot, UK on the I03 beamline at 100 K. Data processing was done by Mr Max Renner using XIA2 (Winter et al. 2013, 2) in the space group $P3_121$ (see Table).

The following was also performed by Mr Max Renner: The phase problem was solved by molecular replacement using PHASER (McCoy et al. 2007) and the Nrd1-CID domain (PDBID 3CLJ) was used as a search model. Model building and refinement were done in an iterative manner using COOT (Emsley and Cowtan 2004) and AUTOBUSTER (Smart et al. 2012) with TLS parameters. Quality assessment of the final geometry was performed using MolProbity (Davis et al. 2007) (see Table 4 for refinement and model statistics).

7.5.4.2 Seb1-RRM$_{388-540}$

For the Seb1-RRM$_{388-540}$ domain, we obtained rod-shaped crystals also after one day in the following conditions:

- 1 M ammonium formate, 100 mM sodium cacodylate pH 6.5, 8% (w/v) poly-γ-glutamic acid polymer (PGA-LM, 200-400 kDa low molecular weight polymer)
- 0.3 M Na-malonate, 0.1 M Na-cacodylate, pH 6.5, 8% (w/v) PGA-LM,
- 0.8 M Na/K hydrogen phosphate pH 7.5
- 1 M Lithium Sulfate, 0.1 M 2-(N-Morpholino)ethanesulfonic Acid (MES) pH 6.5
- 0.1 M NaCl, 0.1 M Tris pH 8.0, 8% (w/v) PEG-20000
- 0.2 M Ammonium Acetate, 0.1 M HEPES pH 7.5, 25% (w/v) PEG-3350

The crystals were frozen by Mr Max Renner and Dr Guido Paesen (postdoctoral researcher at the laboratory of Prof Jonathan Grimes) as described in section 7.5.4.1. Native diffraction data was collected at DLS on the beamline I04 at 100 K and was also processed by Max Renner using XIA2 in the space group C121 (see Table 5). To obtain sulphur single-wavelength anomalous dispersion (S-SAD) data, the diffraction pattern of

22 $RRM_{388-540}$ crystals was collected at DLS on beamline I03 at a wavelength of 1.77 Å using the inverse beam method (at 5° wedges).

For determination of the phases experimentally, the S-SAD data was merged using XIA2 and the sulphur sites located with HKL2MAP (Sheldrick 2010). The main-chain was traced iteratively in SHELXE, the density was modified and the phases extended to native resolution (1.0 Å). This allowed the generation of an initial model using PHENIX autobuild (Adams et al. 2011) and subsequent manual building and refinement was iteratively performed using COOT and PHENIX, respectively. Towards completion of model refinement, anisotropic B-factor refinement was used. For secondary structure element assignment, PHENIX ksdssp was used and the assessment of model geometry was performed using MolProbity (see Table 5 for refinement and model statistics). The phase problem was solved by Dr Kamel El Omari (DLS, Didcot, UK) and model refinement was performed by Mr Max Renner.

7.5.5 Small angle X-ray scattering (SAXS)

SAXS data of the Seb1-$RRM_{388-540}$ was collected at DLS on beamline B21. This was done with the help of Mr Max Renner and all data processing was performed per him exclusively. During data collection at a wavelength of 1 Å and a detector distance of 4.09 m, the sample was kept at 16°C. Fractions from size-exclusion chromatography which did not contain any protein were used for buffer subtraction. Data merging and analysis was performed using the software package ScÅtter (www.bioisis.net/scatter). The crystal structure of the domain was subsequently fitted to the SAXS curve using the FoXS server (Schneidman-Duhovny et al. 2016). ScÅtter was also used to determine the pair distribution function P(r) which was utilized to independently calculate 23 *ab initio* bead models with DAMMIF (Franke and Svergun 2009), that were subsequently averaged with DAMAVER (Volkov and Svergun 2003) and a final refinement run was performed using DAMMIN (Svergun 1999). The $RRM_{388-540}$ crystal structure was then fitted into the resulting *ab initio* bead model using SUPCOMB (Kozin and Svergun 2001).

7.5.6 Peptide binding assay

Recombinantly expressed and purified full-length Seb1 (see section 7.5.3.1) with various point mutations were used to determine their binding specificity to CTD peptides *in vitro*. For this, four repeats of the CTD heptad $(Y_1S_2P_3T_4S_5P_6S_7)_4$ that were either not phosphorylated or phosphorylated on all repeats at position S2, S5, or S7 and that contained a biotin tag at the N-terminus, ordered from the company peptides&elephants GmbH (Potsdam, Germany) were utilized. 5 µg of peptides dissolved in TBE (10 mM Tris pH 8.0, 1 mM ethylenediaminetetraacetic acid (EDTA)) were incubated with 60 µl of equilibrated Streptavidin-coated Dynabeads (Invitrogen) for 1-4 h at 4°C. Unbound peptides were removed by 3 washes with 500 µl OBB (25 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM DTT, 0.03% Triton X-100, 5% glycerol). Next ~ 9.5 µg of recombinant protein in a final volume of 500 µl (protein solution + OBB) were added to the beads and incubated at 4°C overnight on a rotator. The beads were washed 5 times with 500 µl OBB for 5 min at 4°C. Protein was eluted by addition of 1x Lämmli buffer (Laemmli 1970) and incubation at 95°C for 5 min and analysed by SDS-PAGE (see section 7.5.9).

7.5.7 Fluorescence anisotropy assay

7.5.7.1 Seb1-CID$_{1-152}$

Recombinant protein was used (see section 7.5.3.2) and serially diluted with 25 mM HEPES pH 7.3, 100 mM NaCl and 1 mM EDTA. Two repeat CTD peptides $(Y_1S_2P_3T_4S_5P_6S_7)_2$ that were either unphosphorylated or phosphorylated on both repeats at position S2, S5 or S7 and contained an N-terminal 5'-fluorescein amidite (FAM)-tag were obtained from peptides&elephants (Potsdam, Germany) and dissolved in 25 mM HEPES pH 7.3, 300 mM NaCl and 1 mM EDTA. 15 µl of protein solution was mixed with 15 µl of 150 nM peptide and incubated for 20 min at RT in black 96-well half-area plates (Greiner). The samples were excited with linearly polarised light at a wavelength of 485 nm and emission at 520 nm was recorded in the planes parallel and perpendicular to the emission plane at 25°C using a FLUOstar-Omega microplate reader (BMG-Labtech). Five

measurements were performed for each well and averaged, each samples was measured at least in duplicates and repeated with independent protein preparations. The fluorescence anisotropy without protein was subtracted from the sample anisotropies and the K_d values for each peptide was determined by plotting the fluorescence anisotropy signal against the protein concentration (in µM) and curve fitting to the following equation from (Reinstein et al. 1990) which assumes 1:1 stoichiometry:

$$A = A_0 + \left(A_{max} - A_0\right) \cdot 0.5 \cdot \left\{ \left[1 + \frac{P_0 + K_d}{L_0} \right] - \left[\left(1 + \frac{P_0 + K_d}{L_0} \right)^2 - \frac{4 \cdot P_0}{L_0} \right]^{0.5} \right\}$$

Here, A is the fluorescence anisotropy signal (i.e. the y-value of the graph), A_{max} the maximum anisotropy, A_0 is the fluorescence anisotropy at a protein concentration of 0 µM (i.e. the starting value), P_0 is the protein concentration (i.e. the x-value of the graph), K_d the equilibrium dissociation constant (which is to be determined) and L_0 the ligand concentration (in this case the peptide) which was held constant at 0.075 µM. Curve fitting was performed using the program QtiPlot v0.9.9.3.

7.5.7.2 Seb1-RRM$_{388-540}$

Fluorescence anisotropy measurements and data analysis for the Seb1-RRM$_{388-540}$ domain and various point mutants were essentially performed as described above (see section 7.5.7.1) by Mr Oliver Adams and Mr Miles Huseyin under my supervision (both students at the Vasilieva lab). Binding of different concentrations of recombinant RRM$_{388-540}$ (see section 7.5.3.3) was determined to 40 nM FAM-labelled RNA of the sequence AUUAGUAAAA (obtained from Eurofins). Measurements were performed in a final concentration of 25 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, 16.7% (v/v) glycerol, 0.1% IGEPAL, 0.1 mg/ml tRNA and 2.5% (v/v) RNasin (Promega). All RNA used was denatured at 95°C for ~ 2 min and rapidly transferred on ice for cooling before addition to the buffer.

7.5.8 Electrophoretic mobility shift assay (EMSA)

For EMSAs, 8 different 55 nt long RNAs (obtained from Eurofins) with the following sequences were used:

RNA1:

AGUCAAGGAAAACUAGGCAUGCAAAAGAAAAGCCUUAUGUUAAAUUUUAAAUGUU

RNA2:

AUGUUAAAUUACAAUUAGAUUCAUAGUUUGGCUUGCUAUACCUUAGAAAGUUAUU

RNA3:

UUAUUUUAUAGAGCAUGCGCGAAUUUAAAUUUAUGGUACGAGUGUUUUCAUUGAU

RNA4:

UUGAUUAACUGUUUGUUGCUUUCUUUCUACUUAUUAUCUCCAUUUGGAUAAGAAA

RNA5:

AGAAAGCAUAAGCAUUCGUUUUAUACAAAUCAGAGUGACAUAUAGUCUACUUAAA

RNA6:

UAAAUUUACAAUUGAUUAGACUUUGAUAAAUGUAACUGAAAAUUUUUGUUCAUAU

RNA7:

AUAUCGAAGGUCAGCGUCUAAAAUUUUUAGCAUGUUGACUGAAUUCUUUAAACCU

RNA8:

AACCUUUUGAGUUAAACUGAAAGCUGAUUAAUUAAACAAUUAUAUCUUGGUCUGG

40 pmol of RNA were radioactively labelled at the 5'end for 1 h at 37°C by T4 polynucleotide kinase (NEB) as recommended by the manufacturer using 0.37 MBq P^{32}-γ-ATP (370 MBq/ml). The RNA was subsequently purified using illustra Micro-Spin G-25 columns (GE Healthcare) according to the manufacturers instructions. The RNA was stored at -20°C until use and was denatured at 95°C for 2-3 min followed by rapid transfer to ice until cool before addition to the protein solution.

0-50 µM of recombinant $RRM_{388-540}$ (see section 7.5.3.3) was used to determine the affinity of the protein to the above RNAs. The same buffer conditions were used as for the fluorescence anisotropy measurements described in section 7.5.7.2. After incubation of

the protein with RNA for 20 min at RT, the samples were analysed on a pre-run 4-20% TBE gel (Invitrogen) at 100 V for 3 h at 4°C. Gels were dried at 60°C for at least 2 h under vacuum and exposed to phosphor imaging plates (Fujifilm, Japan). Visualisation of the radioactive signal was achieved with an Image Analyser (FLA-600, Fujifilm, Japan) and quantified using the software Aida Advanced Image Data Analyzer (version 4.50.010). The fraction of RNA bound by protein (i.e. (shifted signal)/(total signal)) was plotted against the protein concentration and the K_d value was determined by fitting of the data to the Hill equation shown below using QtiPlot v0.9.9.3, where f is the fraction of RNA bound (i.e. the y-value), P_t is the total protein concentration in the sample (i.e. the x-value) and K_d is the dissociation constant (i.e. the value to be determined):

$$f = \frac{P_t}{K_d + P_t}$$

7.5.9 SDS-PAGE

Samples were prepared by addition of Lämmli loading buffer (Laemmli 1970) and subsequently denatured at 95°C for 5-10 min. The Mini-PROTEAN (Bio-Rad) gel pouring and electrophoresis system was utilized for all gels. Depending on protein size, between 6% and 15% polyacrylamide resolving gel was used containing a 5% stacking gel. Gels were run at 150 V for 15 min followed by 200 V for 45 min. Staining was performed using InstantBlue (Expedeon) according to the manufacturer's instructions or 0.25% Coomassie Brilliant Blue R-250 in 30% (v/v) ethanol and 10% (v/v) acetic acid for 1 h at RT followed by several rounds of destaining with 30% (v/v) ethanol and 10% (v/v) acetic acid until the background was clear. Silver staining was done using the SilverQuest Silver Staining Kit (Invitrogen) as described by the manufacturer with initial fixation overnight.

7.5.10 Western blotting

In order to detect proteins by Western blot, SDS-PAGE of the samples was performed first as described in section 7.5.9. Directly after electrophoresis, the proteins were transferred onto a polyvinylidene fluoride (PVDF) membrane via wet transfer using

the MiniPROTEAN (Bio-Rad) equipment and 25 mM Tris, 192 mM glycine, 1 mM EDTA and 20% (v/v) methanol as transfer buffer. The transfer was performed at 30 V for 16 h at 4°C. Next, the membrane was blocked with TBST (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween-20) + 5% (w/v) skim milk powder for 30 min at RT. A primary antibody or a horse radish peroxidase (HRP)-coupled antibody were diluted in TBST + 5% skim milk powder as shown in Table 9 and incubated with the membrane for 1 h at RT or overnight at 4°C. The membrane was then washed twice for ~ 30 sec and three times for 10 min with TBST. If necessary, the membrane was incubated with a secondary antibody as described above and the washing was repeated as before. Proteins were subsequently visualised by addition of Clarity Western ECL (Bio-Rad) to the membrane according to the manufacturer's instructions, exposition to Amersham Hyperfilm ECL (GE Healthcare) and development using a SRX-101A developer (Konica Minolta).

7.6 *S. pombe in vivo* methods

7.6.1 Media and growth conditions for *S. pombe*

If not stated otherwise, cells were grown in yeast extract broth + supplements (YES) medium (5 g/l yeast extract, 30 g/l glucose and 225 mg/l of leucine, adenine, histidine, uracil, lysine) at 30°C and 200 rpm until OD_{600} ~ 0.5-0.9 was reached. If cells were used that have *seb1* under the control of an *nmt1* promoter, they were first grown in Edinburgh's minimal medium + glutamate (EMMG, 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) which does not contain thiamine before being shifted to YES (which contains thiamine) for 24 h at 30°C and 200 rpm.

For cloning, YES plates containing 2% (w/v) agar supplemented with 100 mg/l Geneticin or 100 mg/l Clonnat (depending on resistance) were used. To use auxotrophic markers for selection, EMMG plates containing 2% (w/v) agar and lacking the corresponding amino acid were utilised. After transformation, cells were grown at 30°C until visible colonies appeared (usually 2 days on YES and ~ 4-5 days on EMMG). For spot testing, cells were grown on YES plates without antibiotics or EMMG plates containing all amino acids listed above.

7.6.2 Genetic manipulation and transformation

For endogenous epitope tagging or change of promoter, standard 2-step PCR methods were used as described (Bähler et al. 1998). For introduction of the mutated *seb1* gene into an ectopic location, the pDUAL plasmid was stably integrated into the *leu1-32* locus as described (Matsuyama et al. 2004). Transformation of *S. pombe* cells was performed using the Lithium acetate methods as described by Hyun Soo Kim and Michael Keogh. In brief, 5 ml of cells in exponential growth phase were harvested and washed with sterile water, followed by LiTE (0.1 M Lithium acetate, 10 mM Tris-HCl pH 7.5, 1 mM EDTA) and finally resuspended in 100 µl LiTE. 20 µg of denatured ssDNA and 1 µg of the DNA to be integrated were added and the cells were incubated at RT for 10 min. After addition of 260 µl PLATE (40% (v/v) PEG 4000, 0.1 M Lithium acetate, 10 mM Tris-HCl pH 7.5, 1 mM EDTA), cells were incubated for 1 h at 30°C. 43 µl of dimethyl sulfoxide (DMSO) was added and a heat shock step for 5 min at 42°C followed. The cell pellet was resuspended in 200 µl YES (for antibiotic selection) or EMMG (for auxotrophic selection). If antibiotics were used, cells were allowed to recover overnight at 30°C before plating on antibiotic-containing YES agar. In the case of EMMG, cells were plated directly after transformation.

7.6.3 Protein extraction for Western blot (TCA method)

In order to assess protein levels *in vivo*, whole cell extracts of *S. pombe* cells were generated using the denaturing trichloroactic acid (TCA) method thereby preventing any protein degradation after cell lysis. For this, the method described in (Keogh et al. 2006a, 2006b) was used.

7.6.4 Spot test

To test cell growth in different mutants, cells were grown in liquid culture until an OD_{600} of 0.5 was reached. 1 ml was harvested and the cells were resuspended in 500 µl of sterile water. 5 serial dilutions (1:10) were made and 5 µl (YES) or 7 µl (EMMG) of each dilution was pipetted onto a dry plate. The cells were cultivated at 30°C and the plates were imaged every 12-24 h for 1 week.

7.6.5 RNA extraction

Whole-cell steady-state RNA was extracted from *S. pombe* cells for various downstream methods (see sections 7.6.6, 7.6.7, and 7.7.3). For this, 50 ml of cells were harvested and washed with 1 ml H_2O before resuspension in cold AE buffer (50 mM sodium acetate pH 5.0, 10 mM EDTA, 1% (w/v) SDS). An equal volume of phenol:chloroform (1:1) pH 6.0 was added and the mixture was vortexed for 15 sec and incubated at 65°C for 5 min. After centrifugation at 25,000 *g* for 15 min at 4°C, the upper aqueous phase was transferred into a new tube and mixed with an equal volume of phenol:chloroform and vortexed as before. After centrifugation as described above, the upper phase was removed and transferred into a new tube. 2.5 volumes of ethanol and 1/10 volume of 3 M sodium acetate (pH 6.5) were added and the RNA was left to precipitate at -80°C for 1 h or -20°C overnight. The precipitated RNA was spun at 25,000 *g* for 30 min at 4°C and the pellets were washed with -20°C cold 80% ethanol before air-drying for ~ 3 min. The RNA was subsequently resuspended in H_2O, the concentration was determined using a NanoDrop Spectrophotometer ND-1000 (PeqLab).

7.6.6 Northern blotting

8 µg of RNA extracted as described in section 7.6.5 was separated using a 1.2% (w/v) agarose formaldehyde gel. After the electrophoresis was complete, the separated RNA was fragmented by submerging the gel in 10 mM NaOH for 20 min at RT while shaking. The gel was subsequently rinsed with H_2O and washed with 20x SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0) for 5 min and 10x SSC for 40 min. The RNA was blotted onto a nylon membrane (Roche diagnostics) by capillary transfer using 10x SSC as a buffer for at least 12 h at RT. The RNA was subsequently crosslinked to the membrane by UV irradiation (using 120 mJ on a Stratagene crosslinker). The membrane was then incubated with 3% (v/v) acetic acid for 10 min while shaking at RT and the rRNA was stained with 0.04% methylene blue in 0.5 M sodium acetate pH 5.2. The membrane was washed with water until the background was complete destained and imaged.

Digoxigenin (DIG)-labelled RNA Probes were generated by *in vitro* transcription using the Ambion Maxiscript Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The membrane was pre-hybridised with DIG-Easy-Hyb solution (Roche diagnostics) at 60°C for 1 h before 100 ng/ml of denatured strand-specific probes were added and further incubated overnight. After hybridisation, the membrane was washed 3x for 15 min and 2x for 10 min with 2x SSC containing 0.1% (w/v) SDS at 60°C. This was followed by a wash with 1x PBS (140 mM NaCl, 10 mM Na_2HPO_4, 2.7 mM KCl, 1.8 mM KH_2PO_4, pH 7.4) for 2 min at RT and blocking of the membrane using 1x blocking solution (Roche diagnostics) in 100 mM maleic acid, 150 mM NaCl, 0.3% Tween-20 for 30 min at RT. α-DIG alkaline-phosphatase-conjugated antibody (Fab-fragment) was added at a final dilution of 1:10,000 and incubated for 1 h. The membrane was subsequently washed 5x for 5 min with 1x PBS at RT and 2x for 10 min with 100 mM Tris-HCl pH 9.5. To visualise the RNA signal, CDP-Star (Sigma-Aldrich) was added and the membrane was exposed to Amersham Hyperfilm ECL (GE Healthcare) and developed using a SRX-101A developer (Konica Minolta). This experiment was performed by Mr Oliver Adams (Part II student) under the supervision of Dr Cornelia Kilchert (postdoctoral researcher, Vasilieva lab).

7.6.7 Reverse transcription coupled with quantitative PCR (RT-qPCR)

2 µg of RNA extracted from *S. pombe* cells as described in section 7.6.5 were digested with 2 U of DNase RQ1 (Promega) in a total volume of 50 µl for 1 h at 37°C according to the manufacturer's instructions and the reaction was terminated using the provided stop solution. 2.5 µl (corresponding to 100 ng of RNA) were used for subsequent reverse transcription with gene specific primers (see Table 6) using the Transcriptor reverse transcriptase (Roche diagnostics) in a total reaction volume of 25 µl. cDNA levels per 5 µl RT reaction were then determined by qPCR using the SensiMix SYBR (Bioline) reaction mix and the result of technical duplicates was averaged. Some of the samples were processed by Mr Oliver Adams (Part II student, Vasilieva lab) under my supervision.

7.6.8 Chromatin immunoprecipitation (ChIP) with detection by qPCR

ChIP analysis was performed by Dr Beth Watts (postdoctoral researcher, Vasilieva lab) essentially as described in (Keogh and Buratowski 2004; Kim et al. 2004b, 2004a). Summarised, 200 ml of cells were crosslinked for 20 min at RT using a 11% (v/v) formaldehyde solution. The reaction was quenched with 30 ml of 3 M glycine and 20 mM Tris. After harvesting, the cell pellets were washed with cold TBS and FA lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS) and the pellet was stored at -80°C until usage.

The cell pellet was resuspended in FA lysis buffer containing 0.5% (w/v) SDS and the cells were lysed by vortexing for 30 cycles with 30 sec vortexing and 30 sec on ice. After ultracentrifugation at 150,000 *g* for 20 min at 4°C, the pellet was crushed using a wooden stick and resuspended in FA lysis buffer. The centrifugation and resuspension steps were repeated, before the chromatin was sonicated for 80 min with cycles of 15 sec on/ 45 sec off at maximum intensity using a Biorupter sonicator (Diagenode). After centrifugation at 25,000 *g* for 20 min at 4°C, the chromatin can be found in the supernatant. The DNA concentration was measured using a NanoDrop Spectrophotometer ND-1000 (PeqLab) and adjusted between samples. The NaCl

concentration was increased to 275 mM and the samples were aliquoted and stored at -80°C until use.

1 ml of shared chromatin was incubated with 5 µl of either packed rabbit IgG agarose (Sigma-Aldrich) or protein-G Dynabead slurry (Life Technologies) couple to the appropriate antibody (see Table 9) on a rotator at 4°C overnight. The beads were extensively washed and DNA:protein crosslinks were reverted using 20 µg of pronase (Roche diagnostics) for 1 h at 42°C followed by incubation at 65°C overnight. Immunoprecipitated DNA was extracted with phenol:chloroform:isoamyl alcohol (25:24:1), ethanol precipitated as described in section 7.6.5 and quantified using qPCR (as in section 7.6.7). All IPs were normalised to input and all primer pairs normalised to a control gene (*fbp1*).

7.6.9 Protein purification and mass spectrometry

Protein purification of Ysh1-HA-TAP was performed essentially as described (Puig et al. 2001). 16 l of *S. pombe* culture were grown to an OD_{600} of about 0.7 and harvested by centrifugation at 3000 *g* and 4°C for 4 min. The combined pellets were washed with water and resuspended in 4 ml TMN buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 10% glycerol) 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)). The cells were frozen by dripping the resuspended pellets into liquid nitrogen and stored at -80°C.

Cell lysis was performed by grinding the frozen cell beads in liquid nitrogen using mortar and pestle for 45 minutes. The cells were thawed and an equal volume of TMN with protease inhibitors was added. All following steps were carried out at 4°C using pre-cooled equipment. Aliquots of 7 ml solution were vortexed 5x for 40 sec in Pyrex tubes containing 7 ml acid-washed glass beads. The lysates were combined and the Pyrex tubes were cleaned with 3 ml of the same buffer. After centrifugation at 3,000 *g* for 8 min, the supernatant was cleared at 75,000 *g* for 1.5 h.

The supernatant was treated with 1.0 µl/ml Benzonase (Sigma-Aldrich, ≥ 90% purity), 10 µg/ml RNase A (Sigma-Aldrich), 60 U/ml RNase T1 (Fisher Scientific), and 0.25 U/ml DNase RQ1 (Promega) for 2 h at RT. The untagged control was incubated under the same conditions. Samples before and after digestion were removed and the success of the treatment was confirmed by RNA/DNA extraction and analysis of the fragment sizes on a 10% (v/v) polyacrylamide/7 M urea gel followed by ethidium bromide staining.

50 µl of a 1:1 mixture of washed IgG agarose (Sigma Aldrich) and IgG Sepharose (GE Healthcare) beads (packed) was added per 100 µg protein in the lysate (as determined by Bio-Rad protein assay using bovine serum albumin as a standard). Binding of Ysh1-HA-TAP to the beads was performed at 4°C overnight on a rotator. The beads were washed 2x for 5 min and 3x for 15 min with 20 ml TMN buffer (without protease inhibitors) and once for 5 min with 20 ml and once with 1 ml TEV cleavage buffer (150 mM NaCl, 10 mM Tris-HCl pH 8.0, 0.5 mM DTT, 0.05% (v/v) IGEPAL). The beads were resuspended in an equal volume of the buffer and 2 µl TEV protease (NEB) per 100 µl of beads was added and cleavage was performed at 4°C overnight on a rotator.

The cleaved protein was eluted using Mini Bio-Spin Chromatography Columns (Bio-Rad) by centrifugation at 100 *g* for 5 sec at 4°C. The beads were washed 2x with TEV cleavage buffer and the flow-throughs were combined with the eluate. $CaCl_2$ was added to a final concentration of 2 mM and the volume was increased to 10 ml with Calmodulin binding buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM $CaCl_2$, 1 mM magnesium acetate, 1 mM Imidazol, 1 mM β-mercaptoethanol, 10% (v/v) glycerol, 0.02% (v/v) IPEGAL) containing protease inhibitors. The same amount of washed Calmodulin beads (GE Healthcare) was added as IgG beads in the previous step and incubated on a rotator at 4°C overnight.

The beads were washed 2x for 2 min and 2x for 20 min with Calmodulin wash buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.1 mM $CaCl_2$, 1 mM magnesium acetate, 1 mM Imidazol, 1 mM β-mercaptoethanol, 10% (v/v) glycerol) containing 0.02% (v/v)

IPEGAL and protease inhibitors and 2x for 5 min with Calmodulin wash buffer containing no IPEGAL. The purified complex still bound to the beads was used for mass spectrometric analyses. For this, the samples were run 1 cm into the resolving gel of an SDS gels (see section 7.5.9) which was subsequently cut into 1 mm pieces using a new razor blade. The gel pieces were subjected to tryptic digestion and 2 h LC-MS/MS Orbitrap which was performed by the Central Proteomics Facility at the Dunn School of Pathology, University of Oxford.

7.7 Genome-wide and bioinformatics methods

7.7.1 ChIP-sequencing (ChIP-Seq)

Chromatin was prepared as described in section 7.6.8, however, double the amount was used per repeat. After decrosslinking of the DNA:protein complexes, the samples were treated with 0.02 mg RNase A (Roche diagnostics) for 1 h at 37°C. DNA was subsequently purified using the ChIP DNA Clean & Concentrator kit (ZymoResearch, USA) according to the manufacturer's instructions. For construction of the sequencing library, NEBNext Fast DNA Library Prep Set for Ion Torrent™ Kit (NEB, USA) was utilised. Sequencing was performed using Ion PITM ChIPs v2 BC and the Ion Chef™ and Ion Torrent Proton instruments (Life Technologies, USA). These experiments were performed by Dr Beth Watts (postdoctoral research, Vasilieva lab), Mr Miles Huseyin (rotation student, Vasilieva lab), and Dr Dong-Hyuk Heo (postdoctoral researcher, Vasilieva lab).

For the data analysis, low quality reads (Phred score less than 20) and reads shorter than 20 nt were removed using Trimmomatic (version 0.36) (Bolger et al. 2014). The resulting reads were aligned to the *S. pombe* genome (version ASM294v2.28) using Bowtie2 (version 2.2.6) (Langmead and Salzberg 2012). This was done by Mr Miles Huseyin (rotation student, Vasilieva lab) and Miss Tea Kecman (*DPhil* student, Vasilieva lab).

7.7.2 Photoactivatable ribonucleotide-enhanced crosslinking and immuno-precipitation (PAR-CLIP)

The PAR-CLIP experiment and data analysis was performed by Dr Carlo Bäjen (Max Planck Institute for Biophysical Chemistry, Göttingen, Germany) at the laboratory of Dr Patrick Cramer essentially as described in (Baejen et al. 2014). In brief, cells grown in CSM minimal medium supplemented with 10 mg/l uracil, 100 µM 4-thiouracil and 4% glucose to an OD_{600} of 0.5 were supplemented with an additional 900 µM 4-thiouracil and further incubated for another 4 h. The cells were irradiated with 12 J/cm² of UV light at a wavelength of 365 nm. After lysis by bead beating, the sample was sonicated using a Covaris S220 system and TAP-tagged Seb1 was IPed overnight at 4°C using rabbit IgG-conjugated protein-G Dynabeads (Sigma-Aldrich and Invitrogen). RNA was partially digested using RNase T1 (Thermo Fisher Scientific), adapters were ligated and the RNA was recovered by Proteinase K digestion (NEB). After phenol:chloroform extraction, reverse transcription was performed using the SuperScript II RTase kit (Invitrogen). The resulting cDNA was amplified, size-selected and quantified on a 2200 TapeStation System (Agilent Technologies). The sequencing reaction was performed on a HiSeq 1500 (Illumina).

Data analysis was performed using the *S. pombe* genome version ASM294v2.25. The identified crosslinking sites were normalised to total RNA abundance which was determined in a separate experiment as follows:

S. pombe WT cells were treated and crosslinked with 1 J/cm² as for PAR-CLIP described above. After cell lysis, the total RNA was extracted using acid phenol:chloroform (Roti-Phenol, Carl Roth) and purified utilising the RNA Clean Concentrator-5 kit (ZymoResearch). rRNA depletion was performed using the Ribo-Zero rRNA removal kit (Epicenter) and the resulting RNA was used for library preparation with the Encore Complete RNA-Seq Library Systems kit (NuGEN). The libraries were quantified and sequenced as before.

7.7.3 RNA-Sequencing (RNA-Seq)

RNA extraction was performed as described in section 7.6.5 and DNase treated as detailed in section 7.6.7. rRNA depletion and library preparation was performed by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (Oxford). Paired-end sequencing in 75 cycles was done on a HiSeq 4000 (Illumina). The resulting reads were quality trimmed (Phred score less than 20) and reads shorter than 20 nt were removed using Trimmomatic (Galaxy version 0.32.3) (Bolger et al. 2014). The remaining reads were subsequently aligned to the *S. pombe* genome (ASM294v2.28) with TopHat (Galaxy version 0.9) (Kim et al. 2013). Quality control of the aligned reads was performed by principle component analysis and Spearman correlation factor calculation between replicates using deepTools (version 2.2.2) (Ramírez et al. 2014).

Read-through transcription was determined by counting all reads in a window from the pA site to 50 nt after the pA site using HTSeq (version 0.6.1.p1) (Anders et al. 2015). and for all gene bodies using the annotations from (Eser et al. 2016). Read-through transcription was normalised to expression levels (gene-body counts) and averaged over different replicates. For each gene, the normalised read-through in each mutant was then divided by the read-through detected in WT, giving a fold-change in read-through. The significance of the overall difference between WT and mutants was calculated in R using the Wilcoxon-Mann-Whitney test (also known as Mann-Whitney-U test). For individual genes, the significance in read-through was determined by the Kruskal-Wallis test using the matR package in R (Braithwaite and Keegan 2014). Genes that show significant read-through in multiple mutants were visualised using Venn diagrams which were plotted by the Venn Diagram Plotter version 1.5.4798.30840.

7.7.4 Correlation between different genome-wide datasets

7.7.4.1 Metagene profile plots

Metagene profiles were generated using an in-house R script which utilises the Bioconductor (Huber et al. 2015; Gentleman et al. 2004) package rtracklyer (Lawrence et al. 2009). In short, mean aligned reads were calculated for every nucleotide in the indicated window and plotted for the indicated genes using ggplot2 (Wickham 2009). For PAR-CLIP data, the averaged Seb1 occupancy signal was smoothed using the Gaussian smoothing function provided in the R package smoother (Hamilton 2015) and multiplied by 2000 (if plotted together with ChIP-Seq data). In order to find enrichment of the UGUA motif on the sense and antisense strands, the motif was located using the Bioconductor packages Biostrings (Pages et al.) and BSgenome (Pages). Metagene profiles were calculated as described for the PAR-CLIP data.

7.7.4.2 Determination of binding events in certain region

For all analyses, annotations from (Eser et al. 2016) were used. For ChIP-Seq data, peak calling was performed with MACS2 (version 2.1.1.20160309) (Zhang et al. 2008) which was used to determine binding in indicated regions. For PAR-CLIP, presence of crosslinking events were used and for the UGUA motif, occurrence of the motif was used, which were determined using Bioconductor in R. The number of genes that show overlapping binding by different methods were visualised as Venn diagrams (plotted as described in section 7.7.3).

7.7.4.3 Boxplots of transcriptional read-through in dependence of binding

In order to assess whether transcriptional read-through (as determined by RNA-Seq) correlates with Seb1 binding events in PAR-CLIP or ChIP-Seq, all genes were split into two subsets depending on whether they show crosslinks in PAR-CLIP or peak summits in ChIP-Seq (as determined using MACS2) in indicated regions. The change in transcriptional read-through as calculated in section 7.7.3 was then plotted for each

mutant for both subsets of genes separately using R. Significant differences were determined using the Wilcoxon-Mann-Whitney test.

7.7.4.4 Visualisation of aligned reads

To visualise aligned reads of any dataset, the integrated genome browser (version 9.0.0) was used (Nicol et al. 2009).

7.7.5 Protein bioinformatics

S. pombe gene names and nucleotide sequences were taken from www.pombase.org. UniProt (www.uniprot.org) was used to get protein sequences for alignment and to look for the presence and regions of domains. Sequence alignments were done using PROMALS3D (Pei et al. 2008), taking published structures and structures determined in this thesis into account. Homologous proteins were searched using NCBI PSI-BLAST in multiple iterations, manually sorting hits by length of aligned region. The resulting proteins were aligned as before and the resulting file was converted to a nexus file. Phylogenetic trees and evolutionary distances were subsequently calculated using SplitsTree4 (version 4.14.2) (Huson and Bryant 2006). Disorder prediction of Seb1 was determined using the DisMeta server (Huang et al. 2014).

7.7.6 Data access

All raw (fastq) and processed sequencing data can be accessed via the NCBI Gene Expression Omnibus (GEO) server using the accession number GSE89914. The structure coordinate files have been deposited to the Protein Data Bank (PDB) and can be viewed using the accession codes 5MDT (CID) and 5MDU (RRM).

8 REFERENCES

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Echols N, Headd JJ, Hung L-W, Jain S, Kapral GJ, Grosse Kunstleve RW, et al. 2011. The Phenix software for automated determination of macromolecular structures. *Methods* **55**: 94–106.

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

Albulescu L-O, Sabet N, Gudipati M, Stepankiw N, Bergman ZJ, Huffaker TC, Pleiss JA. 2012. A Quantitative, High-Throughput Reverse Genetic Screen Reveals Novel Connections between Pre–mRNA Splicing and 5′ and 3′ End Transcript Determinants. *PLoS Genet* **8**: e1002530.

Alexander RD, Innocente SA, Barrass JD, Beggs JD. 2010. Splicing-Dependent RNA Polymerase Pausing in Yeast. *Mol Cell* **40**: 582–593.

Allain FH, Gilbert DE, Bouvet P, Feigon J. 2000. Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target. *J Mol Biol* **303**: 227–241.

Allen BL, Taatjes DJ. 2015. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**: 155–166.

Allison LA, Moyle M, Shales M, Ingles CJ. 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**: 599–610.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Arigo JT, Eyler DE, Carroll KL, Corden JL. 2006. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell* **23**: 841–851.

Arndt KM, Reines D. 2015. Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annu Rev Biochem* **84**: 381–404.

Bacikova V, Pasulka J, Kubicek K, Stefl R. 2014. Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res* **42**: 8024–8038.

Baejen C, Torkler P, Gressel S, Essig K, Söding J, Cramer P. 2014. Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition. *Mol Cell* **55**: 745–757.

Bagchi DN, Iyer VR. 2016. The Determinants of Directionality in Transcriptional Initiation. *Trends Genet* **32**: 322–333.

Bagga PS, Ford LP, Chen F, Wilusz J. 1995. The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. *Nucleic Acids Res* **23**: 1625–1631.

Bähler J, Wu J-Q, Longtine MS, Shah NG, Mckenzie III A, Steever AB, Wach A, Philippsen P, Pringle JR. 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast* **14**: 943–951.

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

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

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

Barillà D, Lee BA, Proudfoot NJ. 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **98**: 445–450.

Barnwal RP, Lee SD, Moore C, Varani G. 2012. Structural and biochemical analysis of the assembly and function of the yeast pre-mRNA 3' end processing complex CF I. *Proc Natl Acad Sci U S A* **109**: 21342–21347.

Bataille AR, Jeronimo C, Jacques P-É, Laramée L, Fortin M-È, Forest A, Bergeron M, Hanes SD, Robert F. 2012. A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. *Mol Cell* **45**: 158–170.

Becker R, Loll B, Meinhart A. 2008. Snapshots of the RNA Processing Factor SCAF8 Bound to Different Phosphorylated Forms of the Carboxyl-terminal Domain of RNA Polymerase II. *J Biol Chem* **283**: 22659–22669.

Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**: 163–175.

Bienkiewicz EA, Moon Woody A-Y, Woody RW. 2000. Conformation of the RNA polymerase II C-terminal domain: circular dichroism of long and short fragments1. *J Mol Biol* **297**: 119–133.

Bienroth S, Keller W, Wahle E. 1993. Assembly of a processive messenger RNA polyadenylation complex. *EMBO J* **12**: 585–594.

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

Bintu L, Ishibashi T, Dangkulwanich M, Wu Y-Y, Lubkowska L, Kashlev M, Bustamante C. 2012. Nucleosomal Elements that Control the Topography of the Barrier to Transcription. *Cell* **151**: 738–749.

Bird LE. 2011. High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. *Methods* **55**: 29–37.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Booth GT, Wang IX, Cheung VG, Lis JT. 2016. Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Res* **26**: 799–811.

Braithwaite D, Keegan K. 2014. *matR: Metagenomics Analysis Tools for R*. https://CRAN.R-project.org/package=matR.

Brannan K, Kim H, Erickson B, Glover-Cutter K, Kim S, Fong N, Kiemele L, Hansen K, Davis R, Lykke-Andersen J, et al. 2012. mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Mol Cell* **46**: 311–324.

Brodsky AS, Silver PA. 2000. Pre-mRNA processing factors are required for nuclear export. *RNA* **6**: 1737–1749.

Bühler M, Spies N, Bartel DP, Moazed D. 2008. TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. *Nat Struct Mol Biol* **15**: 1015–1023.

Cagas PM, Corden JL. 1995. Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II. *Proteins Struct Funct Bioinforma* **21**: 149–160.

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

Calvo O, Manley JL. 2005. The transcriptional coactivator PC4/Sub1 has multiple functions in RNA polymerase II transcription. *EMBO J* **24**: 1009–1020.

Carrillo Oesterreich F, Preibisch S, Neugebauer KM. 2010. Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. *Mol Cell* **40**: 571–581.

Carroll KL, Ghirlando R, Ames JM, Corden JL. 2007. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* **13**: 361–373.

Carroll KL, Pradhan DA, Granek JA, Clarke ND, Corden JL. 2004. Identification of cis Elements Directing Termination of Yeast Nonpolyadenylated snoRNA Transcripts. *Mol Cell Biol* **24**: 6241–6252.

Cassart C, Drogat J, Migeot V, Hermand D. 2012. Distinct requirement of RNA polymerase II CTD phosphorylations in budding and fission yeast. *Transcription* **3**: 231–234.

Castello A, Fischer B, Frese CK, Horos R, Alleaume A-M, Foehr S, Curk T, Krijgsveld J, Hentze MW. 2016. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**: 696–710.

Castelnuovo M, Zaugg JB, Guffanti E, Maffioletti A, Camblong J, Xu Z, Clauder-Münster S, Steinmetz LM, Luscombe NM, Stutz F. 2014. Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res* **42**: 4348–4362.

Chan S, Choi E-A, Shi Y. 2011. Pre-mRNA 3′-end processing complex assembly and function. *Wiley Interdiscip Rev RNA* **2**: 321–335.

Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR, Ule J, Manley JL, Shi Y. 2014. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. *Genes Dev* **28**: 2370–2380.

Chapman RD, Heidemann M, Hintermair C, Eick D. 2008. Molecular evolution of the RNA polymerase II CTD. *Trends Genet* **24**: 289–296.

Chathoth KT, Barrass JD, Webb S, Beggs JD. 2014. A Splicing-Dependent Transcriptional Checkpoint Associated with Prespliceosome Formation. *Mol Cell* **53**: 779–790.

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

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

Churchman LS, Weissman JS. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**: 368–373.

Cléry A, Blatter M, Allain FH-T. 2008. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **18**: 290–298.

Connelly S, Manley JL. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2**: 440–452.

Corden JL. 2013. RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template. *Chem Rev* **113**: 8423–8455.

Corden JL, Cadena DL, Ahearn JM, Dahmus ME. 1985. A Unique Structure at the Carboxyl Terminus of the Largest Subunit of Eukaryotic RNA Polymerase II. *Proc Natl Acad Sci U S A* **82**: 7934–7938.

Coy S, Volanakis A, Shah S, Vasiljeva L. 2013. The Sm complex is required for the processing of non-coding RNAs by the exosome. *PloS One* **8**: e65606.

Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ, Corden JL. 2011. Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet* **7**: e1002329.

Darby MM, Serebreni L, Pan X, Boeke JD, Corden JL. 2012. The Saccharomyces cerevisiae Nrd1-Nab3 transcription termination pathway acts in opposition to Ras signaling and mediates response to nutrient depletion. *Mol Cell Biol* **32**: 1762–1775.

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM. 2006. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A* **103**: 5320–5325.

Davidson L, Muniz L, West S. 2014. 3′ end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev* **28**: 342–356.

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, et al. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**: W375–W383.

Dengl S, Cramer P. 2009. Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* **284**: 21270–21279.

Deo RC, Bonanno JB, Sonenberg N, Burley SK. 1999. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**: 835–845.

Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* **43**: 853–866.

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

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

Egloff S, Dienstbier M, Murphy S. 2012. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet* **28**: 333–341.

Egloff S, O'Reilly D, Murphy S. 2008. Expression of human snRNA genes from beginning to end. *Biochem Soc Trans* **36**: 590–594.

Eick D, Geyer M. 2013. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* **113**: 8456–8490.

Elela SA, Igel H, Ares Jr. M. 1996. RNase III Cleaves Eukaryotic Preribosomal RNA at a U3 snoRNP-Dependent Site. *Cell* **85**: 115–124.

Emsley P, Cowtan K. 2004. Coot : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126–2132.

Enriquez-Harris P, Levitt N, Briggs D, Proudfoot NJ. 1991. A pause site for RNA polymerase II is associated with termination of transcription. *EMBO J* **10**: 1833–1842.

Eser P, Wachutka L, Maier KC, Demel C, Boroni M, Iyer S, Cramer P, Gagneur J. 2016. Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. *Mol Syst Biol* **12**: 857–857.

Ezeokonkwo C, Zhelkovsky A, Lee R, Bohm A, Moore CL. 2011. A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation. *RNA N Y N* **17**: 652–664.

Ezzeddine N, Chen J, Waltenspiel B, Burch B, Albrecht T, Zhuo M, Warren WD, Marzluff WF, Wagner EJ. 2011. A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation. *Mol Cell Biol* **31**: 328–341.

Fabrega C, Shen V, Shuman S, Lima CD. 2003. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Mol Cell* **11**: 1549–1561.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.

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

Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, Nguyen T, Karp S, Bentley DL. 2015. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell* **60**: 256–267.

Franke D, Svergun DI. 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**: 342–346.

Fusby B, Kim S, Erickson B, Kim H, Peterson ML, Bentley DL. 2016. Coordination of RNA Polymerase II Pausing and 3′ End Processing Factor Recruitment with Alternative Polyadenylation. *Mol Cell Biol* **36**: 295–303.

Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. 2014. Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. *Cell* **156**: 812–824.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: 1–16.

Ghazal G, Gagnon J, Jacques P-E, Landry J-R, Robert F, Elela SA. 2009. Yeast RNase III triggers polyadenylation-independent transcription termination. *Mol Cell* **36**: 99–109.

Glover-Cutter K, Kim S, Espinosa J, Bentley DL. 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol* **15**: 71–78.

Graber JH. 2003. Variations in yeast 3'-processing cis-elements correlate with transcript stability. *Trends Genet TIG* **19**: 473–476.

Graber JH, Cantor CR, Mohr SC, Smith TF. 1999. Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic Acids Res* **27**: 888–894.

Greger IH, Aranda A, Proudfoot N. 2000. Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. *Proc Natl Acad Sci* **97**: 8415–8420.

Gromak N, West S, Proudfoot NJ. 2006. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* **26**: 3986–3996.

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

Grzechnik P, Gdula MR, Proudfoot NJ. 2015. Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev* **29**: 849–861.

Grzechnik P, Kufel J. 2008. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol Cell* **32**: 247–258.

Gudipati RK, Villa T, Boulay J, Libri D. 2008. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol* **15**: 786–794.

Gudipati RK, Xu Z, Lebreton A, Séraphin B, Steinmetz LM, Jacquier A, Libri D. 2012. Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol Cell* **48**: 409–421.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* **130**: 77–88.

Guo J, Price DH. 2013. RNA Polymerase II Transcription Elongation Control. *Chem Rev* **113**: 8583–8603.

Guo Z, Sherman F. 1996. 3'-end-forming signals of yeast mRNA. *Trends Biochem Sci* **21**: 477–481.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-C, Munschauer M, et al. 2010. PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp JoVE* **41**: 2034.

Hajheidari M, Koncz C, Eick D. 2013. Emerging roles for RNA polymerase II CTD in Arabidopsis. *Trends Plant Sci* **18**: 633–643.

Hallais M, Pontvianne F, Andersen PR, Clerici M, Lener D, Benbahouche NEH, Gostan T, Vandermoere F, Robert M-C, Cusack S, et al. 2013. CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat Struct Mol Biol* **20**: 1358–1366.

Hamilton N. 2015. *smoother: Functions Relating to the Smoothing of Numerical Data*. https://CRAN.R-project.org/package=smoother.

Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, et al. 2012. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**: 768–779.

Hanes SD. 2014. The Ess1 prolyl isomerase: Traffic cop of the RNA polymerase II transcription cycle. *Biochim Biophys Acta BBA - Gene Regul Mech* **1839**: 316–333.

Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* **15**: 2147–2158.

Hazelbaker DZ, Marquardt S, Wlotzka W, Buratowski S. 2013. Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Mol Cell* **49**: 55–66.

He X, Khan AU, Cheng H, Pappas DL Jr, Hampsey M, Moore CL. 2003. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes Dev* **17**: 1030–1042.

He X, Moore C. 2005. Regulation of yeast mRNA 3' end processing by phosphorylation. *Mol Cell* **19**: 619–629.

Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR, Wilson SM, Minvielle-Sebastia L, Swanson MS. 2002. Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *EMBO J* **21**: 1800–1810.

Heidemann M, Hintermair C, Voß K, Eick D. 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta BBA - Gene Regul Mech* **1829**: 55–62.

Heidmann S, Obermaier B, Vogel K, Domdey H. 1992. Identification of pre-mRNA polyadenylation sites in Saccharomyces cerevisiae. *Mol Cell Biol* **12**: 4215–4229.

Heo D, Yoo I, Kong J, Lidschreiber M, Mayer A, Choi B-Y, Hahn Y, Cramer P, Buratowski S, 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. *J Biol Chem* **288**: 36676–36690.

Hobor F, Pergoli R, Kubicek K, Hrossova D, Bacikova V, Zimmermann M, Pasulka J, Hofr C, Vanacova S, Stefl R. 2011. Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem* **286**: 3645–3657.

Holbein S, Scola S, Loll B, Dichtl BS, Hübner W, Meinhart A, Dichtl B. 2011. The P-Loop Domain of Yeast Clp1 Mediates Interactions Between CF IA and CPF Factors in Pre-mRNA 3′ End Formation. *PLOS ONE* **6**: e29139.

Hollingworth D, Noble CG, Taylor IA, Ramos A. 2006. RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. *RNA N Y N* **12**: 555–560.

Holm L, Rosenström P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**: W545-549.

Hrossova D, Sikorsky T, Potesil D, Bartosovic M, Pasulka J, Zdrahal Z, Stefl R, Vanacova S. 2015. RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs. *Nucleic Acids Res* **43**: 4236–4248.

Hsin J-P, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**: 2119–2137.

Hu J, Lutz CS, Wilusz J, Tian B. 2005. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA N Y N* **11**: 1485–1493.

Huang Y, Acton T, Montelione G. 2014. DisMeta: A Meta Server for Construct Design and Optimization. In *Structural Genomics* (ed. Y.W. Chen), *Methods in Molecular Biology*, pp. 3–16, Humana Press.

Huang Y, Wimler KM, Carmichael GG. 1999. Intronless mRNA transport elements may affect multiple steps of pre-mRNA processing. *EMBO J* **18**: 1642–1652.

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12**: 115–121.

Humphrey T, Sadhale P, Platt T, Proudfoot N. 1991. Homologous mRNA 3' end formation in fission and budding yeast. *EMBO J* **10**: 3503–3511.

Huson DH, Bryant D. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Mol Biol Evol* **23**: 254–267.

Hutchins LN, Murphy SM, Singh P, Graber JH. 2008. Position-dependent motif characterization using non-negative matrix factorization. *Bioinforma Oxf Engl* **24**: 2684–2690.

Hyman AA, Weber CA, Jülicher F. 2014. Liquid-Liquid Phase Separation in Biology. *Annu Rev Cell Dev Biol* **30**: 39–58.

Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ, Corden JL. 2011. Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA N Y N* **17**: 2011–2025.

Jamros MA, Aubol BE, Keshwani MM, Zhang Z, Stamm S, Adams JA. 2015. Intra-domain Cross-talk Regulates Serine-arginine Protein Kinase 1-dependent Phosphorylation and Splicing Function of Transformer 2β1. *J Biol Chem* **290**: 17269–17281.

Järvelin AI, Noerenberg M, Davis I, Castello A. 2016. The new (dis)order in RNA regulation. *Cell Commun Signal CCS* **14**: 9.

Jasnovidova O, Stefl R. 2013. The CTD code of RNA polymerase II: a structural view. *Wiley Interdiscip Rev RNA* **4**: 1–16.

Jeronimo C, Bataille AR, Robert F. 2013. The Writers, Readers, and Functions of the RNA Polymerase II C-Terminal Domain Code. *Chem Rev* **113**: 8491–8522.

Jeronimo C, Collin P, Robert F. 2016. The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. *J Mol Biol* **428**: 2607–2622.

Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, et al. 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**: 753–767.

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

Keogh M-C, Buratowski S. 2004. Using Chromatin Immunoprecipitation to Map Cotranscriptional mRNA Processing in Saccharomyces cerevisiae. In *mRNA Processing and Metabolism* (ed. D. Schoenberg), *Methods in Molecular Biology™*, pp. 1–16, Humana Press.

Keogh M-C, Kim J-A, Downey M, Fillingham J, Chowdhury D, Harrison JC, Onishi M, Datta N, Galicia S, Emili A, et al. 2006a. A phosphatase complex that dephosphorylates γH2AX regulates DNA damage checkpoint recovery. *Nature* **439**: 497–501.

Keogh M-C, Mennella TA, Sawa C, Berthelet S, Krogan NJ, Wolek A, Podolny V, Carpenter LR, Greenblatt JF, Baetz K, et al. 2006b. The Saccharomyces cerevisiae histone H2A variant Htz1 is acetylated by NuA4. *Genes Dev* **20**: 660–665.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: 1–13.

Kim K-Y, Levin DE. 2011. Mpk1 MAPK Association with the Paf1 Complex Blocks Sen1-Mediated Premature Transcription Termination. *Cell* **144**: 745–756.

Kim M, Ahn S-H, Krogan NJ, Greenblatt JF, Buratowski S. 2004a. Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. *EMBO J* **23**: 354–364.

Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S. 2004b. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**: 517–522.

Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S. 2006. Distinct pathways for snoRNA and mRNA termination. *Mol Cell* **24**: 723–734.

Kozin MB, Svergun DI. 2001. Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34**: 33–41.

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

Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, Hofr C, Vanacova S, Stefl R. 2012. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* **26**: 1891–1896.

Kuehner JN, Pearson EL, Moore C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* **12**: 283–294.

Kwak H, Lis JT. 2013. Control of Transcriptional Elongation. *Annu Rev Genet* **47**: 483–508.

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. 2005. RNA Degradation by the Exosome Is Promoted by a Nuclear Polyadenylation Complex. *Cell* **121**: 713–724.

Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi E-A, Xie X, Wan J, Xing Y, Freudenberg JM, et al. 2014. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J* **33**: 878–889.

Laemmli UK. 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**: 680–685.

Laitem C, Zaborowska J, Isa NF, Kufs J, Dienstbier M, Murphy S. 2015. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II–transcribed genes. *Nat Struct Mol Biol* **22**: 396–403.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lawrence M, Gentleman R, Carey V. 2009. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**: 1841–1842.

Lee NN, Chalamcharla VR, Reyes-Turcu F, Mehta S, Zofall M, Balachandran V, Dhakshnamoorthy J, Taneja N, Yamanaka S, Zhou M, et al. 2013. Mtr4-like Protein Coordinates Nuclear RNA Processing for Heterochromatin Assembly and for Telomere Maintenance. *Cell* **155**: 1061–1074.

Lemay J-F, Larochelle M, Marguerat S, Atkinson S, Bähler J, Bachand F. 2014. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat Struct Mol Biol* **21**: 919–926.

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 Dev* **30**: 1558–1572.

Lemieux C, Bachand F. 2009. Cotranscriptional recruitment of the nuclear poly(A)-binding protein Pab2 to nascent transcripts and association with translating mRNPs. *Nucleic Acids Res* **37**: 3418–3430.

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

Levine M. 2011. Paused RNA Polymerase II as a Developmental Checkpoint. *Cell* **145**: 502–511.

Libri D, Dower K, Boulay J, Thomsen R, Rosbash M, Jensen TH. 2002. Interactions between mRNA export commitment, 3’-end quality control, and nuclear degradation. *Mol Cell Biol* **22**: 8254–8266.

Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL. 2002. Functional interaction of yeast pre-mRNA 3’ end processing factors with RNA polymerase II. *Mol Cell* **9**: 1101–1111.

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

Loya TJ, Reines D. 2016. Recent advances in understanding transcription termination by RNA polymerase II. *F1000Research* **5**: 1478.

Lubas M, Andersen PR, Schein A, Dziembowski A, Kudla G, Jensen TH. 2015. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* **10**: 178–192.

Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falkenby LG, Lykke-Andersen S, Andersen JS, Dziembowski A, Jensen TH. 2011. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell* **43**: 624–637.

Luehr S, Hartmann H, Söding J. 2012. The XXmotif web server for eXhaustive, weight matriX-based motif discovery in nucleotide sequences. *Nucleic Acids Res* **40**: W104-109.

Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **17**: 1195–1201.

Luo W, Johnson AW, Bentley DL. 2006. The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric–torpedo model. *Genes Dev* **20**: 954–965.

Lykke-Andersen S, Jensen TH. 2007. Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* **89**: 1177–1182.

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

Mandel CR, Bai Y, Tong L. 2008. Protein factors in pre-mRNA 3′-end processing. *Cell Mol Life Sci* **65**: 1099–1122.

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**: 953–956.

Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. 2006. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci* **15**: 182–189.

Marina DB, Shankar S, Natarajan P, Finn KJ, Madhani HD. 2013. A conserved ncRNA-binding protein recruits silencing factors to heterochromatin through an RNAi-independent mechanism. *Genes Dev* **27**: 1851–1856.

Maris C, Dominguez C, Allain FH-T. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272**: 2118–2131.

Marquardt S, Hazelbaker DZ, Buratowski S. 2011. Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription* **2**: 145–154.

Martin G, Gruber AR, Keller W, Zavolan M. 2012. Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep* **1**: 753–763.

Martin-Tumasz S, Brow DA. 2015. Saccharomyces cerevisiae Sen1 Helicase Domain Exhibits 5'- to 3'-Helicase Activity with a Preference for Translocation on DNA Rather than RNA. *J Biol Chem* **290**: 22880–22889.

Mason PB, Struhl K. 2005. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell* **17**: 831–840.

Mata J. 2013. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol* **10**: 1407–1414.

Matsuyama A, Shirai A, Yashiroda Y, Kamata A, Horinouchi S, Yoshida M. 2004. pDUAL, a multipurpose, multicopy vector capable of chromosomal integration in fission yeast. *Yeast* **21**: 1289–1305.

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

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* **40**: 658–674.

McDevitt MA, Hart RP, Wong WW, Nevins JR. 1986. Sequences capable of restoring poly(A) site function define two distinct downstream elements. *EMBO J* **5**: 2907–2913.

McNeil JB, Agah H, Bentley D. 1998. Activated transcription independent of the RNA polymerase II holoenzyme in budding yeast. *Genes Dev* **12**: 2510–2521.

Mei K, Jin Z, Ren F, Wang Y, Chang Z, Wang X. 2014. Structural basis for the recognition of RNA polymerase II C-terminal domain by CREPT and p15RS. *Sci China Life Sci* **57**: 97–106.

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

Milligan L, Huynh-Thu VA, Delan-Forino C, Tuck A, Petfalski E, Lombraña R, Sanguinetti G, Kudla G, Tollervey D. 2016. Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol Syst Biol* **12**: 874.

Milligan L, Torchet C, Allmang C, Shipman T, Tollervey D. 2005. A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Mol Cell Biol* **25**: 9996–10004.

Minvielle-Sebastia L, Beyer K, Krecic AM, Hector RE, Swanson MS, Keller W. 1998. Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. *EMBO J* **17**: 7454–7468.

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

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

Mischo HE, Proudfoot NJ. 2013. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta* **1829**: 174–185.

Mitsuzawa H, Kanda E, Ishihama A. 2003. Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res* **31**: 4696–4701.

Morales JC, Richard P, Rommel A, Fattah FJ, Motea EA, Patidar PL, Xiao L, Leskov K, Wu S-Y, Hittelman WN, et al. 2014. Kub5-Hera, the human Rtt103 homolog, plays dual functional roles in transcription termination and DNA repair. *Nucleic Acids Res* **42**: 4996–5006.

Moreira A, Wollerton M, Monks J, Proudfoot NJ. 1995. Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *EMBO J* **14**: 3809–3819.

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

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

Murthy KG, Manley JL. 1992. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J Biol Chem* **267**: 14804–14811.

Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet* **39**: 1507–1511.

Nabavi S, Nazar RN. 2010a. Cleavage-induced termination in U2 snRNA gene expression. *Biochem Biophys Res Commun* **393**: 461–465.

Nabavi S, Nazar RN. 2008. Nonpolyadenylated RNA Polymerase II Termination Is Induced by Transcript Cleavage. *J Biol Chem* **283**: 13601–13610.

Nabavi S, Nazar RN. 2010b. Pac1 endonuclease and Dhp1p 5′ → 3′ exonuclease are required for U3 snoRNA termination in Schizosaccharomyces pombe. *FEBS Lett* **584**: 3436–3441.

Nag A, Narsinh K, Martinson HG. 2007. The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nat Struct Mol Biol* **14**: 662–669.

Natarajan M, Lester GMS, Lee C, Missra A, Wasserman GA, Steffen M, Gilmour DS, Henderson AJ. 2013. Negative Elongation Factor (NELF) Coordinates RNA Polymerase II Pausing, Premature Termination, and Chromatin Remodeling to Regulate HIV Transcription. *J Biol Chem* **288**: 25995–26003.

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

Ni Z, Xu C, Guo X, Hunter GO, Kuznetsova OV, Tempel W, Marcon E, Zhong G, Guo H, Kuo W-HW, et al. 2014. RPRD1A and RPRD1B Are Human RNA Polymerase II C-Terminal Domain Scaffolds for Ser5 Dephosphorylation. *Nat Struct Mol Biol* **21**: 686–695.

Nicol JW, Helt GA, Blanchard SG, Raja A, Loraine AE. 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25**: 2730–2731.

Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A. 2005. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* **12**: 144–151.

Noble CG, Walker PA, Calder LJ, Taylor IA. 2004. Rna14–Rna15 assembly mediates the RNA-binding capability of Saccharomyces cerevisiae cleavage factor IA. *Nucleic Acids Res* **32**: 3364–3375.

Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**: 526–540.

Ntini E, Järvelin AI, Bornholdt J, Chen Y, Boyd M, Jørgensen M, Andersson R, Hoof I, Schein A, Andersen PR, et al. 2013. Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* **20**: 923–928.

O'Reilly D, Kuznetsova OV, Laitem C, Zaborowska J, Dienstbier M, Murphy S. 2014. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res* **42**: 264–275.

O'Rourke TW, Loya TJ, Head PE, Horton JR, Reines D. 2015. Amyloid-like assembly of the low complexity domain of yeast Nab3. *Prion* **9**: 34–47.

Pages H. *BSgenome: Infrastructure for Biostrings-based genome data packages and support for efficient SNP representation*.

Pages H, Aboyoun P, Gentleman R, DebRoy S. *Biostrings: String objects representing biological sequences, and matching algorithms*.

Pauws E, Kampen AHC van, Graaf SAR van de, Vijlder JJM de, Ris-Stalpers C. 2001. Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Res* **29**: 1690–1694.

Pei J, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* **36**: 2295–2300.

Pei Y, Schwer B, Shuman S. 2003. Interactions between Fission Yeast Cdk9, Its Cyclin Partner Pch1, and mRNA Capping Enzyme Pct1 Suggest an Elongation Checkpoint for mRNA Quality Control. *J Biol Chem* **278**: 7180–7188.

Pérez-Cañadillas JM. 2006. Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. *EMBO J* **25**: 3167–3178.

Peterson ML, Bertolino S, Davis F. 2002. An RNA polymerase pause site is associated with the immunoglobulin mus poly(A) site. *Mol Cell Biol* **22**: 5606–5615.

Pinto PAB, Henriques T, Freitas MO, Martins T, Domingues RG, Wyrzykowska PS, Coelho PA, Carmo AM, Sunkel CE, Proudfoot NJ, et al. 2011. RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J* **30**: 2431–2444.

Porrua O, Hobor F, Boulay J, Kubicek K, D'Aubenton-Carafa Y, Gudipati RK, Stefl R, Libri D. 2012. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J* **31**: 3935–3948.

Porrua O, Libri D. 2013. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol* **20**: 884–891.

Porrua O, Libri D. 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16**: 190–202.

Potashkin J, Frendewey D. 1990. A mutation in a single gene of Schizosaccharomyces pombe affects the expression of several snRNAs and causes defects in RNA processing. *EMBO J* **9**: 525–534.

Preker P, Almvig K, Christensen MS, Valen E, Mapendano CK, Sandelin A, Jensen TH. 2011. PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res* **39**: 7179–7193.

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

Prescott EM, Proudfoot NJ. 2002. Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci* **99**: 8796–8801.

Proudfoot NJ. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14**: 105–110.

Proudfoot NJ, Brownlee GG. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**: 211–214.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Séraphin B. 2001. The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification. *Methods* **24**: 218–229.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187-191.

Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**: 1231–1245.

Receveur-Brechot V, Durand D. 2012. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci* **13**: 55–75.

Reinstein J, Vetter IR, Schlichting I, Rösch P, Wittinghofer A, Goody RS. 1990. Fluorescence and NMR investigations on the ligand binding properties of adenylate kinases. *Biochemistry (Mosc)* **29**: 7440–7450.

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**: 651–657.

Rodríguez-Molina JB, Tseng SC, Simonett SP, Taunton J, Ansari AZ. 2016. Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. *Mol Cell* **63**: 433–444.

Roguev A, Shevchenko A, Schaft D, Thomas H, Stewart AF, Shevchenko A. 2004. A Comparative Analysis of an Orthologous Proteomic Environment in the Yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Mol Cell Proteomics* **3**: 125–132.

Rondón AG, Mischo HE, Kawauchi J, Proudfoot NJ. 2009. Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. *Mol Cell* **36**: 88–98.

Rotondo G, Frendewey D. 2001. Pac1 ribonuclease of Schizosaccharomyces pombe. *Methods Enzymol* **342**: 168–193.

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

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

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

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

Saldi T, Cortazar MA, Sheridan RM, Bentley DL. 2016. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol* **428**: 2623–2635.

Schäffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* **29**: 2994–3005.

Schlackow M, Marguerat S, Proudfoot NJ, Bähler J, Erban R, Gullerova M. 2013. Genome-wide analysis of poly(A) site selection in Schizosaccharomyces pombe. *RNA N Y N* **19**: 1617–1631.

Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. 2016. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* **44**: W424–W429.

Schönemann L, Kühn U, Martin G, Schäfer P, Gruber AR, Keller W, Zavolan M, Wahle E. 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev* **28**: 2381–2393.

Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, Passmore LA. 2014. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol* **21**: 175–179.

Schrödinger L. 2015. *The PyMOL Molecular Graphics System*.

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

Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J, Cramer P. 2013. Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell* **155**: 1075–1087.

Schwer B, Shuman S. 2011. Deciphering the RNA Polymerase II CTD Code in Fission Yeast. *Mol Cell* **43**: 311–318.

Shah S, Wittmann S, Kilchert C, Vasiljeva L. 2014. lncRNA recruits RNAi and the exosome to dynamically regulate pho1 expression in response to phosphate levels in fission yeast. *Genes Dev* **28**: 231–244.

Shatkin AJ. 1976. Capping of eucaryotic mRNAs. *Cell* **9**: 645–653.

Sheets MD, Ogg SC, Wickens MP. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* **18**: 5799–5805.

Sheldrick GM. 2010. Experimental phasing with *SHELXC* / *D* / *E* : combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**: 479–485.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates III JR, Frank J, Manley JL. 2009. Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. *Mol Cell* **33**: 365–376.

Shi Y, Manley JL. 2015. The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site. *Genes Dev* **29**: 889–897.

Sikorski TW, Buratowski S. 2009. The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* **21**: 344–351.

Singh N, Ma Z, Gemmill T, Wu X, Defiglio H, Rossettini A, Rabeler C, Beane O, Morse RH, Palumbo MJ, et al. 2009. The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. *Mol Cell* **36**: 255–266.

Sipiczki M. 1995. Phylogenesis of fission yeasts. Contradictions surrounding the origin of a century old genus. *Antonie Van Leeuwenhoek* **68**: 119–149.

Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42**: 794–805.

Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, Bricogne G. 2012. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in *BUSTER*. *Acta Crystallogr D Biol Crystallogr* **68**: 368–380.

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

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

Steinmetz EJ, Conrad NK, Brow DA, Corden JL. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. *Nature* **413**: 327–331.

Stiller JW, Hall BD. 2002. Evolution of the RNA polymerase II C-terminal domain. *Proc Natl Acad Sci* **99**: 6091–6096.

St-Pierre B, Liu X, Kha L-CT, Zhu X, Ryan O, Jiang Z, Zacksenhaus E. 2005. Conserved and specific functions of mammalian ssu72. *Nucleic Acids Res* **33**: 464–477.

Sugano S, Shobuike T, Takeda T, Sugino A, Ikeda H. 1994. Molecular analysis of the dhp1+ gene of Schizosaccharomyces pombe: an essential gene that has homology to the DST2 and RAT1 genes of Saccharomyces cerevisiae. *Mol Gen Genet MGG* **243**: 1–8.

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

Svergun DI. 1999. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys J* **76**: 2879–2886.

Takagaki Y, Manley JL. 1997. RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol* **17**: 3907–3914.

Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. 2006. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell* **23**: 853–864.

Tian B, Graber JH. 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* **3**: 385–396.

Tian B, Hu J, Zhang H, Lutz CS. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* **33**: 201–212.

Torchet C, Bousquet-Antonelli C, Milligan L, Thompson E, Kufel J, Tollervey D. 2002. Processing of 3′-Extended Read-Through Transcripts by the Exosome Can Generate Functional mRNAs. *Mol Cell* **9**: 1285–1296.

Tuck AC, Tollervey D. 2013. A Transcriptome-wide Atlas of RNP Composition Reveals Diverse Classes of mRNAs and lncRNAs. *Cell* **154**: 996–1009.

Tudek A, Candelli T, Libri D. 2015. Non-coding transcription by RNA polymerase II in yeast: Hasard or nécessité? *Biochimie* **117**: 28–36.

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. *Mol Cell* **55**: 467–481.

Ursic D, Chinchilla K, Finkel JS, Culbertson MR. 2004. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res* **32**: 2441–2452.

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

Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. 2008a. The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **15**: 795–804.

Vasiljeva L, Kim M, Terzi N, Soares LM, Buratowski S. 2008b. Transcription Termination and RNA Degradation Contribute to Silencing of RNA Polymerase II Transcription within Heterochromatin. *Mol Cell* **29**: 313–323.

Verghese J, Abrams J, Wang Y, Morano KA. 2012. Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (Saccharomyces cerevisiae) as a Model System. *Microbiol Mol Biol Rev* **76**: 115–158.

Volanakis A, Passoni M, Hector RD, Shah S, Kilchert C, Granneman S, Vasiljeva L. 2013. Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast. *Genes Dev* **27**: 2025–2038.

Volkov VV, Svergun DI. 2003. Uniqueness of *ab initio* shape determination in small-angle scattering. *J Appl Crystallogr* **36**: 860–864.

Walter TS, Diprose JM, Mayo CJ, Siebold C, Pickford MG, Carter L, Sutton GC, Berrow NS, Brown J, Berry IM, et al. 2005. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* **61**: 651–657.

Wang H, Zeng F, Liu Q, Liu H, Liu Z, Niu L, Teng M, Li X. 2013. The structure of the ARE-binding domains of Hu antigen R (HuR) undergoes conformational changes during RNA binding. *Acta Crystallogr D Biol Crystallogr* **69**: 373–380.

Watts B. 2015. Investigating mechanisms of transcriptional interference in Schizosaccharomyces pombe. DPhil dissertation, University of Oxford, Linacre College.

Webb S, Hector RD, Kudla G, Granneman S. 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* **15**: R8.

Weber SC, Brangwynne CP. 2012. Getting RNA and Protein in Phase. *Cell* **149**: 1188–1191.

Werner-Allen JW, Lee C-J, Liu P, Nicely NI, Wang S, Greenleaf AL, Zhou P. 2011. cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *J Biol Chem* **286**: 5717–5726.

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

Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, NY http://ggplot2.org (Accessed July 12, 2016).

Wilson MD, Harreman M, Svejstrup JQ. 2013. Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochim Biophys Acta* **1829**: 151–157.

Winter G, Lobley CMC, Prince SM. 2013. Decision making in *xia* 2. *Acta Crystallogr D Biol Crystallogr* **69**: 1260–1273.

Wood V, Harris MA, McDowall MD, Rutherford K, Vaughan BW, Staines DM, Aslett M, Lock A, Bahler J, Kersey PJ, et al. 2011. PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res* **40**: D695–D699.

Wyers F, Rougemaille M, Badis G, Rousselle J-C, Dufour M-E, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, et al. 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**: 725–737.

Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, Tong L. 2010. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**: 729–733.

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**: 1033–1037.

Yamamoto J, Hagiwara Y, Chiba K, Isobe T, Narita T, Handa H, Yamaguchi Y. 2014. DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat Commun* **5**: 4263.

Yanagida M. 2002. The model unicellular eukaryote, Schizosaccharomyces pombe. *Genome Biol* **3**: COMMENT2003.

Yang Q, Coseno M, Gilmartin GM, Doublié S. 2011. Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. *Struct Lond Engl 1993* **19**: 368–377.

Yang Q, Gilmartin GM, Doublié S. 2010. Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc Natl Acad Sci U S A* **107**: 10062–10067.

Yao C, Biesinger J, Wan J, Weng L, Xing Y, Xie X, Shi Y. 2012. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci U S A* **109**: 18773–18778.

Yonaha M, Proudfoot NJ. 1999. Specific Transcriptional Pausing Activates Polyadenylation in a Coupled In Vitro System. *Mol Cell* **3**: 593–600.

Yonezawa Y. 2014. Molecular Dynamics Study of the Phosphorylation Effect on the Conformational States of the C-Terminal Domain of RNA Polymerase II. *J Phys Chem B* **118**: 4471–4478.

Zaborowska J, Egloff S, Murphy S. 2016. The pol II CTD: new twists in the tail. *Nat Struct Mol Biol* **23**: 771–777.

Zamft B, Bintu L, Ishibashi T, Bustamante C. 2012. Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc Natl Acad Sci U S A* **109**: 8948–8953.

Zeitlinger J, Stark A, Kellis M, Hong J-W, Nechaev S, Adelman K, Levine M, Young RA. 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet* **39**: 1512–1516.

Zhang DW, Mosley AL, Ramisetty SR, Rodríguez-Molina JB, Washburn MP, Ansari AZ. 2012. Ssu72 Phosphatase-dependent Erasure of Phospho-Ser7 Marks on the RNA Polymerase II C-terminal Domain Is Essential for Viability and Transcription Termination. *J Biol Chem* **287**: 8541–8551.

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

Zhang Y, Kim Y, Genoud N, Gao J, Kelly JW, Pfaff SL, Gill GN, Dixon JE, Noel JP. 2006. Determinants for Dephosphorylation of the RNA Polymerase II C-Terminal Domain by Scp1. *Mol Cell* **24**: 759–770.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Zhang Z, Fu J, Gilmour DS. 2005. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. *Genes Dev* **19**: 1572–1580.

Zhang Z, Klatt A, Henderson AJ, Gilmour DS. 2007. Transcription termination factor Pcf11 limits the processivity of Pol II on an HIV provirus to repress gene expression. *Genes Dev* **21**: 1609–1614.

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

Zhou D, Frendewey D, Lobo Ruppert SM. 1999. Pac1p, an RNase III homolog, is required for formation of the 3' end of U2 snRNA in Schizosaccharomyces pombe. *RNA N Y N* **5**: 1083–1098.

Zhou Y, Zhu J, Schermann G, Ohle C, Bendrin K, Sugioka-Sugiyama R, Sugiyama T, Fischer T. 2015. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun* **6**: 7050.

Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**: 182–185.

APPENDIX I: LIST OF PROTEINS IDENTIFIED BY MS

Table 10: List of Mascot hits after mass spectrometric analysis of purified proteins
mw: molecular weight; WT: wild-type control; replicates: Ysh1-HA-TAP purification; OS: organism; GN: gene name; PE: protein existence; SV: sequence version

Protein	Description	mw (Da)	Spectral Counts		
			WT	replicate 1	replicate 2
P00761	SWISS_PROT_P00761 TRYP_PIG Trypsin _ Sus scrofa (Pig).	24429.80	3.00	15.00	18.00
sp|P79013| RS11_SCHPO	40S ribosomal protein S11 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps11a PE=2 SV=2	17507.40		6.00	1.00
sp|O74538| RL27B_SCHPO	60S ribosomal protein L27-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl27b PE=2 SV=1	15313.50	2.00	7.00	8.00
sp|P79057| RS23_SCHPO	40S ribosomal protein S23 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps23a PE=2 SV=2	15663.60		6.00	3.00
P35527	SWISS_PROT_P35527 Tax_Id=9606 Gene_Symbol=KRT9 Keratin, type I cytoskeletal 9	62200.00	17.00	37.00	31.00
sp|P79063| ERF1_SCHPO	Eukaryotic peptide chain release factor subunit 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sup45 PE=1 SV=1	48844.40		4.00	1.00
sp|O74548| ARGD_SCHPO	Probable acetylornithine aminotransferase, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=arg1 PE=2 SV=1	47551.90		3.00	3.00
sp|O14055| SYYC_SCHPO	Tyrosine--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tys1 PE=2 SV=1	44619.20		7.00	4.00
sp|P79071| RL6_SCHPO	60S ribosomal protein L6 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl6 PE=1 SV=2	21252.60	7.00	32.00	18.00
sp|O74653| POB1_SCHPO	Protein pob1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pob1 PE=1 SV=1	93860.20		2.00	1.00
sp|Q09826| SDS23_SCHPO	Protein sds23/moc1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sds23 PE=1 SV=1	44332.90		3.00	2.00
sp|O94396| PUS1_SCHPO	tRNA pseudouridine synthase 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pus1 PE=1 SV=1	60376.40		2.00	2.00
P60712	SWISS_PROT_P60712 (Bos taurus) Actin, cytoplasmic 1	41763.80	1.00	4.00	5.00
sp|O94415| SUCB_SCHPO	Probable succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC1620.08 PE=2 SV=1	47329.50		5.00	
sp|Q09845| GPD2_SCHPO	Glycerol-3-phosphate dehydrogenase [NAD(+)] 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=gpd2 PE=1 SV=1	40902.00		10.00	3.00
sp|O74718| ERF3_SCHPO	Eukaryotic peptide chain release factor GTP-binding subunit OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sup35 PE=1 SV=2	72619.00		14.00	9.00
sp|O14069| RL28E_SCHPO	Probable 60S ribosomal protein L28e OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl28e PE=1 SV=1	14764.00	2.00	22.00	8.00
sp|O94444| RPN10_SCHPO	26S proteasome regulatory subunit rpn10 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpn10 PE=1 SV=1	27158.70		2.00	1.00
sp|Q09847| NU146_SCHPO	Nucleoporin nup146 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=nup146 PE=1 SV=1	145921.00		3.00	2.00
sp|P79083| EIF3I_SCHPO	Eukaryotic translation initiation factor 3 subunit I OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sum1 PE=1 SV=1	36842.30		14.00	9.00
sp|O42870| SYFA_SCHPO	Phenylalanine--tRNA ligase alpha subunit OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=frs2 PE=2 SV=3	56302.90		3.00	1.00
Q86YZ3	SWISS_PROT_Q86YZ3 Tax_Id=9606 Gene_Symbol=HRNR Hornerin	282732.00	2.00		2.00
sp|P15567| PUR6_SCHPO	Phosphoribosylaminoimidazole carboxylase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ade6 PE=2 SV=1	60064.90		3.00	2.00
sp|O14075| YEAA_SCHPO	Putative 2-hydroxyacid dehydrogenase UNK4.10 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC2E11.10 PE=2 SV=1	36731.90		2.00	1.00

sp|O42893| TREA_SCHPO	Neutral trehalase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ntp1 PE=1 SV=1	84676.60		3.00	
sp|Q9UUI4| YIY9_SCHPO	Uncharacterized protein C22F8.09 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC22F8.09 PE=1 SV=1	48172.70		3.00	1.00
sp|Q10425| EIF3B_SCHPO	Eukaryotic translation initiation factor 3 subunit B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC25G10.08 PE=1 SV=1	84108.70	1.00	22.00	18.00
sp|O14081| TPSX_SCHPO	Putative alpha,alpha-trehalose-phosphate synthase [UDP-forming] 106 kDa subunit OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC2E11.16c PE=1 SV=1	106913.00		3.00	2.00
sp|P17610| YPT3_SCHPO	GTP-binding protein ypt3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ypt3 PE=1 SV=1	23882.00		1.00	2.00
sp|Q09864| HSP60_SCHPO	Heat shock protein 60, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=hsp60 PE=1 SV=1	62218.80	1.00	40.00	22.00
sp|O74733| CFT1_SCHPO	Protein cft1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cft1 PE=2 SV=1	160452.00		28.00	27.00
sp|O94489| EF3_SCHPO	Elongation factor 3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tef3 PE=1 SV=1	115909.00		63.00	38.00
sp|O13370| DED1_SCHPO	ATP-dependent RNA helicase ded1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ded1 PE=1 SV=1	69823.90	1.00	27.50	22.50
sp|O60063| YBB1_SCHPO	Putative 37S ribosomal protein S4-like OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC13G1.01c PE=2 SV=1	37695.70			5.00
sp|O42923| CIP1_SCHPO	RNA-binding post-transcriptional regulator cip1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cip1 PE=1 SV=1	53517.00		10.00	11.00
sp|Q9UUI5| SEC24_SCHPO	Protein transport protein sec24 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec24 PE=2 SV=1	102002.00		4.00	4.00
sp|P87048| RPN1_SCHPO	26S proteasome regulatory subunit rpn1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpn1 PE=1 SV=2	98080.90		7.00	5.00
sp|Q9URZ5| VPS1_SCHPO	Vacuolar protein sorting-associated protein 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=vps1 PE=2 SV=2	75851.80		6.00	2.00
sp|P17937| RL7A_SCHPO	60S ribosomal protein L7-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl7a PE=2 SV=1	29183.00	1.00	8.00	2.00
sp|O14104| LYS12_SCHPO	Probable homoisocitrate dehydrogenase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=lys12 PE=1 SV=1	39285.50		9.00	8.00
sp|P39964| CEF1_SCHPO	Pre-mRNA-splicing factor cdc5 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cdc5 PE=1 SV=1	86917.10		1.00	2.00
sp|P40370| ENO11_SCHPO	Enolase 1-1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=eno101 PE=1 SV=2	47478.50	1.00	2.00	3.00
sp|O94438| RS3A2_SCHPO	40S ribosomal protein S1-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps102 PE=1 SV=3	28539.60		9.00	4.00
sp|Q9UUM7| HOB3_SCHPO	Protein hob3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=hob3 PE=1 SV=1	30111.20		3.00	2.00
sp|O60064| YBB2_SCHPO	Probable mannose-1-phosphate guanyltransferase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC13G1.02 PE=2 SV=1	46138.70		8.00	5.00
sp|Q9US13| RPN9_SCHPO	Probable 26S proteasome regulatory subunit rpn9 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpn9 PE=2 SV=1	43483.60		4.00	1.00
sp|P18869| LEU3_SCHPO	3-isopropylmalate dehydrogenase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=leu1 PE=1 SV=1	39762.00		4.00	3.00
sp|P40387| TPS1_SCHPO	Alpha,alpha-trehalose-phosphate synthase [UDP-forming] OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tps1 PE=1 SV=2	58546.70		3.00	2.00
sp|P87140| COPG_SCHPO	Probable coatomer subunit gamma OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec21 PE=1 SV=1	101035.00		9.00	3.00
sp|O42931| PRS7_SCHPO	26S protease regulatory subunit 7 homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpt1 PE=1 SV=2	49006.10		14.00	9.00
sp|Q9Y700| EFTU_SCHPO	Elongation factor Tu, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tuf1 PE=2 SV=1	48323.80		2.00	1.00

sp|P21565| SRP54_SCHPO	Signal recognition particle 54 kDa protein homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=srp54 PE=1 SV=1	56751.10		3.00	1.00
sp|Q9UTD7| RRP15_SCHPO	Ribosomal RNA-processing protein 15 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rrp15 PE=2 SV=1	23434.00		3.00	
sp|Q9US43| UPP3_SCHPO	Putative uracil phosphoribosyltransferase urg2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=urg2 PE=2 SV=2	22105.60		7.00	5.00
sp|O60068| ISD11_SCHPO	Protein isd11 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=isd11 PE=2 SV=1	12014.30		2.00	1.00
sp|O14124| DPEH1_SCHPO	Uncharacterized dipeptidase C3A11.10c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC3A11.10c PE=1 SV=1	46090.60	1.00	8.00	7.00
sp|P40921| EF1G_SCHPO	Elongation factor 1-gamma OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tef3 PE=1 SV=1	45829.50		3.00	1.00
sp|Q9UU99| YJX4_SCHPO	CRAL-TRIO domain-containing protein C23B6.04c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC23B6.04c PE=1 SV=1	110838.00		1.00	1.00
sp|Q9Y7L8| RSSA1_SCHPO	40S ribosomal protein S0-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps0a PE=1 SV=1	31776.20		3.04	2.52
sp|O13622| MS116_SCHPO	ATP-dependent RNA helicase mss116, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=mss116 PE=2 SV=1	59502.40		2.00	
sp|P87142| YDMC_SCHPO	Heat shock protein 70 homolog C57A7.12 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC57A7.12 PE=1 SV=3	61338.40		2.00	
sp|P22068| ATPB_SCHPO	ATP synthase subunit beta, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=atp2 PE=1 SV=1	56930.80		10.00	4.00
sp|O60084| TIM44_SCHPO	Mitochondrial import inner membrane translocase subunit tim44 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tim44 PE=2 SV=1	48591.80		15.00	5.00
sp|P40976| LYS2_SCHPO	L-aminoadipate-semialdehyde dehydrogenase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=lys1 PE=1 SV=3	157049.00		4.00	6.00
sp|P05219| TBB_SCHPO	Tubulin beta chain OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=nda3 PE=1 SV=2	49512.10	1.00	8.00	8.00
sp|Q9USL7| RSM18_SCHPO	37S ribosomal protein rsm18, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rsm18 PE=2 SV=1	19036.80		2.00	1.00
sp|O13637| SEC31_SCHPO	Protein transport protein sec31 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec31 PE=1 SV=1	132633.00		11.00	5.00
sp|Q10475| IF4G_SCHPO	Eukaryotic translation initiation factor 4 gamma OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif471 PE=1 SV=1	154194.00		19.00	32.00
sp|O59794| RL17B_SCHPO	60S ribosomal protein L17-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl17b PE=2 SV=1	20832.10		3.00	4.00
sp|O74740| CFT2_SCHPO	Cleavage factor two protein 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cft2 PE=1 SV=1	89520.80		19.00	21.00
sp|O42937| COPB2_SCHPO	Probable coatomer subunit beta' OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec27 PE=2 SV=2	89872.50		3.00	
sp|Q9USP0| EI2BA_SCHPO	Translation initiation factor eIF-2B subunit alpha OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif221 PE=2 SV=1	37803.40		4.00	
sp|O14011| CWF8_SCHPO	Pre-mRNA-splicing factor cwf8 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cwf8 PE=1 SV=1	54229.70		1.00	2.00
sp|Q9Y7Z8| MYO1_SCHPO	Myosin-1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=myo1 PE=1 SV=1	135842.00		6.00	3.00
sp|O14126| PRS6A_SCHPO	26S protease regulatory subunit 6A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tbp1 PE=2 SV=1	48877.10		7.00	5.00
sp|Q9USG6| RL33B_SCHPO	60S ribosomal protein L33-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl33b PE=1 SV=1	12197.50		3.00	
sp|Q9UT08| 2AAA_SCHPO	Protein phosphatase PP2A regulatory subunit A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=paa1 PE=2 SV=1	66664.80		3.00	
sp|P41878| RPN11_SCHPO	26S proteasome regulatory subunit rpn11 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpn11 PE=1 SV=2	34603.40	1.00	3.00	3.00

sp|Q09868| CIP2_SCHPO	RNA-binding post-transcriptional regulator cip2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cip2 PE=1 SV=1	62121.10		7.00	5.00
sp|O13639| SAHH_SCHPO	Adenosylhomocysteinase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pi047 PE=2 SV=1	47425.20		6.00	6.00
sp|P87144| SYTC_SCHPO	Threonine--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ths1 PE=2 SV=1	80213.40		20.00	8.00
sp|O60167| YHC3_SCHPO	ENTH domain-containing protein C19F8.03c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC19F8.03c PE=1 SV=1	73048.00		2.00	3.00
sp|O60096| RT35_SCHPO	37S ribosomal protein S35, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=bot1 PE=2 SV=1	35784.30	1.00	3.00	6.00
sp|Q9UT21| YFY7_SCHPO	Uncharacterized GTP-binding protein C9.07c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC9.07c PE=2 SV=1	40404.40		3.00	1.00
sp|O42938| K6PF_SCHPO	6-phosphofructokinase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pfk1 PE=1 SV=1	102653.00	1.00	30.00	21.00
sp|P05733| RL37B_SCHPO	60S ribosomal protein L37-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl37b PE=1 SV=4	10080.10	1.00	5.00	3.00
sp|Q9Y823| HOSM_SCHPO	Homocitrate synthase, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=lys4 PE=1 SV=1	46335.50		19.00	14.00
sp|P22774-2| HSP7M_SCHPO	Isoform Short of Heat shock 70 kDa protein, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ssp1	73040.00		12.00	9.00
sp|P05734| RL19A_SCHPO	60S ribosomal protein L19-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl19a PE=1 SV=3	22727.60		12.00	4.00
sp|O14134| ELF1_SCHPO	mRNA export factor elf1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=elf1 PE=1 SV=1	116807.00		12.00	6.00
sp|O74975| PCY1_SCHPO	Probable choline-phosphate cytidylyltransferase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC1827.02c PE=1 SV=3	41779.10			2.00
sp|Q9UT32| RL1DB_SCHPO	Putative ribosome biogenesis protein C8F11.04 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC8F11.04 PE=2 SV=1	41386.40		4.00	
sp|P41887| HSP90_SCHPO	Heat shock protein 90 homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=swo1 PE=1 SV=2	80672.20	1.00	21.00	15.00
sp|O60100| IMB4_SCHPO	Probable importin subunit beta-4 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=kap123 PE=2 SV=1	117939.00		3.00	2.00
sp|O94666| RPC1_SCHPO	DNA-directed RNA polymerase III subunit rpc1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpc1 PE=2 SV=1	157714.00		2.00	
sp|Q09878| MET10_SCHPO	Probable sulfite reductase [NADPH] flavoprotein component OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC584.01c PE=2 SV=2	111464.00		14.00	4.00
sp|O13672| RL8_SCHPO	60S ribosomal protein L8 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl8 PE=1 SV=3	28625.00	3.00	19.00	26.00
sp|Q09748| YB68_SCHPO	Dynamin-like protein C12C2.08 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC12C2.08 PE=2 SV=1	87080.80		2.00	1.00
sp|Q9UUJ1| ARP2_SCHPO	Actin-related protein 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=arp2 PE=1 SV=1	44256.70		2.00	2.00
sp|Q9UT48| EIF3H_SCHPO	Eukaryotic translation initiation factor 3 subunit H OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=eif3h PE=1 SV=1	39810.10	1.00	5.00	4.00
sp|O60102| YOX6_SCHPO	Uncharacterized ABC transporter ATP-binding protein C14F5.06 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC14F5.06 PE=2 SV=1	66637.50		3.00	
sp|Q9USP8| IDH2_SCHPO	Isocitrate dehydrogenase [NAD] subunit 2, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=idh2 PE=1 SV=2	41216.30		1.00	2.00
sp|O94501| TCPA_SCHPO	T-complex protein 1 subunit alpha OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cct1 PE=2 SV=1	60099.60		11.00	9.00
sp|P36586| YPT5_SCHPO	GTP-binding protein ypt5 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ypt5 PE=1 SV=1	23006.90		2.00	
sp|P87147| NACA_SCHPO	Nascent polypeptide-associated complex subunit alpha OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=egd2 PE=1 SV=1	18791.70	1.00	6.00	4.00

sp|O60128| RS3_SCHPO	40S ribosomal protein S3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps3 PE=1 SV=1	27571.60		8.00	3.00
sp|O74752| MAS5_SCHPO	Mitochondrial protein import protein mas5 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=mas5 PE=2 SV=1	44864.60		12.00	5.00
sp|P56329| IF2B_SCHPO	Probable eukaryotic translation initiation factor 2 subunit beta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif212 PE=1 SV=2	35992.20		1.00	1.00
sp|O74462| TAF6_SCHPO	Transcription initiation factor TFIID subunit 6 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=taf6 PE=1 SV=1	50313.10		1.00	1.00
sp|O14150| RS14_SCHPO	40S ribosomal protein S14 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps14a PE=1 SV=1	14707.80		12.00	7.00
sp|Q9UTE3| SEB1_SCHPO	Rpb7-binding protein seb1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=seb1 PE=1 SV=3	66506.70		2.00	
sp|Q12705| PP2B_SCHPO	Serine/threonine-protein phosphatase 2B catalytic subunit OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ppb1 PE=2 SV=2	63763.80		3.00	1.00
sp|P24487| ATPA_SCHPO	ATP synthase subunit alpha, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=atp1 PE=2 SV=1	58641.10		5.00	1.00
P02769	SWISS_PROT_P02769 (Bos taurus) Bovine serum albumin precursor	69356.50	4.00	1.00	1.00
sp|Q9UTP0| RL30B_SCHPO	60S ribosomal protein L30-2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl30b PE=2 SV=1	12252.60		1.00	1.00
sp|P0C016| RS27A_SCHPO	Ubiquitin-40S ribosomal protein S27a OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ubi3 PE=1 SV=2	17222.30		6.00	2.00
sp|Q9HGP5| CAPZB_SCHPO	F-actin-capping protein subunit beta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=acp2 PE=1 SV=1	29862.90		3.00	2.00
sp|Q09900| RL38B_SCHPO	60S ribosomal protein L38-2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl38b PE=2 SV=1	8333.71		5.50	3.50
sp|Q9UTG4| RS26B_SCHPO	40S ribosomal protein S26-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps26b PE=2 SV=1	13871.50		3.00	4.00
sp|P87153| TCPH_SCHPO	Probable T-complex protein 1 subunit eta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cct7 PE=1 SV=1	60738.40		6.00	4.00
sp|Q9Y808| YN91_SCHPO	Uncharacterized protein C146.01 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC146.01 PE=1 SV=2	118570.00			4.00
sp|O60143| RL7C_SCHPO	60S ribosomal protein L7-C OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl7c PE=1 SV=1	28748.60		12.50	9.50
sp|P78920| YGNB_SCHPO	Putative nucleosome assembly protein C2D10.11C OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC2D10.11c PE=2 SV=2	43398.10		2.00	1.00
sp|Q1K9C2| MET5_SCHPO	Sulfite reductase [NADPH] subunit beta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sir1 PE=3 SV=1	163995.00		11.00	5.00
sp|O13775| SYEC_SCHPO	Probable glutamate--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC17A5.15c PE=1 SV=1	80825.20		6.00	
sp|Q10101| RS7_SCHPO	40S ribosomal protein S7 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps7 PE=1 SV=2	21951.90		10.00	3.00
sp|O14164| EIF3C_SCHPO	Eukaryotic translation initiation factor 3 subunit C OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=nip1 PE=1 SV=2	104465.00	1.00	17.00	17.00
sp|P24782| DBP2_SCHPO	ATP-dependent RNA helicase dbp2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=dbp2 PE=2 SV=2	61600.10		1.00	3.00
sp|O13863| YDU2_SCHPO	Uncharacterized kinase C1B1.02c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC1B1.02c PE=1 SV=1	60760.50			2.00
sp|P42656| RAD24_SCHPO	DNA damage checkpoint protein rad24 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rad24 PE=1 SV=2	30099.90	0.50	9.50	7.50
sp|O42943| YBP8_SCHPO	Uncharacterized ABC transporter ATP-binding protein C16H5.08c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC16H5.08c PE=1 SV=1	69277.80		9.00	3.00
sp|Q9UTJ4| CGR1_SCHPO	rRNA-processing protein cgr1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cgr1 PE=2 SV=1	13522.40		3.00	2.00

sp|Q09892| HOG1_SCHPO	Mitogen-activated protein kinase sty1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sty1 PE=1 SV=1	40250.50		2.00	2.00
sp|O94303| HIS5_SCHPO	Imidazole glycerol phosphate synthase hisHF OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=his4 PE=2 SV=3	59451.90		2.00	
sp|O13783| BRO1_SCHPO	Vacuolar protein-sorting protein bro1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=bro1 PE=2 SV=1	86694.20		2.00	
sp|P25457| RL7B_SCHPO	60S ribosomal protein L7-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl7b PE=2 SV=1	28467.60		14.50	6.50
sp|O94515| TCPZ_SCHPO	T-complex protein 1 subunit zeta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cct6 PE=2 SV=1	58601.70		2.00	2.00
sp|Q9UTK5| ALM1_SCHPO	Abnormal long morphology protein 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=alm1 PE=1 SV=1	198046.00		1.00	3.00
sp|Q10105| GCN1_SCHPO	Translational activator gcn1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=gcn1 PE=2 SV=1	297620.00		2.00	1.00
sp|O13788| SSR1_SCHPO	SWI/SNF and RSC complexes subunit ssr1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ssr1 PE=1 SV=1	59691.60		2.00	
sp|O74436| SRP68_SCHPO	Signal recognition particle subunit srp68 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=srp68 PE=2 SV=1	67907.50		2.00	
sp|P87157| YB0G_SCHPO	Probable ADP-ribosylation factor-binding protein C25H2.16c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC25H2.16c PE=1 SV=1	59534.60		2.00	
sp|P28189| RS13_SCHPO	40S ribosomal protein S13 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps13 PE=1 SV=2	16960.30	3.00	14.00	14.00
sp|O13784| RS24A_SCHPO	40S ribosomal protein S24-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps24a PE=2 SV=3	15478.40		1.00	1.00
sp|O94567| SYNC_SCHPO	Probable asparagine--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC1773.10c PE=2 SV=1	63919.50		5.00	3.00
sp|Q10134| FEP1_SCHPO	Iron-sensing transcription factor 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=fep1 PE=1 SV=1	60663.60			3.00
sp|O74760| EIF3A_SCHPO	Eukaryotic translation initiation factor 3 subunit A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif32 PE=1 SV=1	107167.00		35.00	36.00
sp|Q7Z9I2| YCP9_SCHPO	Uncharacterized oxidoreductase C663.09c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC663.09c PE=2 SV=1	27965.10		8.00	4.00
sp|O14233| RNA14_SCHPO	mRNA 3'-end-processing protein rna14 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rna14 PE=2 SV=1	83999.70		2.00	
sp|P42657| RAD25_SCHPO	DNA damage checkpoint protein rad25 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rad25 PE=1 SV=1	30386.00		7.00	4.00
sp|O42947| YBPC_SCHPO	Uncharacterized protein C16H5.12c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC16H5.12c PE=1 SV=3	82595.50		10.00	9.00
sp|O13792| SUB2_SCHPO	ATP-dependent RNA helicase uap56 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=uap56 PE=1 SV=3	49272.20		3.00	2.00
sp|Q9UT44| PROA_SCHPO	Probable gamma-glutamyl phosphate reductase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pro1 PE=2 SV=1	48748.40		2.00	
sp|Q9UTL3| RPIA_SCHPO	Ribose-5-phosphate isomerase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rki1 PE=2 SV=1	29346.40		3.00	1.00
sp|P87228| SERA_SCHPO	Putative D-3-phosphoglycerate dehydrogenase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC364.07 PE=1 SV=1	50923.30		7.00	3.00
sp|O60155| YHY6_SCHPO	Putative proline--tRNA ligase C19C7.06 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC19C7.06 PE=2 SV=1	78951.60		11.00	7.00
P10096	SWISS_PROT_P10096 (Bos taurus) Glyceraldehyde_3_phosphate dehydrogenase	35899.30		2.00	
sp|O14236| NOG2_SCHPO	Nucleolar GTP-binding protein 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=nog2 PE=1 SV=1	59966.90		4.00	3.00
sp|O74340| DCA13_SCHPO	Protein sof1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sof1 PE=2 SV=1	50035.60		2.00	2.00

sp|Q10147| TCPB_SCHPO	Probable T-complex protein 1 subunit beta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cct2 PE=2 SV=1	56731.50		6.00	6.00
sp|O94616| RPAC1_SCHPO	DNA-directed RNA polymerases I and III subunit RPAC1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpc40 PE=2 SV=1	39184.20		2.00	3.00
sp|Q9UTM3| RPN5_SCHPO	26S proteasome regulatory subunit rpn5 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpn5a PE=1 SV=1	51658.00		7.00	7.00
sp|P41836| PRS8_SCHPO	26S protease regulatory subunit 8 homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=let1 PE=2 SV=1	45113.00		3.00	2.00
sp|Q9P7Q4| SEC18_SCHPO	Vesicular-fusion protein sec18 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec18 PE=1 SV=1	87624.80		1.00	3.00
sp|Q8TFH1| RL18B_SCHPO	60S ribosomal protein L18-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl18b PE=1 SV=1	21189.00		6.00	1.50
sp|Q01475| SAR1_SCHPO	Small COPII coat GTPase sar1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sar1 PE=1 SV=1	21284.00		2.00	2.00
sp|O42952| RS30_SCHPO	40S ribosomal protein S30 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps30a PE=2 SV=2	6905.84		1.00	7.00
sp|O13794| YSH1_SCHPO	Endoribonuclease ysh1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ysh1 PE=2 SV=2	86061.90		17.00	17.00
sp|P47943| IF4A_SCHPO	ATP-dependent RNA helicase eIF4A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif1 PE=1 SV=2	44461.60		20.00	10.50
sp|O74762| RPN2_SCHPO	26S proteasome regulatory subunit rpn2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpn2 PE=1 SV=1	107371.00		1.00	
sp|O94413| KPR2_SCHPO	Ribose-phosphate pyrophosphokinase 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC1620.06c PE=1 SV=1	34811.30		2.00	
sp|P28748| SPI1_SCHPO	GTP-binding nuclear protein spi1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=spi1 PE=1 SV=1	24575.60	1.00		2.00
sp|O74782| YGBA_SCHPO	Universal stress protein A family protein C25B2.10 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC25B2.10 PE=1 SV=1	34114.70		1.00	
sp|O94634| THDH_SCHPO	Threonine dehydratase, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC1677.03c PE=2 SV=1	66591.90		6.00	5.00
sp|Q09130| IF2G_SCHPO	Eukaryotic translation initiation factor 2 subunit gamma OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif213 PE=2 SV=1	48821.90		3.00	
sp|P30666| MCM3_SCHPO	DNA replication licensing factor mcm3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=mcm3 PE=1 SV=2	97564.80		2.00	
sp|O94672| SC232_SCHPO	Protein transport protein sec23-2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec232 PE=1 SV=1	85453.20		2.00	0.50
sp|P36579| ARF1_SCHPO	ADP-ribosylation factor 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=arf1 PE=2 SV=2	20623.60		2.00	2.00
sp|O42964| IWS1_SCHPO	Transcription factor iws1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=iws1 PE=2 SV=1	49156.50		9.00	4.00
sp|O14289| LEUC_SCHPO	3-isopropylmalate dehydratase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=leu2 PE=1 SV=1	82856.70		14.00	8.00
sp|O60163| AK_SCHPO	Probable aspartokinase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC19F5.04 PE=1 SV=1	56866.60		11.00	10.00
sp|Q92345| PDC2_SCHPO	Probable pyruvate decarboxylase C1F8.07c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC1F8.07c PE=1 SV=3	61770.30		20.00	16.00
sp|Q09171| ODPB_SCHPO	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pdb1 PE=2 SV=1	39652.50		2.00	4.00
sp|Q10157| RL11_SCHPO	60S ribosomal protein L11 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl11a PE=1 SV=1	19896.60	1.00	10.00	13.00
sp|Q9UTM4| TCPE_SCHPO	T-complex protein 1 subunit epsilon OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cct5 PE=2 SV=1	59430.20		6.00	1.00
sp|O74522| ADN3_SCHPO	Adhesion defective protein 3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=adn3 PE=2 SV=1	104688.00		1.00	1.00
sp|Q09669| ADH4_SCHPO	Alcohol dehydrogenase 4 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=adh4 PE=2 SV=1	45570.10		2.00	

sp|P31209| PABP_SCHPO	Polyadenylate-binding protein, cytoplasmic and nuclear OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pab1 PE=1 SV=2	71575.70		11.00	4.00
sp|Q09181| DHH1_SCHPO	Putative ATP-dependent RNA helicase ste13 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ste13 PE=2 SV=1	54832.30		2.00	1.00
sp|O94641| HS104_SCHPO	Heat shock protein 104 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=hsp104 PE=2 SV=1	100608.00		10.00	2.00
sp|O74873| SC231_SCHPO	Protein transport protein sec23-1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec231 PE=2 SV=1	84917.10		2.00	1.00
P02662	SWISS_PROT_P02662 Alpha_S1_casein _ Bos taurus (Bovine).	22978.50	1.00	1.00	1.00
sp|O43001| SYJ1_SCHPO	Inositol-1,4,5-trisphosphate 5-phosphatase 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=syj1 PE=1 SV=1	121926.00		2.00	
sp|Q9UT07| UPS1_SCHPO	Protein ups1 homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAP8A3.10 PE=2 SV=1	19914.50		1.00	1.00
sp|Q9USR7| RL38A_SCHPO	60S ribosomal protein L38-1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl38a PE=2 SV=1	8486.76	1.00	2.00	
sp|Q10161| CLH_SCHPO	Probable clathrin heavy chain OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=chc1 PE=1 SV=1	190186.00		2.00	1.00
sp|P50999| TCPD_SCHPO	T-complex protein 1 subunit delta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cct4 PE=2 SV=2	57022.70		7.00	4.00
sp|Q09668| RL22_SCHPO	60S ribosomal protein L22 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl22 PE=1 SV=3	13330.20	3.00	8.00	3.00
sp|Q9UTN4| PFS2_SCHPO	Polyadenylation factor subunit 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pfs2 PE=2 SV=1	57708.30		12.00	9.00
sp|O60164| PESC_SCHPO	Pescadillo homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ppp1 PE=2 SV=1	69254.50		5.00	2.00
sp|O13801| YE04_SCHPO	Uncharacterized RNA-binding protein C17H9.04c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC17H9.04c PE=1 SV=1	66462.10		2.00	1.00
sp|O43026| G3P2_SCHPO	Glyceraldehyde-3-phosphate dehydrogenase 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=gpd3 PE=1 SV=1	35706.30		5.00	2.00
P20930	SWISS_PROT_P20930 Tax_Id=9606 Gene_Symbol=FLG Filaggrin	435642.00	1.00		
P02666	SWISS_PROT_P02666 Beta_casein _ Bos taurus (Bovine).	23604.30	2.00		2.00
sp|P31318| ARG56_SCHPO	Protein arg11, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=arg11 PE=2 SV=2	97788.80		9.00	7.00
sp|O14293| YF19_SCHPO	Putative aldehyde dehydrogenase-like protein C9E9.09c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC9E9.09c PE=1 SV=1	54823.70		3.00	
sp|O94658| RL36B_SCHPO	60S ribosomal protein L36-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl36b PE=2 SV=1	11256.50	1.00	12.00	8.00
sp|O13831| TRP_SCHPO	Tryptophan synthase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=trp2 PE=2 SV=1	75496.70		14.00	7.00
sp|O13823| IMP4_SCHPO	U3 small nucleolar ribonucleoprotein protein imp4 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=imp4 PE=2 SV=1	33410.60		2.00	
sp|O74781| SYRC_SCHPO	Probable arginine--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC25B2.09c PE=2 SV=1	70753.40		8.00	1.00
sp|Q09676| YA03_SCHPO	Uncharacterized protein C5H10.03 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC5H10.03 PE=2 SV=1	24736.80		3.00	
sp|P52808| RL30A_SCHPO	60S ribosomal protein L30-1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl30a PE=2 SV=2	11602.20		5.00	1.00
Q5D862	SWISS_PROT_Q5D862 Tax_Id=9606 Gene_Symbol=FLG2 Filaggrin_2	248342.00	1.00	1.00	1.00
sp|O60198| METK_SCHPO	S-adenosylmethionine synthase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sam1 PE=2 SV=1	41859.40	1.00	19.00	12.00
sp|Q10179| SYG_SCHPO	Putative glycine--tRNA ligase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=grs1 PE=2 SV=1	74066.60		20.00	9.00
sp|Q92370| TRPG_SCHPO	Anthranilate synthase component 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=trp1 PE=2 SV=2	83114.20		4.00	2.00
sp|O94619| ADN2_SCHPO	Adhesion defective protein 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=adn2 PE=1 SV=1	81235.50		1.00	2.00
sp|O13774| GCH1_SCHPO	GTP cyclohydrolase 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC17A5.13 PE=1 SV=1	26511.70		5.00	

sp|Q9UTQ6| RS15B_SCHPO	40S ribosomal protein S15-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps15b PE=1 SV=1	17640.50		6.00	4.00
sp|O14339| RL17A_SCHPO	60S ribosomal protein L17-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl17a PE=2 SV=1	20819.20		11.50	9.50
sp|Q92372| RFA1_SCHPO	Replication factor A protein 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ssb1 PE=1 SV=1	68262.30		2.00	
sp|Q09722| EIF3M_SCHPO	Eukaryotic translation initiation factor 3 subunit M OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=eif3m PE=1 SV=2	45129.20		16.00	4.00
sp|P52890| ATF1_SCHPO	Transcription factor atf1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=atf1 PE=1 SV=1	59764.90		3.00	2.00
sp|P31406| VATA_SCHPO	V-type proton ATPase catalytic subunit A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=vma1 PE=2 SV=1	68870.60		7.00	5.00
sp|P87312| CLF1_SCHPO	Pre-mRNA-splicing factor cwf4 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cwf4 PE=1 SV=1	80897.30		1.00	2.00
sp|O74864| RLA0_SCHPO	60S acidic ribosomal protein P0 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpp0 PE=2 SV=1	33598.20		11.00	11.00
sp|Q09127| RL10A_SCHPO	60S ribosomal protein L10-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl10a PE=2 SV=2	25393.30		1.00	1.00
sp|P56287| EI2BE_SCHPO	Probable translation initiation factor eIF-2B subunit epsilon OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif225 PE=1 SV=1	76386.00		1.00	1.00
sp|Q9C0X6| RUVB1_SCHPO	RuvB-like helicase 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rvb1 PE=2 SV=1	50093.10		6.00	4.00
sp|O94692| RUVB2_SCHPO	RuvB-like helicase 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rvb2 PE=1 SV=1	51601.50		10.00	6.00
sp|O42897| RPN3_SCHPO	Probable 26S proteasome regulatory subunit rpn3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpn3 PE=2 SV=2	57382.00		3.00	
sp|P50522| EF1A1_SCHPO	Elongation factor 1-alpha-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tef1a PE=1 SV=2	49701.00	5.98	88.43	61.58
sp|P56288| EI2BG_SCHPO	Probable translation initiation factor eIF-2B subunit gamma OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif223 PE=1 SV=2	50502.10		4.00	1.00
sp|O13864| IMB1_SCHPO	Importin subunit beta-1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=kap95 PE=2 SV=1	94830.70		3.00	2.00
sp|P32235| GTP1_SCHPO	GTP-binding protein 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=gtp1 PE=2 SV=2	40876.60		3.00	1.00
sp|O74330| RS27_SCHPO	40S ribosomal protein S27 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps27 PE=2 SV=1	9165.57		4.00	
sp|Q09747| DBP5_SCHPO	ATP-dependent RNA helicase dbp5 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=dbp5 PE=1 SV=1	56359.90		6.00	6.00
sp|Q9UU76| PUF6_SCHPO	Pumilio homology domain family member 6 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=puf6 PE=1 SV=1	72697.00		36.00	8.00
P04264	SWISS_PROT_P04264 Tax_Id=9606 Gene_Symbol=KRT1 Keratin, type II cytoskeletal 1	66086.00	27.00	56.48	42.98
P35908	SWISS_PROT_P35908 Tax_Id=9606 Gene_Symbol=KRT2 Keratin, type II cytoskeletal 2 epidermal	65933.50	16.00	34.99	22.50
P13647	SWISS_PROT_P13647 Tax_Id=9606 Gene_Symbol=KRT5 Keratin, type II cytoskeletal 5	62430.00	5.00	11.50	9.50
P19013	SWISS_PROT_P19013 Tax_Id=9606 Gene_Symbol=KRT4 keratin 4	63961.40	2.00	2.98	0.97
P48668	SWISS_PROT_P48668 Tax_Id=9606 Gene_Symbol=KRT6C Keratin, type II cytoskeletal 6C	60078.40	1.50	1.00	1.50
P02538	SWISS_PROT_P02538 Tax_Id=9606 Gene_Symbol=KRT6A Keratin, type II cytoskeletal 6A	60098.30	3.00		1.00
Q5XQN5	SWISS_PROT_Q5XQN5 (Bos taurus) Keratin, type II cytoskeletal 5	63006.40			0.98
O43790	SWISS_PROT_O43790 Keratin, type II cuticular Hb6 (Hair keratin, type II Hb6) (ghHb6) _ Homo sapiens (Human).	55025.80		1.00	
sp|O13867| UPP1_SCHPO	Uracil phosphoribosyltransferase 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC1B3.01c PE=2 SV=1	24486.70		3.00	1.00
sp|Q9C102| GLT1_SCHPO	Putative glutamate synthase [NADPH] OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAPB1E7.07 PE=2 SV=1	233080.00		5.00	1.00
sp|P57728| RL28B_SCHPO	60S ribosomal protein L28-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl28b PE=1 SV=1	16693.10		3.00	2.00

sp|O43027| RGA6_SCHPO	Probable Rho-GTPase-activating protein 6 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rga6 PE=1 SV=1	80853.00		2.00	2.00
sp|O74884| RL24B_SCHPO	60S ribosomal protein L24-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl24b PE=1 SV=1	16844.40	1.00	16.00	13.00
sp|O13640| YGWH_SCHPO	Uncharacterized alpha-1,2-galactosyltransferase C8D2.17 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pi048 PE=2 SV=2	37029.00		2.00	2.00
sp|O94715| RS15A_SCHPO	40S ribosomal protein S15-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps15a PE=2 SV=1	17526.50		3.00	4.00
sp|Q10192| RL18A_SCHPO	60S ribosomal protein L18-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl18a PE=1 SV=1	21194.90		7.00	4.50
sp|O74341| TCPG_SCHPO	T-complex protein 1 subunit gamma OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cct3 PE=1 SV=1	58533.50		12.00	7.00
sp|P35679| RL4A_SCHPO	60S ribosomal protein L4-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl4a PE=1 SV=2	39796.80	1.00	12.50	8.00
sp|O13907| ACL2_SCHPO	Probable ATP-citrate synthase subunit 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC22A12.16 PE=1 SV=1	53981.00		3.00	3.00
sp|O43060| EIF3F_SCHPO	Eukaryotic translation initiation factor 3 subunit F OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=eif3f PE=1 SV=1	33284.10		7.00	3.00
sp|P79015| RL32A_SCHPO	60S ribosomal protein L32-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl32a PE=1 SV=2	14466.90		4.00	2.00
sp|O14340| YB35_SCHPO	Oxysterol-binding protein homolog C2F12.05c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC2F12.05c PE=1 SV=2	148618.00		33.00	35.00
sp|Q09755| AROF_SCHPO	Putative phospho-2-dehydro-3-deoxyheptonate aldolase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC24H6.10c PE=2 SV=1	39789.40		3.00	3.00
sp|Q9P3A7| CDC48_SCHPO	Cell division cycle protein 48 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cdc48 PE=1 SV=2	90212.80		4.00	1.00
sp|P58234| RS19A_SCHPO	40S ribosomal protein S19-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps19a PE=2 SV=1	16070.30		3.00	7.00
sp|Q10202| DBP3_SCHPO	ATP-dependent RNA helicase dbp3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=dbp3 PE=2 SV=1	63963.20		6.00	2.00
sp|O94754| RS18_SCHPO	40S ribosomal protein S18 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps18a PE=1 SV=1	17442.50		12.00	5.00
sp|Q09769| LONM_SCHPO	Lon protease homolog, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pim1 PE=2 SV=1	118749.00		3.00	2.00
sp|O13914| SYA_SCHPO	Alanine--tRNA ligase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ala1 PE=1 SV=1	107472.00		7.00	
sp|O74351| NFS1_SCHPO	Probable cysteine desulfurase, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC21D10.11c PE=2 SV=2	55598.20		2.00	1.00
sp|O59678| RRS1_SCHPO	Ribosome biogenesis regulatory protein homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC29A3.16 PE=1 SV=1	19050.20		5.00	
sp|Q9P3W6| YLI2_SCHPO	Uncharacterized protein C458.02c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC458.02c PE=1 SV=1	53775.20		3.00	2.00
sp|P78773| ACSA_SCHPO	Probable acetyl-coenzyme A synthetase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC191.02c PE=1 SV=2	73097.40		7.00	4.00
sp|O74352| HOB1_SCHPO	Protein hob1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=hob1 PE=1 SV=1	51431.80		3.00	2.00
sp|O74892| RS2_SCHPO	40S ribosomal protein S2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps2 PE=1 SV=1	27655.90		3.00	2.00
sp|P00332| ADH_SCHPO	Alcohol dehydrogenase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=adh1 PE=1 SV=2	37426.00		12.00	5.00
sp|Q10208| KPYK_SCHPO	Pyruvate kinase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pyk1 PE=1 SV=1	55569.50	2.00	54.00	27.00
sp|Q09794| PYR1_SCHPO	Protein ura1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ura1 PE=1 SV=1	248548.00		47.00	20.00
sp|O59704| ELP1_SCHPO	Elongator complex protein 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=iki3 PE=2 SV=1	141734.00		2.00	3.00

sp|O13936| SPT5_SCHPO	Transcription elongation factor spt5 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=spt5 PE=1 SV=1	109110.00		4.00	4.00
sp|O74402| HSP78_SCHPO	Heat shock protein 78, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=hsp78 PE=2 SV=1	90242.80		4.00	2.00
sp|O74895| RL15A_SCHPO	60S ribosomal protein L15-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl15a PE=2 SV=1	23843.80		2.00	4.00
sp|Q9P6K7| YLF4_SCHPO	Uncharacterized tRNA-binding protein C30C2.04 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC30C2.04 PE=2 SV=1	50646.30		5.00	
sp|P78811| UGPA1_SCHPO	Probable UTP--glucose-1-phosphate uridylyltransferase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC1322.04 PE=1 SV=2	56484.40		4.00	2.00
sp|O59711| LYS9_SCHPO	Saccharopine dehydrogenase [NADP(+), L-glutamate-forming] OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=lys9 PE=2 SV=1	49978.40		12.00	
sp|O13998| YEI3_SCHPO	Uncharacterized GTP-binding protein C27E2.03c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC27E2.03c PE=1 SV=1	44359.10		9.00	6.00
sp|P36584| RL3B_SCHPO	60S ribosomal protein L3-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl3b PE=2 SV=2	43829.30		30.00	10.00
sp|Q9P6M1| GST3_SCHPO	Glutathione S-transferase 3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=gst3 PE=1 SV=3	28131.30		4.00	
sp|P04688| TBA1_SCHPO	Tubulin alpha-1 chain OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=nda2 PE=2 SV=2	51191.70		4.00	3.00
sp|O74407| SYDC_SCHPO	Aspartate--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=dps1 PE=1 SV=1	66719.00		11.00	
sp|O14351| YB45_SCHPO	Uncharacterized oxidoreductase C30D10.05c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC30D10.05c PE=2 SV=1	26664.20		4.00	1.00
sp|O74904| RL35_SCHPO	60S ribosomal protein L35 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl35 PE=1 SV=1	14311.20	4.00	30.00	14.00
sp|Q9UU79| NSA2_SCHPO	Ribosome biogenesis protein nsa2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=nsa2 PE=1 SV=1	29723.00		14.00	19.00
sp|O59771| UMPK_SCHPO	Probable uridylate kinase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC1795.05c PE=2 SV=1	21720.00		3.00	
sp|P04689| TBA2_SCHPO	Tubulin alpha-2 chain OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tub1 PE=1 SV=2	50575.60		3.00	1.00
sp|O14018| SYSC_SCHPO	Serine--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC29A4.15 PE=2 SV=1	51682.10		9.00	1.00
sp|Q9P769| RL10B_SCHPO	60S ribosomal protein L10-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl10b PE=2 SV=1	25379.30		6.00	6.00
sp|P78820| ACAC_SCHPO	Acetyl-CoA carboxylase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cut6 PE=1 SV=2	257078.00		21.00	5.00
sp|O14370| BCA1_SCHPO	Branched-chain-amino-acid aminotransferase, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=eca39 PE=1 SV=3	47896.40		7.00	5.00
sp|O74476| IMB3_SCHPO	Importin subunit beta-3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sal3 PE=2 SV=1	121972.00		3.00	1.00
sp|O59778| BIOB_SCHPO	Biotin synthase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=bio2 PE=1 SV=1	40678.90		1.00	1.00
sp|P04913| H2B1_SCHPO	Histone H2B-alpha OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=htb1 PE=1 SV=3	13828.30	2.00	4.00	2.00
sp|O74484| MPG1_SCHPO	Mannose-1-phosphate guanyltransferase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=mpg1 PE=2 SV=1	39747.90		7.00	4.00
sp|P36585| RL28A_SCHPO	60S ribosomal protein L28-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl28a PE=2 SV=3	16661.10	0.50	7.00	2.50
sp|O14023| ELP3_SCHPO	Probable elongator complex protein 3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=elp3 PE=2 SV=1	61900.60		2.00	1.00
sp|O59779| YCY1_SCHPO	Uncharacterized protein C1235.01 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC1235.01 PE=2 SV=1	69758.60	1.00	5.00	7.00
sp|O14388| RL27A_SCHPO	60S ribosomal protein L27-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl27a PE=2 SV=2	15397.60	1.00	9.50	9.50
sp|Q9UU81| AP1G1_SCHPO	AP-1 complex subunit gamma-1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=apl4 PE=2 SV=1	96126.20		4.00	2.00

sp|O74919| RNC1_SCHPO	RNA-binding protein rnc1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rnc1 PE=1 SV=1	43404.20		11.00	6.00
sp|Q9P782| YNY8_SCHPO	Uncharacterized protein C1711.08 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC1711.08 PE=2 SV=1	37450.70		10.00	8.00
sp|O14026| SET2_SCHPO	Histone-lysine N-methyltransferase, H3 lysine-36 specific OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=set2 PE=1 SV=1	90748.70		3.00	5.00
sp|Q9UUE1| PYC_SCHPO	Pyruvate carboxylase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pyr1 PE=2 SV=1	130976.00		2.00	
sp|O74504| YJD1_SCHPO	UPF0590 protein C594.01 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC594.01 PE=1 SV=2	88820.40		3.00	1.00
sp|O59787| SRP72_SCHPO	Signal recognition particle subunit srp72 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=srp72 PE=2 SV=1	63332.00		1.00	1.00
sp|P05732| RL20_SCHPO	60S ribosomal protein L20 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl20a PE=1 SV=3	20604.20		31.00	19.00
sp|P36591| DYR_SCHPO	Dihydrofolate reductase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=dfr1 PE=2 SV=2	51562.00		5.00	1.00
sp|O74526| KJ45_SCHPO	Probable serine/threonine-protein kinase C70.05c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC70.05c PE=1 SV=1	85981.50			2.00
sp|Q9UUF2| SYMC_SCHPO	Probable methionine--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pi042 PE=2 SV=1	88950.80		16.00	2.00
sp|O74928| PUR9_SCHPO	Bifunctional purine biosynthesis protein ade10 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ade10 PE=2 SV=1	64173.00		9.00	4.00
sp|O59810| VGL1_SCHPO	Vigilin 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=vgl1 PE=1 SV=2	142742.00		10.00	5.00
sp|Q10209| YAY1_SCHPO	Uncharacterized J domain-containing protein C4H3.01 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC4H3.01 PE=1 SV=1	44795.00	2.00	52.00	37.00
sp|Q09801| FIP1X_SCHPO	Pre-mRNA polyadenylation factor fip1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC22G7.10 PE=2 SV=1	37321.50		4.00	4.00
sp|O74536| SNF1_SCHPO	SNF1-like protein kinase ssp2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ssp2 PE=1 SV=1	65957.60		2.00	1.00
sp|P78827| ILV5_SCHPO	Probable ketol-acid reductoisomerase, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ilv5 PE=1 SV=2	45232.30		9.00	5.00
sp|P36602| RIR1_SCHPO	Ribonucleoside-diphosphate reductase large chain OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cdc22 PE=1 SV=2	92084.30		4.00	3.00
sp|Q9UUF7| COPB_SCHPO	Coatomer subunit beta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec26 PE=2 SV=1	104796.00		6.00	1.00
sp|Q9P784| RL4B_SCHPO	60S ribosomal protein L4-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl4b PE=2 SV=1	39939.80		4.00	2.00
sp|O74960| PIL1_SCHPO	Probable sphingolipid long chain base-responsive protein pil1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pil1 PE=1 SV=1	39834.70		14.00	13.00
sp|O59865| RS24B_SCHPO	40S ribosomal protein S24-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps24b PE=2 SV=3	15547.40		9.00	2.00
sp|P36604| GRP78_SCHPO	78 kDa glucose-regulated protein homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=bip1 PE=2 SV=2	73289.60	1.00	3.00	
sp|O14460| EF2_SCHPO	Elongation factor 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=eft201 PE=1 SV=2	93315.70	2.00	55.50	41.00
sp|P78921| TCPQ_SCHPO	Probable T-complex protein 1 subunit theta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cct8 PE=2 SV=3	60008.20		11.00	10.00
sp|O60059| YG58_SCHPO	Pumilio domain-containing protein C56F2.08c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC56F2.08c PE=1 SV=1	72884.60		3.00	2.00
sp|P36612| PRS4_SCHPO	26S protease regulatory subunit 4 homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=mts2 PE=2 SV=1	50100.90		5.00	4.00
sp|P05752| RS6A_SCHPO	40S ribosomal protein S6-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps6a PE=1 SV=2	27494.10		4.00	3.00
sp|Q9UUG0| FAS1_SCHPO	Fatty acid synthase subunit beta OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=fas1 PE=1 SV=1	230776.00		19.00	3.00

sp|O75000| RL12_SCHPO	60S ribosomal protein L12 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl1201 PE=1 SV=1	17673.50	2.00	5.00	4.00
sp|O60060| CARA_SCHPO	Carbamoyl-phosphate synthase arginine-specific small chain OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=arg5 PE=2 SV=1	45705.00		5.00	4.00
sp|O94238| RL14_SCHPO	60S ribosomal protein L14 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl14 PE=2 SV=1	15218.40	1.00	25.00	10.00
sp|Q9P7D4| ACON2_SCHPO	Probable aconitate hydratase 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBP4H10.15 PE=2 SV=3	100000.00		2.00	2.00
sp|Q10222| PTA1_SCHPO	mRNA cleavage and polyadenylation specificity factor complex subunit pta1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pta1 PE=1 SV=2	76293.20		3.00	6.00
sp|P08093| RL2_SCHPO	60S ribosomal protein L2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl2a PE=2 SV=2	27119.40	1.00	13.00	10.00
sp|Q9P7D7| LCF2_SCHPO	Long-chain-fatty-acid--CoA ligase 2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=lcf2 PE=2 SV=1	75851.00		4.00	
sp|P36620| ILVB_SCHPO	Acetolactate synthase, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ilv1 PE=2 SV=2	73255.40		20.00	13.00
sp|Q9P7I2| KCC1_SCHPO	Calcium/calmodulin-dependent protein kinase type I OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cmk1 PE=1 SV=1	38193.40		7.00	5.00
sp|O94260| G3BP_SCHPO	Putative G3BP-like protein OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=nxt3 PE=1 SV=1	48014.90		2.00	1.00
sp|Q10241| CMB1_SCHPO	Mismatch-binding protein cmb1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cmb1 PE=1 SV=1	26323.90	2.00	18.00	12.00
sp|P78946| RL26_SCHPO	60S ribosomal protein L26 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl26 PE=2 SV=1	14350.10	2.00	28.00	40.00
sp|Q9UUG1| RPF2_SCHPO	Ribosome production factor 2 homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC926.08c PE=1 SV=1	36103.10		6.00	
sp|O94284| HMT2_SCHPO	Sulfide:quinone oxidoreductase, mitochondrial OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=hmt2 PE=2 SV=1	51614.40	1.00	11.00	2.00
sp|P09322| H4_SCHPO	Histone H4 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=hhf1 PE=1 SV=2	11434.50		1.00	1.00
sp|Q9P7X7| ARP4_SCHPO	Actin-related protein 4 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=alp5 PE=2 SV=1	48813.90		2.00	3.00
sp|P36621| CAP_SCHPO	Adenylyl cyclase-associated protein OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cap1 PE=1 SV=1	60295.60		6.00	2.00
sp|P78795| EIF3G_SCHPO	Eukaryotic translation initiation factor 3 subunit G OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif35 PE=1 SV=2	31511.00	1.00	8.00	7.00
sp|P09627| PMA1_SCHPO	Plasma membrane ATPase 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pma1 PE=1 SV=1	99982.50		1.00	1.00
sp|Q9UQW6| THIL_SCHPO	Acetyl-CoA acetyltransferase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=erg10 PE=2 SV=1	41026.10		2.00	1.00
sp|O94316| SN114_SCHPO	Pre-mRNA-splicing factor cwf10 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cwf10 PE=1 SV=2	111398.00		3.00	2.00
sp|Q10284| HSP75_SCHPO	Heat shock protein sks2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sks2 PE=1 SV=2	67272.00	1.00	29.00	26.00
sp|Q9URV1| PLG7_SCHPO	Putative phospholipase A2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=plg7 PE=2 SV=1	49887.30		6.00	3.00
sp|O36018| BUD22_SCHPO	BUD22 family protein C4F10.06 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC4F10.06 PE=2 SV=1	44910.30		9.00	6.00
sp|Q10232| RL9A_SCHPO	60S ribosomal protein L9-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl9a PE=2 SV=1	21517.40		4.00	2.00
sp|P78954| IF4E1_SCHPO	Eukaryotic translation initiation factor 4E-1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif451 PE=2 SV=1	24978.50		3.00	3.00
sp|O94362| YHOF_SCHPO	Uncharacterized GTP-binding protein C428.15 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC428.15 PE=2 SV=1	44506.10		2.00	1.00
sp|Q9URX6| RL31_SCHPO	60S ribosomal protein L31 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl31 PE=2 SV=1	13271.10	2.00	12.00	10.00
sp|P10989| ACT_SCHPO	Actin OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=act1 PE=1 SV=1	41791.90	5.50	45.00	38.00

sp|O94374| IMA2_SCHPO	Importin subunit alpha-2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=imp1 PE=1 SV=1	59905.20		3.00	2.00
sp|Q9P3T6| RS5B_SCHPO	40S ribosomal protein S5-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps5b PE=2 SV=1	22379.00		7.00	2.00
sp|O42644| PYRG_SCHPO	CTP synthase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ura7 PE=2 SV=1	66915.20		6.00	7.00
sp|O94378| YG0C_SCHPO	Uncharacterized protein C1604.12 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC1604.12 PE=1 SV=1	93374.10		2.00	3.00
sp|O42846| RL34A_SCHPO	60S ribosomal protein L34-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl34a PE=2 SV=1	12810.10		5.00	3.00
sp|P78958| G3P1_SCHPO	Glyceraldehyde-3-phosphate dehydrogenase 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tdh1 PE=1 SV=1	35901.30	5.00	84.00	77.00
sp|Q10289| FAS2_SCHPO	Fatty acid synthase subunit alpha OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=fas2 PE=1 SV=2	202367.00		20.00	2.00
sp|O42849| SYFB_SCHPO	Phenylalanine--tRNA ligase beta subunit OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=frs1 PE=2 SV=1	66896.70		4.00	
sp|Q9USW5| RS4B_SCHPO	40S ribosomal protein S4-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps4b PE=1 SV=1	29676.20		1.50	1.50
sp|P87158| RS4A_SCHPO	40S ribosomal protein S4-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps4a PE=2 SV=2	29706.30			1.00
sp|P15398| RPA1_SCHPO	DNA-directed RNA polymerase I subunit rpa1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpa1 PE=1 SV=2	189415.00		3.00	
sp|O42867| RL23_SCHPO	60S ribosomal protein L23 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl23a PE=2 SV=1	14890.90		9.00	4.00
sp|Q10330| RL25A_SCHPO	60S ribosomal protein L25-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl25a PE=1 SV=1	15843.90		8.12	5.59
sp|Q10349| SERC_SCHPO	Putative phosphoserine aminotransferase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC1F12.07 PE=2 SV=1	42824.80		3.00	1.00
sp|Q10353| RLP24_SCHPO	Ribosome biogenesis protein rlp24 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rlp24 PE=2 SV=1	22283.50		2.00	3.00
sp|O42935| RL32B_SCHPO	60S ribosomal protein L32-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl32b PE=2 SV=1	14453.90		5.50	5.00
sp|O42958| PAT1_SCHPO	DNA topoisomerase 2-associated protein pat1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC19G7.10c PE=2 SV=2	84190.80		1.00	2.00
sp|O13282| TAF5_SCHPO	Transcription initiation factor TFIID subunit 5 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=taf5 PE=1 SV=1	72443.00			1.00
sp|O60144| RS16_SCHPO	40S ribosomal protein S16 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps16a PE=2 SV=1	15403.40		3.00	
sp|Q09692| SYWC_SCHPO	Tryptophan--tRNA ligase, cytoplasmic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=wrs1 PE=1 SV=1	44935.70		2.00	
sp|O74399| ERB1_SCHPO	Ribosome biogenesis protein erb1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=erb1 PE=1 SV=2	83019.60		2.00	
sp|O74751| YHFA_SCHPO	Uncharacterized protein C1734.10c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC1734.10c PE=2 SV=1	37438.90		2.00	
sp|Q10270| MET16_SCHPO	Probable phosphoadenosine phosphosulfate reductase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC13G7.06 PE=2 SV=1	30577.60		1.00	1.00
sp|O94268| NOP2_SCHPO	Putative ribosomal RNA methyltransferase nop2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBP8B7.20c PE=1 SV=1	69000.70		2.00	1.00
sp|O94512| OBP1_SCHPO	Oxysterol-binding protein-like protein 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=obp1 PE=1 SV=1	58745.60		3.00	2.00
sp|Q9UT56| RS26A_SCHPO	40S ribosomal protein S26-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps26a PE=2 SV=1	13885.50		2.00	1.00
DECOY2_sp| O74991| POF3_SCHPO	reversed: F-box/TPR repeat protein pof3 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pof3 PE=1 SV=1	66286.20		3.00	2.00
sp|O94686| RL43B_SCHPO	60S ribosomal protein L43-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl43b PE=2 SV=1	10373.50		3.50	3.00

sp|Q9P7D1| YOFI_SCHPO	Uncharacterized protein P4H10.18c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBP4H10.18c PE=2 SV=1	35911.80	1.00	1.00
sp|O94305| MGLL_SCHPO	Putative monoglyceride lipase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPCC5E4.05c PE=1 SV=1	41651.90	2.00	1.00
sp|Q9USK6| CBP3_SCHPO	CBP3-like protein OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cbp3 PE=2 SV=1	32987.10		2.00
sp|P05767| RL39_SCHPO	60S ribosomal protein L39 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl39 PE=2 SV=3	6304.50	3.00	3.00
sp|Q9P7Y7| P5CR_SCHPO	Pyrroline-5-carboxylate reductase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=pro3 PE=2 SV=1	29891.30		1.00
sp|P59289| RL37A_SCHPO	60S ribosomal protein L37-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl37a PE=1 SV=2	9954.05	10.50	17.50
sp|G2TRN6| YLZ1_SCHPO	Putative aryl-alcohol dehydrogenase C750.01 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC750.01 PE=3 SV=1	39943.40	1.00	1.00
sp|O14179| SDO1_SCHPO	Ribosome maturation protein sdo1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sdo1 PE=2 SV=1	28093.70	2.00	1.00
sp|O14295| PLR1_SCHPO	Pyridoxal reductase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=plr1 PE=1 SV=1	36845.10	1.00	1.00
sp|Q10137| SEC14_SCHPO	Sec14 cytosolic factor OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec14 PE=2 SV=1	32761.40	1.00	2.00
sp|P40375| MAOX_SCHPO	NAD-dependent malic enzyme OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=mae2 PE=1 SV=1	62585.40	1.00	
sp|O14181| YDS5_SCHPO	Putative 60S ribosomal protein C4F8.05c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC4F8.05c PE=2 SV=1	13802.10	2.00	
sp|Q10334| ALAT_SCHPO	Putative alanine aminotransferase OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC582.08 PE=2 SV=1	56681.40	4.00	
sp|O74791| GRN1_SCHPO	GTPase grn1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=grn1 PE=1 SV=1	52474.30	1.00	
sp|O94699| YG1G_SCHPO	Mitochondrial outer membrane protein C83.16c OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC83.16c PE=2 SV=1	63534.60	1.00	
sp|P87317| YB2G_SCHPO	Uncharacterized protein C31F10.16 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBC31F10.16 PE=2 SV=2	74155.60	1.00	
sp|Q9P546| RSSA2_SCHPO	40S ribosomal protein S0-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps0b PE=1 SV=1	31437.00	1.00	
sp|Q9HDW6| VATC_SCHPO	V-type proton ATPase subunit C OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAPB2B4.05 PE=1 SV=1	44387.50	1.00	1.00
sp|P35551| FBRL_SCHPO	rRNA 2'-O-methyltransferase fibrillarin OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=fib1 PE=1 SV=1	32074.80	1.00	
sp|O74364| ADN1_SCHPO	Adhesion defective protein 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=adn1 PE=2 SV=1	43863.70		1.00
sp|Q96WV5| COPA_SCHPO	Putative coatomer subunit alpha OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPBPJ4664.04 PE=1 SV=1	136499.00	2.00	
sp|Q9HGL8| RL43A_SCHPO	60S ribosomal protein L43-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl43a PE=2 SV=1	10415.50	2.50	3.00
sp|P39825| PROF_SCHPO	Profilin OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=cdc3 PE=1 SV=1	13420.90	1.00	1.00
sp|Q09904| NU124_SCHPO	Nucleoporin nup124 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=nup124 PE=1 SV=1	124098.00		1.00
sp|P31411| VATB_SCHPO	V-type proton ATPase subunit B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=vma2 PE=1 SV=2	55896.50	1.00	2.00
sp|O74391| RL25B_SCHPO	60S ribosomal protein L25-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl25b PE=2 SV=1	15775.80	1.00	
sp|Q10055| FAL1_SCHPO	ATP-dependent RNA helicase fal1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=tif412 PE=1 SV=1	44731.80	1.00	
sp|O94582| TRPE_SCHPO	Probable anthranilate synthase component 1 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=trp3 PE=1 SV=1	54997.00	1.00	1.00
sp|Q10281| GBLP_SCHPO	Guanine nucleotide-binding protein subunit beta-like protein OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rkp1 PE=1 SV=3	34883.30		1.00

P13645	SWISS_PROT_P13645 Tax_Id=9606 Gene_Symbol=KRT10 Keratin, type I cytoskeletal 10	59565.00	26.00	43.98	24.99
P08779	SWISS_PROT_P08779 Tax_Id=9606 Gene_Symbol=KRT16 Keratin, type I cytoskeletal 16	51308.20	3.50	10.50	2.50
P02533	SWISS_PROT_P02533 Tax_Id=9606 Gene_Symbol=KRT14 Keratin, type I cytoskeletal 14	51661.50	1.00	8.00	4.00
Q9QWL7	SWISS_PROT_Q9QWL7 Tax_Id=10090 Gene_Symbol=Krt17 Keratin, type I cytoskeletal 17	48204.10			3.00
P13646_1	SWISS_PROT_P13646_1 Tax_Id=9606 Gene_Symbol=KRT13 Isoform 1 of Keratin, type I cytoskeletal 13	49627.40	1.00	1.00	
Q15323	SWISS_PROT_Q15323 Keratin, type I cuticular HA1 (Hair keratin, type I HA1)	47279.10		0.98	
Q1RMN8	TREMBL_Q1RMN8 (Bos taurus) Similar to Immunoglobulin lambda_like polypeptide 1	24557.00	2.00	6.00	2.00
sp|O13972| GLYD_SCHPO	Probable serine hydroxymethyltransferase, cytosolic OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=SPAC24C9.12c PE=2 SV=1	51900.20		32.00	18.00
sp|O14049| RS8A_SCHPO	40S ribosomal protein S8-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps8a PE=2 SV=1	22597.00		3.00	
sp|O14112| RS10A_SCHPO	40S ribosomal protein S10-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps10a PE=2 SV=1	16283.50		5.00	
sp|O42706| RL21B_SCHPO	60S ribosomal protein L21-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl21b PE=2 SV=2	18359.70		1.00	4.00
sp|O42984| RS17A_SCHPO	40S ribosomal protein S17-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps17a PE=2 SV=1	15522.40		5.00	6.00
sp|O42991| RL16B_SCHPO	60S ribosomal protein L16-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl16b PE=1 SV=1	22174.30	0.50	5.50	5.00
sp|O42848| RL16A_SCHPO	60S ribosomal protein L16-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl16a PE=1 SV=1	22009.20		4.00	1.00
sp|O59675| RS9B_SCHPO	40S ribosomal protein S9-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps9b PE=1 SV=3	22249.20		1.00	2.00
sp|O59855| HSP72_SCHPO	Probable heat shock protein ssa2 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=ssa2 PE=1 SV=3	70297.80	1.00	41.00	36.00
sp|O74175| RL13_SCHPO	60S ribosomal protein L13 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl13 PE=1 SV=1	23550.80	3.00	37.00	42.00
sp|O74306| RL5B_SCHPO	60S ribosomal protein L5-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl5b PE=2 SV=1	33501.20	1.00	16.00	7.00
sp|O74445| PRS10_SCHPO	Probable 26S protease subunit rpt4 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpt4 PE=2 SV=2	43635.00		8.00	2.00
sp|O74894| PRS6B_SCHPO	26S protease regulatory subunit 6B homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpt3 PE=2 SV=1	43579.80		5.00	1.00
sp|O74836| RL1B_SCHPO	60S ribosomal protein L1-B OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl1b PE=1 SV=1	23938.10		5.25	3.25
sp|O14363| RL1A_SCHPO	60S ribosomal protein L1-A OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rpl1a PE=1 SV=1	23905.10		1.00	
sp|O74893| RS20_SCHPO	40S ribosomal protein S20 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=rps20 PE=1 SV=1	13553.40		4.00	4.00
sp|O94407| SEC65_SCHPO	Signal recognition particle sec65 subunit OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) GN=sec65 PE=2 SV=2	23521.40	1.00	2.00	2.00

APPENDIX II: LIST OF PUBLICATIONS

1. Watts BR, Wittmann S, Kilchert C, Vasiljeva L. 2016. Intergenic non-coding transcription utilises an RNA-based mechanism to drive transcriptional interference of neighbouring genes. (manuscript in preparation)

2. Wittmann S, Renner M, Watts BR, Adams O, Huseyin M, Baejen C, El Omari K, Kilchert C, Heo DH, Kecman T, Cramer P, Grimes JM, Vasiljeva L. 2016. The conserved termination factor Seb1 bridges RNA polymerase II and nascent RNA. *Nat Commun* (revised and resubmitted)

3. Kilchert C, Wittmann S, Vasiljeva L. 2016. The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol* **17**: 227–239.

4. Kilchert C, Wittmann S, Passoni M, Shah S, Granneman S, Vasiljeva L. 2015. Regulation of mRNA Levels by Decay-Promoting Introns that Recruit the Exosome Specificity Factor Mmi1. *Cell Rep* **13**: 2504–2515.

5. Shah S, Wittmann S, Kilchert C, Vasiljeva L. 2014. lncRNA recruits RNAi and the exosome to dynamically regulate pho1 expression in response to phosphate levels in fission yeast. *Genes Dev* **28**: 231–244.