

TRANSCRIPTION TERMINATION BY RNA POLYMERASE II

Eleanor White

University of Oxford
Sir William Dunn School of Pathology
and
Pembroke College

Thesis submitted for the degree of Doctor of Philosophy at the University of
Oxford

Trinity Term 2011

ABSTRACT

RNA Polymerase II (Pol II) is responsible for the transcription of all protein-encoding genes. Pol II termination is dependent on RNA processing signals (both terminal intron splice sites, and cleavage and polyadenylation signals) as well as specific terminator elements located downstream of the poly(A) site. Detailed analysis of the human β -globin gene terminator has shown that it contains a sequence-specific region that promotes rapid Co-Transcriptional Cleavage (CoTC) of the nascent transcript – an essential but not well understood step in the human β -globin gene termination process. In the first part of this thesis, the role of sequences within this CoTC-mediated terminator element in the termination process is investigated. Analysis of mutant terminator sequences indicate that homopolymer A tracts are important for Pol II termination. The second part of this study focuses on identifying the activity responsible for CoTC, by using the yeast *S. pombe* as a tool for genetic analysis. The results indicate that the human β -globin gene terminator is inefficient in *S. pombe*, suggesting that a mammalian specific factor(s) are required. In the final part of this study, I describe an investigation into the possibility that the exosome subunit Dis3 or the RNase III enzyme Dicer are involved in CoTC mediated transcription termination. While Dis3 is not involved in the CoTC process my results on Dicer may imply a significant role. Lastly, I present a preliminary investigation into the effect of pre-mRNA processing and the Carboxyl Terminal Domain (CTD) of Pol II on CoTC activity.

ACKNOWLEDGEMENTS

I would first like to express my profound gratitude to my supervisors, Professor Nicholas Proudfoot and Dr. Michael Dye, for giving me the opportunity to carry out this project, and for all their help and support: I learned many valuable lessons over the past four years, and for this I will be forever grateful. Special thanks go also to Dr Monika Gullerova for her help with the yeast work (not to mention all the fun we had in the back bay!) and to all the members of the NJP lab, both past and present, who helped to make my DPhil such an enjoyable experience.

A big thank you to my wonderful friends for their encouragement when things were tough, and, most of all, for making my time in Oxford so much fun. To my best friend Andreas, I also owe an apology for ruining the ‘grand plan’ – for what it is worth, I intend to make up for it soon!

To Benjamin, a very special thank you for always being there for me, supporting me, and, teaching me the importance of the Oxford comma! I could not have done this without you.

Finally, I would like to thank my family, especially my Mum, Dad and Theia, without whom I would never have made it this far. I am grateful for the love, support, and, most of all, the many sacrifices you made for me. I hope one day you will think that it was all worth it!

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iii
GLOSSARY OF ABBREVIATIONS	vii
1. INTRODUCTION	1
1.1 Pre-mRNA 3' end processing – cleavage and polyadenylation	4
1.1a Mammalian 3' end processing	4
1.1b <i>S. cerevisiae</i> 3' end processing	8
1.2 The role of the Pol II CTD in 3' end formation	9
1.2a CTD kinases: Ser5 and Ser2 specific	9
1.2b CTD phosphatases: Ser5 and Ser2 specific	11
1.2c CTD Ser7 phosphorylation and other modifications	11
1.3 Pol II termination	14
1.3a Models for termination	14
1.3b Termination and anti-termination factors	15
1.3c The molecular torpedo	18
1.3d Hybrid model	21
1.3e Terminator sequence elements	22
1.3e (i) Pause elements	22
1.3e (ii) CoTC elements	24
1.3e (iii) Other terminator elements	28
1.4 Connections between capping, splicing, 3' end processing and termination	29
1.5 Transcription termination of other eukaryotic Pol II transcribed genes	31
1.5a Histone gene 3' end processing and termination	31
1.5b snRNA gene termination	34
1.5c snoRNA gene termination	36
1.5d miRNA gene termination	41
1.6 Transcription termination and chromatin structure	43
1.6a Nucleosome positioning	43

1.6b Chromatin remodelling	45
1.7 The exosome	46
1.8 RNase III enzymes	49
1.8a Dicer	50
1.9 Transcription termination by RNA Pol I	50
1.10 Transcription termination by RNA Pol III	56
1.11 Transcription termination in bacteria	58
1.11a Intrinsic terminators	59
1.11b Factor-dependent terminators	59
1.11c General transcription termination lessons learned from bacteria	60
1.12 Common features of transcriptional termination	62
1.12a Fail-safe termination	62
1.12b Pausing and cleaving	64
1.13 Aims of thesis	66
2. MATERIALS AND METHODS	67
2.1 Reagents	68
2.2 Enzymes	69
2.3 Antibodies	70
2.4 Yeast strain	71
2.5 Nucleic Acid markers	71
2.6 Tissue Culture solutions	71
2.7 Yeast Media	71
2.8 Solutions and Buffers	72
2.9 Plasmids and probes	80
2.10 Table of oligonucleotides	84
2.11 DNA manipulations	87
2.12 Tissue Culture	89
2.13 Yeast cultivation	90
2.14 Steady state RNA analysis	91
2.15 Nascent RNA analysis	96
2.16 Protein analysis	98

3. INVESTIGATION OF <i>CIS</i> ELEMENT FUNCTION IN COTC-MEDIATED TRANSCRIPTIONAL TERMINATION	99
3.1 Introduction	100
3.2 Deletion analysis of the Terminator	102
3.3 Homopolymer tracts in the Terminator sequence are important for transcriptional termination	112
3.4 The homopolymer tracts are important for CoTC activity of the human β -globin Terminator	118
3.5 Multiple cleavages of the human β -globin Terminator transcript	122
3.6 Analysis of stable β -globin integrants	125
3.7 A stretch of 30 A residues is not sufficient to cause termination	128
3.8 Discussion	130
4. DESIGN OF A GENETIC SCREEN TO IDENTIFY CANDIDATES RESPONSIBLE FOR COTC ACTIVITY	133
4.1 Introduction	134
4.2 A reporter gene designed to identify factors involved in CoTC mediated transcriptional termination	135
4.3 The Terminator does not terminate efficiently in <i>S. pombe</i>	139
4.4 The human β -globin termination signals invoke antisense transcription in <i>S. pombe</i>	146
4.5 RT-PCR of the mature β -globin message is RT-dependent and primer-independent	149
4.6 Discussion	153
5. INVESTIGATION INTO THE REQUIREMENTS FOR COTC ACTIVITY	159
5.1 Introduction	160
5.2 Dis3 as a possible candidate for CoTC activity	160
5.3 Dicer and its involvement in termination of the human β -globin gene	164
5.4 Pol II CTD and its involvement in termination of the human β -globin gene	169
5.5 Polyadenylation and splicing are required for CoTC activity of the human	

β-globin gene terminator	175
5.6 Discussion	177
6. GENERAL DISCUSSION	180
7. REFERENCES	189

GLOSSARY OF ABBREVIATIONS

A	Adenine
ATP	Adenosine Triphosphate
ARE	AU Rich Element
bp	base pairs
C	Cytosine
CBC	Cap Binding Complex
Cdk	Cyclin-dependent kinase
cDNA	complementary DNA
CFIA	Cleavage Factor IA
CFIB	Cleavage Factor IB
CFIm	mammalian Cleavage Factor I
CFIIm	mammalian Cleavage Factor II
ChIP	Chromatin Immuno-Precipitation
Ci	Curie
CID	CTD Interacting Domain
CLIP	Crosslinking and Immunoprecipitation
CoTC	Co-Transcriptional Cleavage
CPF	Cleavage and Polyadenylation Factor
cps	counts per second
CPSF	Cleavage and Polyadenylation Specificity Factor
CstF	Cleavage stimulatory Factor
CTD	Carboxyl-Terminal Domain

CTP	Cytosine Triphosphate
CUTs	Cryptic Unstable Transcripts
DNA	Deoxyribonucleic Acid
DNA-PK	DNA- activated Protein Kinase
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide Triphosphate
DSE	Downstream Sequence Element
DTT	Dithiothreitol
DUE	Downstream U-rich Element
EC	Elongation Complex
EDTA	Ethylenediaminetetraacetic Acid (disodium salt)
EE	Efficiency Element
EMM	Edinburgh Minimal Medium
G	Guanine
g	g-force
GMP	Guanine Monophosphate
GTP	Guanosine Triphosphate
HCl	Hydrochloric acid
HDE	Histone Downstream Element
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HIV	Human Immunodeficiency Virus
HLF	Heat Labile Factor
hnRNP	heterogeneous nuclear Ribonucleo Protein

HSF	Heat Shock Factor
hscRACE	hybrid selection circular Rapid Amplification of Complementary deoxyribonucleic acid Ends
IP	Immuno-Precipitation
IVS	Intervening Sequence (intron)
kb	kilobase
LTR	Long Terminal Repeat
MEM	Minimal Essential Media
miRNA	micro RNA
ml	millilitre
mRNA	messenger RNA
MSA	Mouse Serum Albumin
Nab	Nuclear polyadenylated RNA binding protein
NELF	Negative Elongation Factor
Nrd1	Nuclear pre-mRNA down regulation
NRO	Nuclear Run-On
nt(s)	nucleotide(s)
NTP	Ribonucleoside Triphosphate
OD	Optical Density
ORF	Open Reading Frame
PABP	Poly(A) Binding Protein
PAGE	Polyacrylamide Gel Electrophoresis
PAP	Poly(A) Polymerase

PBS	Phosphate Buffered Saline
Pcf11	Polyadenylation cleavage factor 11
PCR	Polymerase Chain Reaction
PE	Positioning Element
PEG	Polyethylene Glycol
PET	PNS/EDTA/Trypsin
PIC	Pre-Initiation Complex
PIPES	Piperazine-N, N'-bis 2-ethanesulphonic acid
PMSF	Phenylmethanesulfonyl Fluoride
Pol	Polymerase
Poly(A)	Polyadenylation
PSF	Protein-associated Splicing Factor
P-TEFb	Positive Transcription Elongation Factor b
Py(A)n	Pyrimidine rich stretch at poly(A) site
Rai1	Rat interacting protein 1
rDNA	ribosomal DNA
RPA	RNase Protection Assay
rpm	revolutions per minute
RNA	Ribonucleic Acid
RNAi	RNA interference
RNase	Ribonuclease
rNTPs	Ribonucleotide Triphosphates

rRNA	ribosomal RNA
RT	Reverse Transcription
rut	Rho utilisation site
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	Sodium Dodecyl Sulphate
Ser2P	CTD Ser 2 phosphorylated
Ser5P	CTD Ser 5 phosphorylated
Ser7P	CTD Ser 7 phosphorylated
SETX	Senataxin
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear Ribonucleoprotein
<i>S. pombe</i>	<i>Schizosaccharomyces pombe</i>
ssDNA	single stranded DNA
Ssu72	Suppressor of Sua7, gene 2
T	Thymine
Tat	Trans-activator of transcription
TF	Transcription Factor
TRAMP	Trf4/Air2/Mtr4 Polyadenylation complex
tRNA	transfer RNA
USE	Upstream Sequence Element
UTP	Uridine Triphosphate

UTR	Untranslated Region
UUE	Upstream U-rich Element
UV	Ultraviolet light
WT	Wild Type
YES	Yeast Extract with Supplements

CHAPTER 1

INTRODUCTION

In eukaryotes, transcription is performed by three structurally related, DNA dependent RNA polymerases, namely RNA polymerase (Pol) I, Pol II and Pol III (Cramer et al. 2008). Pol I transcribes ribosomal RNA (rRNA) genes and Pol III transcribes small non-coding RNAs including transfer RNAs (tRNAs), U6 small nuclear RNA (snRNA), and 5S rRNA (reviewed in Paule and White, 2000). Pol II is responsible for the transcription of all protein-encoding genes and many non-coding RNAs, and is the focus of this thesis.

The transcription cycle is a highly complex process composed of three steps: initiation, elongation and termination. In the initiation step, Pol II, along with specific transcription initiation factors, is recruited to the promoter of a gene where the first few nucleotides of RNA are synthesised. Pol II may then escape the promoter and enter the transcription elongation phase in the second stage of the cycle, producing an RNA copy of the gene being transcribed. Finally, termination occurs when RNA synthesis stops and the nascent transcript and Pol II are released from the DNA template.

Transcription of protein-encoding genes by Pol II, and processing of the pre-messenger RNA (pre-mRNA), results in a functional messenger RNA (mRNA) that is translated into protein. The majority of pre-mRNAs undergo three processing reactions. The first involves the addition of a trimethyl cap structure at the 5' end of the pre-mRNA (reviewed by Shuman, 2001). The pre-mRNA is usually made up of exons and introns, which are coding and non-coding sequences, respectively. In the second processing reaction, introns are removed by splicing (reviewed by Matlin and Moore, 2007). Finally, at the 3' end of most pre-mRNAs, a cleavage event occurs

which is followed by addition of a poly(A) tail to the 5' cleavage product. Processing of the pre-mRNA is essential to protect the mRNA from degradation, target it to the cytoplasm, and promote its translation (Bernstein and Ross, 1989; Tarun and Sachs, 1996; Huang and Carmichael, 1996; Sachs et al. 1997; Bousquet-Antonelli et al. 2000; Hilleren et al. 2001; Saguez et al. 2008; Vodala et al. 2008). Pre-mRNA processing and transcription are interconnected and tightly regulated, allowing correct gene expression to occur (Proudfoot et al. 2002).

The main interest of this thesis is the termination step of the transcription cycle. Transcription termination is of vital importance for many reasons: it prevents interference with transcription of downstream genes (Greger et al. 1998) and is thought to allow recycling of Pol II for new rounds of transcription. Indeed, Mapendano et al. (2010) have demonstrated that termination increases transcription re-initiation. Moreover, transcriptional termination enhances mRNA and protein levels by enhancing splicing and preventing degradation of the nascent transcript (West and Proudfoot, 2009). In recent years, it has been shown that with some genes, transcriptional readthrough may generate antisense RNAs that can induce heterochromatin formation, which in turn affects gene expression (Camblong et al. 2007; Gullerova and Proudfoot, 2008). Pol II termination is tightly linked to 3' end processing and the linkage between these two processes is discussed in more detail below.

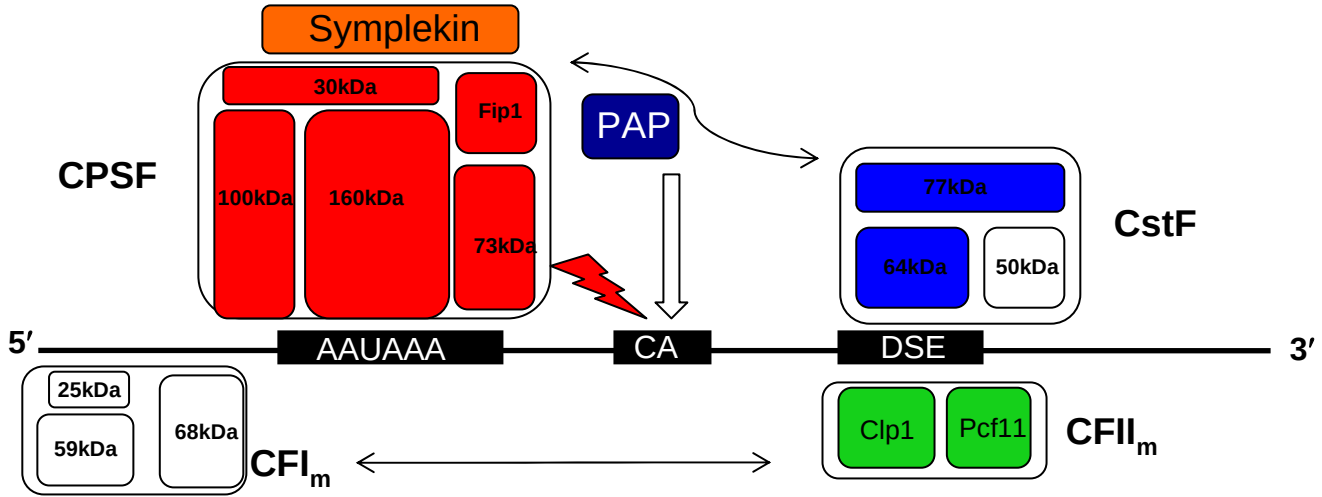
1.1 Pre-mRNA 3' end processing - cleavage and polyadenylation

1.1a Mammalian 3' end processing

The 3' end cleavage and polyadenylation reactions are directed by *cis* elements found in the untranslated region of all pre-mRNAs with the exception of the replication-dependent histone genes. In mammals, there are three sequence elements involved: the poly(A) signal, which is composed of a conserved hexanucleotide AAUAAA (Proudfoot and Brownlee, 1976), the cleavage site, which usually contains the dinucleotide CA (Sheets et al. 1990) at approximately 10-30 nucleotides downstream of the poly(A) signal, and a more variable GU/U rich downstream sequence element (DSE) found up to approximately 30 nucleotides downstream of the cleavage site (reviewed in Mandel et al. 2008) (Figure1.1a). In some genes, a U-rich upstream sequence element (USE) is also present (Hu et al. 2005; Moreira et al. 1998).

A seemingly simple two step process, transcript cleavage followed by the addition of a poly(A) tail, requires highly complex machinery, which consists of more than 14 polypeptides in mammals and 20 in yeast (reviewed in Mandel et al. 2008). A recent analysis of the factors associated with the mammalian core machinery identified approximately 85 proteins, many of which mediate crosstalk with other processes (Shi et al. 2009). The core mammalian 3' end processing complex contains the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CFIm), cleavage factor II (CFIIm), poly(A) polymerase (PAP), poly(A) binding protein (PABP), and symplekin (Figure1.1a).

A) Mammals



B) *S. cerevisiae*

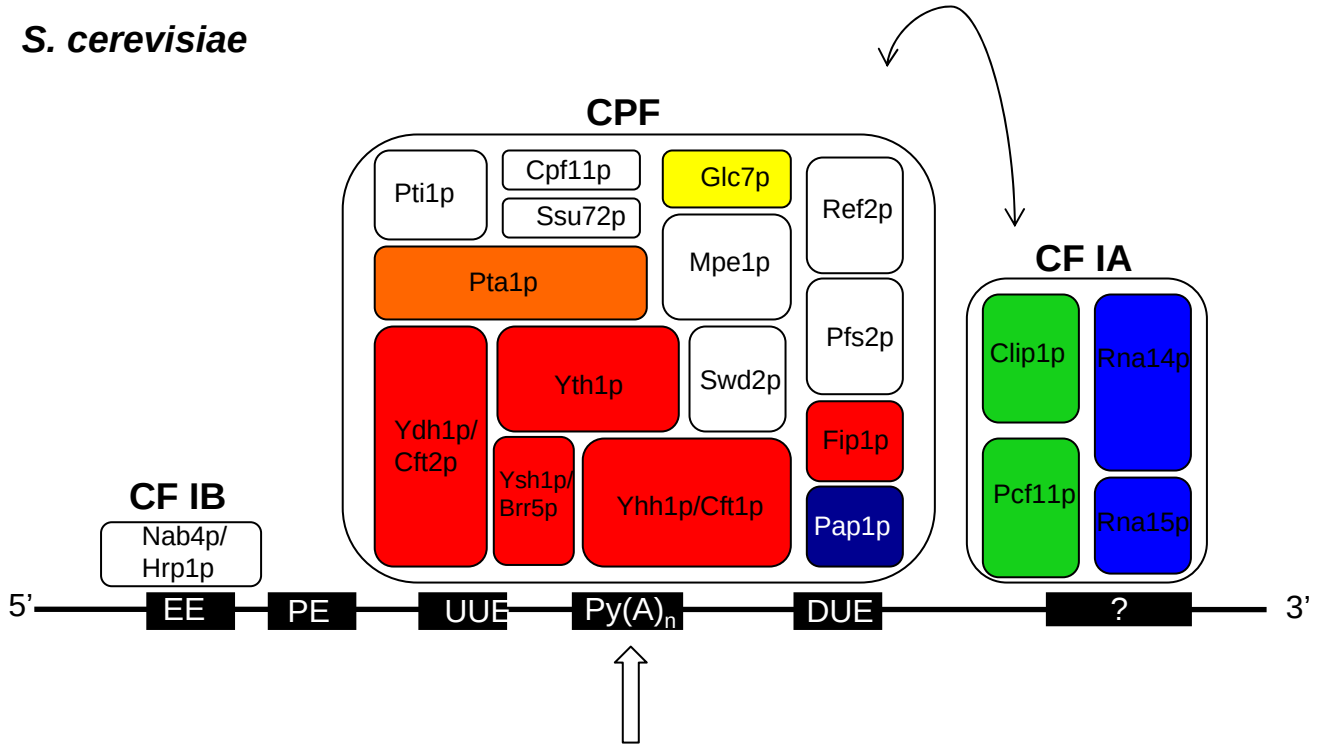


Figure 1.1: Mammalian and *S. cerevisiae* 3' end processing complexes (adapted from Proudfoot., 2004). The pre-mRNAs (black lines) and *cis*-acting sequence elements (labelled black boxes, and referred to in the text) are indicated. *Trans*-acting factors are positioned above their points of contact with pre-mRNA. Contacts between different complexes are emphasised by curved arrows. Sites of cleavage and poly(A) addition are indicated by vertical arrows. Homologous factors between mammalian and *S. cerevisiae* complexes are denoted by the same colours with the exception of non-homologous factors shown by white boxes. Yhh1, Ydh1, Ysh1 and Yth1 are homologous to CPSF160, CPSF100, CPSF73 and CPSF30 respectively. Rna14 and Rna15 are homologous of CstF77 and CstF64 respectively. Symplekin is homologous to Pta1. Diagram not drawn to scale.

CPSF contains 5 subunits, CPSF160, CPSF100, CPSF73, CPSF30, and Fip1 (Murthy and Manley, 1992; Kaufmann et al. 2004), all of which are required for both cleavage and polyadenylation (Christofori and Keller, 1988). The AAUAAA signal is recognised and bound by CPSF160, and it has been suggested that all the subunits of CPSF contribute to the efficiency of binding (Keller et al. 1991; Murthy and Manley, 1995). CstF is composed of three subunits, CstF77, CstF50, and CstF64, the latter of which makes direct contact with the GU rich DSE (Beyer et al. 1997; Takagaki and Manley, 1997). CstF is necessary for cleavage but not polyadenylation (Takagaki et al. 1989). The protein symplekin is thought to act as a scaffold by bridging between CPSF160 and CstF64 (Hofmann et al. 2002; Takagaki and Manley, 2000). An additional factor binding to the pre-mRNA at GU-rich sites upstream of the poly(A) site is CFIm, which has been implicated in the regulation of poly(A) site selection (Ruegsegger et al. 1996; Venkataraman et al. 2005; Brown and Gilmartin, 2003).

CFIIm contains the highly conserved proteins Pcf11 and Clp1, the latter of which interacts with CFIm and CPSF complexes (de Vries et al. 2000). In addition, Clp1 has a kinase activity, which Weitzer and Martinez (2007) propose maintains a 5'-phosphate on the downstream cleavage product aiding termination. PAP is responsible for polyadenylating the cleaved pre-mRNA, whilst PABP binds to the poly(A) tail protecting the pre-mRNA from exonucleolytic degradation (Bernstein and Ross, 1989). PABP also enhances PAP activity and determines the length of the poly(A) tail (Kerwitz et al. 2003; Kuhn et al. 2009). Both proteins are required for polyadenylation but only PAP is required for cleavage (Colgan and Manley, 1997; Minvielle-Sebastia et al. 1997).

Assembly of the 3' end processing machinery on the pre-mRNA is a cooperative process. CPSF160, which directly recognises AAUAAA, binds CstF77, stabilising the association of both CPSF and CstF with the poly(A) signal. In addition, the interaction between the two complexes via symplekin further stabilises their association (Takagaki and Manley, 2000). CPSF160 also interacts with CPSF100, Fip1, CFIm and PAP, reinforcing the interaction of CPSF with the pre-mRNA and recruiting PAP to the complex (Colgan and Manley, 1997; Kaufmann et al. 2004; Murthy and Manley, 1995).

Endonucleolytic RNA cleavage at the poly(A) addition site is thought to be carried out by CPSF73 (Ryan et al. 2004, Mandel et al. 2006), which is also likely to be responsible for histone pre-mRNA cleavage (Dominski et al. 2005; Kolev et al. 2008; Ryan et al. 2004). CPSF73 has a metallo β lactamase domain (MLD) and is a member of the β -CASP protein superfamily (Callebaut et al. 2002). Interestingly, CPSF100 also contains an MLD, albeit with inactivating mutations in its catalytic centre. Kolev et al. (2008), using an *in vitro* system, have shown that conserved residues of the MLD of CPSF73 and CPSF100 are required for the endonucleolytic cleavage of histone pre-mRNAs, and by extension, this is likely to be the case for all other mRNAs. Following cleavage at the poly(A) addition site, the 5' cleavage product is polyadenylated whilst the 3' cleavage product is degraded.

It has long been established that 3' end processing is required for transcription termination. The first evidence for this came from studies on a rare form of α -thalassemia, where the poly(A) signal of the $\alpha 2$ -globin gene was mutated. In this mutant gene, both 3' end processing and termination are disrupted, providing evidence

that 3' end processing and termination are coupled events (Whitelaw and Proudfoot, 1986). It was later revealed that the largest subunit of Pol II, Rbp1, with its unique carboxyl terminal domain (CTD), is required for pre-mRNA processing and transcription termination. It was suggested that the CTD recruited 3' processing factors to the elongating Pol II complex (McCracken et al. 1997; Zhao et al. 1999). Furthermore, it was shown *in vitro* that 3' processing reactions were indeed enhanced by the Pol II CTD and, hence, it was suggested that the CTD itself acts as a component of the 3' processing machinery (Hirose and Manley, 1998). Moreover, the CTD has been shown to be required for coupling transcription and pre-mRNA processing (Bird et al. 2004) by binding to proteins involved in mRNA biogenesis (reviewed in Meinhart et al. 2005). The role of the CTD is discussed in the next section.

1.1b *S. cerevisiae* 3' end processing

In *S. cerevisiae*, 3' end cleavage and polyadenylation signals appear to be more variable than in mammalian genes. Four sequence elements have been identified: an AU-rich efficiency element (EE), an AU-rich positioning element (PE), the cleavage site found approximately 10-30 nucleotides downstream of PE, and upstream and downstream U-rich elements (UUE and DUE respectively) flanking the cleavage site (reviewed in Proudfoot, 2004; Mandel et al. 2008) (Figure 1.1b).

Although the *cis* elements required for 3' processing are not as well defined as in mammals, the yeast 3' processing complex shares many similarities with the mammalian complex (reviewed in Proudfoot, 2004; Mandel et al. 2008). It is made up of three complexes, namely cleavage and polyadenylation factor (CPF), cleavage

factor IA (CFIA), and cleavage factor IB (CFIB). CPF contains homologues to mammalian CPSF components; CFIA contains homologues to CstF64, CstF77 and CFIm; and CFIB is similar to CFIm (reviewed in Proudfoot, 2004) (Figure 1.1b).

In yeast, the poly(A) tail is added by PAP. *S. cerevisiae* also has non-canonical nuclear poly(A) polymerases called Trf4 and Trf5, which form part of the Trf-Air-Mtr4 polyadenylation (TRAMP) complex, and function in RNA surveillance (Haracska et al. 2005; Eggecioglu et al. 2006; Houseley and Tollervey, 2006). There are a number of Trf4 homologues in humans, possibly with conserved functions (West et al. 2006a).

1.2 The role of the Pol II CTD in 3' end formation

1.2a CTD kinases: Ser5 and Ser2 specific

The CTD of the Rpb1 subunit of Pol II is a flexible domain that extends from the core enzyme close to the RNA exit channel (Cramer et al. 2001). It is composed of tandemly repeated heptapeptides (25 or 26 in yeast and 52 in mammals) with the consensus sequence Tyr₁-Ser₂-Pro₃-Thr₄-Ser₅-Pro₆-Ser₇. A range of reversible modifications of the CTD have been described. Phosphorylation is particularly important for correct gene expression and has been studied extensively (reviewed in Egloff and Murphy, 2008).

The CTD displays different phosphorylation patterns during different stages of the transcription cycle and changes in these patterns alter the affinity of CTD binding proteins, suggesting that Pol II may be regulated by a CTD code (Egloff and Murphy,

2008; Buratowski, 2003; 2009). Initially, Pol II is recruited to the promoter in an unphosphorylated form (Pol IIA). Serine 5 is then phosphorylated by cyclin-dependent kinase 7 (Cdk7) (Kin28 in yeast), a subunit of the general transcription factor TFIIF. ChIP experiments have shown that Serine 5 phosphorylation (Ser5P) levels are high at the 5' end of genes, usually the first few hundred nucleotides, and then drop further downstream (Komarnitsky et al. 2000). Ser5P facilitates recruitment of capping enzymes to the complex (Komarnitsky et al. 2000; Schroeder et al. 2000) and the methyltransferase Set1 (Ng et al. 2003), which trimethylates Lysine 4 on histone 3, a mark seen near promoters of active genes (reviewed in Hampsey and Reinberg, 2003). As Pol II progresses along the gene, Serine 2 and Serine 7 are phosphorylated. Serine 2 is phosphorylated by the Cdk9 kinase subunit of the positive elongation factor P-TEFb, which also phosphorylates the elongation factor DSIF (Spt4/Spt5) and NELF (reviewed in Peterlin and Price, 2006). In *S. cerevisiae* the kinase Ctk1 phosphorylates Ser2 with the help of Bur1, the latter being responsible for DSIF phosphorylation (Cho et al. 2001; Zhou et al. 2009; Liu et al. 2009; Qiu et al. 2009). Ser2 phosphorylation (Ser2P) increases towards the 3' end of the gene and couples transcription and 3' end formation by enhancing interaction of the CTD with 3' end processing factors. Indeed, inhibition of P-TEFb or deletion of Ctk1 leads to 3' processing defects (Ahn et al. 2004; Ni et al. 2004). The *S. cerevisiae* CFIA complex contacts the CTD directly and one of its subunits, polyadenylation cleavage factor 11 (Pcf11), binds to Ser2P CTD *in vitro* (Barilla et al. 2001; Licatalosi et al. 2002; Ahn et al. 2004; Lunde et al. 2010), as does the Rat1 interacting factor Rtt103 (Kim et al. 2004b; Lunde et al. 2010). Yhh1, the CPSF160 homologue, and Ydh1, the CPSF100 homologue, also bind to phosphorylated CTD (Dichtl et al. 2002; Kyburz et al. 2003) but to which form they bind – Ser2 or Ser5 – has yet to be determined. Similarly,

CPSF and CstF have both been shown to bind to the CTD (McCracken et al. 1997). In addition, the H3K36 methyltransferase Set2 binds to Ser5 and Ser2 phosphorylated CTD (Hampsey and Reinberg, 2003; Kizer et al. 2005).

1.2b CTD phosphatases: Ser5 and Ser2 specific

Phosphatases are also involved in establishing the correct phosphorylation pattern on the CTD. Ssu72 (SCP1 in humans), a component of the yeast 3' end processing complex, dephosphorylates Ser5 (Meinhart et al. 2005) as does Rtr1 (RPAP2 in humans) (Mosley et al. 2009; Egloff et al. 2011). The specificity of another phosphatase, Fcp1 is unclear as it has been reported to dephosphorylate both Ser2 and Ser5 (Cho et al. 1999; Lin et al. 2002), Ser5 but not Ser2 (Kong et al. 2005) and, lastly, Ser2 but not Ser5 (Cho et al. 2001; Hausmann and Shuman, 2002; Meinhart et al. 2005). ChIP experiments have shown Ssu72 localises to the 3' end of genes (Nedea et al, 2003) but can also be detected at promoter regions (Ansari and Hampsey, 2005) suggesting that Ssu72 may be involved in coupling termination with reinitiation of transcription via gene looping. Dephosphorylated and Ser2P CTD is detected near transcription termination sites, suggesting that CTD dephosphorylation occurs before or during Pol II release (Lian et al. 2008). Dephosphorylation of the CTD is thought to facilitate Pol II recycling.

1.2c CTD Ser7 phosphorylation and other modifications

A recently discovered CTD phosphorylation event is that of Ser7 of the heptapeptide repeat. It has been reported that Ser7 can be phosphorylated by the Cdk7 subunit of the TFIIH general transcription factor (Akhtar et al. 2009; Glover-Cutter et al. 2009; Kim et al. 2009; Boeing et al. 2010). In addition, Bur1 kinase appears to act as a Ser7

kinase in yeast (Tietjen et al. 2010). Recent studies have reported that Cdk9 is able to phosphorylate Ser7 *in vitro* (Glover-Cutter et al. 2009; Kim et al. 2009), and that the DNA-activated kinase DNA-PK is the major Ser7 kinase in HeLa cell nuclear extracts (Egloff et al. 2010).

High levels of Ser7 phosphorylation (Ser7P) were detected on Pol II actively transcribing protein encoding genes and spliceosomal small nuclear RNAs (snRNAs) (Chapman et al. 2007; Egloff et al. 2007). Ser7P has also been detected at high levels at the promoter region of the T-cell receptor β gene and peaks at the 3' end of the gene, which could indicate an involvement of Ser7P in 3' end processing and/or termination (Chapman et al. 2007). In two studies in *S. cerevisiae*, Ser7P was detected at gene promoters and throughout the coding region (Mayer et al. 2010; Tietjen et al. 2010) and, a genome-wide study in *S. cerevisiae* shows Ser7P to be enriched over 5' ends, introns and 3' ends (Kim et al. 2010). Despite these findings, a role for Ser7P in transcription of protein encoding genes has yet to be elucidated. In contrast, it has been shown that the presence of Ser7P on Pol II transcribing snRNA genes facilitates recruitment of the Integrator complex, which is involved in 3' end formation of snRNA genes (Baillat et al. 2005; Egloff et al. 2007).

A series of genome-wide studies in *S. cerevisiae* analysing the phosphorylation patterns of Pol II CTD, as well as the association of Pol II and termination factors with chromatin, indicate that enrichment of Ser5P and Ser7P and a low level of Ser2P, serve as signals for early transcription events (Kim et al. 2010; Mayer et al. 2010; Tietjen et al. 2010). Other CTD modifications and/or termination signals may be used to signal later transcription events.

The less discussed CTD modifications include phosphorylation of Tyrosine 1 and Threonine 4, both of which have been detected *in vivo* (Zhang and Corden, 1991; Baskaran et al. 1993; Hsin et al. 2011). The CTD can also be modified by the addition of monosaccharide N-acetylglucosamine to hydroxyl groups of serine and threonine residues (Kelly et al. 1993), a modification not detected on the phosphorylated form of Pol II. This modification is detected at low levels and a role has not yet been described in transcription. Furthermore, the two Ser-Pro bonds in the heptapeptide repeat can exist in *cis* or *trans* orientation, and they can be modified by the peptidyl-prolyl *cis/trans* isomerases (PPIases) named Pin1 in mammals and Ess1 in *S. cerevisiae*. Serine-Proline bond isomerisation by Pin1 stimulates CTD phosphorylation and inhibits Fcp1 phosphatase *in vitro* (Xu et al. 2003). However, unlike Pin1, Ess1 is thought to promote CTD dephosphorylation (Wu et al. 2000), and Ess1 mutants are defective in transcriptional activation and termination (Krishnamurthy et al. 2009). It has been proposed that Ess1 generates a CTD structural conformation that is recognised by Ssu72, hence facilitating dephosphorylation of the CTD and any other Ssu72-dependent transcriptional events (Krishnamurthy et al. 2009). A recent crystal structure of human symplekin, Ssu72 and CTD phosphopeptide complex reveals that the CTD has a *cis* configuration, the opposite of what was thought to be necessary for dephosphorylation by Ssu72 (Xiang et al. 2010). The isomerisation state of the Ser-Pro bonds affects protein binding to the CTD and hence this adds further complexity to the 'CTD code'. Overall, the apparently simple heptapeptide repeat structure of the CTD enables the large number of dynamic and reversible modifications it can undergo resulting in significant variation in Pol II function.

1.3 Pol II termination

1.3a Models for termination

To explain the connection between 3' end processing and termination, two models for transcription termination were initially proposed, both of which are centred on the role of the poly(A) signal. The 'allosteric' model proposed that transcription of the poly(A) signal triggers conformational changes of the elongation complex leading to transcription termination (Logan et al. 1987). This could be achieved by dissociation of elongation factors and/or association of termination factors. The second model, named the 'torpedo' model, proposed that endonucleolytic cleavage of the pre-mRNA at the poly(A) site produces an uncapped transcript that is rapidly degraded by a 5'-3' exonuclease. When the exonuclease 'catches up' with the polymerase, it destabilises the elongation complex and leads to termination (Connelly and Manley, 1988). The main difference between the two models was that the 'torpedo' model depended upon the identification of a 5'-3' exonuclease. As described below, 5'-3' exonucleases that mediate Pol II termination have since been identified in both yeast and mammalian cells. Each of the original models is supported by a number of studies, however it is now apparent that neither model can fully explain the mechanism of termination. The present view is that transcription termination occurs via a hybrid model, combining features of both the 'allosteric' and 'torpedo' models (Luo et al. 2006; Kaneko et al. 2007; West et al. 2008).

1.3b Termination and anti-termination factors

A functional poly(A) site is required to render Pol II termination competent (Whitelaw and Proudfoot, 1986; Logan et al. 1987; Connelly and Manley, 1988; Birse et al. 1997; Dye and Proudfoot, 1999; West et al. 2008) by an, as yet, only partially resolved mechanism. Interestingly, there is evidence for the importance of two Pol II subunits in termination. In *S. cerevisiae*, mutations of Pol II subunits Rpb3 and Rpb11 resulted in transcriptional readthrough, which prompted the suggestion that factor binding to the Rpb3/Rpb11 heterodimer causes an allosteric change to the elongating polymerase. A 'termination signal' is thought to be transduced through the Rpb3/Rpb11 heterodimer (Steinmetz et al. 2006). Due to the observed effects of deletion of a number of factors involved in 3' end processing on transcription termination, these are also considered as termination factors. In *S. cerevisiae*, depletion and mutational analysis of homologues of CstF64 (Rna15), CstF77 (Rna14), Pcf11, CPSF160 (Yhh1), CPSF73 (Ysh1) and Ssu72 all result in transcriptional readthrough (Birse et al. 1998; Dichtl et al. 2002; Steinmetz and Brow, 2003, Garas et al. 2008). In yeast, ChIP experiments have confirmed that 3' end processing factors become associated with the Pol II elongation complex at the poly(A) site, in a Ser2P CTD dependent manner (Ahn et al. 2004; Kim et al. 2004a; Luo et al. 2006). Consistent with these observations, ChIP analysis in mammals detects components of the 3' processing complex, including CPSF73 and CstF64, at the 3' end of genes (Swinburne et al. 2006; Glover-Cutter et al. 2008).

The 3' processing factor Pcf11 has received special attention for its role in termination. Depletion or mutation of Pcf11 results in 3' end processing and termination defects (Licatalosi et al. 2002; Sadowski et al. 2003; Meinhardt and

Cramer, 2004; Zhang et al. 2005). Pcf11 contains a conserved CTD-interacting domain (CID), is able to bind nascent RNA, and has been shown *in vitro* to act as a bridge between the CTD and pre-mRNA (Licatalosi et al. 2002; Sadowski et al. 2003; Meinhart and Cramer, 2004; Zhang et al. 2005). Many studies have shown Pcf11 binds specifically to Ser2P CTD (Barilla et al. 2001; Licatalosi et al. 2002; Ahn et al. 2004; Lunde et al. 2010) however, in a recent study the occupancy of Pcf11 was measured across mRNA coding genes and was found not to correlate with Ser2P (Kim et al. 2010; Mayer et al. 2010). This result suggests that either Ser2P is masked *in vivo* or Pcf11 is recruited by other 3' end processing factors independently of the CTD (Mayer et al. 2010). Nonetheless, Pcf11 has been shown *in vitro* to dismantle *S. cerevisiae* and *Drosophila* Pol II elongation complexes via a mechanism that involves the CTD and is independent of RNA cleavage activity (Zhang and Gilmour, 2005; Zhang and Gilmour, 2006). Moreover, Pcf11 has been shown to limit Pol II processivity and cause premature termination of Pol II transcribing an HIV provirus template *in vivo* (Zhang et al. 2007). In another study, *S. cerevisiae* strains carrying mutant Pcf11 which is defective in CFIA activity show readthrough transcription, suggesting that RNA cleavage activity is necessary for mediating Pol II termination (Kim et al. 2006). Depletion of Pcf11 in mammalian cells reduces termination efficiency and stabilises the 3' pre-mRNA cleavage product, indicating it has a role in RNA degradation (West and Proudfoot, 2008). It appears that Pcf11 may act in more than one fashion to promote termination; firstly, it could aid the 5'-3' RNA exonuclease degrade the 3' cleavage product and, secondly, it could induce conformational changes in the elongating Pol II complex.

Certain factors have been identified that act as anti-terminators, and as such it would be expected that they disassociate from the elongation complex at the 3' end of the gene. For example, the transcriptional co-activator PC4 (Sub1 in yeast), which can interact with CstF64, was shown to suppress termination in *S. cerevisiae* when overexpressed, suggesting that it possess anti-termination activity via its interaction with CstF64 (Calvo and Manley, 2001). As expected, PC4/Sub1 dissociates from the elongation complex at the poly(A) site (Calvo and Manley, 2001). ChIP analysis shows that the elongation factor Paf1C (Costa and Arndt, 2000; Betz et al. 2002; Squazzo et al. 2002; Rondon et al. 2004; Kim et al. 2010) and the TREX/THO complex, which functions in mRNP biogenesis and maintenance of genome integrity (Garcia-Rubio et al. 2008), are both found throughout the body of genes, but their levels decrease upon passing through the poly(A) site, indicating that they dissociate from the elongation complex, perhaps rendering it termination competent (Ahn et al. 2004; Kim et al. 2004a). The *S. cerevisiae* mRNA binding protein Npl3 has been shown to interact with Ser2P CTD and stimulate the elongation rate of Pol II (Dermody et al. 2008). Furthermore, Npl3 also competes for binding to the pre-mRNA with CstF64 (Rna15 in yeast), and Npl3 binding prevents termination (Bucheli and Buratowski, 2005; Bucheli et al. 2007). Phosphorylation of Npl3 disrupts its interaction with the CTD and the pre-mRNA, exposing the pre-mRNA to CstF64 (Dermody et al. 2008). ChIP assays are in agreement with this model, as they show levels of Npl3 decrease downstream of the poly(A) site (Bucheli and Buratowski, 2005). This interaction of Npl3 with the transcription complex appears to be similar to bacterial termination mechanisms. In particular the anti-terminator λ Q interacts with the transcription protein NusA and so restricts access to the emerging transcript, preventing Rho-dependent termination (Shankar et al. 2007).

1.3c The molecular torpedo

Confirmation of the 'torpedo' model of Pol II termination came with the identification of 5'-3' exonucleases – Xrn2 in humans and Rat1 in yeast – being required for transcription termination (Kim et al. 2004b; West et al. 2004). In yeast, Rat1 is activated by its cofactor, Rai1 (Stevens and Poole, 1995; Xiang et al. 2009), and both proteins are found at the 3' end of genes (Kim et al. 2004b). Kim et al. (2004b) showed that Rat1 with a point mutation in its exonuclease domain is unable to complement termination defects in a mutant *rat1* strain, demonstrating the significance of the exonuclease activity in termination. ChIP analysis shows that Rat1 is recruited to the 3' end of genes, possibly through interactions with 3' end processing factors (Kim et al. 2004b). Indeed, Rat1 recruitment to the 3' end of genes is decreased by inactivation of the 3' end processing factor Pcf11 (Luo et al. 2006). Furthermore, Rat1 is required for recruitment of Pcf11 and Rna15 (CstF64 homologue) to the 3' end of genes, suggesting that there is interdependent recruitment of Rat1 and 3' end processing factors, indicating a role of Rat1 in 3' end processing as well as termination (Luo et al. 2006). The Ser2P CTD binding protein, Rtt103, copurifies with Rat1 and subunits of the cleavage and polyadenylation factors and is also detected at the 3' end of genes (Kim et al. 2004b). It is possible that Rtt103 has a role in termination, as the *C. elegans* homologue, CIDS-1, was recently identified in an RNAi genome-wide screen for factors involved in termination (Cui et al. 2008).

Although Rat1 is implicated in transcription termination, Rat1 activity alone is not sufficient to bring about termination *in vivo* (Kim et al. 2004b). It has also been shown, *in vitro*, that Rat1 is unable to dismantle elongation complexes (Gu et al. 1996; Kireeva et al. 2000). There are two lines of evidence indicating that degradation

of the 3' product of poly(A) site cleavage is not sufficient to terminate Pol II (Luo et al. 2006; Kim et al. 2006). Firstly, in a Rat1 deficient strain, the termination defect is not associated with stabilisation of the 3' cleavage product RNA. Secondly, the yeast cytoplasmic 5'-3' exonuclease Xrn1, which is homologous to Rat1, does not complement the termination defect of a Rat1 deficient strain, even though it can complement the exonucleolytic activity of Rat1 (Luo et al. 2006). Hence, these findings demonstrate that there is more to termination than degradation of the 3' cleavage product; other activities are required to bring about termination.

Kawauchi et al. (2008) provide evidence that Sen1, an RNA/DNA dependent ATPase (Kim et al. 1999), cooperates with Rat1 to promote termination. The N-terminal domain of Sen1 interacts with the Pol II CTD as well as the dsRNase III enzyme Rnt1, which is required for maturation of several non-coding RNAs (Ursic et al. 2004). Mutation of the Sen1 helicase domain results in transcript processing and termination defects, which in turn affect the genome-wide distribution of Pol II over both coding and non-coding genes, as well as a Pol I transcription defect (Ursic et al. 1997; Rasmussen and Culbertson, 1998; Steinmetz et al. 2001; 2006; Kawauchi et al. 2008). A recent study by Mischo et al. (2011) shows that RNA:DNA hybrids known as R-loops, are formed during transcription – particularly downstream of the poly(A) signal. These RNA:DNA hybrids are resolved by the action of Sen1 and accumulate in its absence. This study demonstrates that although the helicase domain of Sen1 is dispensable for 3' end processing, it is required for efficient termination. Hence, in the current model, Sen1 acts to resolve R-loops and, by doing so, exposes the 3' endonucleolytic cleavage product to the action of Rat1 (Kawauchi et al. 2008; Mischo et al. 2011).

Human homologues of Rat1 and Sen1 have been shown to be required for termination (West et al. 2004; Kaneko et al. 2007; Suraweera et al. 2009; Skourti-Stathaki et al. 2011). Depletion of Xrn2, the mammalian Rat1 homologue, leads to a termination defect and accumulation of the 3' cleavage product of the β -globin transcript (West et al. 2004). Furthermore, Xrn2 is detected at the 3' end of the β -actin gene with the highest level seen just after the poly(A) signal (Kaneko et al. 2007), which is consistent with localisation of Rat1 observed in yeast (Kim et al. 2004b; Luo et al. 2006). Recruitment of Xrn2 to the 3' end processing machinery is required for degradation of the 3' cleavage product – but not the cleavage reaction itself (Kaneko et al. 2007). In the case of the β -globin transcript, West et al (2004) demonstrate that Xrn2 entry occurs following endonucleolytic cleavage at the co-transcriptional cleavage (CoTC) terminator element (discussed below). In genes which lack a CoTC terminator, cleavage at the poly(A) site creates an entry site for Xrn2 (Gromak et al. 2006).

Xrn2 has been shown to interact with CstF64 and the multifunctional p54nrb/NonO (p54)-protein-associated splicing factor (PSF) (Kaneko et al. 2007). PSF and p54 bind each other and have been shown to have roles in transcription, splicing and polyadenylation (Rosonina et al. 2005; Liang and Lutz, 2006). Immunodepletion experiments suggest that, at least *in vitro*, p54/PSF, like Xrn2, is required for degradation of the 3' cleavage product (Kaneko et al. 2007). Moreover, p54 depletion adversely affects Xrn2 recruitment and transcription termination, indicating that p54/PSF recruits Xrn2 to the 3' end processing machinery and is required for optimal Xrn2 activity (Kaneko et al. 2007). Furthermore, PSF-1, the *C. elegans* homologue of p54/PSF, has also been implicated in transcription termination, suggesting a

conserved function (Cui et al. 2008). A recent study has implicated Senataxin (SETX), the mammalian Sen1 homologue, in the recruitment of Xrn2 to the 3' end of genes (Skourti-Stathaki et al. 2011).

SETX has been linked to transcription initiation, termination, pre-mRNA processing, and DNA damage repair (Skourti-Stathaki et al. 2011; Suraweera et al. 2007, 2009). Similarly to Sen1, SETX interacts with Pol II and RNA processing factors, and mutation or depletion of SETX leads to a changed Pol II profile, reduced mRNA levels and a termination defect (Suraweera et al. 2009; Skourti-Stathaki et al. 2011). Furthermore, it has been shown that R-loops are formed in mammals in a transcription dependent manner, as in yeast. Skourti-Stathaki et al. (2011) demonstrate that the formation of R-loops is dependent upon a functional poly(A) signal and a termination pause element. Interestingly, the presence of R-loops and their resolution are essential steps in termination. It is thought that R-loops formed over the β -actin gene termination pause element allow Pol II to pause downstream of the poly(A) signal. SETX then helps recruit Xrn2 and resolves the R-loops, which exposes the 3' poly(A) site cleavage product for degradation by Xrn2 (Skourti-Stathaki et al. 2011).

1.3d Hybrid model

There are several pre-requisites for efficient transcriptional termination to occur, including transcription of a functional poly(A) signal; endonucleolytic cleavage of the nascent transcript; an exonuclease (Xrn2/Rat1); an RNA:DNA helicase (SETX/Sen1); the association of termination factors – Ssu72, Pcf11 and 3' processing factors; and dissociation of anti-termination factors – Paf1C, TREX/THO complex, PC4 (Sub1)

and Npl3 – all of which point towards a combined model for Pol II termination (Figure 1.2). Further evidence for a hybrid model comes from a study by West et al. (2008), which shows that transcription of the poly(A) signal is necessary to render Pol II termination competent, as was proposed by the ‘allosteric’ model (Logan et al. 1987), and cleavage of the transcript, either at the poly(A) site or at a downstream CoTC mediated terminator element, creates an entry site for Xrn2 to degrade the 3' cleavage product and lead to termination, as postulated in the ‘torpedo’ model (Connelly and Manley, 1988). It was also shown in this study that Pol II is released from the DNA template following degradation of the 3' cleavage product but only if a poly(A) signal has been transcribed. Thus it appears that aspects of the original ‘allosteric’ and ‘torpedo’ models are involved in Pol II termination.

1.3e Terminator sequence elements

For some genes it has been shown that transcription termination requires sequence elements, termed terminators, located downstream of the poly(A) site. Two major classes of terminator elements, pause and CoTC elements, have been described.

1.3e (i) Pause elements

It has been demonstrated that Pol II accumulates at promoter regions and downstream of poly(A) signals, indicating transcriptional pausing (Glover-Cutter et al. 2008). Indeed, many protein-encoding genes, including the $\alpha 2$ -globin, γ -globin and IgM genes, contain sequences downstream of their poly(A) sites that cause Pol II pausing and their deletion results in a decrease in termination efficiency (Enriquez-Harris et al. 1991; Ashfield et al. 1991; Plant et al. 2005; Gromak et al. 2006). The MAZ element is a G-rich sequence found at the 3' end of the C2 complement gene, which

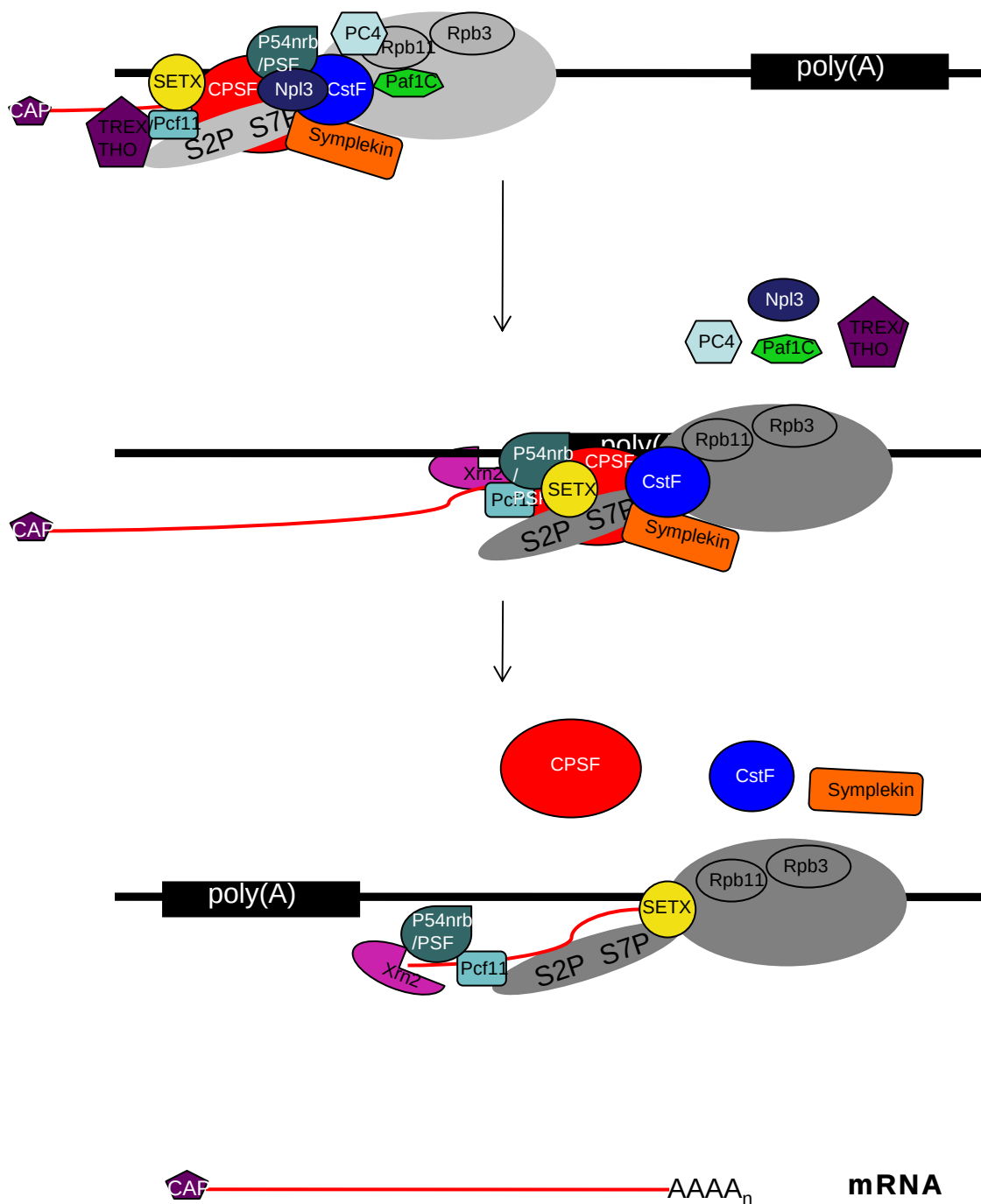


Figure 1.2: Mammalian Pol II termination. Elongating Pol II is represented by a light grey oval and is depicted with several associated factors. Upon recognition of the poly(A) signal (black rectangle), Pol II becomes termination competent (dark grey oval). Factors that may act as anti-terminators – for example Paf1C, PC4, Npl3 and the TREX/THO complex - are released upon transcription of the poly(A) signal. Furthermore, factors necessary for termination are recruited to Pol II- notably Xrn2, which is probably recruited by p54nrb/PSF and Senataxin (SETX). Cleavage at the poly(A) signal leads to polyadenylation and release of the mRNA, and provides an entry site for Xrn2. Xrn2 degrades the 3' RNA, with the help of SENX which resolves formation of R-loops. This leads to release of Pol II from the DNA template. Pol II subunits Rpb3 and Rpb11 are also involved in termination. Other factors which are required for termination and not depicted here include chromatin remodelling factors. Diagram not drawn to scale.

when repeated four times mediates pausing and facilitates termination (Ashfield et al. 1994; Yonaha and Proudfoot, 1999). Another G-rich region located downstream of the human β -actin poly(A) signal has been described and shown to direct termination (Gromak et al. 2006). The efficiency of pause sites as terminators appears to correlate with the strength of the poly(A) signal (Plant et al. 2005; Gromak et al. 2006). Moreover, the poly(A) signal itself – the strength of which determines the speed with which the 3' processing apparatus assembles on the nascent transcript (Chao et al. 1999) – has been suggested to pause Pol II in a manner mediated by CPSF and independent of the CTD and any further pause elements downstream of it (Orozco et al. 2002; Nag et al. 2006; 2007). In *S. pombe*, the *ura4* and *nmt2* genes both possess downstream sequence elements (DSE), located downstream of their poly(A) signals, which function as pause elements and are required for termination (Birse et al. 1997, Aranda and Proudfoot, 1999).

Pause elements are thought to promote termination by slowing transcription elongation and allowing 5'-3' exonuclease more time to degrade the nascent RNA and 'catch up' with transcribing Pol II (Gromak et al. 2006) (Figure 1.3a). Overall, it is likely that pausing is a common feature of termination. It is employed by eukaryotic Pol I, II and III complexes and even bacterial RNA polymerases (discussed below), but a consensus pause sequence has yet to be identified.

1.3e (ii) CoTC elements

The most thoroughly analysed CoTC element is an 850bp AT rich DNA sequence (Figure 1.4A) located approximately 900bp downstream of the human β -globin

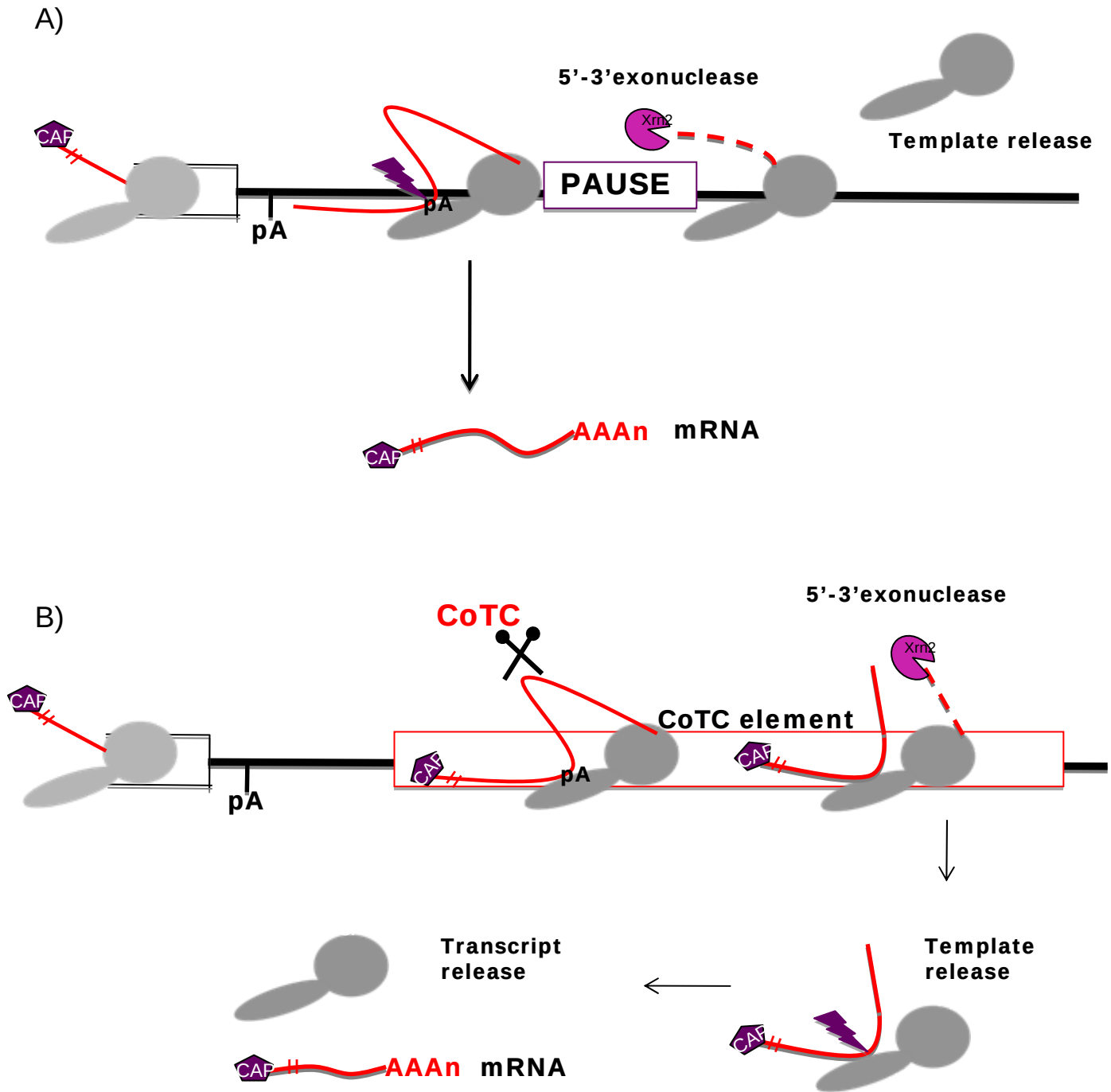


Figure 1.3: Mammalian Pol II termination at pause and CoTC terminator elements. A) Pause-mediated termination. Elongating Pol II (light grey oval), DNA (black line) and nascent RNA transcript (red line) are indicated. Upon recognition of the poly(A) signal (pA), Pol II becomes termination competent (dark grey oval). Cleavage at the poly(A) signal (purple lightning bolt), releases the mRNA and creates an entry site for the exonuclease Xrn2 (pink). At the pause element (white rectangle), the elongation rate decreases allowing more time for Xrn2 to degrade the nascent RNA and 'catch up' with Pol II leading to release of Pol II from the DNA template. **B)** CoTC-mediated termination. As above, recognition of the poly(A) signal causes Pol II to become termination competent. Cleavage of CoTC element (white rectangle) transcript, creates an entry site for the exonuclease Xrn2 (pink) to degrade the downstream transcript and 'catch up' with Pol II leading to release of Pol II from the DNA template with the pre-mRNA still attached to it. Only after cleavage at the poly(A) signal is the mRNA released from Pol II. Diagram is not drawn to scale.

A) GCATAGTGTTACCATCAACCACCTTAACTTCATTTTTCTTATTCAATACCTAGGTAGGTAGATGCTAGATT
CGTATCACAAATGGTAGTTGGTGGAATTGAAGTAAAAAGAATAAGTTATGGATCCATCCATCTACGATCTAA

CTGGAAATAAAATATGAGTCTCAAGTGCTCCTTGCTCCTCTCCCAGTCAAATTCTGAATCTAGTTGGCAA
GACCTTTATTTTATACTCAGAGTTCACCAGGAACAGGAGAGAGGGTCAGTTTAAGACTTAGATCAACCGTT

GATTCTGAAATCAAGGCATATAATCAGTAATAAGTGATGATAGAAGGGTATATAGAAGAATTTTATTATAT
CTAAGACTTTAGTTCCGTATATTAGTCATTATTCACACTACTCTTCCCATATATCTTCTTAAATAATATA

GAGAGGGTGAAACCTAAATGAAATGAAATCAGACCCTTGCTTACACC
CTCTCCCACCTTGGATTTTACTTTACTTTAGTCTGGGAACAGAATGTGG

Region 8

ATAAACAAAAATAAATTTGAATGGGTAAAGAATTAACTAAGACCTAAAACCATAAAAAATTTTAAAGAA
TATTTGTTTTTATTTAACTTACCCAATTTCTTAATTTGATTCTGGATTTTGGTATTTTAAAAATTTCTT

ATCAAAAGAAGAAAATTTCTAATATTCATGTTGCAGCCGTTTTTTGAATTTGATATGAGAAGCAAAGGCAAC
TAGTTTTCTTCTTTAAGATTATAAGTACAACGTCGGCAAAAACTTAACTATACTCTTCGTTTCCGTTG

AAAAGGAAAAATAAAGAAGTGAGGCTACATCAAACTAAAAATTTCCACAAAAAAGAAAACAATGAACA
TTTTCTTTTTTATTTCTCACTCCGATGTAGTTTGATTTTTTAAAGGTGTGTTTTTCTTTTGTTACTTGT

AATGAAAGGTGAACCATGAAATGGCATATTTGCAAACCAAATTTTCTTAAATATTTTGGTTAATATCCAA
TTACTTTCCACTTGGTACTTTACCGTATAAACGTTTGGTTTATAAAGAATTTATAAAACCAATTATAGTT

AATATATAAGAAACACAGATGATT
TTATATATTCTTTGTGTCTACTAA

Region 9

CAATAACAAACAAAAATTTAAAAATAGGAAAAATAAAAAATTTAAAAAGAAGAAATCCTGCCATTTATGCGA
GTTATTGTTTGTTTTTTAATTTTTATCCTTTTATTTTTTAAATTTTCTTCTTTTAGGACGGTAAATACGCT

GAATTGATGAACCTGGAGGATGTAAACTAAGAAAAATAAGCCTGACACAAAAAGACAAATACTACACAACC
CTTAACACTTGGACCTCCTACATTTTGATTCTTTTTATTTCGGACTGTGTTTTTCTGTTTATGATGTGTTGG

TTGCTCATATGTGAACATAAAAAAGTCACTCTCATGGAACAGACAGTAGAGGTATGGTTTCCAGGGGTTG
AACGAGTATACACTTTGTATTTTTTCAGTGAGAGTACCTTTGTCTGTCATCTCCATACCAAGGTCCCAAC

GGGGTGGGAGAATCAGGAACTATTACTCAAAGGGTATAAAATTTTCAGTTATGTGGGATGAATAAATT
CCCCACCCTCTTAGTCCTTTGATAATGAGTTTCCCATATTTTAAAGTCAATACACCCTACTTATTTAA

Region 10

B) TGCTGTGGTGACCTCCCCCTAACCATAAACATTATTTTTGTTGTTACTTCATAATGTTATTTTGATATTGTGT
ACGACACCACTGGAGGGGGATTGGTATTGTAATAAAAAACAACATGAAGTATTACAATAAACTATAACACA

CATAATGTAAATGTGTTTTTAAAGAGGAGGAAGAGAAGGAATATATTTGTGGGTATGTAGTTAGGTATGAGG
GTATTACATTTACACAAAAAATCTCCTCCTTCTCTTCTTATATAAACACCCATACATCAATCCATACTCC

AAAGGGGATGCCTCCGCTGGCCCATGCTGAGACATCCCTTCCCCCTGAGGGACCAGCCACATGATTGTATAA
TTTCCCCACGGAGGCGACCGGGTACGACTCTGTAGGGAAGGGGGACTCCCTGGTCGGTGTACTAACATATT

TATAGAATAGAGTTTATTTGGGACATGGGGAGGGGAGCTAAGAGGGTAGTAGAGACAGAGAAAGGCAGAGAC
ATATCTTATCTCAAATAAACCTGTACCCCTCCCTCGATTCTCCCATCATCTCTGTCTTTCCGTCTCTG

AGAGCAGAGAAATAGAGAGGTAGAGACTGGCCATGAGCACGTGGAGAGAGGGGGAAGAGGTGAGAGACAGA
TCTCGTCTCTTTATCTCTCCATCTCTGACCGGTACTCGTGACCTCTCTCCCCCTTCTCCACTCTCCTGTCT

GCAGTAGCAAGAAGGCAAGAGCAAGAGAGAGCAGAGGGGG
CGTCATCGTTCTTCCGTTCTCGTTCTCTCTCGTCTCCCCC

Figure 1.4: CoTC termination elements. A) The 850bp human β -globin terminator subdivided into regions, labelled 8, 9 and 10. Each region is capable of directing CoTC-mediated termination. **B)** The mouse serum albumin CoTC termination element sequence.

poly(A) site which was shown to be required for Pol II termination (Dye and Proudfoot, 2001). Hybrid selection nuclear run on (NRO) analysis of nascent transcripts of the β -globin CoTC terminator element demonstrated that they are cotranscriptionally cleaved. Hence, this terminator was labelled a co-transcriptional cleavage (CoTC) element (Figure 1.4A) (Dye and Proudfoot, 2001). RNase Protection analysis (RPA) of nascent transcripts indicated that CoTC occurs at multiple positions within the human β -globin gene CoTC terminator transcripts (Dye and Proudfoot, 2001).

Other examples of CoTC termination elements have been described for the ϵ -globin gene (Dye and Proudfoot, 2001) and the mouse serum albumin (MSA) gene (Figure 1.4B) (West et al. 2006). The CoTC termination element of the MSA gene is located approximately 800bp downstream of the poly(A) signal and there are multiple sequence elements within it that promote termination. These elements are cleaved co-transcriptionally at multiple positions analogous to the human β -globin gene (West et al. 2006).

The CoTC terminators described so far are positioned further from the poly(A) site than pause elements. Transcripts of CoTC terminators are highly unstable and have been shown to provide an entry site for Xrn2 to degrade the 3' cleavage product (West et al. 2004). Transcript degradation leads to release of Pol II with the pre-mRNA from the DNA template. Subsequent cleavage at the poly(A) site releases the mRNA from Pol II (West et al. 2008) (Figure 1.3b). Hence, 3' end processing of the β -globin mRNA occurs post-transcriptionally. The 5' cleavage products generated by CoTC in the MSA and the human β -globin gene terminators are oligo-adenylated, probably by

the human homologue of Trf4, prior to degradation by the exosome complex (West et al. 2006). CoTC-mediated termination has been shown to increase mRNA levels. It is thought that this is a consequence of pre-mRNA release from the site of synthesis where nuclear RNA surveillance occurs (West and Proudfoot, 2009).

The details of how CoTC cleavage events occur have yet to be revealed. A previous study on the human β -globin CoTC element suggested the importance of its secondary structure (Teixeira et al. 2004). In this study, it was proposed that autocatalytic cleavage by this structure resulted in transcriptional termination. However, very low cleavage efficiency is observed *in vitro* suggesting that an endonuclease is likely to be required *in vivo*.

1.3e (iii) Other terminator elements

There are other terminator sequences where the mechanism of action has yet to be determined, for example the adenovirus major late promoter contains a sequence element, the CCAAT box, which functions in an orientation-dependent manner and is required for termination (Connelly and Manley, 1989a). It was suggested that a protein factor may bind to this sequence element and lead to termination, but replacing this region by other factor binding sites does not lead to termination (Connelly and Manley, 1989b), and hence, it is unlikely that, even if such a factor exists, it acts as a 'roadblock'. Another example comes from the mouse β -major globin gene. Analysis of the 3' flanking region of this gene identified a number of sequence elements, including a 69bp AT-rich sequence, that cause termination in the presence of a functional poly(A) signal. How this element works is not yet clear but it

appears to be dependent on its position relative to the poly(A) signal (Tantravahi et al. 1993).

1.4 Connections between capping, splicing, 3' end processing and termination

Pre-mRNA processing – capping, splicing and 3' end cleavage/polyadenylation – acts in a coordinated manner and each step is connected with each other. The first processing event to occur is capping, which consists of the co-transcriptional addition of a trimethyl cap structure to the 5' end of the nascent transcript. The cap binding complex (CBC) composed of two subunits, CBC20 and CBC80, binds the cap structure (Visa et al. 1996; McCracken et al. 1997). Nuclear extracts depleted of CBC80 exhibit reduced 3' end cleavage but only a modest effect is observed on polyadenylation (Flaherty et al. 1997). Flaherty et al (1997) suggest the reason for this could be a loss of physical interaction between the CBC and the 3' end processing machinery. Hence, it is likely that CBC could indirectly affect Pol II termination. Indeed, a recent study in *S. cerevisiae* by Jimeno-Gonzalez (2010) demonstrates that a defect in capping leads to premature Rat1 dependent termination, confirming a link between capping and transcription termination.

It has been shown that pre-mRNA splicing and termination are also connected. One example of this is the case of terminal exon definition which has been shown to co-transcriptionally enhance poly(A) site efficiency and, thereby, enhance termination (Niwa and Berget, 1991; Niwa et al. 1992; Dye and Proudfoot, 1999).

Molecular evidence for interconnections between splicing and 3' end formation comes from a study showing that protein components of U2 snRNP, which associates with the 3' splice site and the branch point, enhance 3' end processing through interactions with CPSF (Kyburz et al. 2006). In addition, a subunit of the splicing factor U2AF, U2AF65, co-purifies with CFI_m and CFII_m complexes of the 3' end processing machinery. Mechanistic evidence comes from observations that U2AF associates with the 3' splice site and contacts PAP, thus enhancing the poly(A) site function (Vagner et al. 2000). Furthermore, the interaction with PAP stimulates U2AF65 to bind to introns and, thus, enhance splicing (Vagner et al. 2000).

The splicing factor SRp20 has been identified in a genome wide RNAi screen in *C. elegans* for factors involved in termination (Cui et al. 2008). Previous studies showed that SRp20 co-immunoprecipitates with CFI_m (Dettwiler et al. 2004) and affects polyadenylation in Rous sarcoma virus (Maciolek and McNally, 2007). It has been proposed that SRp20 functions either in degradation of the 3' cleavage product or in release of Pol II from the DNA template (Cui et al. 2008). Another factor involved in splicing, the polypyrimidine tract-binding protein (PTB), is detected at the 3' end of genes (Swinburne et al. 2006), and has been demonstrated to inhibit 3' end processing (Castelo-Branco et al. 2004). It has been suggested that PTB competes for binding to the GU/U rich DSE region of mammalian poly(A) sites with CstF64. Binding by PTB inhibits 3' end processing. Lastly, SETX, which is required for termination, has been shown to affect splicing efficiency and alternative splice site selection, creating another link between splicing and termination (Suraweera et al. 2009). The mechanism by which SETX affects splicing has not yet been identified, but it is

possible that this is due to inhibition of the interaction of splicing factors with their cognate sequences or, due to a reduction in Pol II transcriptional elongation rate.

An important consequence of efficient transcriptional termination is that it can enhance gene expression by promoting pre-mRNA processing (West and Proudfoot, 2009). The authors demonstrated that termination can have a positive post-transcriptional effect on pre-mRNA splicing. Furthermore, another study has demonstrated a link between 3' end processing/termination and transcriptional initiation (Mapendano et al. 2010). In this study, it was shown that 3' end formation stimulates re-initiation and the authors suggest that the recycling of factors from the terminator to the promoter is necessary for continued transcription. In this fashion, termination may act to release transcription factors.

1.5 Transcription termination of other eukaryotic Pol II transcribed genes

1.5a Histone gene 3' end processing and termination

Mammalian replication-dependent histone genes encode the only mRNAs that do not possess a poly(A) tail but instead have a stem-loop structure at their 3' ends (Marzluff et al. 2008). The mature 3' end of histone mRNAs is formed by an endonucleolytic cleavage after a CA dinucleotide in the pre-mRNA. Two *cis*-acting sequence elements are required for processing. The first is a conserved 16 nucleotide stem loop structure – situated 5 nucleotides upstream of the cleavage site – which interacts with the stem loop binding protein (SLBP). The second is a purine rich sequence element, labelled histone downstream element (HDE), which is located 6-12 nucleotides downstream of the cleavage site (Dominski and Marzluff, 1999).

The HDE is recognised by base pairing to the 5' end of U7 snRNA, thus recruiting U7 snRNP (reviewed by Dominski and Marzluff, 2007). Binding of SLBP to the stem loop helps recruitment of U7 snRNP by ensuring accessibility of downstream sequences for base pairing of U7 snRNA with the HDE (Jaeger et al. 2006). ZFP100, a zinc finger protein that interacts with the SLBP (Dominski et al. 2002) and Lsm11 (Pillai et al. 2003; Azzouz et al. 2005; Wagner et al. 2006), a component of the U7 snRNP, acts as a bridge between SLBP and U7snRNP. Gick et al (1987) have shown that a Heat Labile Factor (HLF) is required for histone transcript cleavage. Isolation of this factor revealed that it contains components of the pre-mRNA 3' end cleavage/polyadenylation machinery including symplekin, CPSF, and CstF subunits (Kolev and Steitz, 2005). Following evidence that symplekin is involved in mRNA processing (Takagaki and Manley, 2000), it was suggested that symplekin acts as a scaffold that brings the other subunits of HLF to the histone pre-mRNA (Kolev and Steitz, 2005). UV crosslinking experiments lead to the identification of the endonuclease responsible for cleavage of histone pre-mRNAs as CPSF73 (Dominski et al, 2005; Kolev et al. 2008), which is also responsible for cleavage of polyadenylated pre-mRNAs (Ryan et al. 2004; Mandel et al. 2006). The cleavage reaction also requires CPSF100 and symplekin, which are also components of the pre-mRNA 3' end cleavage/polyadenylation machinery (Kolev et al. 2005; Kolev et al. 2008). Furthermore, an essential U7 snRNP-associated protein, FLASH, has been shown to stimulate the cleavage reaction (Yang et al. 2011).

Similar to standard mRNA genes, transcription of histone genes does not terminate until it is some way beyond the 3' end processing site. An early report suggested that termination of the sea urchin H2a gene occurs approximately 200bp downstream of

the 3' end of the gene (Birchmeier et al. 1982) and further studies identified a DNA element at this location as a transcriptional terminator (Briggs et al. 1989; Johnson et al. 1986). Studies of the mouse H2a gene have shown that termination occurred approximately 600bp downstream of the 3' end of the gene and was dependent on 3' end processing signals (Chodchoy et al. 1991; Gu and Marzluff, 1996). These results show that 3' end processing and transcription termination of mouse H2a histone pre-mRNAs are coupled events, as is the case for 3' end processing and transcription termination of other protein-encoding genes (Proudfoot et al. 2002; Bentley, 2005; Buratowski, 2005). A transcriptional pause site has been identified downstream of some histone 3' processing signals (Adamson and Price, 2003) suggesting a similar arrangement of poly(A) and pause signals identified in other protein encoding genes. However, it is not known whether the histone pause promotes more efficient termination. It has also been shown that the negative elongation factor NELF, which inhibits release of Pol II into transcriptional elongation (Peterlin and Price, 2006), is needed for histone pre-mRNA termination (Narita et al. 2007).

The mature histone mRNA is stabilised by the terminal stem loop, whilst the 3' cleavage product is degraded by a 5'-3' exonuclease *in vitro* (Walther et al. 1998). It is possible that transcription termination is enhanced by a 5'-3' exonuclease, via a 'torpedo' mechanism. Yang et al (2009), demonstrate that the 3' cleavage product is degraded 5'-3' by a U7-dependent exonuclease activity, which is not Xrn2. The authors of this study suggest that the 5'-3' exonuclease responsible is CPSF73. However, there is no direct evidence that CPSF73 possesses exonucleolytic activity. Overall, despite the fact that the *cis*-regulatory elements differ between histone pre-

mRNAs and other protein-encoding pre-mRNAs, the enzymatic machinery is similar and it is possible that the same principles for transcription termination are employed.

1.5b snRNA gene termination

Pol II transcribed spliceosomal U1, U2, U4 and U5 snRNAs are nonpolyadenylated noncoding RNAs involved in the splicing process (Matlin and Moore, 2007; Hallegger et al. 2010). The 3' end of mature snRNA genes is generated by endonucleolytic cleavage of the primary transcript 9-19 nucleotides upstream of a conserved *cis*-acting 3' box element (Uguen and Murphy, 2003). This cleavage event generates an extended 3' end that is trimmed after export to the cytoplasm (Egloff et al. 2008).

It has recently been shown that the Integrator complex, which contains homologues of CPSF73 and CPSF100 termed Int9 and Int11 respectively, plays a role in pre-snRNA 3' end formation (Baillat et al. 2005; Egloff et al. 2007). Subunits of the Integrator complex associate with U1 and U2 snRNA genes, and RNAi mediated knockdown of Int9 or Int11 cause a defect in pre-snRNA production (Baillat et al. 2005). Due to its homology to CPSF73 and its requirement for snRNA transcription, it is possible that Int9 is the endonuclease involved in 3' end formation.

snRNA 3' end formation shows promoter specificity indicating a link between transcription initiation and 3' end processing (de Vegvar et al. 1986; Hernandez and Weiner 1986). Further to this, it has been shown that Pol II CTD and its phosphorylation state affect 3' end formation of snRNA genes, indicating that processing occurs co-transcriptionally (Medlin et al. 2003; Jacobs et al. 2004), as

shown for protein-encoding genes. When CTD phosphorylation at Ser2 was blocked, recognition of the 3' box was abolished, causing a reduction in 3' box dependent processing (Medlin et al. 2003; Medlin et al. 2005; Egloff et al. 2007). Ser7 of the CTD is phosphorylated during transcription of snRNA genes (Egloff et al. 2007) and substitution of serine with alanine has been shown to affect recognition of the 3' box (Egloff et al. 2007; Chapman et al. 2007). It has also been demonstrated that both Ser2 and Ser7 phosphorylation are required for recruitment of Integrator to U1 and U2 snRNA genes (Egloff et al. 2010).

Despite the ubiquity of the 3' box it appears that different snRNA genes employ different termination mechanisms. NRO analysis shows that the U1 and U2 gene transcription terminates approximately 200 and 800bp downstream of the 3' box respectively (Cuello et al. 1999; Medlin et al. 2003). An unidentified termination factor appears to bind a G-tract in U1 precursor RNA located between the 3' box and a T-stretch (Cuello et al. 1999). The terminator sequence is also able to direct efficient termination of the U2 gene when placed after the 3' box even though no such sequence has been identified in the U2 termination region. It has also been shown that NELF is required for termination of the U2 snRNA gene, as is the case for histone genes (Egloff et al, 2009).

Common features can be identified between snRNA and protein-encoding genes. The 3' box is in some way analogous to the poly(A) signal as they are both processing signals and are required for transcription termination (Cuello et al. 1999). To date, there is no evidence for the requirement of a 5'-3' exonuclease in snRNA termination.

However, it is possible that transcription termination is enhanced via a ‘torpedo’-like mechanism. Xrn2, or perhaps Int9, are possible candidates for such an activity.

In *S. cerevisiae*, 3' end formation of snRNAs requires an endonucleolytic cleavage but no 3' box sequences are present (Allmang et al. 1999; Morlando et al. 2002). A stem loop structure is formed at the 3' end of these snRNAs, which is cleaved by Rnt1, a member of the RNase III family (Chanfreau et al. 1997). The 5' Rnt1 cleavage product is trimmed by the exosome to form the mature 3' end of the snRNA (Allmang et al. 1999; van Hoof et al. 2000a; 2000b). Co-transcriptional association of specific factors combined with snRNP formation protects the mature snRNA 3' end from further trimming (Morlando et al. 2004; Houalla et al. 2006). However, mature snRNAs are formed in strains depleted of Rnt1 (Abou Elela and Ares, 1998; Chanfreau et al. 1997) suggesting there is redundancy in the system. The termination process of U4 snRNA has been shown to be similar to snoRNAs. Here transcription is terminated by the Nrd1 complex, with Nrd1 binding downstream of the Rnt1 cleavage site (Steinmetz et al. 2001). In *S. pombe*, transcription termination of the U2 snRNA gene involves an endonucleolytic cleavage by dsRNase III family member Pac1 (Zhou et al. 1999) followed by 5'-3' exonucleolytic degradation by the exonuclease Dhp1 (Nabavi and Nazar, 2010a), in a ‘torpedo’-like mechanism.

1.5c snoRNA gene termination

There are two classes of snoRNAs, namely the C/D box snoRNAs, which direct 2'-O-methylation of rRNAs, and the H/ACA box snoRNAs which direct pseudouridylation of rRNAs (reviewed by Kiss, 2002). The majority of mammalian snoRNAs guide modifications of pre-rRNAs and are encoded within introns. Therefore, if they are

transcribed as part of an intron sequence, it is unlikely that they are associated with transcription termination (Richard and Kiss 2006). Others, which are required for pre-rRNA processing, are transcribed from their own promoters. Possibly due to the lack of introns in *S. cerevisiae*, the majority of snoRNAs are transcribed from their own promoters, and a great deal more research has been conducted into this system. As a result, the following discussion is based on work carried out in *S. cerevisiae*. For some snoRNAs, 3' end processing occurs via endonucleolytic cleavage by Rnt1, followed by exosome-directed exonucleolytic trimming (Chanfreau et al. 1998). It has been shown that co-transcriptional formation of snoRNPs protects the mature 3' end from further trimming (Morlando et al. 2004; Ballarino et al. 2005; Yang et al. 2005; Houalla et al. 2006).

The major pathway for snoRNA 3' end formation and termination in *S. cerevisiae* is the Nrd1-dependent pathway (Steinmetz et al. 2001; Kim et al. 2006). Nrd1 is an RNA binding protein that interacts with Sen1 (Steinmetz and Brow 1996; Steinmetz et al. 2001; Ursic et al. 2004; Vasiljeva and Buratowski, 2006) and the RNA-binding protein Nab3 (Yuryev et al. 1996), and together they form the Nrd1 complex. Nrd1 and Nab3 bind to specific RNA sequences located downstream from the mature 3' end of snoRNAs (and snRNAs) (Steinmetz and Brow, 1996; 1998; Conrad et al. 2000; Steinmetz et al. 2001; Carroll et al. 2004; 2007). Mutations of these binding sites, or components of the Nrd1 complex, cause termination defects (Carroll et al. 2004; Kim et al. 2006). The Nrd1 complex associates with the Pol II CTD and co-precipitates with several snoRNAs and some snRNAs (Ursic et al. 1997; 2004). In addition, Nrd1 has been shown to interact with the Pol II CTD, the exosome, and TRAMP, thus linking termination with snoRNA 3' end processing (Kim et al. 2006; Steinmetz et al.

2001; Vasiljeva and Buratowski, 2006). This 'Nrd-1' termination pathway is also involved in the synthesis and termination of Cryptic Unstable Transcripts (CUTs), which are non-coding RNAs (approximately 200-600 nucleotides long), first discovered in yeast, and originating from intergenic regions (Kapranov et al. 2005; Wyers et al. 2005).

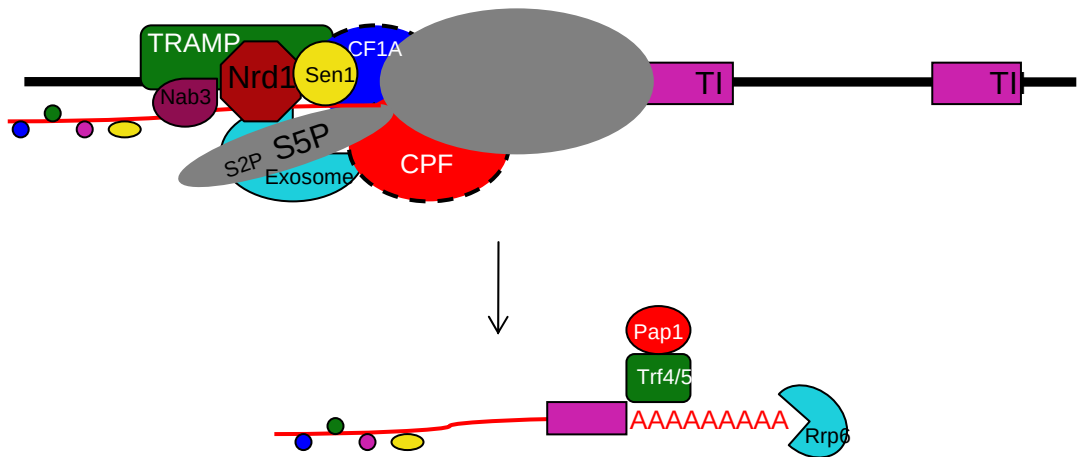
Even though snoRNAs are not polyadenylated, their 3' ends are formed using factors from the pre-mRNA 3' end processing machinery (Fatica et al. 2000; Morlando et al. 2002). These factors are the Pti1-associated CPF subcomplex APT, which includes Ref2, Ssu72, Pta1, Glc7 and Swd2; subunits of CFIA including Rna14, Rna15 and Pcf11; and the CPSF73 homologue Ysh1 (Dheur et al. 2003; Ganem et al. 2003; Nedeia et al. 2003; 2008; Steinmetz and Brow, 2003; Cheng et al. 2004; Kim et al. 2006). Mutations in components of the Paf1C elongation complex lead to a transcription termination defect of snoRNAs, indicating the importance of coordinating elongation with 3' processing and termination (Sheldon et al. 2005). It has been suggested that Paf1C may recruit Nrd1 and thus affect termination (Sheldon et al. 2005). Overall, mutations or deletions of factors involved in termination of transcription of mRNA genes appear to affect only a subset of snoRNA genes and, hence, it is likely that different pathways are used to direct transcription termination.

Some snoRNA genes appear to have two terminator elements; the major one being an Nrd1-dependent terminator followed by another which contains features similar to mRNA cleavage and polyadenylation signals (Fatica et al. 2000; Morlando et al. 2002; Steinmetz and Brow, 2003; Steinmetz et al. 2006). Surprisingly, both sequences are required for efficient termination. It has been shown that when placed in the 3'

UTR of a protein-encoding gene, the snoRNA terminator can direct polyadenylation (Steinmetz et al. 2006). In agreement with this finding, polyadenylation of some snoRNAs (and snRNAs) has been reported in exosome defective strains (van Hoof et al. 2000; Eggecioglu et al. 2006; Wyers et al. 2005). Defects in transcription termination and Nrd1 recruitment are observed in TRAMP-defective strains, suggesting that the TRAMP complex is necessary for Nrd1 recruitment to snoRNA genes (Grzechnik and Kufel, 2008). Furthermore, it has been shown that the noncanonical poly(A) polymerases Trf4 and Trf5 add adenines to the 3' end of the RNA product of the major terminator, whilst Pap1 polyadenylates snoRNA precursors generated at both terminators. To explain these observations, Grzechnik and Kufel (2008) suggest a model (Figure 1.5) based on the association of the Nrd1 and cleavage/polyadenylation complexes with the CTD. Here, Nrd1 and Nab3, assisted by TRAMP, recognise the first terminator sequence. Termination at this site leads to polyadenylation of the snoRNA precursor by Pap1 and Trf4, which is then trimmed by Rrp6, an exosome subunit, to form the mature snoRNA. If transcription fails to terminate at the first site and proceeds to the second site, pre-snoRNAs are processed by the 3' end cleavage/polyadenylation complex and polyadenylated by Pap1. These polyadenylated snoRNAs are then processed by the core exosome together with Rrp6. It is thought that this might be a failsafe mechanism of termination.

Studies in *S. cerevisiae* highlight the many common factors involved in the mRNA and snoRNA termination pathways and indicate that strict regulation exists to discriminate between the two. It has been suggested that one of the key determinants of terminator use is the distance travelled by Pol II when it reaches the terminator and that Nrd1 dependent termination is associated with termination of short transcripts

A)



B)

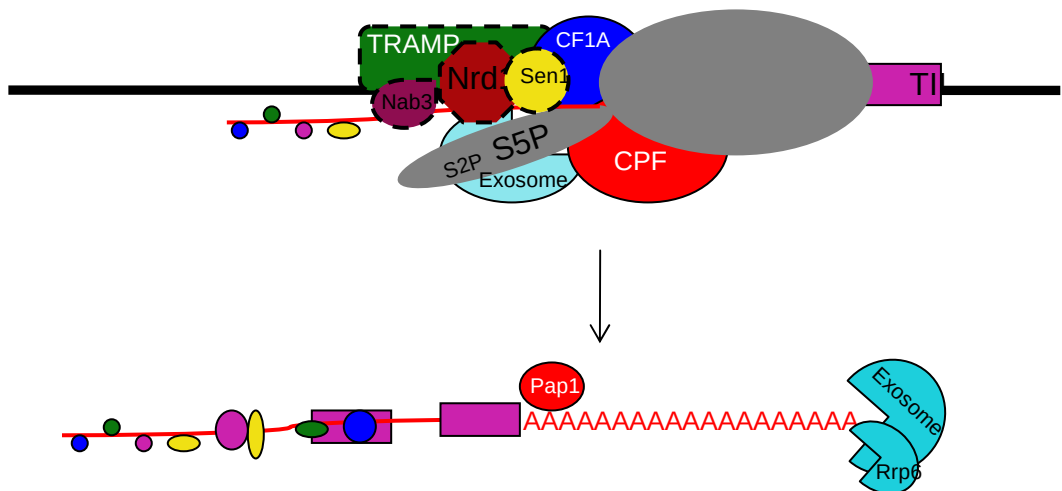


Figure 1.5: Pol II termination on *S. cerevisiae* snoRNA genes (adapted from Grzechnik and Kufel, 2008). DNA is shown as a black line. Pre-snoRNAs are shown as red lines with proteins (coloured circles) binding to them to form the snoRNPs. **A)** Termination at the major termination site (TI). Termination at TI occurs in an Nrd1-Nab3-Sen1 dependent manner. Nrd1 interacts with the Ser5 phosphorylated (S5P) CTD, the TRAMP and exosome complexes. The cleavage and polyadenylation factors CF1A and CPF are inactive and are indicated with dashed lines. Pre-snoRNAs released at TI are adenylated by the TRAMP components Trf4/5, polyadenylated by Pap1 and exonucleolytically digested by Rrp6. **B)** Failure of Pol II to terminate at TI leads to fail-safe CF1A and CPF dependent termination. The Nrd1-dependent complex is inactive and indicated with dashed lines. Pre-snoRNAs released at TII are polyadenylated by Pap1 and processed by the core exosome with Rrp6. Diagram not drawn to scale.

(Gudipati et al. 2008; Steinmetz et al. 2006). Further evidence highlighting the difference between the snoRNA and mRNA termination pathway is the observation that snoRNA termination is not affected in Rat1-defective strains, suggesting that a ‘torpedo’ mechanism is not used to terminate transcription, as is the case in protein-encoding genes (Kim et al. 2006). It would be interesting to see if a similar mode of snoRNA termination exists in mammals with an Nrd1 analogous factor, perhaps something resembling the U1 snRNA ‘termination factor’.

1.5d miRNA gene termination

MicroRNAs (miRNAs) are short (approximately 22 nucleotide-long) RNA molecules that act as post-transcriptional regulators by binding to complementary sequences on mRNA, usually leading to translational inactivation or mRNA degradation (Bartel, 2009). miRNA genes are usually transcribed by Pol II (Lee et al. 2004). Many miRNA genes are found within introns of coding and non-coding genes or even in exons (Rodriguez et al. 2004) and transcription of these miRNAs is coupled to their processing. Other miRNAs, found in intergenic regions, have their own promoters (Lau et al. 2001), and are transcribed as monocistronic or polycistronic units. miRNAs are processed by cleavage of the primary miRNA (pri-miRNA) by Drosha, a dsRNaseIII like enzyme, which is part of the Microprocessor complex. The Drosha cofactor DGCR8 is also part of this complex. The cleavage event results in the release of 60 to 90 nucleotide-long RNA hairpins called pre-miRNAs (Gregory et al. 2006), which are subsequently exported to the cytoplasm where they are further processed to mature miRNAs by another RNase III enzyme called Dicer (Lund and Dahlberg, 2006).

In the case of intron-encoded miRNAs, termination is not necessary and indeed is undesirable, as this could disrupt the host gene transcription. Coupling between transcription and processing has been demonstrated with the Microprocessor complex being recruited to sites of host pre-mRNA to cleave pri-miRNA transcripts during ongoing transcription (Kim and Kim, 2007; Morlando et al. 2008).

It has recently been shown that co-transcriptional processing of intergenic miRNAs is coupled to Pol II termination in a manner similar to mRNA genes (Ballarino et al. 2009). It has been demonstrated that termination occurs after Drosha cleavage of pre-miRNAs and that Xrn2-depletion leads to transcriptional readthrough (Ballarino et al. 2009). This data suggests that, similar to the ‘torpedo’ model of Pol II termination of mRNAs, Drosha cleavage provides an entry site for Xrn2, which degrades the 3' cleavage product, leading to termination. It has also been shown that Drosha recruitment to miRNA genes depends upon the presence of a miRNA promoter (Ballarino et al. 2009), a situation reminiscent of snRNA genes where an snRNA promoter is required for Integrator recruitment, which in turn affects termination.

Many miRNA genes are grouped in clusters, so termination due to Drosha cleavage could affect expression of downstream miRNAs. Indeed, in 80% of endogenous miRNA clusters, separate miRNA units are less than 1.5kb apart. An investigation into the role of spacing in miRNA transcription indicates that as spacing increases, the possibility of inter-cluster termination increases (Ballarino et al. 2009), indicating the importance of spacing. Poly(A) signals are found at the end of many miRNA transcription units (Saini et al. 2007). It has been proposed that these poly(A) signals are used as failsafe terminators to prevent readthrough if correct processing has not

occurred, similar to the situation observed with snoRNA transcription (Ballarino et al. 2009).

1.6 Transcription termination and chromatin structure

In the nucleus, DNA is packaged into chromatin. The basic, repeating unit of chromatin is the nucleosome, which consists of approximately 147bp of DNA wrapped around a histone protein core, itself consisting of two copies of each of the four core histone proteins; H2A, H2B, H3 and H4 (Kornberg, 1974; Luger et al. 1997). Chromatin can exist in two different forms: heterochromatin and euchromatin, which are associated with transcriptionally silent or active regions of the genome respectively.

Chromatin structure and nucleosome positioning are important in gene expression (Kornberg and Lorch, 1999). Chromatin structure is influenced by a number of factors, including chromatin remodelling activities (Vignali et al. 2000) and the DNA sequence binding preference of nucleosomes (Segal et al. 2006; Ioshikhes et al. 2006; Lee et al. 2007; Kaplan et al. 2009).

1.6a Nucleosome positioning

Studies of nucleosome positioning have revealed that promoter regions of constitutively active genes are generally nucleosome-depleted (reviewed by Cairns, 2009). A recent study in *S. cerevisiae* has shown that the 3' ends of genes, near poly(A) signals, are also nucleosome-free regions (Mavrich et al. 2008). In human cells, nucleosome depleted regions have been detected approximately 100bp upstream

and downstream of poly(A) signals (Spies et al. 2009). Nucleosome depletion does not appear to depend on expression levels as both active and inactive genes show the same profile (Spies et al. 2009). It is possible that nucleosome depletion at the 3' end of genes facilitates termination either by allowing access of 3' processing or termination factors to the pre-mRNA, or by enhancing Pol II release from the DNA template.

In a recent study, Kaplan et al. (2009) assessed the importance of DNA sequences in nucleosome positioning by comparing nucleosome occupancy at transcription start and end sites on yeast genomic DNA *in vitro* and *in vivo*. The results were similar, suggesting that the DNA sequence is a major determinant of nucleosome occupancy *in vivo*. However, a recent study has implicated the action of ATP-dependent *trans*-acting factors in determining nucleosome positioning at the 5' end of yeast genes, suggesting that the sequence alone is not sufficient in determining nucleosome positioning (Zhang et al. 2011). However, it has been known for some time that poly(dA):poly(dT) tracts do not associate efficiently with nucleosomes (Nelson et al. 1987). In search of sequence motifs in nucleosome-depleted promoters, it was revealed that there were numerous poly(dA):poly(dT) elements (Bernstein et al. 2004). This finding was confirmed when, in a screen of over 1000 5bp length sequences, the sequence with the lowest average nucleosome occupancy was AAAAA (Kaplan et al. 2009). Hence, if chromatin structure is important in termination, this might explain the high frequency of A-rich sequences at the end of genes, for example the human β -globin and the mouse β -major globin genes, which utilise such sequences to direct efficient termination.

1.6b Chromatin remodelling

Chromatin remodelling occurs either via covalent modifications of the histone tails (reviewed in Bannister and Kouzarides, 2011) or via ATP-dependent chromatin remodelling factors (reviewed by Narlikar et al. 2002; Racki and Narlikar, 2008) of which there are three families; SWI/SNF, ISWI, and CHD/Mi2. There are numerous possible histone modifications but methylation of K36 on H3, which is enriched in the body of genes and decreases near poly(A) sites, has been proposed as being involved in termination (Lian et al. 2008).

The *S. cerevisiae* Isw1 chromatin remodelling complex has been shown to play a role in all the steps of the transcription cycle (Morillon et al. 2003). There are two forms of the Isw1 complex; Isw1a, which contains Ioc3 and Isw1b, which contains either Ioc2 or Ioc 4. Isw1b has been detected in the terminator region of the actively transcribed MET16 gene. Furthermore, in cells lacking Ioc2, Pol II density increases over the 3' end of the gene suggesting a termination defect and that Ioc2 has a role in termination (Morillon et al. 2003). In addition, CTD phosphorylation is affected by Ioc2 depletion, which in turn affects 3' end processing factor recruitment and, thus, termination (Morillon et al. 2003).

A CHD/Mi2 chromatin remodelling family member, Hrp1 (Chd1 in *S. Cerevisiae*), was identified as a transcription termination factor in a genetic screen in *Schizosaccharomyces pombe* (*S. pombe*) (Alen et al. 2002). Chd1 is a conserved protein, which has a role in the maintenance of chromatin structure at the 3' end of genes. Yeast strains deficient in Hrp1/Chd1 exhibit termination defects as well as alterations in chromatin structure (Alen et al. 2002). One of the genes tested in their

study, GAL10, was not affected and appeared to require the combined action of Chd1, Isw1 and Isw2.

Further evidence that changes in chromatin structure can affect transcription termination comes from a recent study in *S. pombe* where it was demonstrated that a chromatin binding protein can recruit factors that are directly involved in termination (Gullerova and Proudfoot, 2008). The RNAi associated H3K9 mark over convergent genes is recognised by the heterochromatic protein Swi6 (HP1 in humans), which then recruits cohesin to the intergenic region. This in turn acts as a roadblock causing elongating Pol II to terminate.

1.7 The exosome

The exosome is a multisubunit 3'-5' exonucleolytic complex first discovered in studies of 5.8S rRNA 3' processing which has diverse functions (Mitchell et al. 1997). In the cytoplasm, it is involved in mRNA decay after deadenylation (Anderson and Parker, 1998) as well as turnover of defective mRNAs detected during translation (reviewed by Doma and Parker, 2007). It degrades mRNAs that contain AU-rich destabilisation elements (Chen et al. 2001; Mukherjee et al. 2002) and mRNAs that are targeted and cleaved in the process of RNAi (Orban and Izaurralde, 2004). The yeast nuclear exosome has been shown to be involved in 3' end processing of non-coding RNAs including snRNAs, snoRNAs, tRNAs and rRNAs (Allmang et al. 1999). Nuclear exosome is also involved in rRNA processing in mammalian cells (Schilders et al. 2005).

Several studies have shown that mRNAs produced in yeast strains with 3' end processing or mRNP packaging defects are rapidly degraded by the exosome. Thus the exosome is a key player in mRNA quality control (Bousquet-Antonelli et al. 2000; Hilleren et al. 2001; Libri et al. 2002; Rougemaille et al. 2007; Saguez et al. 2008). Furthermore, the exosome plays a major role in controlling the levels of different classes of non-coding RNAs, including the yeast cryptic unstable transcripts (CUTs), human promoter-upstream transcripts (PROMPTs) as well as antisense transcripts (Wyers et al. 2005; Arigo et al. 2006; Thiebaut et al. 2006; Preker et al. 2008). Finally, studies in *S. cerevisiae* demonstrate that the nuclear exosome is involved in maintaining heterochromatic silencing of rDNA repeats and telomeric loci (Houseley et al. 2007; Vasiljeva et al. 2008) and has been shown to contribute to RNAi-mediated gene silencing (Camblong et al. 2007).

The exosome consists of nine conserved proteins, which form the catalytically inactive 'core exosome'. The nucleolytic activity of the exosome complex derives from additional subunits and co-factors that assemble on to this core (Dziembowski et al. 2007; Liu et al. 2006). In plants there is a slight difference in that one of the 'core exosome' subunits displays some phosphorolytic exonuclease activity (Chekanova et al. 2000). The exact composition of the complex depends on its cellular location, and the organism it derives from (reviewed in Houseley et al. 2006).

The *S. cerevisiae* exosome gains its catalytic activity from the Dis3 (also known as Rrp44) subunit, which is found in both the nuclear and cytoplasmic exosome complexes (Mitchell et al. 1997; Allmang et al. 1999). Dis3 contains an RNase II catalytic domain, which is responsible for 3'-5' exonucleolytic activity (Frazao et al.

2006), and an N-terminal PIN domain, which exhibits endonucleolytic activity (Lebreton et al. 2008; Schaeffer et al. 2009; Schneider et al. 2009). It has been suggested that the endonucleolytic activity of Dis3 may be important for degradation of RNA substrates with secondary structures (Lebreton et al. 2008). The nuclear exosome contains another catalytic subunit, Rrp6, which possesses 3'-5' exonucleolytic activity (Briggs et al. 1998; Mitchell et al. 1997).

In humans, there are two different orthologues of the yeast Dis3 subunit, labelled hDis3 and hDis3L. Although both orthologues interact with the exosome, hDis3L is entirely cytoplasmic and lacks endonucleolytic activity (Tomecki et al. 2010; Staals et al. 2010). The human homologue of Rrp6, PM/Sc1100 (hRrp6), is predominantly associated with the nuclear exosome (Allmang et al. 1999; Tomecki et al. 2010) with only small amounts detected in the cytoplasm (Chen et al. 2001; Lejeune et al. 2003).

The exosome complex shows only very weak activity *in vitro* and from this it was inferred that there must be activating co-factors associated with the exosome *in vivo*. Indeed, it has been shown that in *S. cerevisiae*, the TRAMP complex activates the nuclear exosome *in vitro* and *in vivo* (Wyers et al. 2005; LaCava et al. 2005; Vanacova et al. 2005). TRAMP contains a non-canonical PAP, Trf4 or Trf5; an RNA-binding protein, Air1 or Air2; and an RNA helicase Mtr4. Trf4 and Trf5 have been shown to oligo-adenylate substrate RNAs, which are thereby labelled as a substrate for the exosome to initiate 3'-5' RNA degradation (LaCava et al. 2005; Vanacova et al. 2005; Wyers et al. 2005). Furthermore, it has also been shown that binding of the helicase Mtr4 is enhanced by the oligo(A) tails (Bernstein et al. 2008). There is also evidence that the TRAMP complex can also function independently of this

oligoadenylation activity (Houseley et al. 2007; LaCava et al. 2005; Rougemaille et al. 2007; Wyers et al. 2005). The existence of mammalian homologues of the TRAMP complexes suggests that a similar mechanism of activation of transcript degradation exists in humans. The *S. cerevisiae* cytoplasmic exosome is activated by a different co-factor, the SKI complex. SKI homologues are present in both the cytoplasm and nucleus of human cells and have even been detected on transcribed genes, suggesting additional roles for this complex in higher eukaryotes (Zhu et al. 2005).

1.8 RNase III enzymes

The RNase III family of enzymes are endoribonucleases, which cleave double stranded RNA. RNase III enzymes are divided into three classes based on their domain composition. Class 1 RNase III enzymes are composed of a ribonuclease domain and a double stranded RNA-binding domain (dsRBD). The best-studied eukaryotic class 1 enzyme is the *S. cerevisiae* Rnt1 enzyme. Class 2 RNase III enzymes have 2 ribonuclease domains and a dsRBD. Class 3 RNase III enzymes, also known as the Dicer family, have 2 ribonuclease domains, a dsRBD and an N-terminal DExD/H-box helicase domain, a PAZ domain and a small domain of unknown function (DUF283).

The dimeric catalytic domain structure of RNase III enzymes allows each domain to act on one strand of the substrate dsRNA to catalyse complete double stranded (ds) RNA cleavage. Cleavage by RNase III enzymes produces a 5' phosphate group and a two nucleotide overhang at the 3' end (Robertson et al. 1968). The different RNase III

classes have different substrates, selection of which is mediated by the constituent enzyme domains and interacting co-factors (reviewed by MacRae and Doudna, 2007). RNase III enzymes have been shown to be involved in transcriptional termination and/or processing of rRNAs, yeast snRNAs and snoRNAs, miRNAs and small interfering (si)RNAs. In yeast, it has been demonstrated that 3' cleavage products of RNase III enzymes are substrates for the 5'-3' exonuclease Rat1, which is involved in transcription termination (Kawauchi et al. 2008; Rondon et al. 2009; Ghazal et al. 2009).

1.8a Dicer

Dicer is involved in the biogenesis of miRNAs and siRNAs, both of which are key components of RNA silencing pathways (Jinek and Doudna, 2009). In humans, there is one species of Dicer, which processes short siRNAs (21-23 nucleotide long) from longer dsRNA precursors. These siRNAs guide the RNA-Induced Silencing Complex (RISC) to cognate sites on target mRNAs (reviewed in Hannon, 2002). RISC then carries out endonucleolytic cleavage of the target transcripts (Hammond, 2005). Similarly, miRNAs are processed to their mature form from longer pri-miRNAs in the cytoplasm by Dicer. The mature miRNAs then direct target specific gene silencing by blocking translation (reviewed by Carrington and Ambros, 2003) or by causing degradation of the target mRNA.

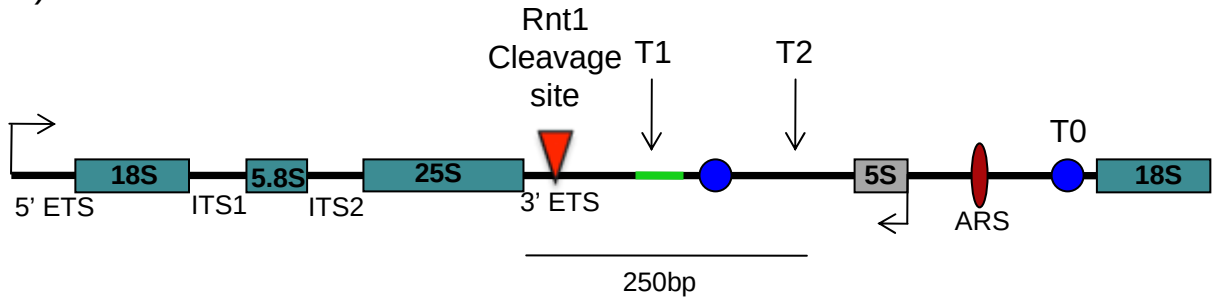
1.9 Transcription termination by RNA Pol I

RNA Pol I is responsible for transcription of rRNA, which represents more than 50% of the total RNA in growing cells. Ribosomal RNA genes are organised in a tandem

array of units comprising the 18S, 5.8S and 28S (25S in yeast) rDNA genes separated by two internal and two external transcribed spacer sequences (ITS and ETS respectively) and an intergenic spacer sequence (IGS) between the rDNA clusters (Figure 1.6). In yeast, the IGS is interrupted by the 5S rRNA gene, which is transcribed by Pol III in opposite polarity. Each ribosomal repeat is transcribed in the nucleolus to synthesise the 45S pre-rRNA in mammals (35S in yeast), which is then processed to release the 18S, 5.8S and 28S (25S in yeast) rRNAs.

Transcription termination by mammalian Pol I occurs in the IGS region at a series of terminator elements found within 1kb downstream of the 28S rRNA coding sequence (Grummt et al. 1985; Labhart and Reeder, 1986 and 1987). Pol I transcription terminators consist of a T-rich element followed by an orientation specific factor-binding site that has been shown to be recognised *in vitro* by Reb1 in yeast (Lang et al. 1994; Ju et al. 1990; Lang and Reeder, 1995). The mammalian major terminator element is called the Sal box, it is orientation-specific, and is recognised by TTF-I (Grummt et al. 1985; Kuhn et al. 1988; Bartsch et al. 1988; Evers et al. 1995). In mouse cells, the Sal box is repeated ten times in the IGS region (Bartsch et al. 1987; Evers and Grummt, 1995). The T-rich element is thought to act as a transcript 'release' element and substitution analysis indicates that it is the overall base composition and not specific sequences that are important in the release activity (Lang et al. 1994; Lang and Reeder, 1995). It has been shown that A or T rich elements are more efficient than G or C elements in enabling Pol I release suggesting that the DNA-RNA hybrid may have an important role in polymerase release (Kuhn and Grummt, 1989; Lang and Reeder, 1995).

A) *S. cerevisiae*



B) Mouse

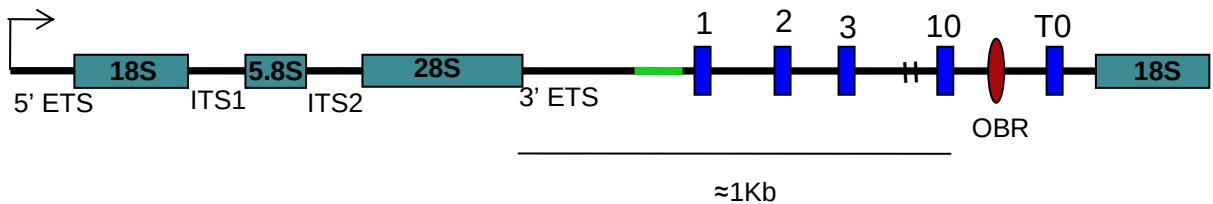


Figure 1.6: rDNA genes in *S. cerevisiae* and mouse. **A)** The yeast rDNA transcription unit. The 35S precursor consists of the 18S, 5.8S and 25S rRNAs separated by two Internal Transcribed Spacers (ITS1 and ITS2) and two External Transcribed Spacers (5' ETS and 3' ETS). The InterGenic Spacer (IGS) region in yeast is interrupted by the 5S rDNA gene, which is transcribed by Pol III in the opposite direction. The Rnt1 cleavage site (red triangle), the T-rich stretch (green line), the Reb-1 binding sites (blue circles), and the Autonomously Replicating Sequence (ARS, red oval) are shown. The second Reb1 binding site (T0) is in the opposite orientation and therefore unable to mediate termination. The major Pol I termination site is T1. **B)** The mouse rDNA transcription unit consists of the 45S precursor that contains the sequence of the 18S, 5.8S and 28S rRNAs, the ITS1, ITS2, 5' ETS and 3' ETS. As above, the T-rich stretch is represented by a green line, the Sal boxes as blue rectangles and the Origin of Bidirectional Replication (OBR) as a red oval shape. There are 10 Sal boxes found in the IGS and an additional one upstream of the transcription start site (T0). Pol I termination occurs predominantly at the Sal box 1 which is recognised by the mTTF-I termination factor. Diagram not drawn to scale.

Early *in vitro* studies of yeast Pol I transcription termination using purified components led to the proposal of a model whereby termination occurs when Pol I encounters DNA-bound Reb1/TTF-I causing it to pause, and this subsequently promotes termination at the upstream T-rich 'release' element (Lang et al. 1994). In mammals, an additional factor called Pol I and Transcript Release Factor (PTRF) is thought to be recruited by TTF-I, and was implicated as being necessary for the release step (Mason et al. 1997). Indeed, when PTRF is added to Pol I in *in vitro* transcription assays, it is sufficient to dismantle the elongation complex. The function of PTRF depends on the T-rich element and it has been suggested that it acts by binding the RNA U-stretch (Jansa et al. 1998; Jansa and Grummt, 1999). In *S. cerevisiae*, 90% of Pol I terminate at the first terminator (T1), at the T-rich 'release' element (Ju et al. 1990; Lang and Reeder, 1995), whereas the rest terminate at a second T-rich tract called the failsafe terminator (T2) (see Figure 1.6) (Reeder et al. 1999; Prescott et al. 2004).

More recent *in vivo* studies in yeast indicate that Pol I termination is more complex and may resemble some aspects of Pol II termination (Prescott et al. 2004; El Hage et al. 2008; Kawauchi et al. 2008; Braglia et al. 2011). Studies in *S. cerevisiae* have shown that pre-rRNA processing is coupled to transcription elongation and termination (Kufel et al. 1999; Reeder et al. 1999; Prescott et al. 2004; Kawauchi et al. 2008). It has been shown that the 3' end of the pre-rRNA transcript is cleaved in the 3' ETS to produce the 35S pre-rRNA by the dsRNA-specific RNase III enzyme, Rnt1 (Elela et al. 1996; Kufel et al. 1999; Henras et al. 2004). Inactivation of Rnt1 leads to readthrough of the T1 termination site, indicating that co-transcriptional cleavage is important in Pol I termination (Reeder et al. 1999; Prescott et al. 2004;

Kawauchi et al. 2008). Thus it appears that, in a similar manner as described for protein-encoding genes, the co-transcriptional cleavage is necessary for Pol I to terminate possibly using a ‘torpedo’ termination mechanism (El Hage et al. 2008; Kawauchi et al. 2008). Further evidence for this model of rRNA termination is that the 5'-3' exonuclease Rat1 was shown by ChIP analysis to associate with the IGS1 region of yeast rDNA and that in a strain lacking Rat1, there is a termination defect that cannot be complemented by catalytically inactive Rat1 (El Hage et al. 2008). Secondly, the RNA/DNA helicase Sen1 has also been detected at the terminator region and double Rat1 and Sen1 inactivation leads to severe termination defects (Kawauchi et al. 2008). The authors of this paper propose a model in which Rat1 is recruited to the 3' Rnt1 cleavage product and cooperates with Sen1 to degrade Pol I associated transcripts. When Rat1 ‘catches up’ with elongating Pol I, it destabilises the complex, promoting Pol I release from the template. The RNA/DNA kinase Grc3, has also been implicated in this process, and is thought to control the phosphorylation status of the 3' Rnt1 cleavage product and, thus, regulate its accessibility to Rat1 exonuclease (Braglia et al. 2010).

To reconcile the *in vitro* data from early experiments with the more recent data indicating that Pol I termination is explained by a ‘torpedo’ model, Braglia et al. (2011) suggest that the original Reb1-mediated pathway is part of a failsafe termination mechanism (Figure 1.7). In the absence of Rnt1, a second co-transcriptional cleavage event occurs at the T1 site (Braglia et al. 2011). Cleavage at T1 provides a ‘failsafe’ entry site for the 5'-3' exonuclease Rat1. Moreover, in the absence of Rnt1, an intact Reb1 binding site is required (presumably for binding by Reb1 or an alternative protein recognising the Reb1-binding site) to pause Pol I and

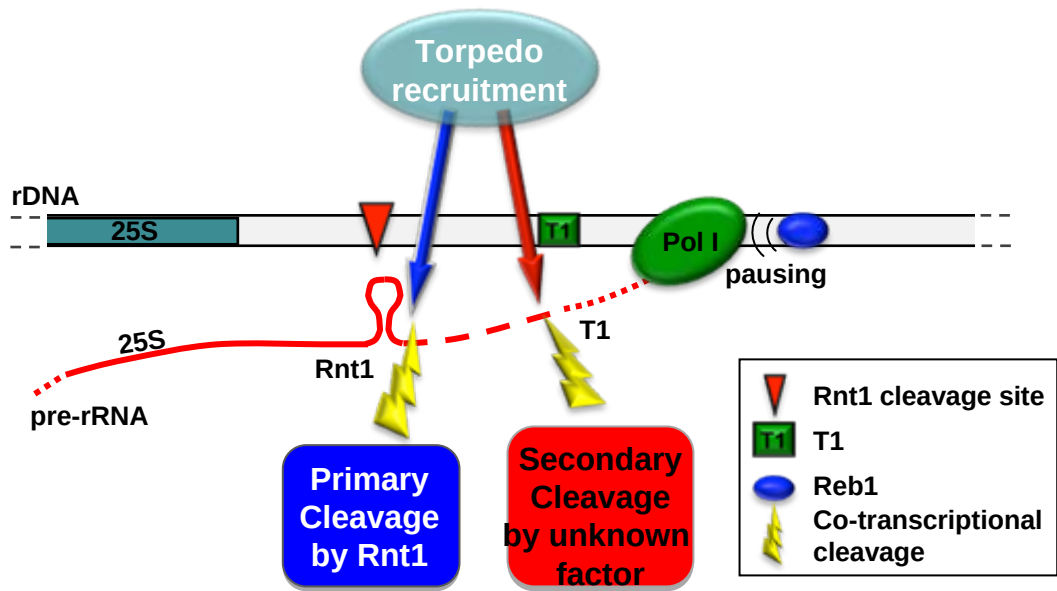


Figure 1.7: *S. cerevisiae* Pol I termination model (taken from Braglia et al., 2010). Two cleavage events occur at the major Pol I terminator site. Rnt1 cleavage provides an entry site for 5' to 3' exonuclease Rat1 to degrade the downstream transcript leading to termination. When Rnt1 is absent, an alternative cleavage event occurs at T1 providing a different entry site for Rat1. Cleavage at T1 requires the binding of Reb1 (or a different unspecified protein) to the Reb1 binding site to pause Pol leading to termination.

promote termination. To explain the data, the authors suggest a combined termination mechanism, involving ‘torpedo’ and ‘pause/release’, whereby pausing works together with cleavage at T1 to enhance an inefficient or kinetically slow event (Braglia et al. 2011). This RNA cleavage event at T1 is reminiscent of the CoTC event observed in the human β -globin gene, drawing a further parallel between Pol I and Pol II termination.

1.10 Transcription termination by RNA Pol III

RNA Pol III is responsible for the transcription of a large variety of short noncoding RNA genes, including tRNAs, 5S rRNA, U6 snRNA and 7SL RNA. Recent studies have demonstrated that the Pol III transcriptome is much larger than had been thought previously (reviewed in Dieci et al. 2007). Pol III transcription is very efficient, and this is thought to be due in part to the shortness of Pol III transcribed genes, and because of rapid recycling of Pol III to promoters which is dependent on Pol III termination signals (Dieci and Sentenac, 1996; Ferrari et al. 2004).

Pol III termination signals usually occur within the first 40bp downstream of the mature 3' end of RNAs. It is composed of a T-stretch, the length of which varies but most commonly is 4Ts in mammals (Braglia et al. 2005) and 5 to 6Ts in *S. cerevisiae* (Allison and Hall, 1985; Hamada et al. 2000). Sequences surrounding the T-stretch have been shown to influence termination (Bogenhagen and Brown, 1981; Mazabraud et al. 1987; Gunnery et al. 1999; Braglia et al. 2005) and it has been suggested that they may influence the pause time or processivity of Pol III (Braglia et al. 2005). Mutational analysis of these sequences leads to termination defects, which indicates

the requirement of this signal for termination (Bogenhagen and Brown, 1981). In an early study, it was shown that Pol III purified from ovarian tissue of *X. laevis* terminated transcription by recognising a T-rich signal (Cozzarelli et al. 1983). This provided evidence that Pol III possesses transcription termination activity without a requirement for additional factors. The Pol III termination process has been divided into 2 steps: firstly, transcriptional pausing when the terminator is recognised; followed by the second step which is polymerase and transcript release (Campbell and Setzer, 1992). However, it was also shown that Pol III pauses at non-terminator T stretches during transcription. This observation led to the suggestion that Pol III release at the terminator is dependent on the nascent transcript being displaced during elongation (Campbell and Setzer, 1992), suggesting the presence of *trans*-acting factors.

One such factor is the RNA binding protein La, which associates with Pol III RNA transcripts by binding to their 3'UUU_{OH} motif, preventing their exonucleolytic degradation (Maraia and Intine, 2001). There has been conflicting evidence as to the function of La. *In vitro* studies have shown a requirement for La in Pol III termination and reinitiation and it was proposed that it functions in transcript release (Gottlieb and Steitz, 1989; Maraia et al. 1994). More recently, La has been detected at Pol III transcribed genes *in vivo* (Fairley et al. 2005). However, the previous observation that Pol III is able to terminate transcription accurately by itself (Cozzarelli et al. 1983) and a more recent study which shows Pol III terminates in *X. laevis* nuclear extracts depleted of La indicate that Pol III transcription and termination is unaffected by its absence (Lin-Marq and Clarkson, 1998).

French et al. (2008) have conducted EM experiments visualising Pol III on *S. cerevisiae* 5S rDNA and detect a higher density of Pol III in a strain deleted for the La protein. In addition, they have shown that La has no effect on the rate of synthesis or the size of 5S rRNA and they propose that the extra Pol III visualised in the absence of La may reflect a defect in Pol III release. Taken together, the data relating to La may indicate a slightly different mode of termination between yeast and mammalian Pol III termination. It is possible that longer terminators found in yeast form less stable complexes and do not require accessory factors to aid in dismantling the Pol III elongation complex, whereas shorter mammalian terminators require accessory factors to terminate (Richard and Manley, 2009). Furthermore, additional factors have been implicated in Pol III transcription termination and reinitiation in mammals. The Pol III general transcription factor TF_{III}C, the Pol II transcription coactivator PC4 and Topoisomerase I have all been shown to interact with the terminator region of tRNA and VA1 genes, and are thought to facilitate Pol III termination and reinitiation (Wang and Roeder, 1996, 1998). It has also been suggested that the nuclear factor 1 (NF1) transcription factor stimulates Pol III termination and recycling (Wang et al. 2000). However, the NF1 binding sequence is not present at the terminator regions of many Pol III transcribed genes, and ChIP analysis failed to detect NF1 at Pol III transcribed gene loci (Fairley et al. 2005).

1.11 Transcription termination in bacteria

Transcription in bacteria is performed by a four subunit RNA Polymerase (RNAP), which has many structural and functional similarities with eukaryotic RNA Polymerases, reflecting their common evolutionary origin (Cramer, 2002). In

addition, transcription in bacteria is regulated by factors that have eukaryotic homologues or orthologues. Hence, considering these similarities, it is likely that studies of transcription termination in bacteria will reveal common features of termination for all RNA Polymerases. RNAP uses *cis*-acting sequence elements and/or *trans*-acting factors to elicit transcription termination. The *cis*-acting termination elements are grouped into two classes: intrinsic and factor-dependent terminators.

1.11a Intrinsic terminators

Intrinsic terminators in *E.coli* are characterised by a GC rich region of approximately 20 nucleotides in the RNA, which forms an RNA hairpin, followed by a run of U residues (d' Aubenton Carafa et al. 1990). The U tract causes RNAP to pause, which in turn provides enough time for the RNA hairpin to form (Gusarov and Nudler, 1999). At the same time, the paused RNA polymerase undergoes an active site rearrangement that inhibits nucleotide addition (Touloukhonov et al. 2007). The transcription protein NusA enhances pausing and termination by interacting with the RNA hairpin and a portion of the RNA exit channel (Touloukhonov et al. 2001; Landick and Yanofsky, 1987; Ha et al. 2010). The hairpin invades the main channel of RNA polymerase causing the RNA:DNA hybrid to melt which leads to dissociation of RNAP from the DNA template and transcript release (Gusarov and Nudler, 1999; Komissarova et al. 2002; Epshtein et al. 2007).

1.11b Factor-dependent terminators

The most widely studied factor-dependent terminators are those of *E.coli* and bacteriophage λ that respond to the termination factor Rho. Rho is a hexameric

ATPase, which binds RNA at Rho utilisation (rut) sites, which are pyrimidine-rich RNA elements over 30 nucleotides long, positioned upstream of the termination site. A model for Rho-dependent termination is that it binds to unstructured and ribosome-free RNAs and uses energy from ATP hydrolysis to travel along the transcript in a 5' to 3' direction, whilst threading the RNA through its central cavity, until it 'catches up' with elongating RNAP (reviewed in Ciampi, 2006). It was postulated that Rho can cause termination by using its translocase activity to push the RNA polymerase forward or pull the nascent RNA out, or use its helicase activity to unwind the RNA:DNA hybrid.

In contrast to the above model of Rho function, a recent study proposes that Rho loads directly on to RNAP early in elongation. Then when the RNA transcript is long enough it loads onto the RNA and traps the complex in an inactive conformation at Rho-dependent termination sites (Epshtein et al. 2010) (Figure 1.8). Furthermore, Rho-mediated termination depends on the connecting RNAP β' -lid and trigger loop domains. This observation led to the proposal of an allosteric model of Rho termination. In this model, Rho pushes against the β' -lid, to mediate RNA:DNA hybrid melting, and perhaps this also transmits an allosteric signal to the trigger loop facilitating catalytic inactivation and dissociation of the elongation complex (Epshtein et al. 2010).

1.11c General transcription termination lessons learned from bacteria

The original 'allosteric' and 'torpedo' models of transcription termination by Pol II were supported by studies of transcription termination in bacteria. The intrinsic terminator model of transcription termination in bacteria favours the allosteric model

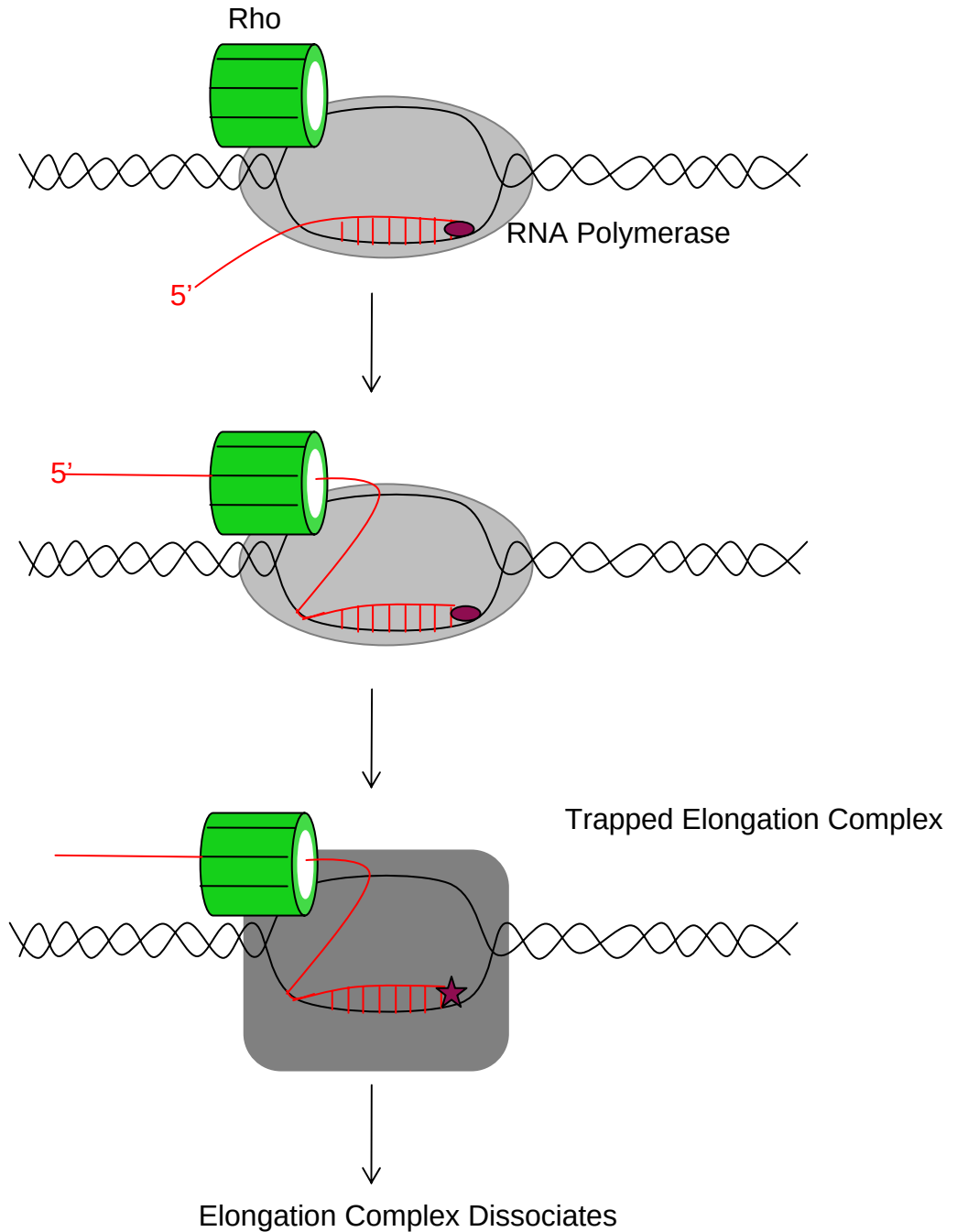


Figure 1.8: *E. coli* RNA Polymerase Rho-mediated termination (taken from Epshtein et al., 2010). DNA (black line), RNA (red line) and RNAP (grey oval) are indicated. The hexameric termination factor Rho (green cylinder) binds to RNAP throughout the transcription cycle. When the RNA is long enough, Rho loads onto the RNA and drags the transcript through its central cavity using energy from ATP hydrolysis. Topological strain causes formation of unstable trapped elongation complex where a conformational change in the trigger loop domain of the active site occurs leading to termination.

of Pol II termination whereby transcription of the poly(A) signal changes the composition or conformation of elongating Pol II making it more unstable and prone to termination (Logan et al. 1987). The original model of Rho-mediated termination, and the fact that *E. coli* Rho is able to dismantle *S. cerevisiae* Pol II elongation complexes *in vitro* (Lang et al. 1998), were used as evidence in support of the ‘torpedo’ model of Pol II termination, whereby cleavage at the poly(A) site enabled 5'-3' exonuclease recruitment which could ‘catch up’ with Pol II leading to termination (Connelly and Manley, 1988). Perhaps the most recent model of Rho-mediated termination outlined in Epshtein et al. (2010) (Figure 1.8) resembles more closely the hybrid model proposed for Pol II transcription termination.

The organisation of RNA and DNA within transcribing polymerases in the elongation complex is conserved in all RNA polymerases, which suggests that the general model of termination may be universal. It has been suggested that for Pol II transcribed genes Sen 1 or Rat 1 (Senataxin or Xrn2 in mammals respectively) may contact the Pol II Rbp1 lid, which could act as an effector domain, leading to termination in a similar fashion to how Rho leads to termination of RPAP in *E. coli* (Kuehner et al. 2011).

1.12 Common features of transcriptional termination

1.12a Fail-safe termination

Transcription termination is an intricate and critical process in gene expression. The yeast *S. cerevisiae* has developed fail-safe mechanisms to ensure Pol II termination on protein-encoding genes when the poly(A) dependent termination pathway is

compromised. In a recent study by Rondon et al. (2009), it was shown that in Rat1 deficient yeast strains, the Nrd1-dependent pathway can terminate elongating Pol II that has read past the poly(A) signal. This pathway is more usually involved in 3' end formation and termination of snoRNA genes and the synthesis and termination of CUTs (see section 1.5c).

Another fail-safe mechanism revealed by Rondon et al (2009), involved cleavage of the nascent transcript by Rnt1, providing an entry site for Rat1 to degrade defective transcripts. Termination by this mechanism also leads to degradation of the nascent transcript by the exosome. However, if there is a poly(A) site near the Rnt1 cleavage site, it recruits 3' processing factors and the mRNA is made (Rondon et al. 2009). Interestingly, it was shown that RNA cleavage alone is unable to mediate termination. This indicates that Rnt1 not only cleaves the transcript but also recruits termination factors. A number of yeast genes have Rnt1 cleavage sites and a genome wide analysis of an Rnt1 deficient strain resulted in increased transcriptional readthrough for a number of genes (Ghazal et al. 2009). In addition, the *S. pombe* RNase III enzyme Pac1 is able to induce termination in a protein-encoding Pol II transcript and can also act as a fail-safe terminator in non-polyadenylated transcripts (Nabavi and Nazar, 2010b).

It has been shown that transcriptional termination prevents transcriptional interference (Greger et al. 1998). Furthermore, termination may also prevent antisense transcripts being generated, which could lead to heterochromatin formation in yeast (Camblong et al. 2007; Gullerova and Proudfoot, 2008). The fail-safe termination mechanisms of Pol II described above are coupled to exosome degradation, suggesting they also act

as RNA quality control pathways. The only suggestion of such a failsafe mechanism occurring in mammalian cells comes from a study of miRNA genes, where it was shown that poly(A) signals are found downstream of many miRNA transcriptional units (Saini et al. 2007). It has been proposed by Ballarino *et al.* (2009), that these signals may act in a fail-safe type mechanism similar to what has been seen with snoRNA genes. However, there is, as yet, no direct evidence to support this. It is interesting to consider whether similar failsafe mechanisms of Pol II termination are present in mammalian cells.

1.12b Pausing and cleaving

Polymerase pausing has been widely implicated in transcriptional termination. As discussed earlier, there are a number of pause elements, found downstream of poly(A) signals. These elements are thought to slow down transcriptional elongation and allow enough time for 5'-3' RNA exonuclease (Xrn2/Rat1) to degrade the nascent transcript, 'catch up' with Pol II and lead to transcription termination. For Pol I, Reb1/TTF-I binding sites are involved in Pol I pausing and are thought to enhance an inefficient or kinetically slow event. In the case of Pol III, the T-rich element, which is part of the termination region of Pol III, is thought to direct pausing of Pol III. Bacterial polymerases pause at intrinsic terminators allowing sufficient time for an RNA hairpin to form which leads to termination. Overall, pausing is likely to be a common feature of termination, its role to enhance the time available for other events, such as RNA hairpin formation in bacteria or 'catching up' of a 5'-3' exonuclease as in the torpedo model of Pol II termination.

Endonucleolytic cleavage of the nascent RNA transcript is central to transcription termination. For most protein-encoding genes, transcript cleavage occurs either at the poly(A) addition site catalysed by CPSF73, or at a downstream CoTC terminator by an as yet unidentified activity. As discussed above, RNase III family members – Rnt1, Pac1 and Drosha – are able to endonucleolytically cleave Pol I and/or Pol II nascent RNA transcripts creating entry sites for the 5'-3' exonuclease Rat1/Xrn2 and thus mediating Pol II termination either as the principal transcription termination mechanism or as a failsafe-mechanism.

It is worth noting that bacterial RNA polymerases as well as eukaryotic Pol I, Pol II, and Pol III, all have an intrinsic ability to direct cleavage of the nascent RNA transcript, and this is associated with backtracking of the RNA polymerase (Whitehall et al. 1994; Kuhn et al. 2007; Cheung and Cramer, 2011). In the case of Pol II, the transcription elongation factor TFIIS is able to bind to Pol II (Sopta et al. 1985; Agarwal et al. 1991) and stimulate Pol II endoribonuclease activity (Izban and Luse, 1993; Guo and Price, 1993). Structural studies of purified Pol II from *S. cerevisiae* have led to a model in which, when Pol II encounters a DNA sequence that affects elongation, it pauses and backtracks by one nucleotide. Pol II then cleaves the displaced dinucleotide at the 3' end of the RNA so that elongation may continue (reviewed in Cheung and Cramer, 2011). However, if Pol II backtracks by more than one nucleotide, then Pol II arrests and elongation is inhibited. TFIIS binds to the arrested Pol II and reactivates it by inducing cleavage and release of the backtracked RNA. This in turn means that the new 3' end of the RNA is in the active site and transcription can resume. Both pausing and cleavage steps are a general feature of transcription termination and it is interesting to speculate whether this intrinsic ability

of Pol II is also used for termination, perhaps within CoTC elements. Structural modelling of purine bases onto the backtracked RNA suggests that these would clash with Pol II (Cheung and Cramer, 2011). This indicates that backtracked RNA prefers to bind to sites containing pyrimidine bases. This could explain why sites of DNA arrest are usually purine-rich (Hawryluk et al. 2004).

1.13 Aims of thesis

A detailed study from this laboratory has shown that the human β -globin gene terminator contains a sequence specific region that promotes rapid Co-Transcriptional Cleavage (CoTC) of the nascent transcript, which is a prerequisite for release of Pol II from the DNA template (Dye and Proudfoot, 2001; West et al. 2008) and is potentially a widespread feature of Pol II termination. Thus, the purpose of my thesis has been to investigate the signals that mediate CoTC activity in more detail to better understand CoTC mediated transcription termination. One of the aims of my project has been to investigate the role of sequences within this element in the termination process and this work is presented in Chapter 3. In Chapter 4, I address a different aspect of my study, which has been to identify the activity responsible for CoTC. To do this, I used the yeast *S. pombe* as a tool for genetic analysis. The results indicate that the human β -globin gene terminator does not direct efficient transcription termination in *S. pombe* and thus I decided to adopt a candidate approach to identify factors involved in CoTC activity. In Chapter 5, I describe an investigation into the possibility that Dis3 or Dicer RNA endonucleases are involved in CoTC mediated transcription termination and present preliminary analyses of the role of the CTD of Pol II and pre-mRNA processing events in CoTC activity.

CHAPTER 2

MATERIALS AND METHODS

2.1 Reagents

Most reagents used were obtained either from Sigma-Aldrich or from VWR for BDH and were of 'AnalaR' grade. Reagents obtained from other sources are listed below:

[$\alpha^{32}\text{P}$]dATP 3000Ci/mmol (10mCi/ml)	GE healthcare
[$\alpha^{32}\text{P}$]rUTP 3000Ci/mmol (10mCi/ml)	GE healthcare
5x Transcription Buffer	Promega
α -amanitin	Sigma
Acrylamide (40%, acrylamide:bis 19:1)	Severn Biotech
Agar Bacteriological	Oxoid
Agarose	Melford
Agar-technical	DIFCO
Biotinylated rUTP	Roche
Blasticidin	Invitrogen
Bradford reagent	Biorad
Deoxyribonucleotide triphosphates (dNTPs)	Invitrogen
Dithiothreitol (DTT)	Invitrogen
Dulbecco's Modified Eagle Medium	PAA
Dulbecco's PBS	PAA
ECL Western Blotting Detection Reagents	GE healthcare
Flp-In T-REX Core Kit	Invitrogen
Foetal Calf Serum	Gibco BRL
Fungizone	Gibco BRL
Glycogen	Roche
HEPES	Melford

Hybond-C extra protein transfer membrane	GE healthcare
Hybond-NX nitrocellulose	GE healthcare
Hygromycin B	Invitrogen
Lipofectamine 2000	Invitrogen
MicroSpin G25 Columns	GE healthcare
Maxi-prep kit	Qiagen
MAXIsript Kit	Ambion
Non Essential Amino Acids	PAA
OptiMEM	Gibco BRL
Penicillin/Streptomycin	PAA
Ribonucleotide triphosphates (rNTPs)	Roche
RNA Millenium Marker Probe Template	Ambion
Streptavidin coated magnetic beads	Promega
Trizol	Invitrogen
RNase-free tRNA	Roche
TOPO TA Cloning Kit	Invitrogen
Trypsin	Gibco BRL
Yeast extract	Oxoid
Zeocine	Invitrogen

2.2 Enzymes

Restriction enzymes were obtained from either Roche or New England Biolabs (NEB). Other enzymes were obtained from the following suppliers:

Shrimp Alkaline Phosphatase	Promega
-----------------------------	---------

T4 RNA Ligase	NEB
T7 RNA Polymerase	Promega
SP6 RNA Polymerase	Promega
Pfu Turbo DNA Polymerase	Stratagene
RNase out	Invitrogen
Proteinase K	Roche
RNase A	Roche
RNase H	NEB
RNase T1	Roche
T4 DNA Ligase	Roche
T4 Polynucleotide Kinase	NEB
SuperScript III reverse transcriptase	Invitrogen
DNase I- RNase free	Roche
Lysozyme	Sigma
GoTaq DNA Polymerase	Promega
Phusion Taq DNA Polymerase	Finnzymes
Sensimix (Real Time PCR reaction mix)	Bioline
KO Hot Start Polymerase	Novagen Toyobo

2.3 Antibodies

Anti-Actin (A2066)	Sigma
Anti-Mouse (M4155)	Sigma
Anti-Rabbit (R5506)	Sigma
Anti-Dis3 (KIAA1008-A01)	Abnova
Anti-Dicer (13D6)	Abcam

2.4 Yeast strain

<u>Strain</u>	<u>Genotype</u>	<u>Reference</u>
A2	<i>h⁺ ade6-M216 his3-D1 leu1-32 ura4-D18</i>	Lab stock

2.5 Nucleic Acid markers

DNA molecular weight markers (1Kb Plus DNA Ladder) for agarose gel electrophoresis were obtained from Invitrogen. Radioactive DNA markers (Hyperladder V) for acrylamide gels were obtained from Invitrogen and labelled with $\gamma^{32}\text{P}[\text{dATP}]$ using PNK. RNA molecular weight markers (RNA Millenium size markers) were obtained from Ambion.

2.6 Tissue Culture solutions

DMEM was supplemented with 10% foetal calf serum, 100units/ml penicillin, and 100units/ml streptomycin.

PBS/EDTA/Trypsin (PET) contained 0.5mM EDTA and 0.05% trypsin in PBS.

2.7 Yeast Media

YES: 5g/L yeast extract and 30g/L glucose. The media was supplemented with 225mg/L adenine, histidine, leucine, and uracil.

For solid media, 2% Difco Agar-technical was also added.

EMM: 3g/L potassium hydrogen phthalate, 2.2g/L sodium hydrogen phosphate, 5g/L ammonium chloride, 20g/L glucose, 2% (v/v) salts, 0.1% (v/v) vitamins, 0.01% (v/v)

minerals. The media can be supplemented with 225mg/L adenine, histidine, leucine and/or uracil as required.

For solid media, 2% Difco Agar-technical was also added.

Salts (50x): 52.5g/L magnesium chloride, 0.735g/L calcium chloride, 50g/L potassium chloride, and 2g/L sodium sulphate.

Vitamins (1000x): 1g/L panthothenic acid calcium salt, 10g/L nicotinic acid, 10g/L inositol, and 10mg/L biotin.

Minerals (10000x): 5g/L boric acid, 4g/L manganese sulphate, 4g/L zinc sulphate, 2g/L iron chloride, 0.4g/L molybdcic acid, 1g/L potassium iodide, 0.4g/L copper sulphate, and 10g/L citric acid.

2.8 Solutions and Buffers

Organic solutions:

Phenol/Chloroform: Phenol (pH4 or 8 for RNA or DNA isolation respectively) and chloroform were mixed in equal volumes.

Electrophoresis buffers:

50x E buffer (for agarose gels):

Sodium Acetate	250mM
EDTA	50mM
Tris	2M

Glacial acetic acid	91ml/L
---------------------	--------

5x TBE buffer (for polyacrylamide gels):

Tris	50mM
EDTA	10mM
Boric acid	50mM

RNA loading buffer (for polyacrylamide gels):

Formamide	80% (v/v)
EDTA	10mM
Bromophenol blue	0.25% (v/v)
Xylene cyanol	0.25% (v/v)

10x Orange G loading dye (for agarose gels):

Ficoll	25% (w/v)
Orange G	2.5mg/ml
EDTA	10mM

Solutions for yeast transformations:

LiOAc mix:

LiOAc	1mM
TE	1x

PEG mix:

LiOAc	1mM
-------	-----

TE	1x
PEG 4000	40% (v/v)

Solutions for DNA analysis:

STET:

Tris HCl, pH8	10mM
Sucrose	10%
EDTA	50mM
Triton X-100	0.5%

TE

Tris HCl, pH8	10mM
EDTA	1mM

Solutions for RNA analysis:

RNA solutions were either treated with 0.02% DEPC or made up with DEPC-treated water.

RNA lysis buffer:

Tris HCl, pH7.5	10mM
Sodium Chloride	140mM
Magnesium Chloride	1.5mM
NP40	0.5%

RNA lysis buffer & sucrose:

24% (w/v) sucrose in RNA lysis buffer.

High Salt Buffer (HSB):

Tris HCl, pH7.5	10mM
Sodium Chloride	500mM
Magnesium Chloride	10mM

DNase buffer:

Tris HCl, pH7.5	10mM
Magnesium Chloride	10mM

R-loop buffer:

Formamide	80% (v/v)
Sodium chloride	400mM
PIPES, pH6.4	40mM
EDTA	1mM

RNA elution buffer:

Ammonium acetate	500mM
EDTA	1mM
SDS	0.1%

RNase protection buffer:

Sodium chloride	300mM
-----------------	-------

Tris HCl, pH7.5	10mM
-----------------	------

EDTA	5mM
------	-----

Prior to use add:

RNase A	40µg/ml
---------	---------

RNase T1	1000U/ml
----------	----------

Proteinase K mix:

Proteinase K	3.3mg/ml
--------------	----------

SDS	7.7%
-----	------

Bead/probe conjugation buffer (for hybrid selection):

RNase protection buffer without RNases.

Bead wash buffer (for hybrid selection):

Tris HCl, pH7.5	10mM
-----------------	------

Sodium chloride	50mM
-----------------	------

EDTA	5mM
------	-----

20x SSC:

Sodium chloride	3M
-----------------	----

Sodium citrate	300mM
----------------	-------

This solution was adjusted to pH7 by addition of hydrochloric acid.

10x MOPS:

Sodium acetate 0.05M

EDTA 0.01M

This solution was adjusted to pH 7.0 by addition of hydrochloric acid.

Northern Hybridisation solution:

Formamide 50%

SDS 0.07%

Sodium chloride 5mM

Sodium phosphate 2.5mM

Northern wash solution 1:

SSC 2x

SDS 0.1%

Northern wash solution 2:

SSC 0.5x

SDS 0.1%

Solutions for NRO:

HLB/NP40:

Tris HCl, pH7.5 10mM

Sodium chloride 10mM

Magnesium chloride 2.5mM

NP40 0.5% (v/v)

HLB/NP40/sucrose:

10% sucrose in HLB/NP40

2x Transcription buffer:

Tris HCl, pH7.9	40mM
Potassium chloride	300mM
Magnesium chloride	10mM
Glycerol	40% (v/v)

Prior to use add:

DTT	2mM
-----	-----

20x SSPE:

Sodium Chloride	3.6M
Sodium dihydrogen phosphate	200mM
EDTA	20mM

This solution was adjusted to pH7.4 by addition of sodium hydroxide.

Denaturing solution:

Sodium chloride	1.5M
Sodium hydroxide	0.5M

Neutralising solution:

Tris HCl, pH7.2	0.5M
Sodium chloride	1.5M
EDTA	1mM

100x Denhardts solution:

Ficoll	2%
Polyvinylpyrrolidone	2%
Bovine serum albumin	2%

Pre-hybridisation solution:

SSPE	6x
Formamide	50% (v/v)
SDS	0.1%
Denhardts	5x
tRNA	100µg/ml

Hybridisation solution:

SSPE	7x
Formamide	58% (v/v)
SDS	0.12%
Denhardts	5.8x
tRNA	100µg/ml

Solutions for protein analysis:

RIPA buffer:

Sodium chloride	15mM
Triton X-100	1%
Sodium deoxycholate	0.5%
SDS	0.1%

Glycerol	10%
----------	-----

Prior to use add:

PMSF	3mM
------	-----

β -mercaptoethanol	5%
--------------------------	----

2x Laemmli:

Tris HCl, pH6.8	62mM
-----------------	------

Glycerol	25%
----------	-----

SDS	2%
-----	----

Bromophenol blue	0.25% (v/v)
------------------	-------------

PBS & Tween:

0.1% Tween in PBS

2.9 Plasmids and probes

DNA constructs were made by restriction digestion or long-range PCR using Pfu or KO Hot Start Polymerase. Details of each construct are given under the chapter in which they first appear. All plasmids were checked via DNA sequencing analysis.

Chapter 3

Plasmid construction

p β T⁺, p β Δ 8-10, p β BCD,p β ACD,p β ABD,p β CD, p β AD, p β TMut1 and pTat were a kind gift from Dr M Dye. The following constructs were made by PCR amplification

of p β T⁺ with the oligonucleotide pairs shown in brackets: p β A (REV/ Δ DF), p β TMut2 (10mut2F/10mut2R), p β TMut3 (10mut1F/10mut2R), p β TMut4 (10mut2F/10mut1R), and p β TMut2Control (10mut2CF/10mut2CR). The following constructs were made by PCR amplification of p β ABC with the oligonucleotide pairs shown in brackets: p β AC (REV/ β 10mid5'B) and p β AB (β 10mid3'B/ Δ DF). p β BC was made by PCR amplification of p β BCD using β 10mid3'B and 10.3. p β ABC and p β BC were made by blunt end ligation of fragments ABC and BC respectively to the linear p β T⁻ product resulting from PCR amplification of p β T⁺ using oligos β 4.3 and Δ DF. The fragments ABC and BC were made by PCR amplification of p β T⁺ using oligonucleotide pairs Δ DR/CoTCf and β 10mid5'A/ Δ DR. p β AA was made by blunt end ligation of fragment Asm to the linear p β A product (REV/ Δ DF). Asm was made by PCR amplification of p β T⁺ using primer pair CoTCf and REV. p β 30A was made by ligation of the annealed oligonucleotide pair A30/T30 to the linear p β T⁻ product.

pcDNA5/FRT/T⁺ and pcDNA5/FRT/Mut2 were made by blunt end ligation of PCR products β T⁺ and β Mut2 to around the vector PCR product of pcDNA5/FRT/TO opened using primer pair pc5f/pc5r. The PCR products β T⁺ and β Mut2 were generated from p β T⁺ and p β TMut2 respectively using primer pair β allf/ β allr.

M13 probes

The β -globin gene specific ssDNA probes have been described in Ashe et al. (1997). Probes P and U3 have been described in Ashfield et al. (1994). Probe A has been described in Dye and Proudfoot (1999).

Hybrid selection probe

The probe used ($\beta 3$) has been described previously in Dye and Proudfoot, 1999.

RPA-RT probe

The HIV-LTR probe used has been described previously in Plant et al. (2005).

Northern Probes

The β -globin exon 2 RNA probe was prepared by T3 transcription of a template produced by PCR amplification of $p\beta T^+$ using oligonucleotides E2FT7/E2RT3.

Chapter 4

Plasmid construction

pUI was a kind gift from Dr A. Aranda. The following constructs were made by blunt end ligation of the insert into a linear pUI vector prepared by *Stu*I restriction digest. The inserts were prepared as follows: pUI_{ura4Term}; the *ura4* terminator signal was amplified from *S. pombe* UI strain (Aranda and Proudfoot, 2001) genomic DNA with primer pair *ura4TermF/ura4TermR*, pUI_{spacer}; spacer was PCR amplified from $p\beta\Delta 8-10$ (described in Dye and Proudfoot, 1999) using oligonucleotides $\beta 5.5/\beta 7.3$, pUI_{Term/pUImutTerm}; Term and mutTerm were PCR amplified from $p\beta T^+$ and $p\beta T^{Mut2}$ respectively using oligonucleotide pair CoTCfCoTCr and pUI_{pATerm}, and pUI_{pAmutTerm} were PCR amplified from $p\beta T^+$ and $p\beta T^{Mut2}$ respectively using oligonucleotide pair 1.1/CoTCr.

Northern Probes

Probe 1, which is complementary to the *ura4* exon 1 sense RNA, was prepared by T3 transcription of a template produced by PCR amplification of pUI using oligonucleotides *ura1aFT7/ura1aRT3*. Similarly, probe 2, complementary to *ura4* exon 2 sense RNA, was prepared by T3 transcription of a template produced by PCR amplification of pUI using oligonucleotides *ura2aFT7/ura2aRT3*. Sense probe 1 and sense probe 2, complementary to antisense *ura4* RNA, were prepared by T7 transcription of the same PCR templates used for the antisense probes 1 and 2. The *adh* RNA probe was prepared by T3 transcription of a template produced by PCR amplification of *S. pombe* genomic DNA using oligonucleotides *adhFT7/adhRT3*. Promoter/exon 1 and exon 3 RNA probes were prepared by T7 transcription of templates produced by PCR amplification of p β T⁺ using oligonucleotide pairs *PromFT7/Ex1R* and *Ex3FT7/Ex3R* respectively. The 28S probe was prepared by T7 transcription of a template produced by PCR amplification of HeLa genomic DNA using oligonucleotides *28Sf/28SrT7*.

Chapter 5

Plasmid construction

p52repCTD and p5repCTD were a kind gift from Dr. S. Murphy and have been described in Gerber et al. (1995). pT+ Δ pA and pT+ Δ SA were made by PCR amplification using primers β 4.3/CoTCf from HIV β p(A)mut and HIV β IVS2SAmut respectively. HIV β p(A)mut and HIV β IVS2SAmut were a kind gift from Dr. M. Dye and have been described in Dye and Proudfoot (1999).

RNA interference

The Dis3 and Dicer siRNA sense sequence were AGGUAGAGUUGUAGGAAUAUU and UGCUUGAAGCAGCUCUGGAdTdT respectively.

2.10 Table of oligonucleotides

The oligonucleotides are listed under the chapter in which they first appear.

Chapter 3

REV	5'CATCAATTCTCGCATAAATGGCAGG
ΔDF	5'TCTAGAGCTTGGCACTGGCCG
10mut1F	5'CTCCAGGTTCATCAATTCTCGCATAAATGGCAGGATGTTCTTCTT GTTAATTTGTTTATTTTCCTATTGTTAATTGTTTGTTTGTTATTG
10mut1R	5'GATGTAACACTAAGAACAATAAGCCTGACACAACAAGACACATA CTACACAACCTTGCTCATATGTGACACATAACAAAGTCACTCT
10mut2F	5'CTCCAGGTTCATCAATTCTCGCATAAATGGCAGGATGGTCTTCT TGGTAATTTGGTTATTTTCCTATGGTTAATTGGTTGTTTGTTATTG
10mut2R	5'GATGTACCACTAAGAACCATAAGCCTGACACAACCAGACACCT ACTACACAACCTTGCTCATATGTGACCCATAACCAAGTCACTCT
10mut2CF	5'CGCCAGGTGCATCAATGCTCGCAGAAATGACAAGATTTTCTGCT TTTTAATTTTTTTTATTTTCCGATTTTTCATTTTTTGTGTTATTAGTG
10mut2CR	5'GATATAAACTCAGAAAAATAAGTCTGCCCCAAAAAGACAAAT ATTACACCACCTTGTTTCCTATGTGAAACATAAAAAAGTCCCTCTCC TGAAAA
β10mid5'B	5'GACAAATACTACACAACCTTGC
β10mid3'B	5'GCAAGGTTGTGTAGTATTTGTC
10.3	5'CAGACAGTAGAGGTATGGTTTCCAGGG
β4.3	5'CTTGAATCCTTTTCTGAGGGATG
ΔDR	5'TTTCCATGAGAGTGACTTTTTTATGTTTCAC

CoTCf	5'CACAGATGATTCAATAACAAAC
β10mid5'A	5'GCCATTTATGCGAGAATTG
A30	5'AAAAAAAAAAAAAAAAAAAAAAAAAAAAA
T30	5'TTTTTTTTTTTTTTTTTTTTTTTTTTTT
pc5f	5'CTGGGGATGCGGTGGGCTCTA
pc5r	5'GCGTATATCTGGCCCGTACATCGC
βallf	5'CCATGATTACGAATTCGAGCTCGGTA
βallr	5'AGAAAATACCGCATCAGGCGCCATT
E2FT7	5'TAATACGACTCACTATAGGGAGACTTGGACCCAGAGGTTCTTTG
E2RT3	5'AATTAACCCTCACTAAAGGGAGAATCCACGTGCAGCTTGTCACA
Int1f	5'GGAGACCAATAGAAACTGGGC
Ex2r	5'GAGCCAGGCCATCACTAAAG
CoTCf	5'CACAGATGATTCAATAACAAAC
CoTCr	5'GTTTCCTTGATTCTCCCACCCCC
28Sf	5'GCAGGAGGTGTCAGAAAAGTTACCACAG
28Sr	5'CTAACCTGTCTCACGACGGTCTAAACCC
rhologo	5'AAATGCACTGACCTCCCACAT
cRACEr	5'ATGTGCATTAGCTGTTTGCAG
E2f	5'CTTGGACCCAGAGGTTCTTTG
E2r	5'ATCCACGTGCAGCTTGTCACA

Chapter 4

ura4TermF	AAAAGGCCTGGGAATAAAAAGTAATTTGC
-----------	-------------------------------

Ura4TermR	AAACTCGAGCTACATGGTATTTTACAT
β5.5	TAGAGGCTTGATTTGGAGG
β7.3	CAATATCAAGCTACAAAAAGC
1.1	5'GAAGGGCCTTGAGCATCTGGATT
ura1aFT7	TAATACGACTCACTATAGGGAGAAGGATTTGACCAGGATATG
ura1aRT3	AATTAACCCTCACTAAAGGGGAGACTTGTATAATACCCTCGCCTG
ura2aFT7	TAATACGACTCACTATAGGGAGAACCTTAGAATGGTTTGAGAAGC
ura2aRT3	AATTAACCCTCACTAAAGGGGAGAAGGAGTACGATATTGCTGTCCC
adhFT7	TAATACGACTCACTATAGGGGAGATTCAAGGTGACTGGCCTCTT
adhRT3	AATTAACCCTCACTAAAGGGGAGACAAGGCACGATAGCAAGTGA
PromFT7	TAATACGACTCACTATAGGGGAGATGGGTCTCTCTGGTTAGACCA
Ex1R	TCACCACCAACTTCATCCACG
Ex3FT7	TAATACGACTCACTATAGGGGAGACTGGGCAACGTGCTGGTCTGT
Ex3R	TTAGTGATACTTGTGGGCCAGG
28SrT7	TAATACGACTCACTATAGGGGAGACTAACCTGTCTCACGACGGTCTA AACCC
U1f	AGGATTTGACCAGGATATG
U1r	CTTGTATAATACCCTCGCCTG
U2f	ACCTTAGAATGGTTTGAGAAGC
U2r	AGGAGTACGATATTGCTGTCCC
SS1	GTCTCTCTGGTTAGACCAGATCTG
SS2	GAACCCACTGCTTAAGCCTC
SS3	CTGAGGAGAAGTCTGCCGTTACT
E3r	TTAGTGATACTTGTGGGCCAGG

EF1 α F	GGCCCTAAATTCTTGAAGTCTGGTGA
EF1 α R	CTTGCCAGCTCCAGCAGCCTTCTT

Chapter 5

β -actinInt1f	CGGGGTCTTTGTCTGAGC
β -actinInt1r	CAGTTAGCGCCCAAAGGAC

2.11 DNA manipulations

Cloning

Cloning was performed according to Sambrook et al. (1989). In brief, vectors were digested with restriction enzymes as per the manufacturer's instructions whenever necessary and phosphatase using Shrimp Alkaline Phosphatase. In most cases, the vectors were prepared by long range PCR (Barnes, 1994; see below) with dephosphorylated primers. The DNA fragments to be used as inserts were also prepared by PCR with phosphorylated primers. In the majority of cases, the DNA vector and insert were gel purified before being ligated using T4 DNA ligase as per the manufacturer's instructions. The ligation mixtures were transformed into *E.coli* strain NM554. The bacteria were made competent using the calcium chloride method and stored at -80°C (kind gift from J. Monks). Plasmid DNA was isolated and analysed by restriction digest or colony PCR. The results were confirmed via DNA sequencing.

Plasmid isolation

Plasmid DNA was isolated from bacterial cultures using the STET boiled lysis method (Birnboim and Dolly, 1979). For large-scale preparations, the Qiagen maxi prep kit was used.

Long range PCR

Long range PCR (Barnes, 1994) was used to delete or mutate sequences within a plasmid or open up a plasmid. This was done by using primers which flanked the region to be deleted or, where the plasmid was going to be opened up at, and for mutations, the primers contained the mutations and covered the region to be mutated. Either *Pfu* or KOD Hot Start polymerase were used according to the manufacturer's instructions, with 20ng of plasmid DNA as template. The cycling conditions varied but were decided upon according to the guidelines from the manufacturer of the Polymerase to be used.

Quantitative PCR

Quantitative PCR (qPCR) was performed using the SensiMix No Ref Kit in a Corbett Rotorgene 3000 or 6000 machine. Reactions were done in triplicates and quantitated using the Rotorgene 3000 software. The cycling conditions were: 95°C 10 minutes, 45x [95°C 10 seconds, 55-60°C 15 seconds, 72°C 20-30 seconds]. A SYBR Green melt curve was performed simultaneously to confirm the quality of the PCR product.

Generation of HEK293 expression cell lines

The Flp-In T-Rex Core Kit was used to create HEK293 Flp-In T-Rex Expression Cell Lines containing stable integrants of the β -globin constructs pcDNA5/FRT/T+ and pcDNA5/FRT/Mut2, as per the manufacturer's instructions. The HEK293 Flp-In T-Rex Host Cell Line (kind gift from Torben Jensen) was used as the parental cell line for integration of these constructs and it contains stable genomic integrants of two plasmids: pFRT/lacZeo and pcDNA6/TR both of which were supplied with the kit. The Host Cell Line was grown in D-MEM media supplemented with 10% FCS, 1% NEAA, 100units/ml penicillin/streptomycin, 100 μ g/ml Zeocin, and 10 μ g/ml Blasticidine S HCl at 37°C with 5% CO₂. On day 1, 7x10⁵ cells on a 6cm dish were co-transfected with 9 μ g of pOG44 (plasmid expressing Flp recombinase) and 1 μ g pcDNA5/FRT/T+ or pcDNA5/FRT/Mut2 in D-MEM media supplemented with 10% FCS and 1% NEAA. At the same time, an untransfected control was processed. On day 2, the cells were split into a 15cm dish in D-MEM media supplemented with 10% FCS and 1% NEAA. On day 3, media was changed to D-MEM media supplemented with 10% FCS, 1% NEAA, 100units/ml penicillin/streptomycin, 10 μ g/ml Blasticidin S HCl and 100 μ g/ml Hygromycin B. Cells were cultured at 37°C with 5% CO₂ for three weeks until Hygromycin B and Blasticidin resistant colonies appeared. Individual colonies were selected and grown. The lines containing pcDNA5/FRT/T+ and pcDNA5/FRT/Mut2 are referred to henceforth as wild type and mutant stable cell lines.

2.12 Tissue Culture

Growth and Maintenance

HeLa cells were grown in D-MEM media supplemented with 10% FCS, 100units/ml penicillin and streptomycin at 37°C with 5% CO₂. For HEK293 cells, the D-MEM

media was also supplemented with 1% non-essential amino acids (NEAA). Cells were passaged when they had reached confluence. They were washed in PBS and treated with PET until they detached from the dish (a few minutes, usually). The cells were subsequently resuspended in D-MEM media and divided as required.

Transient transfections

1×10^6 cells were grown on 9cm plates and transfected with 5-10 μ g of reporter plasmid DNA and/or 1 μ g of Tat plasmid plus 12 μ l of Lipofectamine 2000 as per manufacturer's instructions. The cells were harvested 16-72 hours after transfection.

RNA interference

On day 1, 1×10^5 HeLa cells were grown per well in a 24-well plate. On day 2; 3 μ l of a 20mM stock of siRNA duplex or siRNA buffer and 47 μ l of Opti-MEM were mixed together. In a separate tube, 3 μ l of Lipofectamine 2000 plus 12 μ l of Opti-MEM were incubated at room temperature for 7 minutes. The mixtures were combined and incubated at room temperature for 25 minutes. Following that, 35 μ l of Opti-MEM was added and the mixture was pipetted directly into the well containing 500 μ l of D-MEM plus 10% Foetal Bovine Serum. On day 3, cells were split into a well of a 6-well plate. On day 4, cells were retransfected, as before, with the addition of the reporter plasmids, p β T⁺ and pTat. On day 5 fresh media was added and cells were allowed to recover. On day 6, the cells were either harvested for Western Blotting analysis, RNA isolation, or Nuclear Run-On analysis.

2.13 Yeast cultivation

Growth and Maintenance

Yeast strains were stored at -80°C in YES with glycerol added to a final concentration of 20%. When recovered from the freezer, cultures were grown on YES plates at 30°C and then grown in liquid YES for transformation, or in EMM lacking the appropriate amino acid(s) for further experiments. When in liquid medium, cultures were shaken at 180 rpm.

Transformation

20ml cultures were grown to approximately $0.5-1 \times 10^7$ cells/ml, harvested by centrifugation, and washed in water before being resuspended in 100µl LiOAC mix. Salmon sperm DNA was added as a carrier with 1-5µg of plasmid DNA and incubated at room temperature for 10 minutes prior to the addition of 260µl PEG mixture. The cells were then incubated at 30°C for one hour and then heat shocked at 42°C for 15 minutes. Transformed cells were washed in sterile water and plated on selective medium.

Growth sensitivity screens

Cells grown in EMM media lacking leucine or leucine and uracil, were plated in 10-fold serial dilutions on selective media plates.

2.14 Steady state RNA analysis

Mammalian Nuclear RNA isolation

Cells were washed in cold PBS and then scraped into 10mls of PBS. The cells were transferred to a nunc tube and pelleted by centrifugation at 500g for 5 minutes. The pellet was resuspended in 500µl RNA lysis buffer, underplayed with 500µl RNA lysis buffer/24% sucrose, and then centrifuged at 15800g for 10 minutes at 4°C. The

nuclear pellet was resuspended in 200µl HSB with 10 units of RNase free DNase I and incubated at 37°C for 5 minutes. The protein was then digested by addition of 200µl of proteinase K mix and incubation at 37°C for 30 minutes. Following this, the protein was removed by phenol/chloroform extraction and the nuclear RNA was ethanol precipitated. The RNA was further DNase treated by resuspending the RNA pellet in 200µl of DNase buffer and 10units of RNase free DNase I for 30 minutes at 37°C. DNase was removed by phenol/chloroform extraction and the RNA was ethanol precipitated. The RNA pellet was resuspended in DEPC-treated water, quantitated using a nanodrop, and stored at -20°C.

Mammalian total RNA isolation

Cells were washed with PBS, and 1ml of Trizol reagent was added directly to the cells. The RNA was then extracted following the manufacturer's instructions.

Yeast RNA isolation

Cell cultures growing in logarithmic phase were centrifuged for 3min at 12500g and washed in water. Cell pellets were resuspended in 500µl of TES, and then 500µl of acidic phenol (pH 4.5) was added before incubating at 65°C for one hour. The samples were vortexed every 10-15 minutes throughout the incubation time and centrifuged at 15800g for 10 minutes at room temperature at the end. The phases were separated and the aqueous phase was further purified by two acid phenol/chloroform extractions. The RNA pellet was resuspended in DEPC-treated water, quantitated using a nanodrop, and stored at -20°C.

RNase Protection Assay

A radiolabelled RNA riboprobe was made by transcription of 0.5µg of linearised pGEM containing the HIV-LTR probe (Plant et al. 2005), which was *in-vitro* transcribed in a reaction containing 1x transcription buffer, 10mM DTT, 400µM rATP, 400µM rCTP, 400µM rGTP, 40µCi [α^{32} P]rUTP, 10 units RNase Inhibitor, and 20 units T7 RNA Polymerase. The reaction was incubated for 1-2 hours at 37°C. After transcription the probe was treated with 2 units RNase-free DNase I for 30 minutes. Full-length riboprobe was purified from a 5% polyacrylamide 8.3M urea gel by excising a gel slice and eluting the RNA in elution buffer for 1 hr at 37°C shaking. Eluted RNA was ethanol precipitated in the presence of 10µg tRNA and resuspended in DEPC treated water.

For hybridisation, 500 cps of probe and 15 µg of nuclear RNA were ethanol precipitated and resuspended in 30µl R loop buffer. The hybridisation mixture was heated at 85°C for 10 minutes and then hybridised at 56°C overnight. 300µl RNase Protection buffer was then added and RNase digestion was performed for 2 hours at 18°C. Subsequently, 30µl of proteinase K mix was added to inactivate the RNases and incubated at 37°C for 30 minutes, and this was followed by phenol/chloroform extraction and ethanol precipitation in the presence of 10µg of tRNA. The RNA pellet was resuspended in 10µl RNA loading buffer, denatured at 85°C for 10 minutes, and fractionated by electrophoresis on a 6% polyacrylamide 8.3M urea gel. Protected fragments were visualized using autoradiography and quantitated using a Phosphoimager.

Hybrid selection circular RACE

A biotinylated selection probe was made by transcription of 0.5µg of linearised pGEM containing the β3 insert (Dye and Proudfoot, 1999), which was *in-vitro* transcribed in a reaction containing 1x transcription buffer, 10mM DTT, 1mM rATP, 1mM rCTP, 1mM rGTP, 0.65mM rUTP, 0.35mM biotinylated rUTP, 10 units RNase Inhibitor, and 20 units T7 RNA Polymerase. The reaction was incubated for 1-2 hours at 37°C. 1/10th of the reaction was spiked with [$\alpha^{32}\text{P}$]rUTP before transcription and analysed on 5% polyacrylamide 8.3M urea gel to check for the integrity of the probe. After transcription, the probe was treated with 2 units RNase-free DNase I for 15 minutes before being phenol/chloroform extracted. The biotinylated RNA probe was ethanol precipitated and resuspended in DEPC treated water.

1/20th of the biotinylated probe and 5µg of nuclear RNA were ethanol precipitated and resuspended in 30µl R-loop buffer. The hybridisation mixture was heated at 85°C for 10 minutes and then hybridised at 56°C overnight.

For the selection, 150mg of streptavidin-coated paramagnetic beads were washed four times in 150µl of wash buffer and then resuspended in 100µl of wash buffer. Incubation with 200µg of tRNA for 10 minutes blocks the beads. Once the beads were drained, the hybridisation mixture was added with 300µl of RNase protection buffer and the probe was allowed to bind to the beads within a 15-minute incubation period. The beads were drained and resuspended in 40µl 1x RNase H buffer. 50pmol of the oligonucleotide rholigo was added followed by denaturation at 65°C for 5 minutes. The samples were then cooled at 37°C before RNase H was added and the reactions incubated for 20 minutes. The beads were selected and the supernatant ethanol precipitated in the presence of 20µg of glycogen. The RNA was ligated at

room temperature for one hour using T4 RNA ligase. 1/4th of the reaction was used for cDNA synthesis with primer cRACer using SuperScript III. For PCR amplification, 1/20th of the reaction was used with divergent primers (cRACer and CoTCf) using GoTaq DNA polymerase. PCR products were analysed on agarose gels, cloned into TOPO vector, and then sequenced.

Reverse-Transcription PCR

1µg of RNA was reverse transcribed with either 50pmol of oligo-dT or 5pmol of a gene-specific primer using SuperScript III according to the manufacturer's instructions. For PCR amplification, 1/20th of the reverse transcription reaction was used. The PCR conditions varied but they were decided according to the guidelines from the manufacturer of the Polymerase to be used.

Northern Blot Analysis

15-40µg of RNA were mixed with 4µl 10x MOPS, 7.2µl formaldehyde and 20µl formamide and heated at 65°C for 15 minutes. 4µl of RNA loading buffer was then added to the samples before being separated on 1% formaldehyde agarose gel. Separated RNA was then transferred from the gel to Hybond-N nitrocellulose membrane using the Vacuum transfer machine for 90 minutes. Transferred RNA was cross-linked to the membrane in a UV-crosslinker (Stratalinker, Stratagene) with 700µJ x 100/cm² of energy. The membrane was then placed in hybridisation tubes containing 20ml of hybridisation solution and 50µg of denatured salmon sperm DNA was then added, followed by incubation from 2 hours to overnight at 62°C in a rotating hybridisation oven.

A radiolabelled RNA riboprobe was made by transcription of 0.1µg of PCR amplified template containing a T7 or T3 promoter. The probe was *in-vitro* transcribed in a reaction containing [$\alpha^{32}\text{P}$]rUTP and using the T7/T3 maxiscript kit (Ambion) according to the manufacturer's instructions. The riboprobe was passed through a G25 column to separate labelled products from unincorporated nucleotides, denatured at 95°C for 30 seconds, and then added to the membrane containing fresh hybridisation buffer/salmon sperm DNA mixture. Hybridisation was conducted for 6 hours to overnight at 62°C in a rotating hybridisation oven. The blots were then washed for 5 minutes at 62°C with washing solution 1, followed by 40 minutes at 62°C in wash solution 2. Signals on blots were visualized using autoradiography and quantitated using a PhosphoImager.

2.15 Nascent RNA analysis

Nuclear Run-On filter preparation

NRO filters were prepared using the slot blot apparatus from BIO-RAD. Three layers of Whatman paper were soaked in 5x SSPE buffer and placed in the slot blot underneath a pre-wet nitrocellulose hybridisation membrane. The blotting apparatus was assembled and each slot was washed out with 500µl of 5x SSPE. After this, 5µg of each of the M13 probes, diluted in 250µl of 5x SSPE was added to the appropriate slot. The slots were washed once more with 500µl of 5x SSPE. The filters were then placed in denaturing solution for 5 minutes and then neutralising solution for 2 minutes, before being air-dried. M13 probes were crosslinked to the filters by 2-minute exposure to UV light on a transilluminator. Filters were then placed in 10ml nunc tubes containing 2.5ml of prehybridisation solution and incubated overnight at 42°C in a rotating hybridisation oven.

Nuclear Run-On analysis of transfected HeLa nuclei

In the cold room, cells were rinsed in cold PBS, scraped into 10ml PBS, and pelleted by centrifugation at 500g for 5 minutes. The cell pellet was resuspended in 4ml HLB/NP40, incubated on ice for 5 minutes, and then underlayered with 1ml HLB/NP40/sucrose solution. The lysed cells were then centrifuged at 500g for 5 minutes. The size of the nuclear pellet was estimated and resuspended in an equal volume of 2x transcription buffer before being transferred to a 1.5ml microcentrifuge tube. 300mM rATP/rGTP/rCTP were added followed by 60μCi [α^{32} P]rUTP. The transcription was allowed to proceed for 15 minutes at 30°C. The reaction was stopped by centrifugation at 15800g for 30 seconds. The nuclear pellet was resuspended in 1ml Trizol reagent and RNA was isolated according to the manufacturer's instructions. The RNA was then resuspended in 60μl of DEPC-treated water, partially hydrolysed on ice for 5 minutes by addition of 15μl of 1M sodium hydroxide, and finally, neutralised by addition of 30μl of 0.1M Tris/0.1M HCl. Prehybridisation solution was replaced with 650μl of hybridisation solution and the RNA mix and incubated at 42°C in a rotating hybridisation oven overnight.

Filters were then washed in 1x SSPE/0.1% SDS on a shaking platform at room temperature for 20 minutes. More stringent washes, at 65°C were carried out if necessary. The washing solution could also be made more stringent by decreasing the SSPE down to 0.1x. Hybridisation signals were visualized by autoradiography and quantitated using a PhosphoImager.

2.16 Protein analysis

Western Blot analysis

The cells were washed and harvested in ice cold PBS. After centrifugation at 500g for 5 minutes at 4°C, the cell pellet was resuspended in RIPA buffer, incubated on ice for 10 minutes, and then centrifuged at 15800g for 10 minutes. Protein concentration was estimated by the Bradford assay. 10µg of protein extract was resuspended in 2X Laemmli buffer. The reactions were incubated at 100°C for 5 minutes to denature the proteins fully. The proteins were resolved on a SDS-polyacrylamide gel and transferred to Hybond-C extra nitrocellulose membrane using a semi-dry Transblot apparatus. The membrane was blocked with 5% skimmed milk in PBS with 0.1% Tween (PBST) for one hour on a rocking platform. Incubation with the primary antibody took place either for one hour at room temperature or at 4°C overnight in PBST with 5% skimmed milk. The membrane was washed in PBST (three times for 10 minutes) and probed with the secondary antibody for one hour. The membrane was then washed three times for 10 minutes with PBST and the blot was developed using the ECL Western Blotting Detection Kit according to the manufacturers instructions.

CHAPTER 3

INVESTIGATION OF *CIS* ELEMENT FUNCTION IN CoTC-MEDIATED TRANSCRIPTIONAL TERMINATION

3.1 Introduction

The most closely-examined termination sequence, for genes transcribed by Pol II, is that of the human β -globin gene (Dye and Proudfoot, 2001). This is an 800 base pair (bp) AT rich sequence element located approximately 1kb downstream of the poly(A) site in the β -globin 3' flanking region (Figure 3.1). Nuclear Run-On (NRO) analysis shows that this sequence is required for efficient termination. Furthermore, hybrid-selection NRO analysis reveals that nascent transcripts of the human β -globin terminator element are co-transcriptionally cleaved by an activity termed Co-Transcriptional Cleavage (CoTC) (Dye and Proudfoot, 2001). CoTC generates an unprotected 5' end, which is a substrate for degradation by the 5' to 3' RNA exonuclease, Xrn2 (West et al. 2004). CoTC and subsequent Xrn2 activity, mediates release of Pol II from the DNA template (West et al. 2008).

For the purpose of analysing CoTC and Pol II termination induced by the β -globin terminator sequence, the CoTC element has been divided into 3 sub-sections; labelled 8, 9 and 10 (Figure 3.1). It has been shown that each section (approximately 300 bp) is capable of directing transcription termination on its own and each can mediate CoTC activity (Dye and Proudfoot, 2001). Sequence comparison does not reveal any obvious sequence homology between sections 8, 9 and 10. However, all three terminator regions are AT rich. Deletion analysis of regions 8 and 9 has shown that rather than there being a specific sequence motif present, the entire sequence of each region is required to mediate efficient Pol II termination (personal communication, M. Dye). These observations are supported by RNase Protection Analysis (RPA) of hybrid selected nascent transcripts which indicates that CoTC, which is essential for

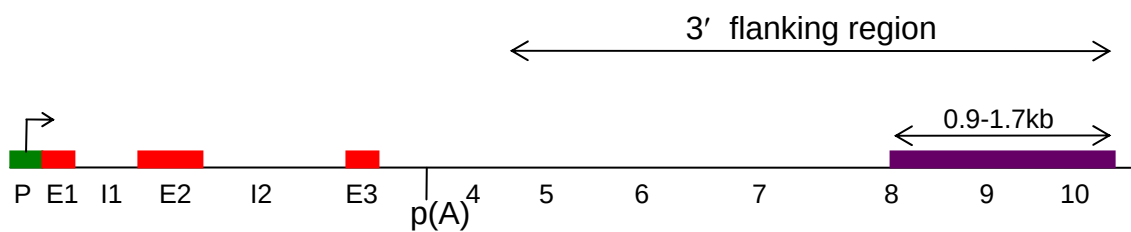


Figure 3.1: Structure of the human β -globin gene and 3' flanking region. DNA is indicated as a black line. The arrow represents the start site of transcription. P indicates the promoter. Exons are shown as red boxes and labelled E1 to E3; the introns are labelled I1 and I2. The poly(A) site is indicated by p(A). Regions 4 to 10, assigned based on restriction enzyme recognition sites, are indicated in the diagram. The terminator element is shown in purple and extends from 0.9 to 1.7kb downstream of the poly(A) sequence.

Pol II termination, is a stochastic process occurring throughout the β -globin terminator transcript (Dye and Proudfoot, 2001).

As detailed deletion analysis of sections 8 and 9 had already been conducted (M. Dye, unpublished data), I chose to examine the 3' section of the terminator, region 10 (henceforth referred to as 'Terminator'). This region of the terminator is particularly interesting because it contains a number of A tracts, which it was thought could represent conserved terminator elements.

3.2 Deletion analysis of the Terminator

To analyse sequences required for efficient transcription termination, the Terminator was divided into four sections labelled A, B, C and D (Figure 3.2B). These sequences were then cloned into a β -globin construct, p $\beta\Delta$ 5-10, which contains the β -globin gene driven by the HIV-LTR and has had the 3' flanking region removed (Figure 3.2A).

Constructs p β A, p β AB and p β ABC contain the β -globin gene, 240bp of post p(A) site sequence (lacking terminator activity) and then section A, A and B, and A, B and C of the Terminator respectively. p β T⁺ contains the β -globin gene, 240bp of post p(A) site sequence and the whole Terminator element. p β T⁻ lacks the Terminator as does p $\beta\Delta$ 8-10, which contains the β -globin gene and 500 bp of post p(A) site sequence (regions 5-7). p $\beta\Delta$ 8-10 was used as a negative control as it has previously been demonstrated that regions 5-7 are incapable of directing efficient termination (Dye and Proudfoot, 2001) and is comparable in sequence length to the full length β -globin

A



B

A
CAATAACAAACAAAAATTTAAAAATAGGAAAATAAAAAATTTAAAAAGAAGAAAATCCT
GTTATTGTTTGTTTTTAATTTTTATCCTTTATTTTTTTAATTTTTCTTCTTTTAGGA

B
GCCATTTATGCGAGAATTGATGAACCTGGAGGATGTAAACTAAGAAAAATAAGCCTGA
CGGTAAATACGCTCTTAACACTTTGGACCTCCTACATTTTGATTCTTTTTATTTCGGACT

C
CACAAAAAGACAAATACTACACAACCTTGCTCATATGTGAAACATAAAAAGTCACTCT
GTGTTTTCTGTTTATGATGTGTTGGAACGAGTATACACTTTGTATTTTTTTCAGTGAGA

D
CATGGAAACAGACAGTAGAGGTATGGTTTCCAGGGGTGGGGGTGGGAGAATCAGGAAA
GTACCTTTGTCTGTCATCTCCATACCAAAGGTCCCCAACCCACCCTCTTAGTCCTTT

CTATTACTCAAAGGGTATAAAATTTTCAAGTTATGTGGGATGAATAAATT
GATAATGAGTTTCCCATATTTTAAAGTCAATACACCCTACTTATTTAA

Figure 3.2: The β -globin plasmid and Terminator sequence. **A)** The $p\beta T+$ β -globin plasmid. DNA is indicated as a black line. The arrow indicates the start site of transcription, the dark green rectangle indicates the gene ORF, pA represents the poly(A) signal and the blue, red, yellow and green boxes represent Terminator sections A, B, C and D respectively. **B)** The Terminator sequence. The sequence is divided into 4 sections: A, B, C and D which are shown in blue, red, yellow and green characters respectively.

Terminator. HeLa cells were transiently transfected with constructs p β A, p β AB and p β ABC, as well as p β T+ as a positive control and, p β T- and p β Δ 8-10, as negative controls (Figure 3.3). In each case, a plasmid expressing the transactivator Tat (pTat) was co-transfected to enable efficient Pol II initiation/elongation from the HIV-LTR (for review see Cullen, 1993). The transcription termination capacity of each sequence was measured using RNase Protection readthrough assay (RPA-RT) (Plant et al. 2005). In this assay, nuclear RNA isolated from transfected HeLa cells is hybridised to a radiolabelled antisense RNA probe that spans the HIV-LTR. Potentially two different protection products can be detected using this probe. In the case of efficient termination one short band (85 nt), labelled P, resulting from protection of the probe complementary to the 5' end of mRNA is detected at the bottom of the gel. In the case of inefficient termination, two bands are detected. The 85 nt band P, and a larger readthrough product (240 nt), labelled RT, that results from protection of a longer part of the probe, due to hybridisation to transcripts that have read past the end of the gene around the plasmid backbone and through the HIV-LTR (Figure 3.3B). The RT product is relatively unstable and not exported to the cytoplasm (Dye, 1999). Consequently, this is not a strictly quantitative assay. However, the ratio of RT/P in general reflects transcription termination efficiency, as will be shown below.

Following RNase digestion, protected products were analysed by PAGE (Figure 3.3B). The abundance of P and RT bands in the presence of the Terminator is shown in lane 1 (p β T+) where there is a very strong band for P, which detects the 5' end of correctly initiated mRNA. The RT band is relatively weak, confirming the previous reports of efficient termination. In the absence of the Terminator, p β T- (lane 2) and

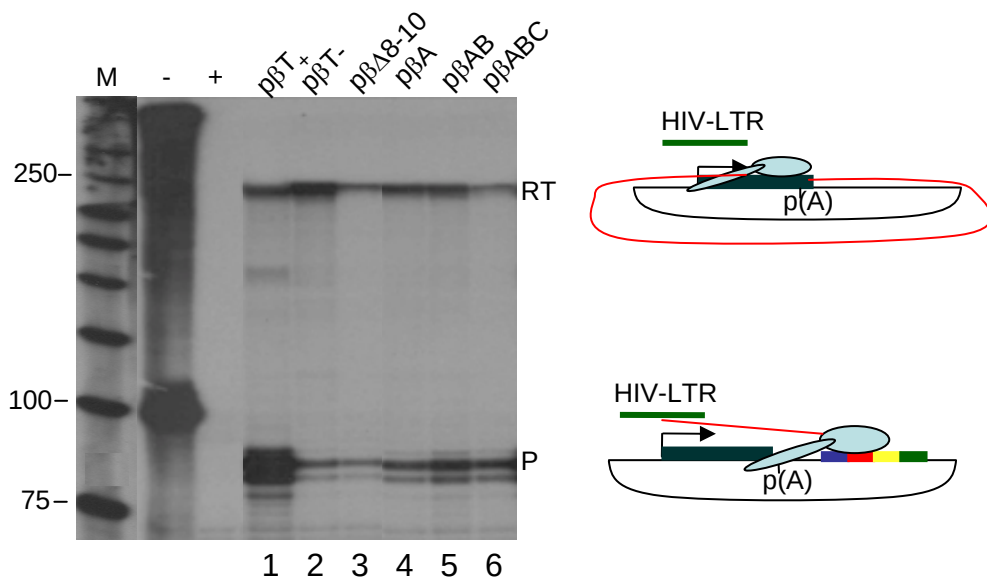
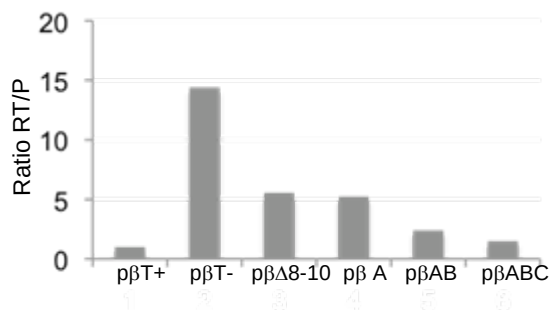
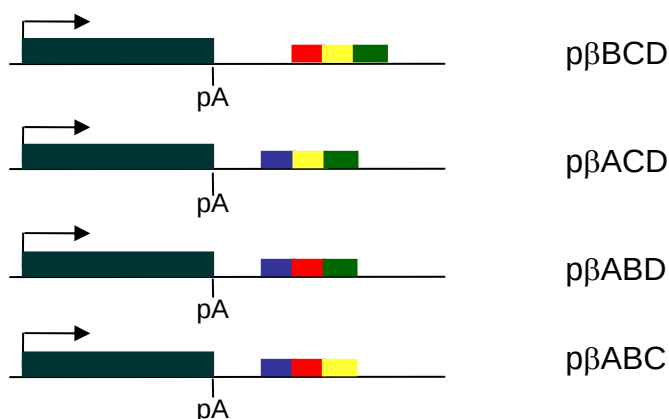
A**B****C**

Figure 3.3: Effect of Terminator element deletions on Pol II termination - I. A) Deletion constructs of the Terminator. The blue, red and yellow boxes indicate Terminator sections A, B and C respectively. **B)** RNase protection readthrough assay (RPA-RT). The rationale behind this experiment is represented in diagrams to the right hand side of the autoradiograph. Nuclear transcripts are hybridised to a radiolabelled antisense RNA probe (HIV-LTR) that spans the transcription start site (arrow) of the β -globin constructs. Band P (85 nt) results from hybridisation of correctly initiated transcripts to the antisense probe. Band RT (240 nt) results from hybridisation of readthrough transcripts to the antisense probe. RPA-RT analysis of HeLa cells transiently transfected with pβA, pβAB and pβABC deletion constructs. pβT+ was used as a positive control. pβT- and pβΔ8-10 were used as negative controls. '+' and '-' indicate positive and negative controls for efficiency of RNA degradation. Size markers are indicated by M. This experiment has been repeated at least three times and the pattern of RT and P bands obtained is reproducible. In this and subsequent RPA-RT analyses, the lanes have been rearranged for presentation purposes. **C)** Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, pβT+, which is set at 1.

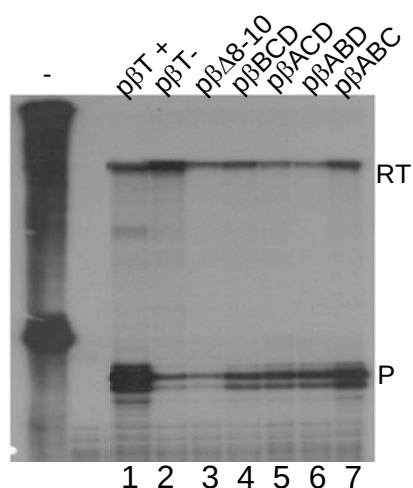
p β Δ 8-10 (lane 3), there is a significant change in the level of P and RT band. Band signal strength was quantified by PhosphoImager and the ratio RT/P calculated for each construct, normalised to the positive control, p β T+, and displayed in the graph below the data panel (Figure 3.3C). In this experiment, the effect of increasing the size of the Terminator element was assessed. In lane 4 (p β A), a strong RT band and a weaker P band are detected, indicating inefficient termination. When the Terminator is increased by including section B (p β AB), the termination defect decreases, as indicated by the stronger P band (lane 5). Finally, in lane 6, when 3 out of 4 sections from the Terminator are included (p β ABC), the level of RT decreases substantially indicating more efficient termination. Overall, the data in Figure 3.3C shows that as the size of the Terminator increases, the ratio RT/P decreases, indicating more efficient transcription termination.

I next investigated the effect of deleting individual sections of the Terminator using the RPA-RT assay. I made 4 constructs, labelled p β BCD, p β ACD, p β ABD and p β ABC, from which sections A, B, C or D were deleted respectively from p β T+ by a long range PCR strategy (Barnes, 1994). Nuclear RNA, isolated from Hela cells transiently transfected with pTat and p β BCD, p β ACD, p β ABD, p β ABC or the control constructs p β T+, p β T- and p β Δ 8-10, was subjected to RPA-RT analysis using the antisense HIV-LTR probe. Following RNase digestion, protected products were analysed by PAGE (Figure 3.4B). Band signal strength was quantified by PhosphoImager and the ratio RT/P calculated for each construct, normalised to the positive control, p β T+, and displayed in the graph below the data panel (Figure 3.4C). This data shows that deletion of a single section of the Terminator has a small effect on termination efficiency. However, this effect does not appear to be very significant

A



B



C

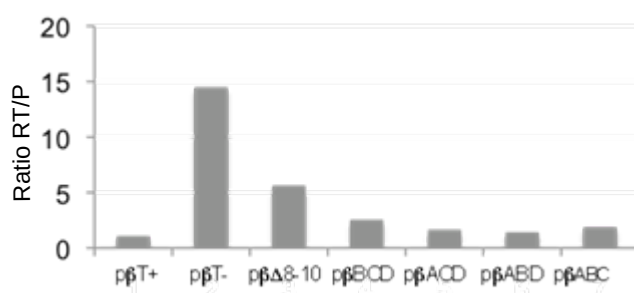


Figure 3.4: Effect of Terminator element deletions on Pol II termination - II. **A)** Deletion constructs of the Terminator. The blue, red yellow and green boxes indicate Terminator sections A, B, C and D, respectively. **B)** RPA-RT analysis of HeLa cells transiently transfected with pβBCD, pβACD, pβABD and pβABC deletion constructs. pβT+ was used as a positive control. pβT- and pβΔ8-10 were used as negative controls. '+' and '-' are as in Figure 3.3B. RT represents readthrough transcripts and P represents correctly initiated transcripts. This experiment has been repeated at least three times and the pattern of RT and P bands obtained is reproducible. **C)** Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, pβT+, which is set at 1.

when compared to removal of the whole Terminator region, p β T- (lane 2). Nonetheless, it appears that there is a hierarchy in which different sections of the Terminator have different strengths (A>D>B>C). In particular, deletion of section A leads to the highest RT/P ratio indicating the biggest termination defect. It is noteworthy that section A has a large number of oligo A tracts (this point is returned to later in this Chapter). Overall, the constructs lacking one section of the Terminator are quite efficient, but the full Terminator is still the most efficient at directing transcription termination. Thus it appears that as with regions 8 and 9 of the β -globin terminator (M. Dye, personal communication), the entire sequence is required for efficient transcriptional termination.

To investigate individual effects more fully, I examined if a combination of two Terminator sections was sufficient for full Terminator efficiency. I made six constructs labelled p β CD, p β BD, p β BC, p β AD, p β AC and p β AB which contain shorter versions of the Terminator (Figure 3.5A), and analysed them by the RPA-RT assay. Nuclear RNA isolated from HeLa cells transiently transfected with pTat and p β CD, p β BD, p β BC, p β AD, p β AC, p β AB or the control constructs p β T+, p β T- and p β Δ 8-10 was subjected to RPA analysis using the antisense HIV-LTR probe. Following RNase digestion, protected products were analysed by PAGE (Figure 3.5B). Band signal strength was quantified by PhosphoImager, and the ratio RT/P calculated for each construct, normalised to the positive control, p β T+, and displayed in the graph below the data panel. As expected from the data in Figure 3.4, which indicates that there is a hierarchy between the different sections (A>D>B>C), p β AD is more efficient at directing termination than p β BC (Figure 3.5C). Interestingly, some combinations are more efficient than others, and this does not always correlate

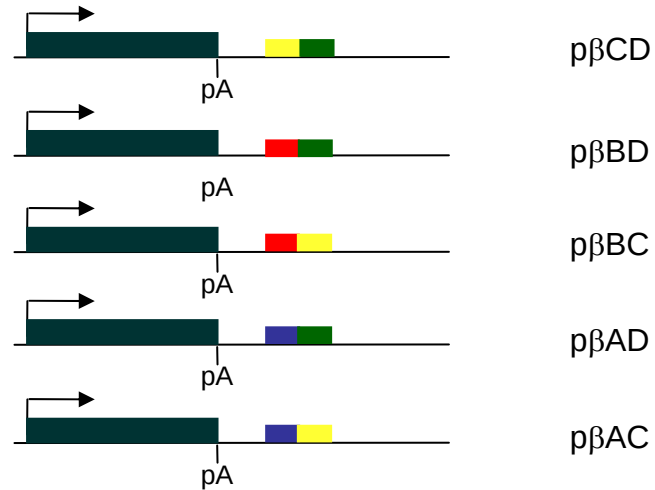
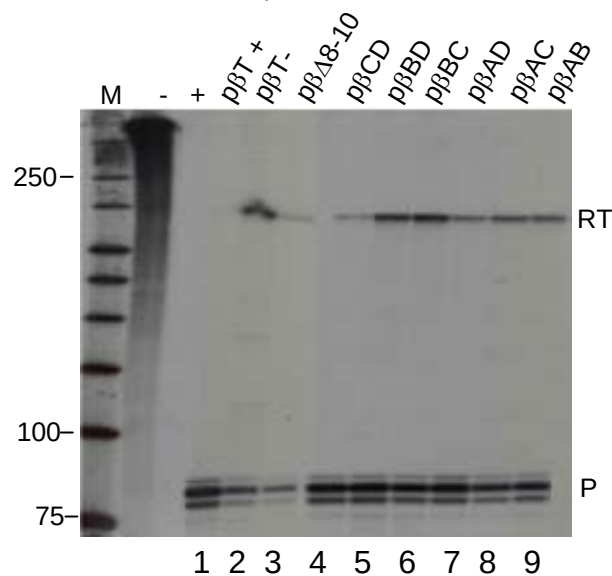
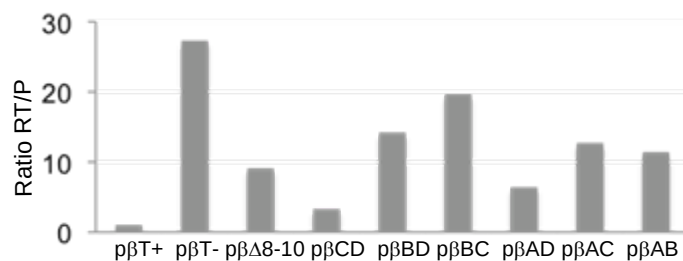
A**B****C**

Figure 3.5: Effect of Terminator element deletions on Pol II termination - III. **A)** Deletion constructs of the Terminator. The blue, red yellow and green boxes indicate Terminator sections A, B, C and D respectively. **B)** RPA-RT analysis of HeLa cells transiently transfected with pβCD, pβBD, pβBC, pβAD, pβAC and pβAB deletion constructs. pβT+ was used as a positive control. pβT- and pβΔ8-10 were used as negative controls. '+' and '-' are as in Figure 3.3B. RT represents the readthrough transcripts and P represents correctly initiated transcripts. This experiment has been repeated at least three times and the pattern of RT and P bands obtained is reproducible. **C)** Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, pβT+, which is set at 1.

with the effectiveness of the individual sections. For example, p β CD is the most efficient followed by p β AD. This could suggest that different sections of the Terminator have different functions, for example cleavage or pausing, and different combinations of these activities affect termination. Furthermore, the structure formed by these elements would vary depending on the sequence of the sections around it and perhaps this could be one of the reasons for this discrepancy. In spite of this, two sections of the Terminator are not sufficient for efficient termination to occur.

In summary, deletion of any region of the Terminator leads to a transcription termination defect (Figure 3.3, 3.4 and 3.5). RPA-RT analysis indicates that different sections of the Terminator possess different capacities to elicit termination (Figure 3.4). I next tested the termination efficiency of two A sections. To this end, I made a construct containing 2 copies of section A (p β AA). Nuclear RNA isolated from HeLa cells transiently transfected with pTat and constructs p β A, p β AA or the control constructs p β T⁺ and p β T⁻ was subjected to RPA-RT analysis using the antisense HIV-LTR probe. Following RNase digestion, protected products were analysed by PAGE (Figure 3.6B). Band signal strength was quantified by PhosphoImager and the ratio RT/P calculated for each construct, normalised to the positive control, p β T⁺, and displayed in the graph below the data panel (Figure 3.6C). This data shows that doubling up section A (p β AA) does enhance Terminator function, but not to the level seen with the full Terminator sequence (Figure 3.6). However, it is notable that region A contains several runs of A residues (Figure 3.2). These A tracts are distributed throughout the Terminator (Figure 3.7). Therefore, I next decided to analyse the role of A tracts in the Terminator.

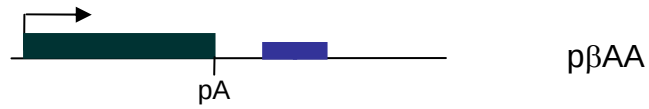
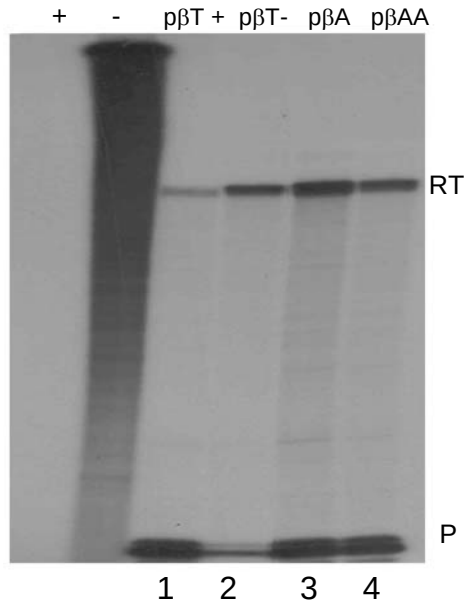
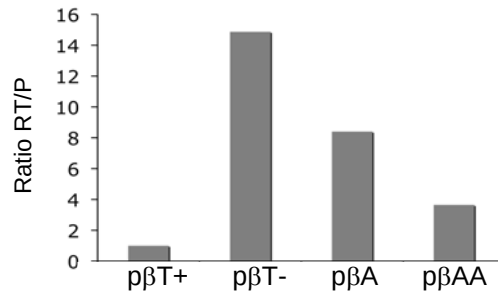
A**B****C**

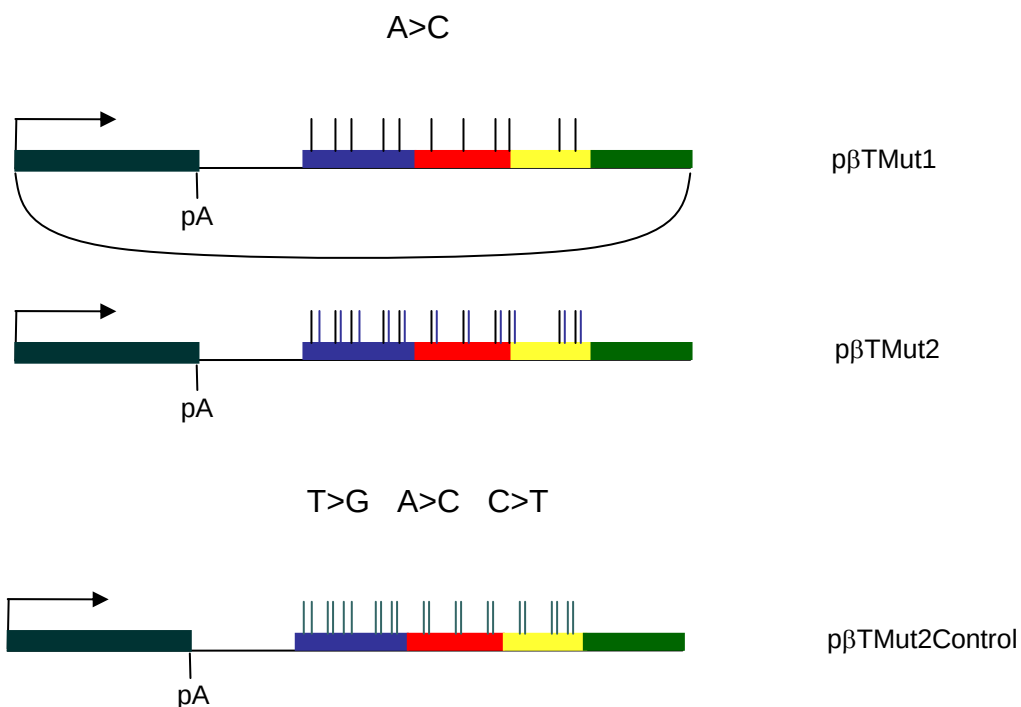
Figure 3.6: Doubling of Section A. **A)** The pβAA construct. The arrow represents the start site of transcription and the ORF is represented by a black rectangle. The poly(A) signal is represented by pA. The Terminator is replaced with two copies of section A represented by blue rectangles. **B)** RPA-RT analysis of HeLa cells transiently transfected with pβA and pβAA deletion constructs. pβT+ was used as a positive control and pβT- as a negative control. '+' and '-' are as in Figure 3.3B. RT represents the readthrough transcripts and P represents correctly initiated transcripts. This experiment has been repeated at least three times and the pattern of RT and P bands obtained is reproducible. **C)** Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, pβT+, which is set at 1.

3.3 Homopolymer tracts in the Terminator sequence are important for transcriptional termination

Models of transcription termination of prokaryotic RNA polymerases and eukaryotic Pol I and Pol III all highlight the importance of homopolymer tracts in the termination process (see Introduction). Therefore, I next examined the possible role of homopolymer tracts dispersed throughout the Terminator in directing mammalian Pol II termination.

To do this, I made 2 constructs – in one, p β TMut1, the A homopolymer tracts were disrupted by substitution of 11 single A residues with single C residues; and in the other, p β TMut2, A tracts were disrupted by substitution of 11 double A residues with double C residues (Figure 3.7B) using long range PCR. Nuclear RNA was isolated from HeLa cells transiently transfected with pTat and p β TMut1, p β TMut2 or the control constructs p β T⁺ and p β T⁻, and analysed using the RPA-RT assay. The protected products were analysed by PAGE (Figure 3.8A). Band signal strength was quantified by PhosphoImager and the ratio RT/P calculated for each construct, normalised to the positive control, p β T⁺, then displayed in the graph below the data panel (Figure 3.8B). In lane 1 (p β T⁺), there is a strong band for P and weak band for RT as expected, and in lane 2 (p β T⁻), there is a weak P and a strong RT. In lane 3 (p β TMut1) there is a significant decrease in P, and therefore a change in RT/P ratio, which indicates that A tracts have a role in transcription termination efficiency. The role of A tracts in Pol II termination is highlighted by examination of lane 4 (p β TMut2) where it can be seen that, when A tract disruption is increased by double mutations, the amount of RT is significantly increased. It appears from this data that

A



B

CAATAACAAACA**CC**AATTA**CC**ATAGGAAAATA**CC**AAATT**CCA**AGAAG**CC**ATCCT
 GTTATTGTTTGTT**GG**TAAATT**GG**TATCCTTTTATT**GG**TTTAAT**GG**TTCTTCT**GG**TAGGA

GCCATTTATGCGAGAATTGAT**GAA**CCTGGAGGATGT**CC**ACTAAGAA**CC**ATAAGCCTGA
 CGGTAAATACGCTCTTA**ACT**ACTTGACCTCCTACAT**GG**TGATTCTT**GG**TATTCGGACT

CACA**CC**AGAC**CC**TACTACACA**CC**TTGCTCATATGTG**ACC**CAT**AA**CC**AA**AGTCACTCT
 GTGTT**GG**TCTGT**GG**ATGATGTGTTGGAACGAGTATACACT**GG**GTATT**GG**TT**CA**GTGAGA

CAT**GG**AA**AC**CAGACAGTAGAGGTATGGTTTCCAGGGGTTGGGGGTGGGAGAATCAGGAAA
 GT**AC**CTTTGTCTGTCATCTCCATACCAAAGGTCCCAACCCCA**CC**CTCTTAGTCCTTT

CTATTACTCAAAGGGTATAAAATTT**CA**GTTATGTGGGATGAATAAATT
 GATAATGAGTTTCCCATATTTTAAAGTCAATACACCCTACTTATTTAA

Figure 3.7: Mutational analysis of the Terminator. **A)** Diagrams of oligo A tract mutants. The Terminator representation is as in Figure 3.2A. pβTMut1 has 11 A to C point mutations. pβTMut2 has 22 point mutations; the original single point mutations were converted to doublet point mutations by conversion of neighboring A residues to Cs. pβTMut2Control has 22 point mutations of residues which do not form part of the oligo A tracts. **B)** Location of AA to CC doublet mutations in pβTMut2.

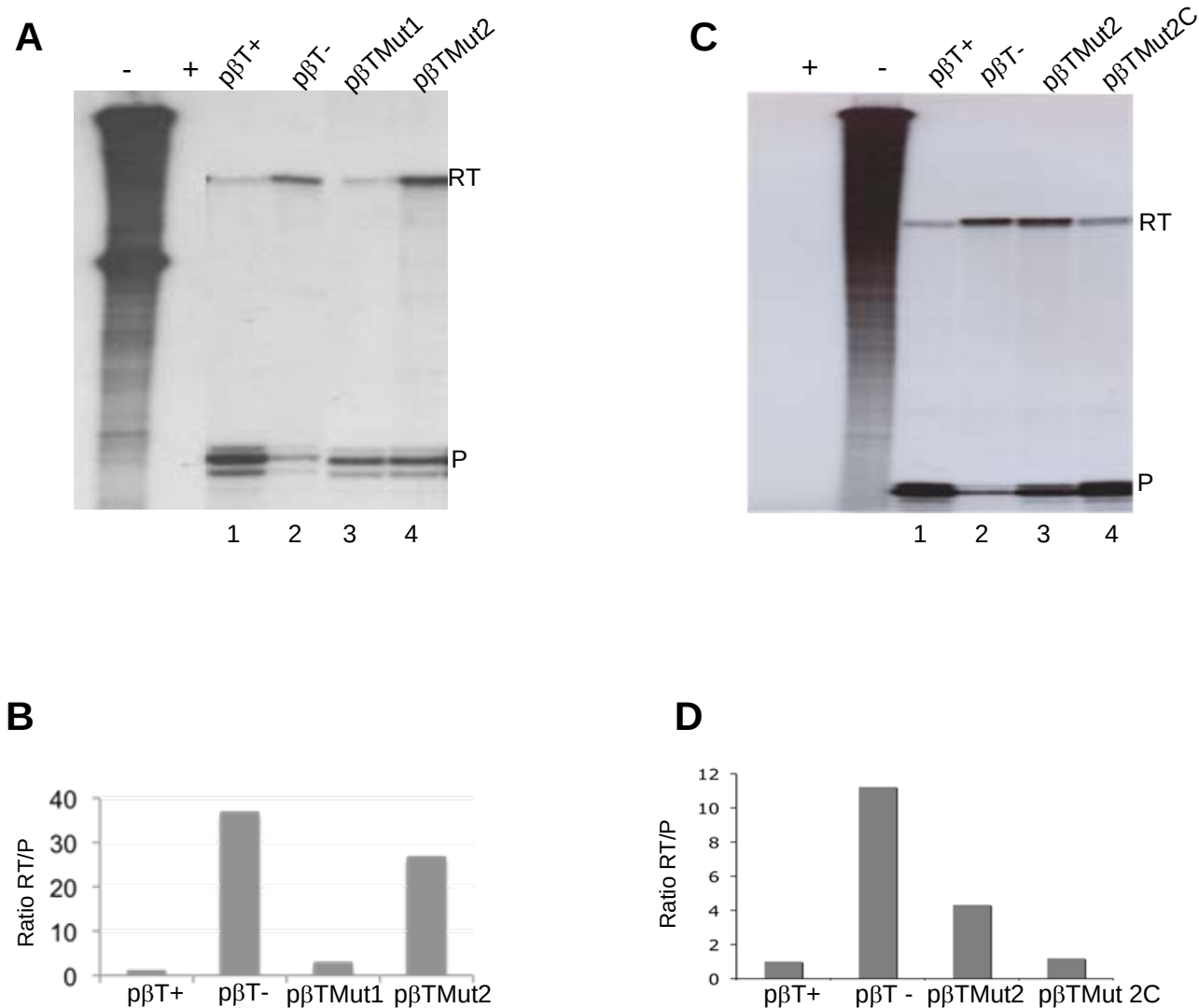


Figure 3.8: Mutational analysis of the Terminator. **A)** RPA-RT analysis of HeLa cells transiently transfected with pβTMut1 and pβTMut2 constructs. pβT+ was used as a positive control and pβT- as a negative control. '+' and '-' are as in Figure 3.3B. RT represents the readthrough transcripts and P represents correctly initiated transcripts. This experiment has been repeated at least three times and the pattern of RT and P bands obtained is reproducible. **B)** Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, pβT+, which is set at 1. **C)** RPA-RT analysis of HeLa cells transiently transfected with pβTMut2 and pβTMut2Control constructs. pβT+ was used as a positive control and pβT- as a negative control. '+' and '-' are as in Figure 3.3B. RT represents the readthrough transcripts and P represents correctly initiated transcripts. This experiment has been repeated at least three times and the pattern of RT and P bands obtained is reproducible. **D)** Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, pβT+, which is set at 1.

while single mutations have only a minor effect on Pol II termination, double mutations of the A tracts cause a significant increase in readthrough transcription. This shows that intact A tracts are required for efficient termination. To determine whether the effect observed with p β TMut2 was specific to the disruption of the A tracts and not due to the large number of mutations introduced into the Terminator sequence, I generated a control construct labelled p β TMut2Control, in which 22 point mutations were incorporated into the Terminator sequence in locations other than the A tracts. To analyse the effect of these mutations, nuclear RNA was isolated from HeLa cells transiently transfected with pTat and p β TMut2, p β TMut2Control or the controls, p β T⁺ and p β T⁻, and analysed by the RPA-RT assay. Following RNase digestion, protected products were analysed by PAGE (Figure 3.8C). Band signal strength was quantified by PhosphorImager and the ratio RT/P calculated for each construct, normalised to the positive control, p β T⁺, then displayed in the graph below the data (Figure 3.8D). In control lanes 1 and 2 (p β T⁺ and p β T⁻), the expected distribution of RT and P bands is seen. In lane 3 (p β TMut2), the strong RT band and weak P band confirms the effect of AA to CC mutations on Pol II termination. In lane 4 (p β TMut2Control), the much weaker RT band shows that introduction of 22 point mutations in other parts of the Terminator does not cause a termination defect.

The design of p β TMut2 results in disruption of A tracts throughout the Terminator. To test whether the A tracts at the 5' end or 3' end of the Terminator were more susceptible to the effect of these mutations, I made two constructs, labelled p β TMut3 and p β TMut4, using the long range PCR strategy. p β TMut3 has double point

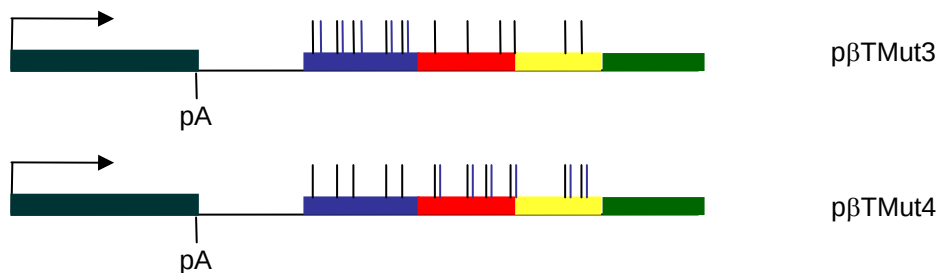
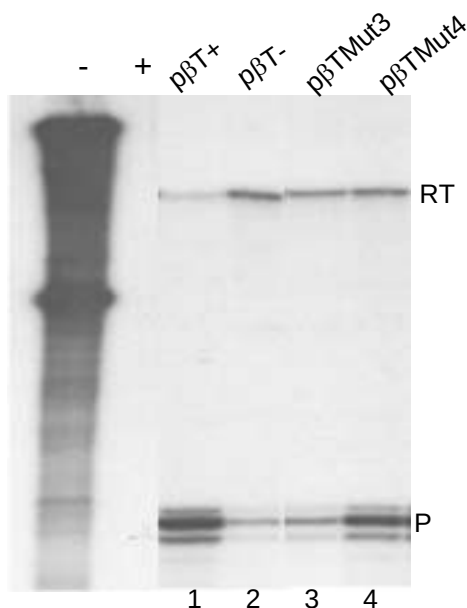
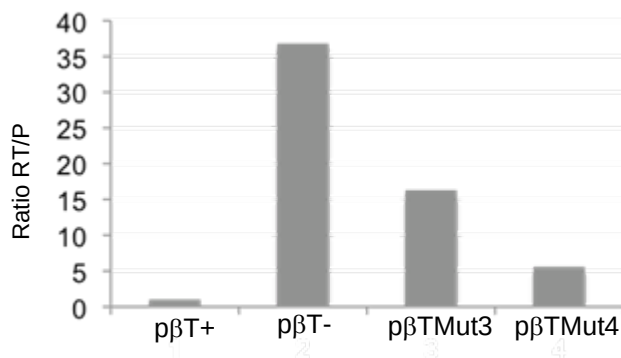
A**B****C**

Figure 3.9: Mutational analysis of the Terminator. **A)** The Terminator representation is as in Figure 3.2A. pβTMut3 has 16 point mutations and pβTMut4 has 17 point mutations; both changing A residues to Cs. **B)** RPA-RT analysis of HeLa cells transiently transfected with pβTMut3 and pβTMut4 constructs. pβT+ was used as a positive control and pβT- as a negative control. '+' and '-' are as in Figure 3.3B. RT represents the readthrough transcripts and P represents correctly initiated transcripts. This experiment has been repeated at least three times and the pattern of RT and P bands obtained is reproducible. **C)** Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, pβT+, which is set at 1.

mutations at the 5' end and single mutations at the 3' end, whereas p β TMut4 has single point mutations at the 5' end and double mutations at the 3' end (Figure 3.9A).

Nuclear RNA isolated from HeLa cells transiently transfected with pTat and p β TMut3, p β TMut4, or the control constructs p β T⁺ and p β T⁻, was analysed in the RPA-RT assay. Following RNase digestion, protected products were analysed by PAGE (Figure 3.9B). Band signal strength was quantified by PhosphorImager and the ratio RT/P calculated for each construct, normalised to the positive control, p β T⁺, then displayed in the graph below the data panel (Figure 3.9C). In control lanes 1 and 2 (p β T⁺ and p β T⁻), the expected distribution of RT and P bands is seen. In lane 3 (p β TMut3), there is a weak signal for P and a strong RT band indicating a significant termination defect. Finally, in lane 4 (p β TMut4) there is an RT band of similar intensity as in p β TMut3 (lane 3), however, the signal for the P band is much higher, indicating more efficient termination than for p β TMut3. Quantitation analysis of RT and P bands is shown in the graph below the data panel, and it can be seen that there is a 3 fold greater termination defect when the homopolymer A tracts are disrupted at the 5' end of the terminator rather than the 3' end. This result indicates that the A tracts at the 5' end of the Terminator sequence are particularly important in termination. This is especially interesting in light of the deletion data showing that section A is the most important section of the Terminator sequence. Furthermore, it confirms that the A tracts are critical to mediate transcription termination.

Having demonstrated that disruption of A homopolymer tracts had a significant effect on the accumulation of readthrough transcripts in the RPA-RT assay, I decided to confirm that it was due to a transcriptional defect by NRO analysis. NRO is a direct analysis of transcription because it maps the position of active polymerases. NRO analysis was performed on nuclei isolated from HeLa cells transiently transfected with pTat and either p β T⁺, as a positive termination control construct, or the

p β TMut2 construct with the A tract disruptions that show the largest termination defect based on the RPA-RT analysis. In this NRO experiment, active Pol II density was analysed over regions that included the promoter (P), the final exon of β -globin (B3) and immediately downstream of the poly(A) site (B4) as shown in the diagram in Figure 3.10A. Single stranded DNA probes homologous to sequence in the vector backbone, probe A, and the U3 region of the HIV-LTR, probe U3, detect transcriptional readthrough. Nascent transcripts were radiolabelled by incubation with α^{32} P-UTP, then partially hydrolysed and hybridised to single stranded DNA probes fixed to the hybridisation filters shown in Figure 3.10B. As expected, hybridisation signals in the p β T+ experiment are high over probes P, B3 and B4 and low over probes A and U3 (Figure 3.10B, top panel) indicating efficient termination. In contrast, hybridisation signals are seen throughout the plasmid in the p β TMut2 NRO experiment. The significant hybridisation signals over A and U3 show that mutations at the homopolymer tracts of the Terminator (p β TMut2) impede the termination process and cause Pol II transcriptional readthrough (Figure 3.10B lower panel). Strength of NRO signals was quantified by PhosphorImager, normalised to the number of U residues in each probe and displayed in the graph next to the autoradiograph (Figure 3.10B, right hand side panel).

3.4 The homopolymer tracts are important for CoTC activity of the human β -globin Terminator

It has previously been shown that the transcriptional termination capacity of fragments of the β -globin terminator corresponds to Co-transcriptional Cleavage (CoTC) of the Terminator transcript (Dye and Proudfoot, 2001). To test whether the

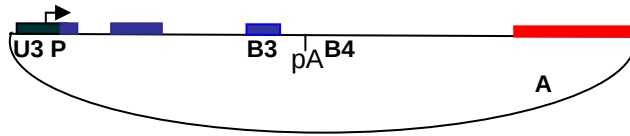
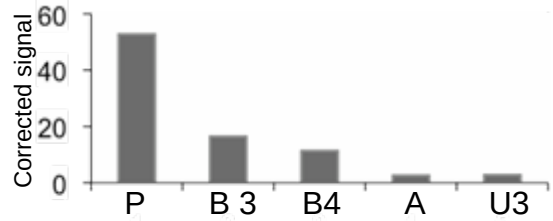
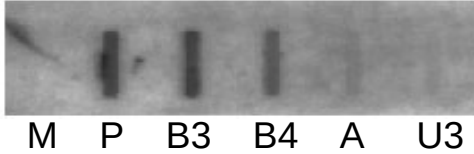
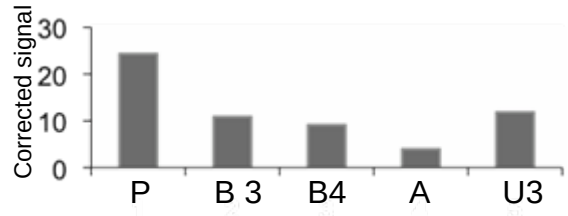
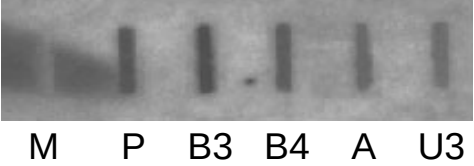
A**B**p β T+p β TMut 2

Figure 3.10: Nuclear Run On analysis of the effect of mutating the human β -globin Terminator on termination. A) Diagram of the β -globin plasmid. The start site of transcription (arrow) and regions of complementarity of antisense ssDNA M13 probes are indicated (bold). **B)** NRO analyses of nascent transcripts of p β T+ and p β TMut2 constructs. NRO probes are indicated below their respective positions on the filters. M represents empty M13 vector and is a background control. Graphs on the right of the NRO data panels show strength of hybridisation signals after normalisation for the number of U residues present in each transcript.

doublet mutations of the Terminator homopolymer tracts introduced in the p β TMut2 construct affects CoTC, a reverse transcription PCR (RT-PCR) assay was carried out on total RNA isolated from HeLa cells transiently transfected with pTat and either the positive control p β T+ or p β TMut2. cDNAs to detect nascent transcripts within the body of the gene and of the Terminator region were synthesised using reverse primers, Ex2r and CoTCr located in Exon 2 or the 3' end of the Terminator region respectively (Figure 3.11A). These cDNAs were then amplified using forward primers located either in intron 1, for the intron 1/exon 2 transcript, or at the 5' end of the Terminator sequence, for the Terminator transcript. In this assay, the level of RT-PCR product obtained from the Terminator region gives an indication of the extent of transcript cleavage over that region. To control for the amount of transcription in each construct, the amount of PCR product over the Terminator was normalised to nascent transcript in the body of the gene represented by the PCR product over the junction of Exon 2 and Intron 1. The results of this analysis are shown in the agarose gels in Figure 3.11B (upper panel). In lane 1 (p β T+), a weak band is seen for the Terminator transcript and a strong band over the intron 1/exon 2 border. This indicates a relatively low abundance of continuous Terminator transcripts. In lane 3 (p β TMut2), strong bands are detected for both the Terminator and intron 1/exon 2 transcripts. 28S rRNA levels were measured as a loading control by RT-PCR (Figure 3.11B; lowest panel). A minus Reverse Transcriptase control was used (lanes 2 and 4) to show that PCR products detected arise from cDNA synthesised during the reverse transcription reaction, and that there was no DNA contamination. The higher level of PCR product detected over the Terminator in p β TMut2 (lane 3) as compared to the smaller amount detected in p β T+ (lane 1) indicates that the Terminator region is not being cleaved in p β TMut2 as efficiently as in p β T+. The level of nascent transcript produced in p β T+

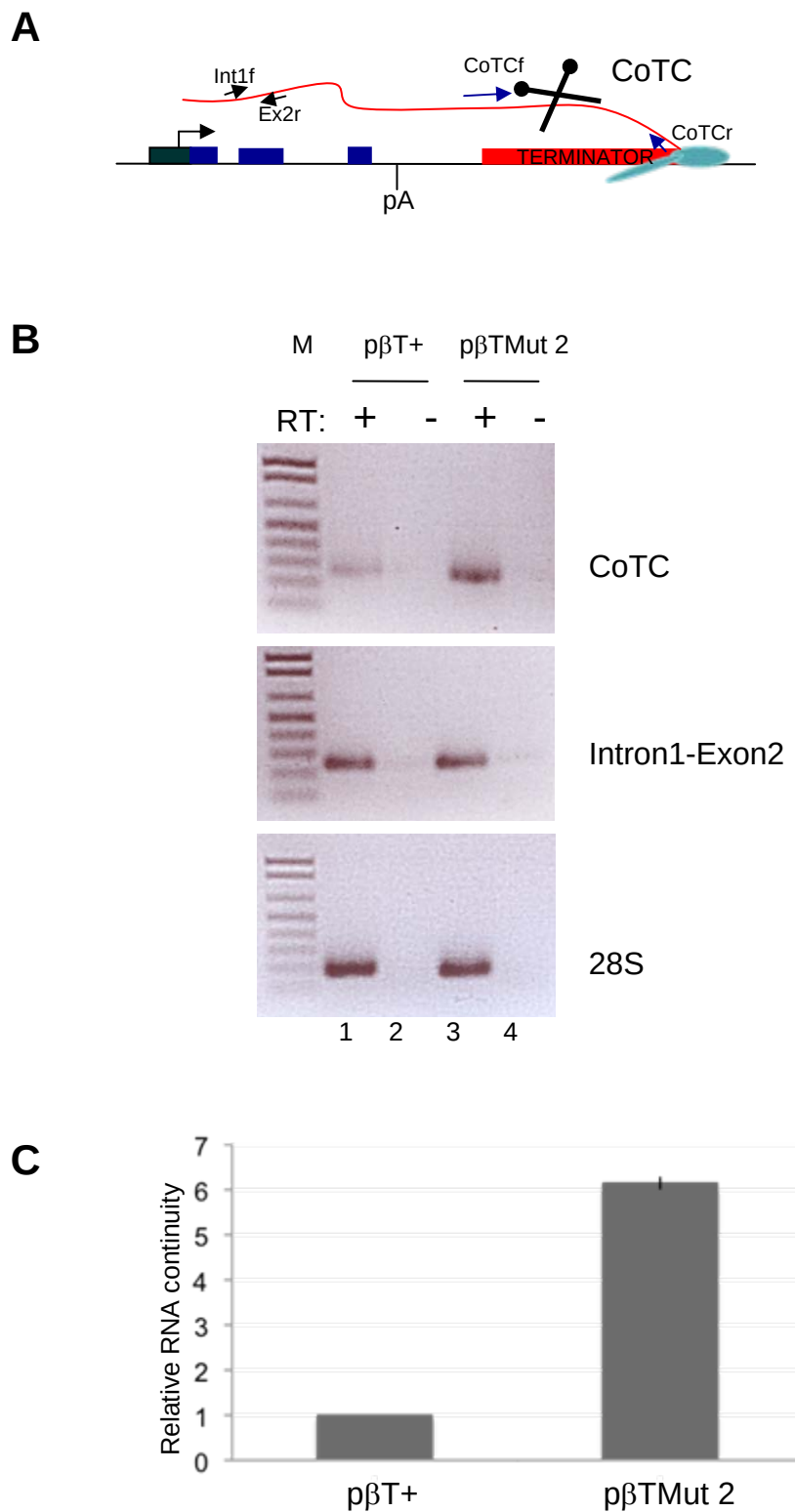


Figure 3.11: RT-PCR analysis of pβT+ and pβTMut2 transcripts. **A)** Diagram showing the position of the PCR amplicons, and primers used for RT-PCR. PCR product detected over the Terminator region (CoTC) is indicative of the level of transcripts that are not cleaved over that region. **B)** RT-PCR analysis of total RNA isolated from HeLa cells transiently transfected with pβT+ or pβTMut2. The levels of nascent message over the Terminator region and intron 1/exon 2 are shown. 28S rRNA is used as a loading control. **C)** Quantitative RT-PCR analysis of total RNA isolated from HeLa cells transiently transfected with pβT+ or pβTMut2. The ratio of Terminator to intron1/exon2 RNA from pβTMut2 is normalised to pβT+. Error bars indicating standard deviation over three biological replicates are shown.

and p β TMut2 are comparable as seen by the equivalent levels of intron 1/exon 2 product in the middle panel of Figure 3.11B. Hence, the ratio of Terminator transcript over nascent message is higher in p β TMut2, indicating that the Terminator region is being cleaved less efficiently when the A tracts are mutated. To gain quantitative data, the RT-PCR cleavage analysis was repeated using quantitative Real Time PCR and the same DNA primers to detect Terminator and intron 1/exon 2 control transcripts as above. Quantitative (q)RT-PCR analysis confirms the effect and shows that disruption of A tracts causes transcript cleavage to decrease by at least 6 fold (Figure 3.11C). However, this may be a conservative estimate because, as mentioned previously, these readthrough RNA transcripts are subject to rapid turnover in the nucleus. These experiments demonstrate that CoTC activity is significantly reduced by disruption of the homopolymer tracts and suggests that it is for this reason that termination efficiency is reduced in the p β TMut2 construct.

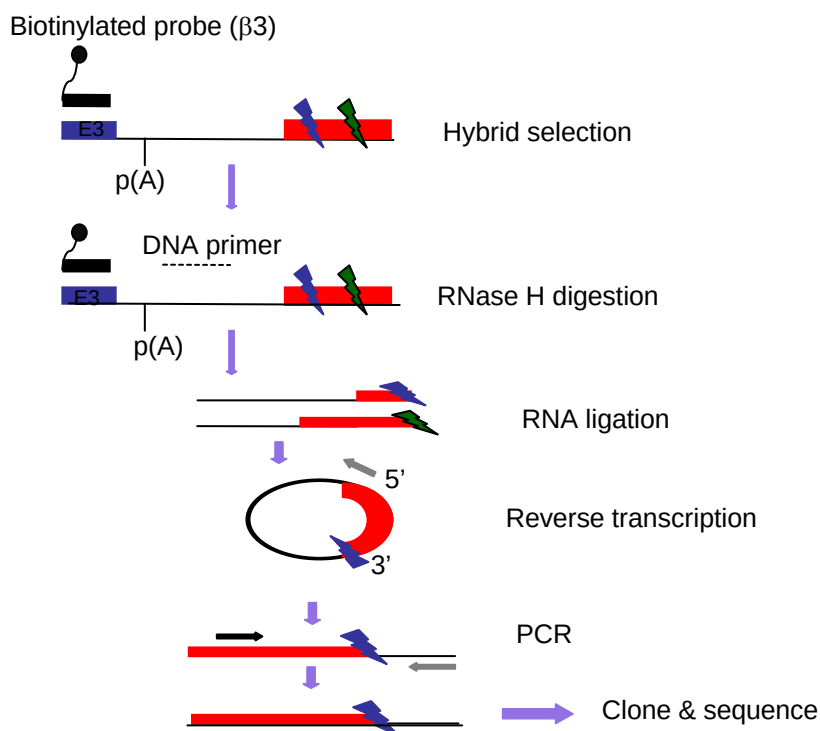
3.5 Multiple cleavages of the human β -globin Terminator transcript

To determine the location and distribution of CoTC cleavage sites in the wild type Terminator transcript, and to analyse the effect of homopolymer tract disruption on the distribution of these sites, I next performed hybrid selection circular (hsc)RACE. This technique, which is a combination of two different methods; cRACE and hybrid selection (Brognia, 1999; Dye and Proudfoot, 2001), was developed and employed to map CoTC sites in the 5' region of the β -globin terminator (West and Proudfoot, 2006). It involves the selection of specific RNAs using an antisense biotinylated probe. These selected linear RNAs are then released from the selection probe using oligonucleotide directed RNase H cleavage then circularised by RNA ligation. The

circular RNAs are then used to synthesise cDNA which, following PCR amplification, can be sequenced. The point where the 5' and 3' ends of the originally linear RNA meet can be used to map CoTC sites because the 3' end is in some cases created by a CoTC event. The details of this analysis are outlined in the diagram of Figure 3.12A.

To conduct this experiment, HeLa cells were transiently transfected with pTat and p β T+ or p β TMut2. Nuclear RNA was isolated from the transfected cells and the β -globin Terminator transcripts were selected with a biotinylated probe (β 3) complementary to exon 3 of the β -globin gene. The Terminator transcripts were released from the selection probe by RNase H directed cleavage using an oligonucleotide complementary to the region immediately downstream of the poly(A) signal. The released RNA was then circularised and used for cDNA synthesis followed by PCR amplification. PCR products were cloned into the TA vector and approximately 50 colonies were selected from each sample and sequenced. In the case of the mutant Terminator (p β TMut2) the experiment was repeated three times more in order to obtain the correct number of colonies. The 3' ends of these selected RNAs were identified by sequencing and are shown in Figure 3.12B. The various species identified are likely to correspond to a combination of primary sites of transcript cleavage, degradation intermediates or possibly positions of Pol II pausing. The ends identified in the wild type Terminator are scattered throughout the Terminator region with a hotspot of activity found at the 5' end of the Terminator as indicated on the diagram in Figure 3.12B. The RNA ends from the mutant Terminator (p β TMut2) have a different distribution. Rather than being spread out across the region they are concentrated at the 5' end of the Terminator. It is evident from comparing the ends

A



B

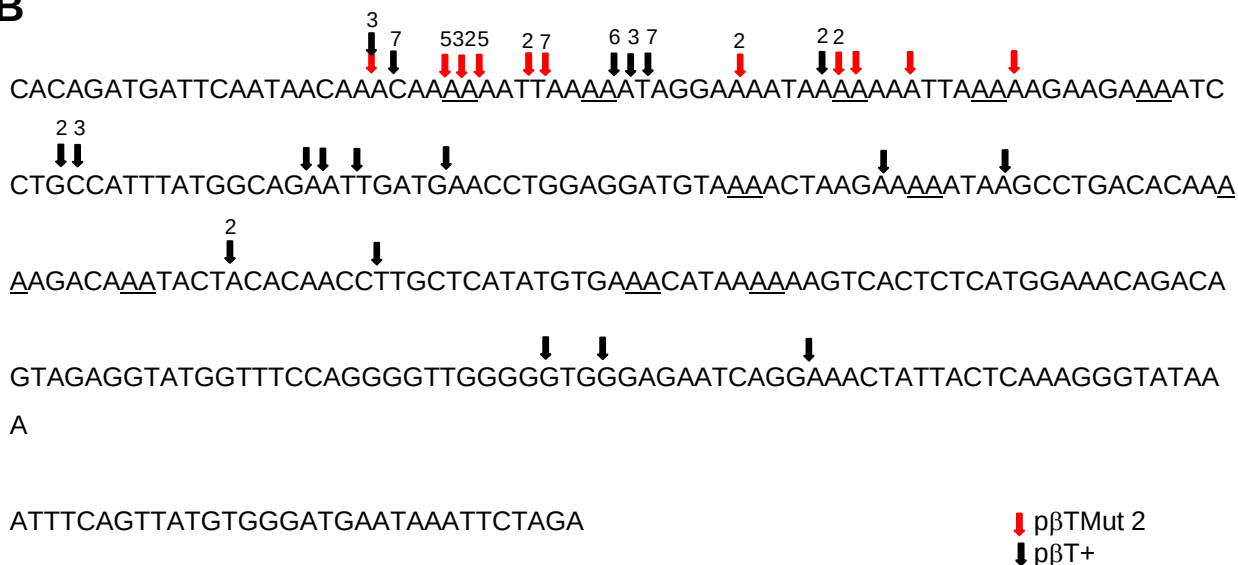


Figure 3.12: Hybrid selection circular RACE analysis of CoTC generated RNA 3' ends. **A)** The hscRACE procedure. The Terminator element (red box), exon 3 (blue box) and cleavage sites (lightning bolts) are shown. A biotinylated RNA probe, β3 (tailed box), was hybridised to nuclear RNA isolated from HeLa cells transiently transfected with pβT+ or pβTMut2 and then isolated through magnetic selection. The selected RNA is subjected to RNaseH digestion after a specific DNA primer (dotted line) is hybridised to it leading to release of the RNA from the beads. Released RNA is ligated and reversed transcribed with a primer directed to amplify across the ligated junction. The cDNA is then amplified by PCR using gene specific primers. The PCR products are analysed on a gel, cloned and sequenced. **B)** Terminator DNA sequence showing the positions of CoTC derived RNA 3' ends in pβT+ (black arrows) and pβTMut2 (red arrows). The numbers above each arrow indicate the number of clones identified for each site. Positions of AA to CC mutations in pβTMut2 are underlined.

produced by p β T⁺ and p β TMut2 that mutation of the A tracts prevents cleavage in the 3' portion of the Terminator transcript (Figure 3.12B). Indeed, the 5' end of the Terminator is the most A rich region of the Terminator, even after the mutations, reinforcing the idea that the A tracts are important in termination.

3.6 Analysis of stable β -globin integrants

It has previously been shown that the entire β -globin terminator operates in the chromosomal context of transgenic mice (Dye and Proudfoot, 2001). However, my mutational analysis of the Terminator undertaken so far has been performed in a plasmid-based system. Next, I decided to test if the Terminator operated when in a chromosomal context. This would allow evaluation of the effect of disturbing the Terminator homopolymer tracts on termination when the Terminator is in a more physiological chromatin environment. To do this, I utilised the HEK293 cells and the Flp-In T-Rex recombinase system to create stable cell lines (see Materials and Methods). Two stable cell lines were made which contain the β -globin gene driven by the HIV-LTR, 240bp of post poly(A) site sequence and either the wild type Terminator or the mutant Terminator from the p β TMut2 construct (henceforth referred to as wild type or mutant stable cell lines respectively). The plasmid expressing the transactivator Tat (pTat) was transfected to allow expression from the HIV-LTR.

I used RT-PCR to confirm that the β -globin gene is expressed in the stable cell lines. Total RNA was isolated then reverse transcribed with oligodT to select for mature mRNA. Amplification of these cDNAs with primer pair E2f/E2r, gives a product

representing the steady state level of message over exon 2. The data indicates that the β -globin gene is expressed in both the wild type and mutant stable cell lines (Figure 3.13A). However, the levels detected in the mutant stable cell line are lower than in the wild type. The levels of 28S rRNA were detected as a loading control. To confirm full gene expression, I next decided to investigate the size of the RNA being produced by Northern Blot analysis. RNA isolated from the wild type and mutant stable cell lines before and after transcriptional activation with Tat, and also the parental HEK293 cell line, was separated on a 1% agarose formaldehyde gel, transferred to a nitrocellulose membrane and hybridised to a uniformly radiolabelled antisense RNA probe complementary to exon 2 of the β -globin gene. The hybridised radiolabelled RNA probe was then visualised by PhosphorImager and shown in Figure 3.13B. In addition, a uniformly radiolabelled antisense RNA probe specific for the 28S rRNA gene was used to detect 28S rRNA as a loading control (Figure 3.13B, lower panel). The bands correspond to the full-length mature β -globin mRNA and confirm that the β -globin gene is expressed in both the wild type and mutant stable cell lines (lanes 1 & 3). Presumably the heterogeneity of this band reflects variation in poly(A) tail length. This analysis confirms the RT-PCR assay above and shows that the mutant cell line produces significantly lower levels of β -globin mRNA than the wild type. Transcription initiates at the HIV-LTR even in the absence of Tat as β -globin mRNA is detected in the minus Tat controls (lanes 2 & 4). It is apparent that the parental HEK293 cell line allows promiscuous HIV-LTR transcription initiation in the absence of Tat. However, higher levels of β -globin mRNA are detected when Tat is present. The lack of a signal in the parental HEK293 cell line confirms that the endogenous β -globin gene is not expressed in these cells.

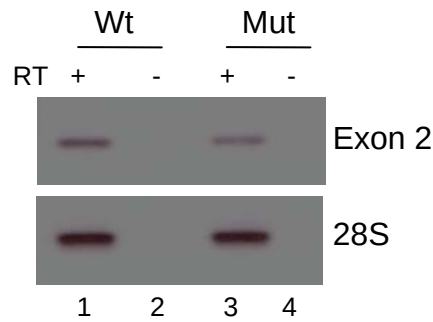
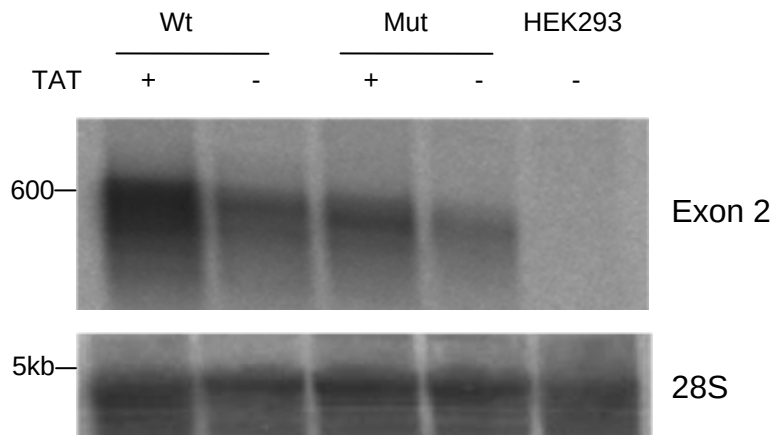
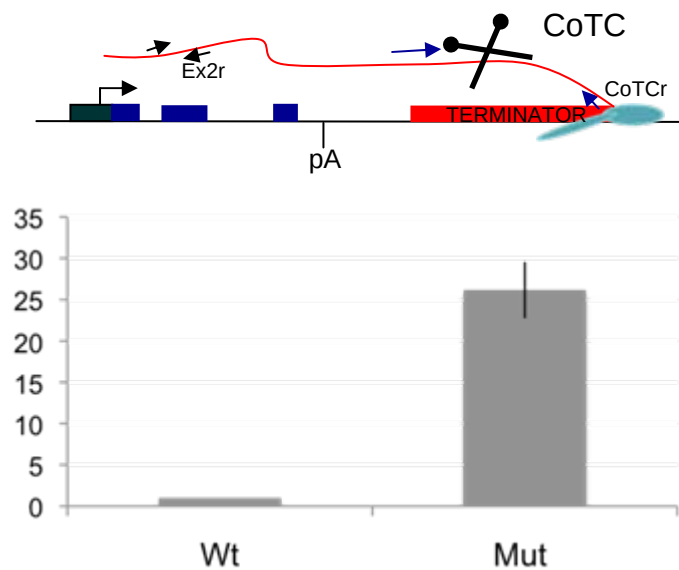
A**B****C**

Figure 3.13: CoTC of transcripts from stably integrated β -globin genes. **A)** RT-PCR analysis of total RNA isolated from HEK293 cells containing stably integrated constructs with the Terminator elements from p β T+ and p β TMut2 (Wt and Mut respectively). cDNA was synthesised using an oligodT primer then PCR amplified using exon2 primers. RT-PCR analysis of 28S rRNA is a loading control. **B)** Northern Blot analysis of nuclear RNA isolated from HEK293 cells containing stably integrated constructs with the Terminator elements from p β T+ and p β TMut2 and from the parental HEK293 line before the integration. An antisense RNA probe complimentary to exon 2 was used. 28S rRNA was detected as a loading control and displayed in the lower panel. **C)** Diagram showing the position of primers used in RT-PCR analysis. The level of PCR product detected over the Terminator region is indicative of the level of transcripts that are not cleaved over that region. This is then compared to the level of transcription as indicated by the nascent message represented by the intron 1/exon 2 amplicon. Quantitative RT-PCR analysis of total RNA isolated from HEK293 cells containing stably integrated constructs with the Terminator elements from p β T+ and p β TMut2. The ratio of Terminator to intron1/exon2 RNA in Mut is normalised to Wt. Error bars indicating standard deviation over three biological replicates are shown.

The main aim of producing these stable cell lines was to analyse the effects of AA mutations in a more physiological chromatin environment. Steady state RNA analysis shows a decrease in β -globin gene expression, which is concordant with West and Proudfoot (2009). I next used qRT-PCR to analyse CoTC cleavage of the Terminator transcripts in the wild type and mutant stable cell lines (section 3.4). Total RNA isolated from the wild type and mutant stable cell lines after transfection of pTat, was reverse transcribed with primers CoTCr and Ex2r. These cDNAs were then amplified using primer pairs CoTCf/CoTCr and Int1f/Ex2r respectively. The level of PCR product detected over the Terminator is indicative of the extent of transcript cleavage over that region. To control for the amount of transcription, the level of PCR product detected over the Terminator is normalised to the nascent message represented by the PCR product detected over the intron 1/exon 2 junctions. The results in Figure 3.13C confirm the results from the plasmid-based system showing that the mutant Terminator is defective in CoTC activity. In this environment it appears that transcript cleavage is reduced by at least 27 fold, which is significantly higher than what was observed in the plasmid-based system. This shows that when the Terminator is placed in a more natural environment, the importance of the homopolymer tracts is more pronounced.

3.7 A stretch of 30 A residues is not sufficient to cause termination

The results described above suggest the possibility that an A rich sequence may be sufficient to promote Pol II termination. To test this possibility, I returned to the transient transfection system and made a construct labelled p β 30A, in which the Terminator is substituted by a stretch of 30 A residues (Figure 3.14). Nuclear RNA

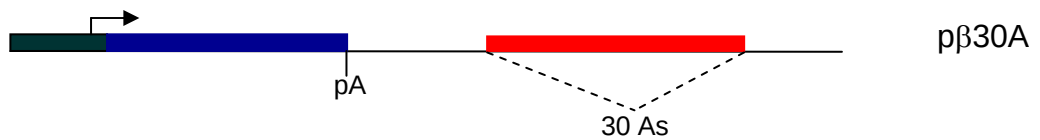
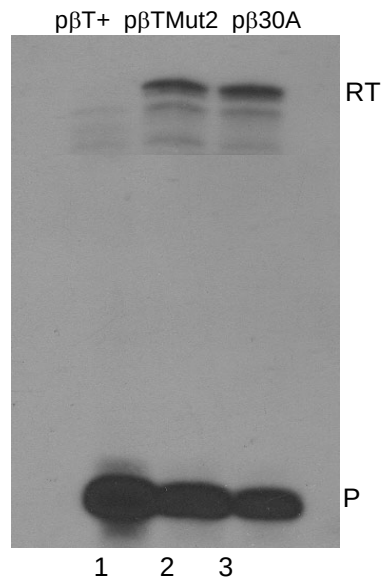
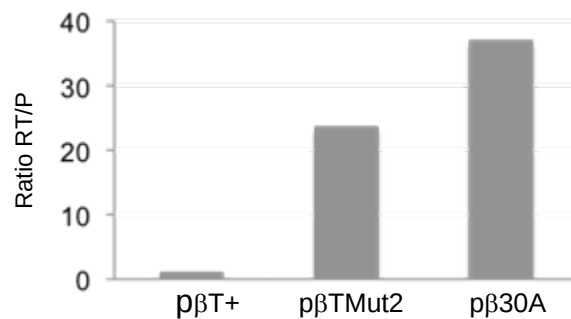
A**B****C**

Figure 3.14: Testing the effect of a long A tract (30A) in Pol II termination. **A)** The pβ30A construct. The arrow represents the start site of transcription, and the ORF is represented by a blue rectangle. The poly(A) (pA) and 30A sequence cassette (red rectangle) are indicated. **B)** RPA-RT analysis of β-globin RNA from HeLa cells transiently transfected with pβ30A. RT represents the readthrough transcripts and P represents correctly initiated transcripts. **C)** Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, pβT+, which is set at 1.

isolated from Hela cells transiently transfected with pTat and constructs p β 30A or p β T⁺ and p β TMut2 as positive and negative controls respectively, was analysed by the RPA-RT assay. The protected products were analysed by PAGE (Figure 3.14B). Band signal strength was quantified by PhosphoImager and the RT/P ratio calculated for each construct, normalised to the positive control, p β T, then displayed in the graph below the data panel (Figure 3.14C). The results in Figure 3.14 show that the ratio of RT/P obtained for p β 30A (lane 3) does not show any increase in termination efficiency as compared to p β TMut2 (lane 2) suggesting that a stretch of 30 A residues is not sufficient to cause termination. Indeed, termination efficiency is reduced below the low levels seen in p β TMut2. This is a further indication of the complexity of the Terminator signal. This also suggests that dispersed A tracts are required.

3.8 Discussion

RPA-RT and PhosphoImager quantitation of Terminator deletion constructs shows that the entire region (285bp) is required for efficient termination (Figure 3.3, 3.4 and 3.5). However, the deletion analysis reveals that not all sections of the Terminator are equally capable of mediating termination. Deletion of region A leads to the highest termination defect followed by D, B and C (Figure 3.4). As a result of this finding, a construct with two copies of section A, namely p β AA, was made (Figure 3.6). RPA-RT analysis quantitation shows that although p β AA is better able to direct termination, it is still almost 4 times less efficient than the whole Terminator (Figure 3.6). This data indicate that the entire Terminator is required and potentially, the order of the sections may be important.

The importance of homopolymer tracts in the β -globin Terminator was investigated by substitution of A residues with Cs, either singly or in doublets. The results of the analysis of the single mutations showed an average two-fold decrease in termination efficiency (Figure 3.8A and 3.8B). Disturbing the homopolymer tracts even further by doublets causes a more significant decrease in termination efficiency (Figure 3.8). Variations of these constructs were created; p β TMut3 and p β TMut4, both of which confirm that double nucleotide mutations of the tracts affect termination (Figure 3.9) and highlight the importance of the 5' end of the Terminator sequence.

Quantitative RT-PCR analysis of nascent Terminator transcripts from HeLa cells transiently transfected with p β T+ or p β TMut2 confirms that CoTC is severely disrupted by mutation of the homopolymer tracts (Figure 3.11). This result gives strong evidence for a role of the A homopolymer tracts in the CoTC process and confirms the role of CoTC in Pol II termination at the human β -globin Terminator. Furthermore, the importance of the homopolymer tracts in termination was confirmed in the stable cell lines (Figure 3.13C). When the Terminator is found in a more natural chromatin environment, the importance of the homopolymer tracts is more pronounced and this is indicated by the dramatic decrease in CoTC activity in the mutant stable cell line. Another interesting observation from these experiments is that when termination is defective (in mutant stable cell line), the level of mRNA is much reduced in comparison to wild type (Figure 3.13A & B). This observation was also made in the plasmid-based system with various constructs defective in termination (data not shown) and confirms previous results showing a link between termination efficiency and mRNA expression (West and Proudfoot, 2009; Mapendano et al. 2010).

Hybrid selection circular RACE reveals that RNA 3' ends are dispersed throughout the Terminator in the presence of intact homopolymer tracts. The distribution of these RNA 3' ends is dramatically altered by the introduction of AA to CC mutations in p β TMut2. The only remaining 3' ends are at the 5' end of the Terminator which is the only remaining A rich region, even after AA to CC mutations are introduced (Figure 3.12B). Interestingly, in p β T⁺ and p β TMut2, most RNA 3' ends map to this region of the Terminator. Following on from these observations, I tested the possibility that a long homopolymer tract of 30A residues might be sufficient to direct CoTC and subsequent Pol II termination. Interestingly, I found that the long (30bp) A tract does not direct transcriptional termination (Figure 3.14). This result suggests that a long stretch of A residues is not sufficient to direct efficient termination and reinforces the hypothesis that short A tracts dispersed within apparently random sequence are strong terminator elements.

Satisfied that I had conducted a thorough analysis of sequence dependence of Pol II termination on the β -globin gene, and had revealed a very interesting role of A tracts in this process, I began to consider the role of *trans* acting factors in the CoTC process. The various approaches used to address this are discussed in the following Chapters.

CHAPTER 4

DESIGN OF A GENETIC SCREEN TO IDENTIFY CANDIDATES RESPONSIBLE FOR CoTC ACTIVITY

4.1 Introduction

Pol II transcription of the human β - and ϵ - globin genes and the mouse serum albumin (MSA) gene terminates hundreds of base pairs past the poly(A) signal and requires cotranscriptional cleavage (CoTC) of 3' flanking RNA sequences (Dye and Proudfoot, 2001; West et al. 2006). Hence, Pol II termination on these genes requires both a functional poly(A) signal and a downstream CoTC terminator sequence.

Analysis of the β -globin gene has shown that the CoTC terminator can have positive effects on gene expression leading to an increase in mRNA and protein levels (West and Proudfoot, 2009). Therefore, as it appears that the CoTC terminator is an enhancer of gene expression, it is likely that other genes employ the CoTC terminator mechanism. One of the aims of my thesis has been to determine the sequence dependence of CoTC and to find factors involved in the CoTC mechanism that could be used to identify candidate CoTC terminators in other genes. In the previous Chapter I presented a detailed dissection of one section (region 10; henceforth referred to as the Terminator) of the human β -globin terminator, which is also used throughout this part of the study. I have shown that mutated versions of the Terminator do not elicit CoTC and are defective in termination, providing further evidence that the transcriptional termination capacity of the β -globin terminator corresponds to the ability to cleave the transcript (Dye and Proudfoot, 2001).

Identification of factor(s) responsible for the CoTC event could give insights into the molecular mechanism of termination and could help to identify other genes, which employ CoTC mediated termination. Thus, in this part of the study, I designed a

system, which could exploit yeast genetics and be used in a genetic screen to help identify potential factor(s) involved in CoTC. The yeast *Schizosaccharomyces pombe* (*S. pombe*) was chosen for this purpose as it is more similar to humans than *Saccharomyces cerevisiae* (*S. cerevisiae*) due to a higher sequence similarity of homologous genes (Wood et al. 2002). I thought this similarity to higher eukaryotes might enhance the possibility that the CoTC Terminator would function in this system.

4.2 A reporter gene designed to identify factors involved in CoTC mediated termination

To test the activity of the CoTC element in *S. pombe* I employed a plasmid based system in which an intron has been inserted into the normally intronless gene *ura4* to generate a construct labelled pUI (Figure 4.1; Aranda and Proudfoot, 1999; Aranda and Proudfoot, 2001). The system is designed such that terminator elements can be placed within the synthetic intron. If, a terminator element within the intron is recognised by Pol II, then a truncated transcript is made (Figure 4.1). If the terminator element is not recognised, then transcription continues through the *ura4* gene, up to the wild type *ura4* termination signals located downstream of the *ura4* poly(A) signal. As shown in Figure 4.1, splicing out of the intron results in the production of a mature *ura4* mRNA, which in turn leads to production of a functional Orotidine 5'-phosphate decarboxylase protein. Terminator elements can be selected by transforming the plasmids into a strain lacking an endogenous *ura4* gene, and thus only cells producing a functional Orotidine 5'-phosphate decarboxylase protein are able to survive in minimal media lacking uracil.

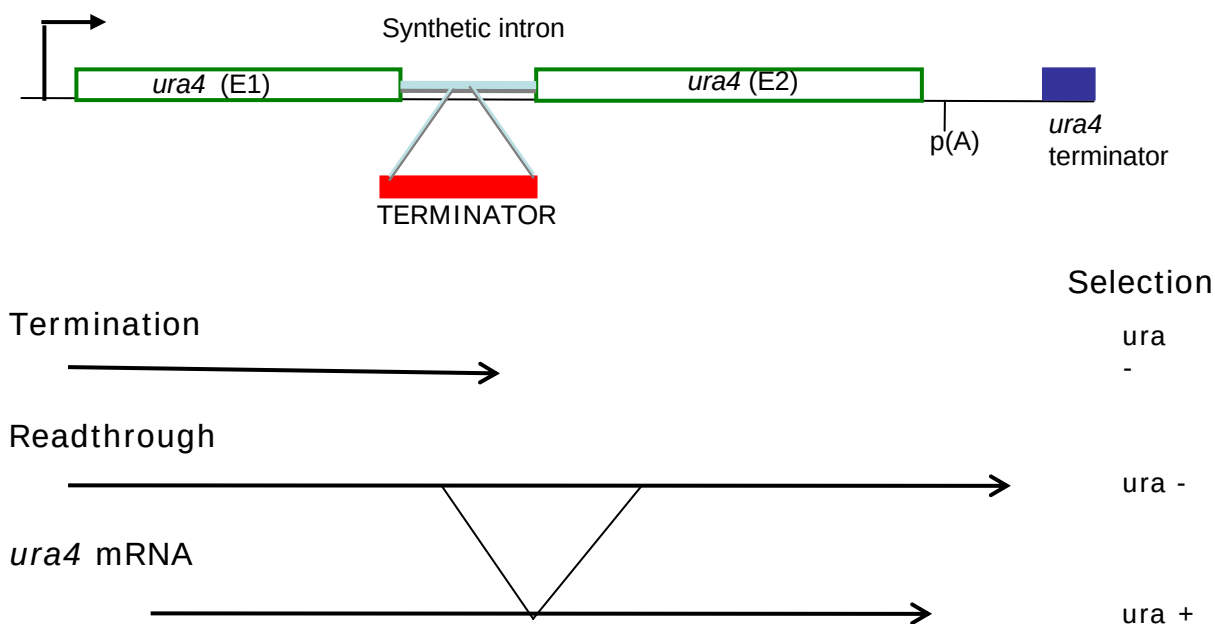


Figure 4.1: *S. pombe* Terminator reporter system. Diagram of the modified *ura4* gene. The synthetic intron is inserted within the ORF of *ura4*. Exon 1 (E1), exon 2 (E2), and the intron (blue line) are shown. The polyadenylation site (pA) and the *ura4* terminator (lilac box) are also indicated. Control and candidate terminator sequences (red box), were inserted within the intron. The constructs were used to transform a *ura4* deletion strain. Termination within the intron produces a truncated transcript. Failure to terminate leads to readthrough transcription. Wild type *ura4* termination signals are then used to produce an unspliced pre-mRNA. Splicing the intron produces *ura4* mRNA allowing cells to survive in minimal media lacking uracil.

A number of derivative constructs, in which various sequences were placed in the synthetic intron, were made from the parental pUI construct that contains the modified *ura4* gene and the *leu1-32* gene coding for leucine (shown in Figure 4.2). As a positive control to confirm that a termination signal can be recognised within the pUI intron, I made pUI*ura4*Term, which contains the *ura4* minor poly(A) signal and termination element, collectively referred to as the *ura4* terminator. Addition of terminator elements within the synthetic intron of pUI increases the size of the intron significantly. Therefore, to control for any effects on splicing efficiency when the introns are larger, I made a construct labelled pUIspacer, which contains a 500bp post poly(A) sequence from the human β -globin gene that has been shown to be unable to direct transcriptional termination (Dye and Proudfoot, 2001). Two more constructs were made where wild type or mutant Terminator sequences – the latter corresponding to the mutated Terminator sequence found in p β TMut2 – were inserted into the synthetic intron and labelled pUITerm and pUImutTerm respectively. Pol II termination at the CoTC terminator is dependent upon upstream poly(A) signals (Dye and Proudfoot, 2001). Therefore, I made two additional constructs in which a DNA fragment containing the human β -globin poly(A) signal and downstream CoTC signal (wild type or mutant) were inserted into the pUI intron (Figure 4.2). These constructs were labelled pUIpATerm and pUIpAmutTerm respectively.

The plasmids were transformed in a strain lacking the *ura4* and *leu1-32* genes and grown in media without leucine. Lack of leucine exerts pressure on the cells to maintain the plasmids. The growth of the strains was analysed by spotting serial dilutions of liquid cultures onto plates. The growth data for the different strains grown on minimal media without leucine (EMM-Leu), or both leucine and uracil (EMM-

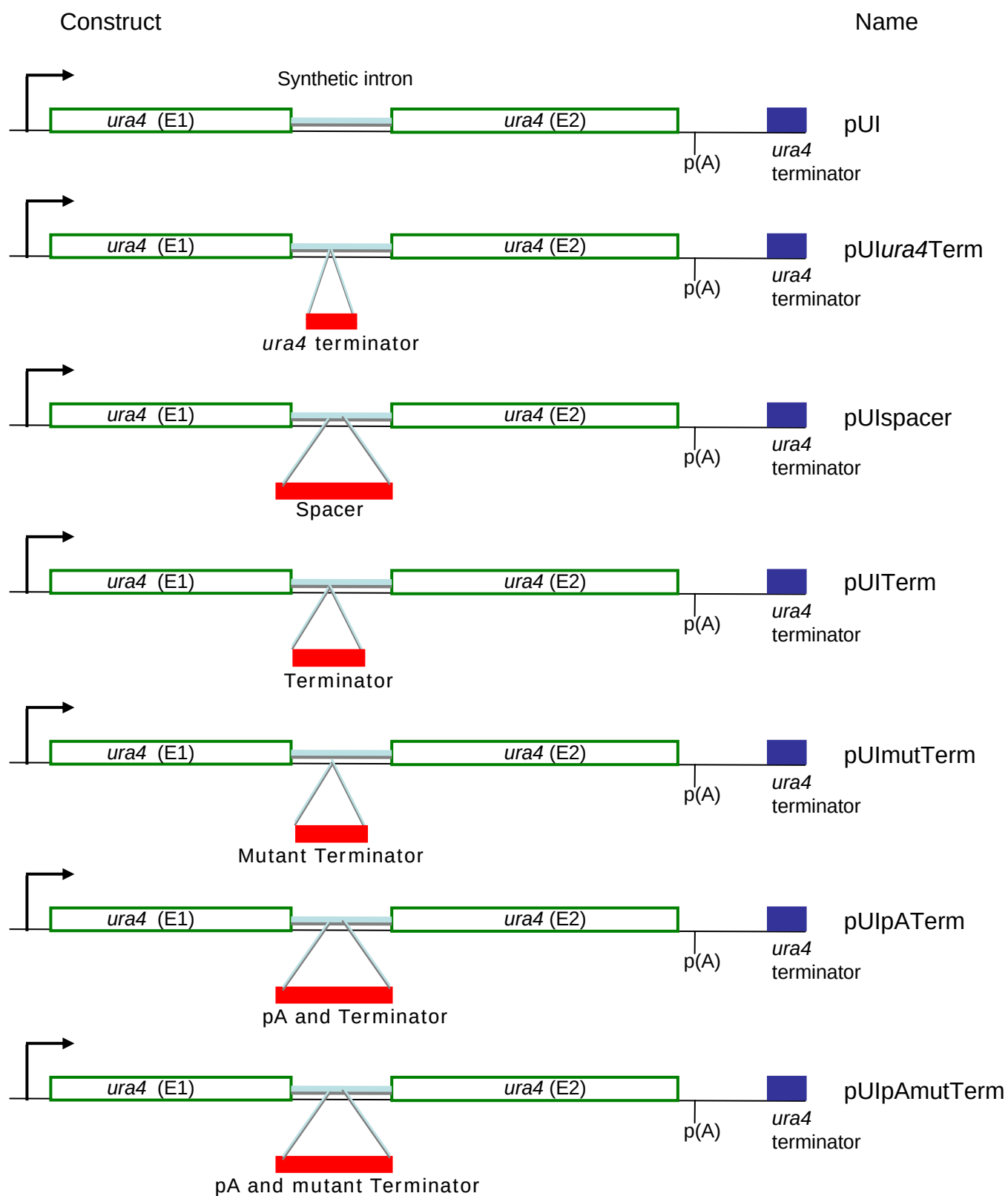


Figure 4.2 Reporter constructs of the modified *ura4* gene. Control and candidate terminator sequences (red boxes) were inserted within the intron.

Leu/-Ura) are shown in Figures 4.3A and 4.3B, respectively. All strains are able to grow on minimal media lacking leucine, showing that they have maintained the plasmid (Figure 4.3A). However, when cells are grown on media lacking both leucine and uracil, it appears that only pUI and pUIspacer (Figure 4.3B, lanes 1 and 3 respectively) are able to grow. Thus from this analysis it is apparent that only these two strains are able to produce functional Ura4 protein. The observation that the positive termination control pUI*ura4*Term does not produce a functional message and is unable to grow (Figure 4.3B, lane 2) indicates that this assay may be used to determine functional Pol II terminator elements. Cells transformed with pUIpATerm (Figure 4.3B, lane 4), pUIpAmutTerm (Figure 4.3B, lane 5), pUITerm (Figure 4.3B, lane 6) and pUImutTerm (Figure 4.3B, lane 7) are all unable to survive in the absence of uracil, indicating that these elements have curtailed *ura4* expression, possibly due to causing termination within the synthetic intron.

4.3 The Terminator does not terminate efficiently in *S. pombe*

The growth data shows that a functional Ura4 protein is not made when the wild type and mutant Terminator sequences, either with or without a poly(A) signal, are placed within the synthetic intron. Two possible reasons that the cells are unable to produce a functional Ura4 protein are; firstly, termination within the intron causes truncation of the *ura4* transcript or secondly, introduction of sequences within the intron affects pre-mRNA splicing leading to production of an aberrant *ura4* mRNA. To test whether the β -globin Terminator element directed transcription termination in *S. pombe*, I measured the abundance of transcripts before and after the Terminator sequences by qRT-PCR analysis of RNA isolated from cells transformed with either pUI,

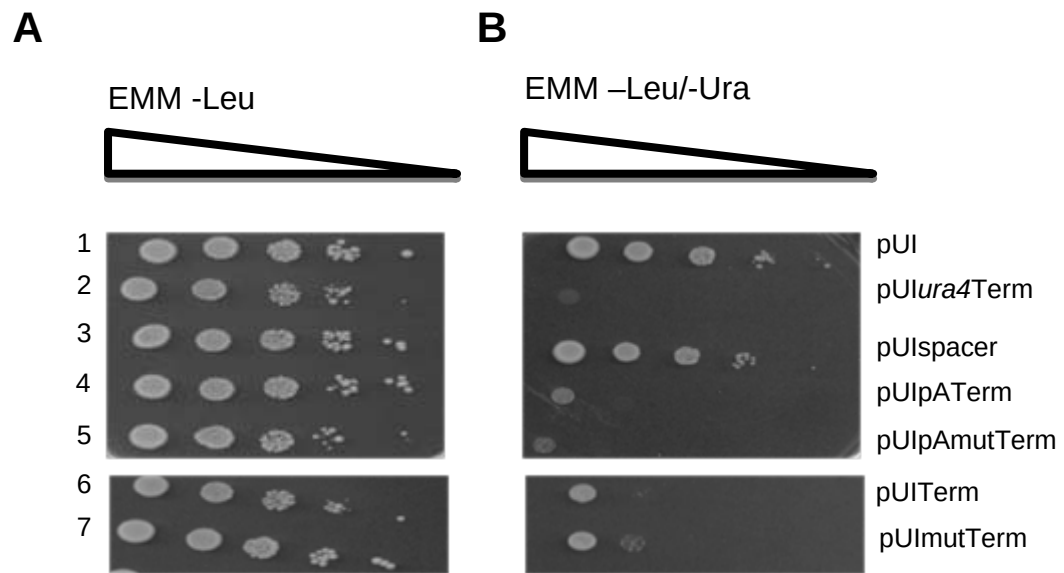


Figure 4.3: Growth on minimal media lacking uracil. A) pUI, pU*lura4*Term, pUIspacer, pUIpATerm, pUIpAmutTerm, pUITerm and pUImutTerm cells were plated in 10-fold serial dilutions on minimal medium lacking leucine. **B)** pUI, pU*lura4*Term, pUIspacer, pUIpATerm, pUIpAmutTerm, pUITerm and pUImutTerm cells were plated in 10-fold serial dilutions on minimal medium lacking leucine and uracil.

pUI*ura4*Term, pUIspacer, pUITerm, pUImutTerm, pUIpATerm or pUIpAmutTerm and grown in media lacking leucine. cDNAs to detect *ura4* exon 1 and exon 2 transcripts were synthesised using reverse primers, U1r and U2r, located in exon 1 or exon 2 of the modified *ura4* gene respectively (Figure 4.4A). The resulting cDNAs were then amplified using primer pair U1f/U1r, for the exon 1 transcript or, U2f/U2r, for the exon 2 transcript. In this assay, the level of PCR product detected over exon 2 (probe 2) was normalised to the amount of PCR product detected over exon 1 (probe 1) and the ratio of probe 2/probe 1 gives an indication of the extent of transcription proceeding through the synthetic intron. The result of this experiment is shown in Figure 4.4B. As anticipated, the high ratio of probe 2/probe 1 seen in construct pUI indicates that Pol II transcribes through the intronic sequence. When the *ura4* terminator is placed in the intron then the ratio of probe 2/probe 1 drops significantly. The level of exon 2 transcript detected is very low indicating very efficient termination. All the constructs containing the wild type or mutant Terminator, with or without the poly(A) signal exhibit a decrease in the ratio of probe 2/probe 1 in comparison to pUI and pUIspacer, suggesting that some termination is taking place. However, the level of termination is not significant when compared to the effect observed when the *ura4* terminator element is present in the intron.

In the qRT-PCR assay described above, the presence of exon 2 transcripts detected in strains transformed with pUITerm, pUImutTerm, pUIpATerm and pUIpAmutTerm, is indicative of transcription proceeding through the various elements located within the intron, and into exon 2. However, when the same constructs are transformed in yeast and grown on minimal media lacking uracil, they are unable to grow, indicating that a functional *ura4* mRNA is not made. Together these results suggest that perhaps there

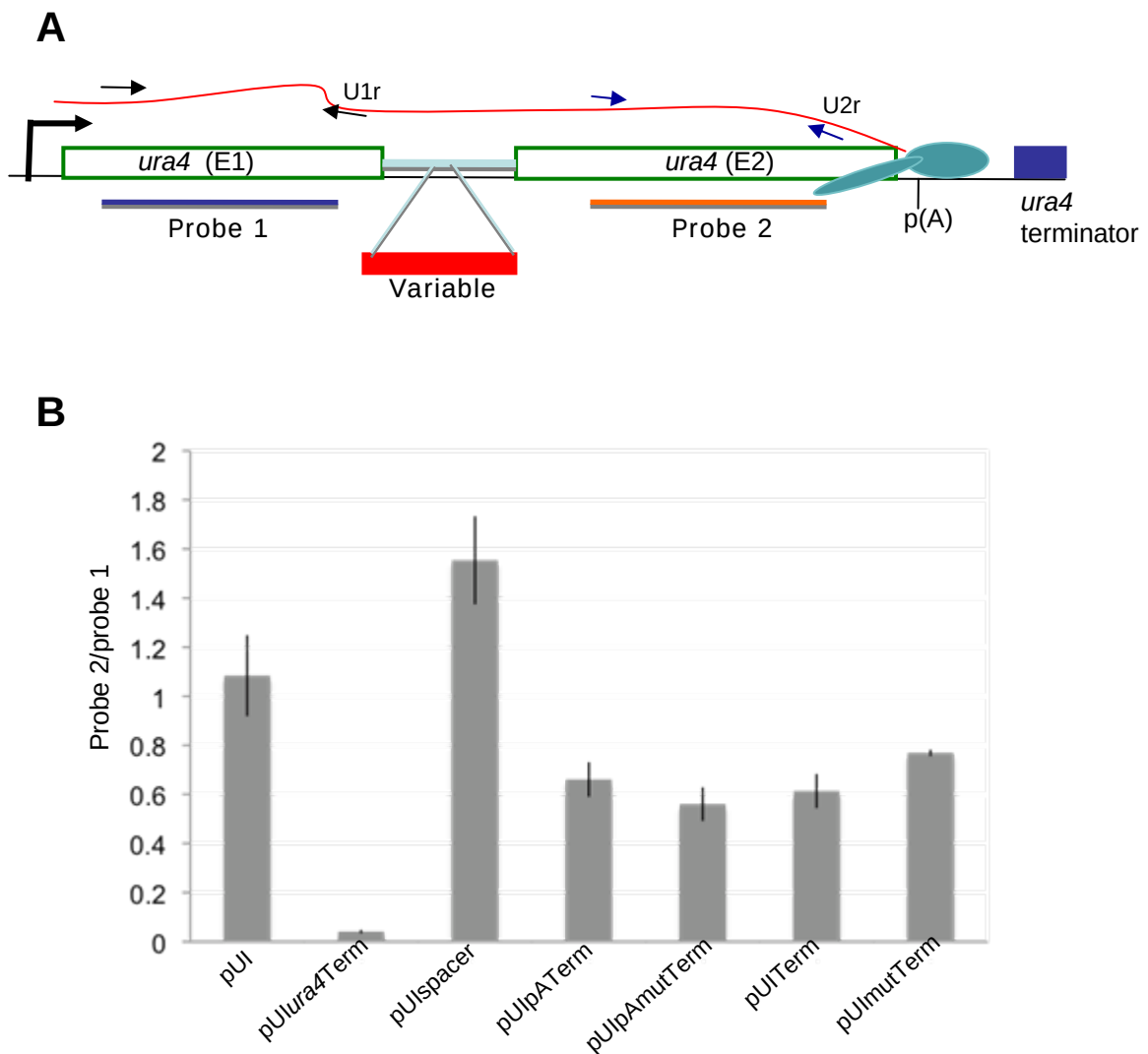


Figure 4.4: Quantitative RT-PCR analysis of transcripts from the reporter constructs. A) Diagram showing the position of PCR primers and amplicons. U1r and U2r were used for Reverse Transcription. The level of PCR product detected over exon 2 is indicative of the level of transcripts reading through the intron. This is then compared to the level of transcription as indicated by the message represented by the exon 1 amplicon. **B)** Graph showing quantitation of RNA from pUI, pUra4Term, pUspacer, pUlpATerm, pUlpAmuTerm, pUITerm and pUImutTerm cells. RNA was analysed by qRT-PCR and the level of message over exon 2 normalised to the level of message over exon 1. Error bars indicating standard deviation over three biological replicates are shown.

is a splicing defect and thus, to determine if the chimeric *ura4* genes were spliced correctly, I next decided to analyse the size of the *ura4* mRNA produced from the various constructs using a Northern Blot assay. RNA isolated from cells transformed with pUI, pUI*ura4*Term, pUIspacer, pUITerm, pUImutTerm, pUIpATerm or pUIpAmutTerm and grown in media lacking leucine, was fractionated on a 1% agarose formaldehyde gel, transferred to a nitrocellulose membrane and hybridised to a uniformly radiolabelled antisense RNA probe complementary to exon 1 (probe 1) of the *ura4* gene. After washing, to remove any excess probe, the hybridised radiolabelled RNA probe was visualised by PhosphoImager (Figure 4.5B). In lane 1, probe 1 detects a single strong band from cells transformed with pUI, which corresponds to the correctly spliced *ura4* mRNA. In lane 2 (pUIspacer), probe 1 detects a longer *ura4* species, indicating a defect in splicing, and a faint band, which corresponds to the band representing correctly spliced *ura4* mRNA in pUI (lane 1). The lower faint band probably represents a small proportion of pUIspacer transcripts that are correctly spliced, which probably explains the growth of cells transformed with the pUIspacer plasmid on minimal media lacking uracil (Figure 4.3B). In lane 3 (pUITerm), probe 1 detects a strong band that is longer than the *ura4* mRNA. This is probably a readthrough band indicating that when the human β -globin Terminator is included in the intron readthrough occurs, and the pre-mRNA does not get spliced. However, the presence of a much shorter and weaker band, which corresponds to *ura4* exon 1 mRNA, is indicative of termination within the intron (see below; lane 7). This shows that the presence of the β -globin Terminator has some effect on transcription in *S. pombe*. In lane 4 (pUImutTerm), the long readthrough band seen with the wild type Terminator is observed, but not the shorter one, indicating that the mutant Terminator does not invoke termination within the intron. In lane 5

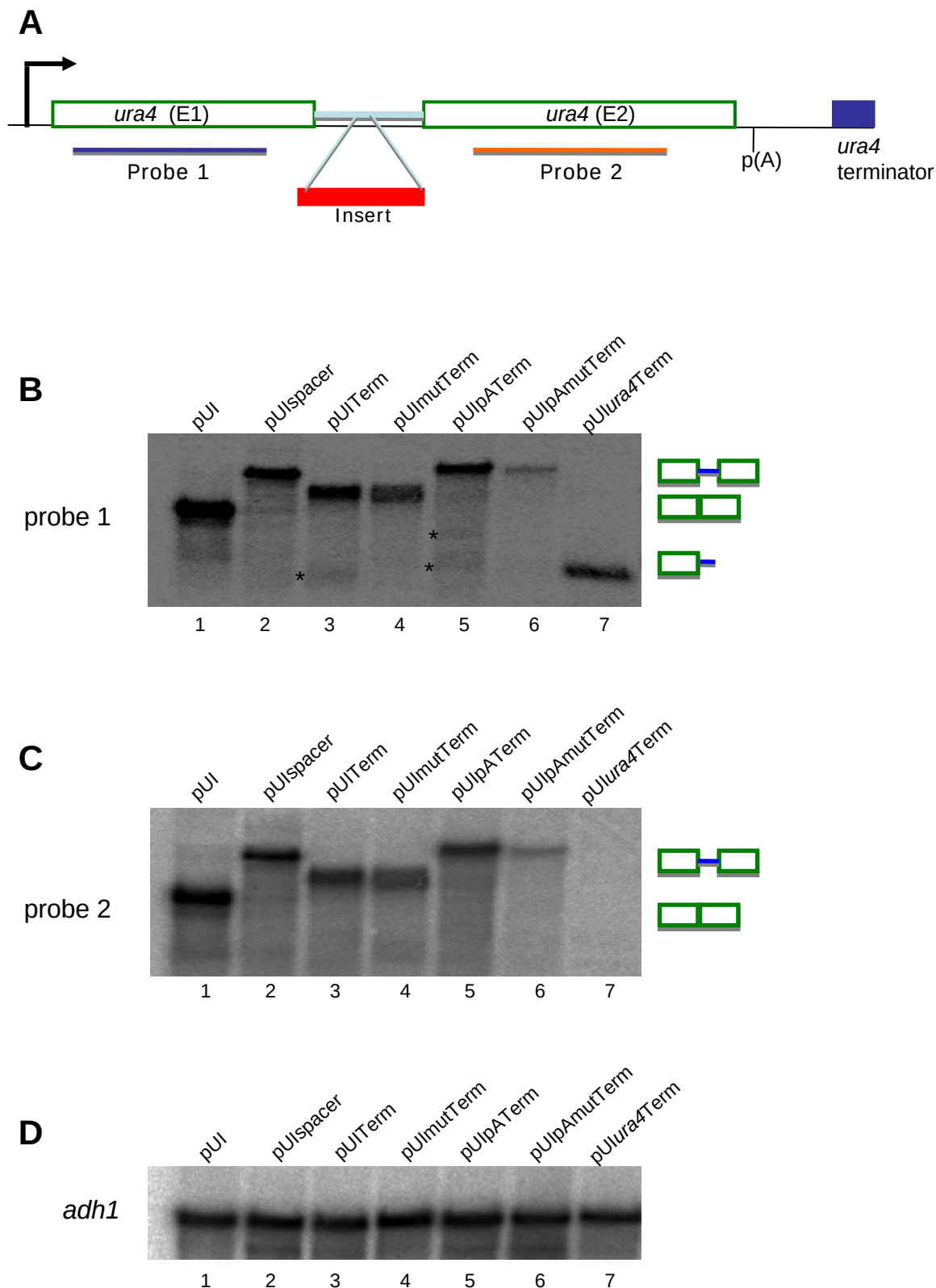


Figure 4.5: Northern Blot analysis of reporter construct transcripts. **A)** Diagram showing the location of radiolabelled antisense RNA probes. **B)** Northern Blot analysis of RNA isolated from pUI, pUIspacer, pUITerm, pUImutTerm, pUIpATerm, pUIpAmutTerm and pUIura4Term cells using probe 1. **C)** Northern Blot analysis of RNA isolated from pUI, pUIspacer, pUITerm, pUImutTerm, pUIpATerm, pUIpAmutTerm and pUIura4Term cells using probe 2. **D)** An antisense RNA probe to *adh1* was used to detect *adh1* mRNA as a loading control.

(pUIpATerm), a strong readthrough band and two smaller sized bands are detected. The strong band indicates that readthrough occurs and the pre-mRNA does not get spliced. However, the smallest band could indicate cleavage at the poly(A) signal and the larger band cleavage at the Terminator element, indicative of a CoTC event taking place. In lane 6 (pUIpAmutTerm), a signal for the unspliced band is visible but no smaller bands are detected suggesting that termination does not occur. Finally, in lane 7, probe 1 detects an abundant species corresponding to exon 1 indicating that efficient termination occurs when the *ura4* terminator is placed within the synthetic intron.

To confirm the results of the Northern Blot with *ura4* probe 1, the membrane was stripped and hybridised to a uniformly radiolabelled antisense RNA probe complementary to exon 2 (probe 2) of the *ura4* gene. After washing, to remove any excess probe, the hybridised radiolabelled RNA probe was visualised by PhosphoImager (Figure 4.5C). In lanes 1-6, the same species are identified as with probe 1 confirming the identity of these RNAs. In lane 7 (pUI*ura4*Term), no signal is observed with probe 2 confirming the results observed with probe 1 (Figure 4.5B, lane 7) that showed termination occurs within the intron. In order to ensure equal loading, the membrane was stripped once more and hybridised to a uniformly radiolabelled antisense RNA probe specific for the *adh1* gene, to detect *adh1* mRNA. The results of this are shown in Figure 4.5D and indicate approximately equal loading of all the samples.

The Northern Blot shows that pUI*ura4*Term terminates very efficiently whilst the other constructs do not. pUI pre-mRNA is spliced to produce *ura4* mRNA and to a

lesser extent, the same occurs with pUIspacer. However, pUImutTerm and pUIpAmutTerm, do not exhibit termination and the pre-mRNA produced is not spliced to produce *ura4* mRNA. Cells transformed with pUITerm and pUIpATerm exhibit a small degree of termination but ultimately, most transcription appears to continue through these sequences producing readthrough transcripts which are not spliced to produce a mature *ura4* mRNA.

4.4 The human β -globin termination signals invoke antisense transcription in *S. pombe*

In addition to analysing sense *ura4* RNA transcripts in the Northern blot discussed above, I also analysed the antisense *ura4* RNA transcripts. RNA isolated from cells transformed with pUI, pUI*ura4*Term, pUIspacer, pUITerm, pUImutTerm, pUIpATerm or pUIpAmutTerm and grown in media lacking leucine, was fractionated on a 1% agarose formaldehyde gel, transferred to a nitrocellulose membrane and hybridised to a uniformly radiolabelled sense RNA probe complementary to exon 1 (probe 1) of *ura4*. After washing, to remove any excess probe, the hybridised radiolabelled RNA probe was visualised by PhosphorImager and shown in Figure 4.6A (top panel). In lane 1, probe 1 detects a weak band from cells transformed with pUI, which is smaller than the full-length *ura4* RNA. It is possible this band corresponds to an antisense transcript, which initiates within the intron. In lane 2 (pUIspacer), the same band plus additional higher size bands are detected, indicating that transcription also initiates from other places within this longer intron. Weak bands are observed in lanes 3 (pUITerm) and 4 (pUImutTerm) suggesting that the presence of the Terminator, whether wild type or mutant, has a repressive effect

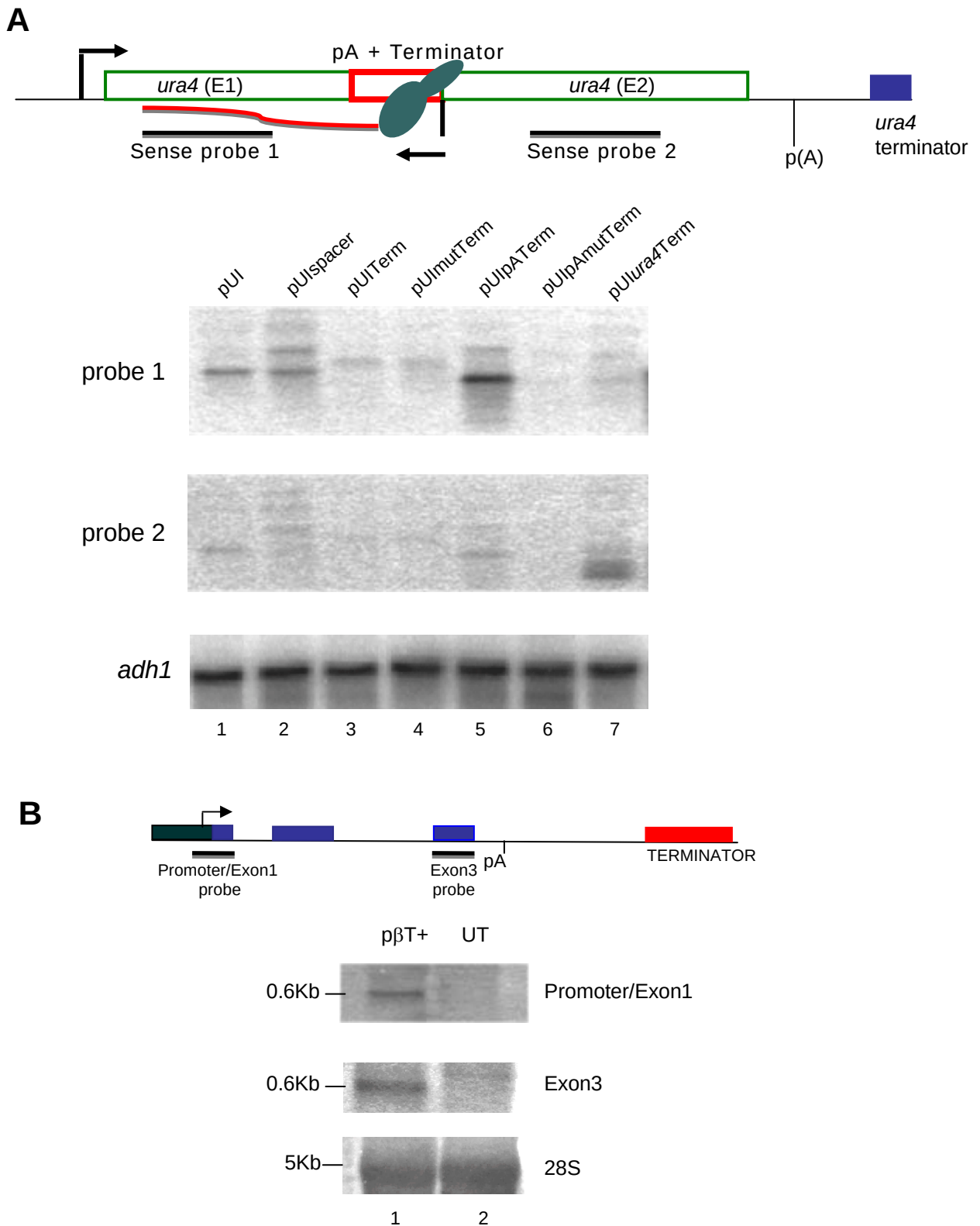


Figure 4.6: Northern Blot analysis of antisense transcripts. A) Top panel: Diagram showing the location of sense strand radiolabelled RNA probes. Middle panel: Northern Blot analysis of RNA isolated from pUI, pUispacer, pUITerm, pUImutTerm, pUIpATerm, pUIpAmutTerm and pUIura4Term cells using sense probes1 and 2. Lower panel shows loading control *adh1* mRNA detected using an antisense *adh1* probe. **B)** Top panel: Diagram of the β -globin gene construct (p β T+). Location of sense RNA probes is indicated. Lower panel: Northern blot analysis of RNA isolated from p β T+ transfected (p β T+) and control untransfected (UT) HeLa cells. 28S rRNA is used as a loading control.

on antisense transcription initiation within the intron. A strong band is visible in lane 5 (pUIpATerm), showing that when the human β -globin poly(A) signal plus the Terminator is included in the intron, this induces a significant level of antisense transcription. The absence of a signal in lane 6 (pUIpAmutTerm), shows that a wild type Terminator sequence is required to initiate antisense transcription. When compared to lane 3 (pUITerm), it becomes clear that the Terminator alone is not sufficient, and both termination signals are required. In lane 7, faint bands indicate low level antisense transcription initiation occurs when the *ura4* minor poly(A) signal and termination element are present within the intron.

In order to confirm that the antisense transcripts detected with probe 1 were initiated within the intron, the membrane was stripped and then hybridised to a uniformly radiolabelled sense RNA probe complementary to exon 2 (probe 2) of the *ura4* gene. After washing, to remove any excess probe, the hybridised radiolabelled RNA probe was visualised by PhosphoImager (Figure 4.6A; middle panel). There are no specific signals detected in lanes 1 to 6 but, very weak bands can be seen, which arise from the previously hybridised probe 1. The lack of specific signals suggests that any antisense transcripts detected were indeed initiated upstream of exon 2. However, in lane 7 (pUI*ura4*Term), a short strong band is detected suggesting that antisense transcription in this construct initiates upstream of exon 2 and terminates before exon 1. In order to ensure equal loading, the membrane was stripped and hybridised to a uniformly radiolabelled antisense RNA probe specific for the *adh1* gene, to detect *adh1* mRNA. The results of this experiment are shown in Figure 4.6A (bottom panel) and indicate approximately equal loading in all lanes. Overall, the results show that the human β -globin gene poly(A) signal and Terminator sequence can act as an initiation site for

antisense transcription in *S. pombe*. It appears that neither element on its own is capable of initiating antisense transcription.

Northern Blot analysis shows that antisense transcription appears to be initiated by the human β -globin poly(A) signal in combination with the Terminator element in *S. pombe*. Thus, I decided to investigate the existence of antisense transcription in transiently transfected HeLa cells by Northern Blot analysis. RNA isolated from untransfected HeLa cells and HeLa cells transiently transfected with p β T+ and pTat, was fractionated on a 1% agarose formaldehyde gel, transferred to a membrane and hybridised to two different uniformly radiolabelled sense RNA probes complementary to promoter/exon 1 and exon 3 of the β -globin gene. In addition, a uniformly radiolabelled antisense RNA probe specific for the 28S rRNA gene was used to detect 28S rRNA as a loading control. Hybridised radiolabelled RNA probes were visualised by PhosphoImager (Figure 4.6B). Antisense transcripts (approximately 600nt) were detected with both promoter/exon 1 and exon 3 probes in cells transfected with p β T+ and pTat, but not in untransfected cells. This result indicates that antisense β -globin gene transcripts are present in transfected HeLa cells.

4.5 RT-PCR of the mature β -globin message is RT-dependent and primer-independent

I next used RT-PCR analysis to further characterise the β -globin antisense transcripts. cDNA was synthesised from nuclear RNA isolated from HeLa cells transiently transfected with p β T+ and pTat, using reverse primers SS1, SS2 and SS3, which are located in the promoter region of the β -globin gene (Figure 4.7A). These cDNAs were

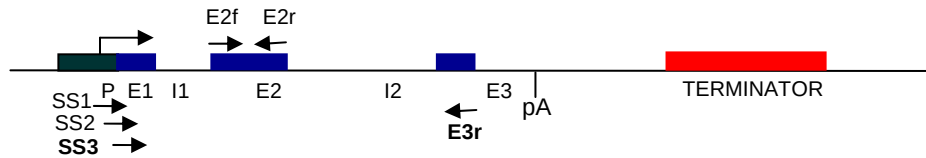
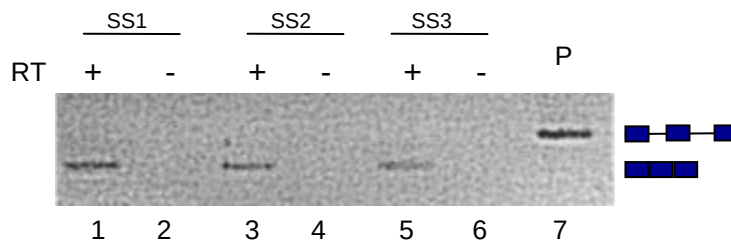
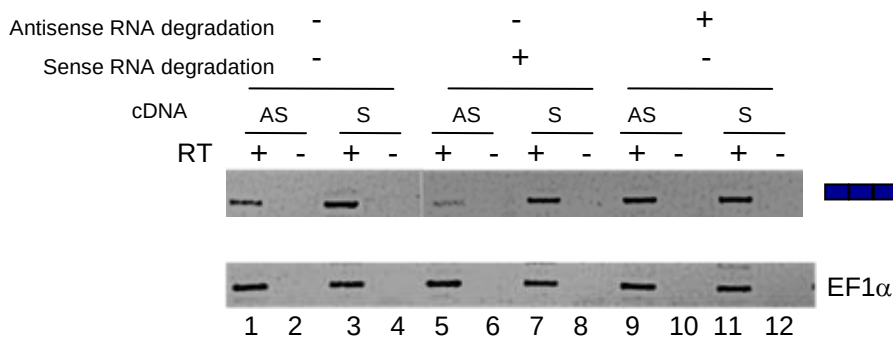
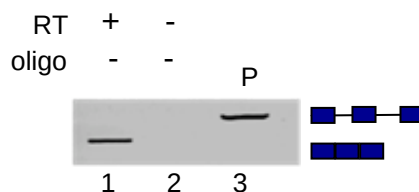
A**B****C****D**

Figure 4.7: RT-PCR analysis of antisense transcripts in pβT+ transfected HeLa cells. **A)** Diagram of the β-globin gene. The arrow represents the start site of transcription and P indicates the promoter. Exons are labelled E1 to E3 and introns are labelled I1 and I2. Primers (arrows) used in RT-PCR analysis are indicated. **B)** RT-PCR analysis of nuclear RNA isolated from HeLa cells transiently transfected with pβT+ and pTat, using three different primers for the RT reaction. A positive PCR control (P) is shown in lane 7. **C)** RT-PCR analysis of nuclear RNA isolated from HeLa cells transiently transfected with pβT+ and pTat, subjected to RNase H treatment with oligonucleotide directed against sense or antisense RNA. Oligos SS1 or E3r were used for RT synthesis of 'sense' (S) or 'antisense' (AS) cDNA respectively. RT-PCR of EF1α (lower panel) was used as a loading control. **D)** Oligonucleotide primer independent reverse transcription. RT-PCR analysis of nuclear RNA isolated from HeLa cells transiently transfected with pβT+ and pTat. No oligo was used in the RT reaction and the level of message over the ORF of the gene is shown.

then PCR amplified using primer E3r located at the 3' end of the gene (Figure 4.7A) and SS3. As a positive control for the PCR reaction, p β T⁺ DNA was used which produces an amplicon the same size as the unspliced message (Figure 4.7B, lane 7). The PCR product obtained from amplification of the three different cDNAs is the same size as the mature spliced β -globin message, indicating the presence of an antisense mRNA. However, this result suggests the unlikely possibility that splicing signals are recognised in reverse and the antisense RNA is spliced in exactly the same positions as the mature mRNA. Suspicious of this unusual result, I decided to check if this RT-PCR product was dependent on the presence of a genuine antisense RNA or was an artifact. To do this, I targeted the destruction of the sense or antisense RNA by oligo-directed RNase H digestion. As a positive control, the RNA was treated only with RNase H before cDNA synthesis. Following RNase H treatment, cDNAs were synthesised using primers SS1 or E3r to reverse transcribe antisense or sense RNA respectively. The cDNAs were then PCR amplified using the SS1/ E3r primer pair and the resulting PCR products are displayed in Figure 4.7C (lanes 1-4). In lane 1, a band is detected from PCR amplification of the 'antisense' cDNA synthesised using primer SS1. In lane 3, there is a stronger band arising from PCR amplification of the 'sense' cDNA indicating that there is a higher abundance of sense RNA. Next, the effect of cleaving the sense mRNA was investigated using primer E2r, which is complementary to exon 2 of the mRNA. This targeted RNase H treatment should allow the antisense RNA to remain intact whilst the sense mRNA is cleaved in the middle. Following this, 'sense' and 'antisense' cDNAs were synthesised using primers E3r or SS1 respectively, then amplified using primer pair SS1/E3r (Figure 4.7C, lanes 5-8). In lane 5, there is a faint band from amplification of the 'antisense' cDNA. Comparison of this band with lane 1 shows that when the sense mRNA is

targeted for degradation, the abundance of the antisense PCR product decreases significantly. This result indicates that the antisense PCR product derives, at least in part, from the sense mRNA. In lane 7, there is a strong band for 'sense' cDNA showing that some sense mRNA remains. However, this band is weaker than in lane 3 showing that oligo-directed RNase H treatment has been partially successful in removing sense mRNA. Considering the fact that not all the mRNA is digested it is likely that the small level of product detected in lane 5 is due to the remaining mRNA. However, it is also possible that this is due to the presence of an antisense RNA. Subsequently, I carried out oligo-directed RNase H digestion to target the antisense RNA in exon 2, with primer E2f. Following this treatment, the putative antisense RNA is destroyed. The results show that destruction of the putative antisense RNA has no effect on production of the RT-PCR product that was reverse transcribed with primer SS1 (lane 9). As a positive control that the sense RNA is not destroyed, cDNA was synthesised with primer E3r and a band is detected in lane 11. This indicates that the PCR product detected after cDNA synthesis with SS1 does not derive from antisense RNA. A minus Reverse Transcriptase control was used throughout this experiment (lanes 2, 4, 6, 8, 10 and 12), to show that the PCR products detected arise from the cDNA synthesised during the RT reaction, and to confirm that there was no DNA contamination. Furthermore, to control for loading, an RT-PCR product for the EF1 α gene is displayed in the lower panel of Figure 4.7C.

Overall, the data presented in Figure 4.7C indicates that the correctly spliced RNA detected with a primer designed to detect antisense RNA is not dependent on the presence of an antisense RNA but does require mRNA. This phenomenon could arise from the SS1 primer being complementary to the 3' end of the mRNA. However, as

the same RT-PCR result was obtained when three different reverse primers (SS1, SS2 and SS3) were used for cDNA synthesis (Figure 4.7B, lanes 1, 3 & 5), it appeared unlikely that this was the reason. A possible explanation for this could be that the mRNA primes cDNA synthesis itself or there is another RNA present that acts as a primer for cDNA synthesis. To investigate this possibility, cDNA was synthesised from nuclear RNA isolated from HeLa cells transiently transfected with p β T⁺ and pTat, without the addition of a reverse primer. PCR amplification of this cDNA using primer pair SS1/E3r produced an amplicon when Reverse Transcriptase was present in the RT reaction but not when the enzyme was absent, showing that there is no contaminating DNA (Figure 4.7D, lanes 1 & 2). As a positive PCR control, p β T⁺ DNA was amplified which gave rise to a product corresponding in size to the unspliced message (lane 3). The results here show that RT-PCR of the mature β -globin message is RT-dependent but primer independent. This artefact makes it very difficult to investigate the β -globin antisense transcripts, but also draws attention to the possibility of self-priming when using this technique. Based on these observations, it would be sensible to carry out minus primer controls during RT-PCR analysis and be especially cautious when studying antisense transcription.

4.6 Discussion

In this Chapter, I developed a system in *S. pombe* to identify potential factor(s) involved in CoTC related termination. The initial growth data (Figure 4.2) was encouraging as pUI and pUIspacer cells could survive in the absence of uracil, whereas pUI*ura4*Term, pUITerm and pUIpATerm could not. At the time it was rather surprising that pUImutTerm and pUIpAmutTerm were also unable to survive.

However, further investigation into the quality of the RNA showed that the growth data was explicable in terms of the production of aberrant mRNA in terminator containing clones, rather than a significant termination effect. The *ura4* terminator was extremely efficient in directing transcriptional termination.

Northern Blot analysis reveals that splicing efficiency of the synthetic intron is greatly reduced as the intron increases in length. This is demonstrated in the pUIspacer construct where most of the RNA produced is unspliced. However, the very small amount of correctly spliced *ura4* mRNA is able to produce sufficient amounts of Orotidine 5'-phosphate decarboxylase for the cells to survive in uracil deficient media (Figure 4.3B). It further suggests that the sequence within the intron affects the levels or efficiency of splicing. In the pUIpATerm and pUIpAmutTerm where the size of the intron is the same as the pUIspacer, there is no splicing taking place and therefore, these cells are unable to grow on media lacking uracil. Further confirmation of this comes from the pUITerm and pUImutTerm constructs where the intron is smaller in size than the pUIspacer and yet again splicing out of the intron does not occur and the cells do not survive in the absence of uracil.

Termination efficiency of the pUITerm and pUIpATerm construct is not significant in comparison to the *ura4* terminator. The Northern Blot analysis reveals a very strong signal for unspliced transcripts in both cases (Figure 4.5, lanes 3 and 5) and smaller bands indicating low level transcription termination. In addition, transcripts from pUImutTerm and pUIpAmutTerm constructs remain unspliced but in this case, there are no smaller bands detected indicating that termination is not taking place (Figure 4.5B, lanes 4 and 6). This is in agreement with data in the previous Chapter

suggesting that the mutant Terminator is unable to elicit CoTC and termination. Thus, I can conclude that the mutant Terminator works neither in HeLa cells nor *S. pombe*.

During this investigation, I found that the human β -globin gene poly(A) signal and Terminator can promote antisense transcription in *S. pombe* when placed within the synthetic intron of the modified *ura4* gene (Figure 4.6A; top panel lane 5). The requirement for both, a poly(A) signal and a Terminator element, allows us to envisage a scenario whereby the Terminator acts as an enhancer and the poly(A) signal as the TATA box, and together they are able to initiate transcription in an antisense direction. Furthermore, the fact that the mutant Terminator is not able to initiate antisense transcription could provide a clue as to the reason why it does not function as a Terminator. It is conceivable that the wild type Terminator DNA sequence is nucleosome depleted due to the sequence being AT rich. However, once the sequence is changed, as is the case in the mutant Terminator, then nucleosomes could associate with the DNA, which could in turn prevent correct termination from occurring. On a similar note, if this region has nucleosomes present, it would be hard for transcription to initiate from there spuriously.

I also observed that the *ura4* terminator is capable of initiating antisense transcription at the 3' end of the gene in the pUI*ura4*Term construct, but not when placed within the intron, which could suggest a role for the strength of the poly(A) signal since a stronger poly(A) signal is present in the wild type location. Furthermore, it is interesting to note that antisense transcription initiation at the *ura4* terminator is only detected from the pUI*ura4*Term construct even though this element is present in all the other constructs. A clue to what is happening here could lie in the fact that

transcription terminates within the intron only in pUI*ura4*Term. Thus there is no transcription over the 3' end *ura4* termination region. This may allow antisense transcription to initiate. Further work will be required to examine this hypothesis.

These interesting observations led me to consider the possibility that termination signals act as promoters in reverse in *S. pombe*. To this end, I tried to detect antisense transcripts from five different *S. pombe* endogenous genes – *nmt1*, *nmt2*, *ura4*, *mep33* and *cdk9* – using Northern blot analysis. However, I did not see any evidence for this effect, suggesting that perhaps this phenomenon is only observed in plasmids. However, another possibility is that antisense transcripts, originating from termination signals, are indeed expressed at low levels and/or are degraded by the exosome. Perhaps a more sensitive approach is needed to detect these transcripts if they exist, and also an exosome depleted *S. pombe* strain may prove more useful.

Investigation into the presence of antisense transcripts of the β -globin gene and whether they have any regulatory role in transcription and termination has proven difficult. Northern Blot analysis reveals the presence of a low level antisense transcript, which is the same size as the mRNA (Figure 4.6B). Volloch et al. (1996) have previously demonstrated the presence of antisense β -globin RNA in mouse. This antisense RNA was entirely complementary to the mRNA and it was claimed originated within the cytoplasm. The size of the β -globin antisense RNA detected in Figure 4.6B is the same size as the β -globin mRNA, however, this has nuclear origin suggesting that this might not be the same species described previously (Volloch et al. 1996). Due to the large amount of RNA needed for detection of this antisense transcript on a Northern Blot, I decided to use RT-PCR analyses, as it is more

sensitive. Characterisation of the antisense transcript by RT-PCR led me to discover that RT-PCR of the human β -globin mRNA is RT-dependent, but primer independent (Figure 4.7D). This was an unexpected observation, which highlights the importance of rigorous controls. RT-PCR is a widely used technique and it would be prudent for minus primer controls to be used alongside the minus reverse transcriptase controls. I aim to continue my investigation into the β -globin antisense transcript to further clarify its potential significance.

Although the data in this Chapter indicates that the human β -globin Terminator mediates some low level termination in *S. pombe*, the effect is not strong enough to use this system in a genetic screen to identify potential factor(s) involved in CoTC related termination. However, some very interesting observations were made in this part of the study. As *S. pombe* Pol II is similar to mammalian Pol II, it appears that CoTC may not be due to an intrinsic Pol II activity. The inefficiency of the Terminator could suggest that although homologues of some of the factor(s) responsible for CoTC related termination exist in *S. pombe*, their primary function may be different. For example, they may not be dedicated termination factors. Another explanation could be that *S. pombe* lacks the factor(s) responsible and different factor(s) can elicit CoTC and subsequent termination to some extent. Furthermore, the CoTC process is sequence dependent and possibly factors in *S. pombe* do not respond to the human sequence. In addition, the presence of an antisense transcript being produced from pUIpATerm, and not from pUIpAmutTerm in *S. pombe*, could be supportive of the idea that nucleosome distribution is an important factor in the CoTC termination process. To this end, I have made stable cell lines containing the human β -globin gene with a wild type or mutant Terminator,

which I plan to use in ChIP analysis to examine the chromatin structure in the Terminator region.

CHAPTER 5

INVESTIGATION INTO THE REQUIREMENTS FOR CoTC ACTIVITY

5.1 Introduction

The aim of my thesis is to understand more about the mechanism of CoTC mediated transcription termination. In Chapter 3, I uncovered a role for the A tracts, found in the β -globin Terminator sequence, in Pol II transcription termination and CoTC activity. In Chapter 4, I described the development of a system that could provide the basis for a genetic screen for mediators of CoTC activity. However, after careful analysis of the results of these experiments, I reasoned that although observations from this system were informative, it did not provide a sensitive or robust enough platform for a genetic screen.

During the course of my studies, I have also analysed candidates that could mediate CoTC. Thus, in this Chapter I describe an investigation into the possibility that Dis3, the endonuclease component of the mammalian exosome or Dicer, a protein involved in the RNAi pathway, are involved in CoTC mediated Pol II transcription termination in mammalian cells. I also describe preliminary analyses of the role of the Carboxy Terminal Domain (CTD) of mammalian Pol II and pre-mRNA processing events in transcription termination and the CoTC activity.

5.2 Dis3 as a possible candidate for CoTC activity

Dis3 is an exosome subunit that has recently been reported to have endonucleolytic activity (Lebreton et al. 2008; Schaeffer et al. 2009; Schneider et al. 2009). To test the possibility that Dis3 is involved in CoTC activity, the protein was depleted from HeLa cells using RNA interference (RNAi). HeLa cells were transfected with a siRNA duplex specifically targeted to Dis3 mRNA. As controls, mock transfections

were performed in parallel where the heteroduplex was omitted. To check for Dis3 protein expression levels, Western Blot analysis using an anti-Dis3 antibody was performed (Figure 5.1A, top panel). In lane 1 (mock treated cells), two bands are detected; the higher sized band represents Dis3 and the lower sized band is an unidentified non-specific band. In lane 2 (Dis3 depleted cells), the same two bands are detected; however, the signal for Dis3 is much lower. Actin was used as a loading control and detected using an anti-actin antibody (Figure 5.1A, lower panel). Band signal strength was quantitated using ImageJ. This representative Western Blot shows that by using RNAi, the level of Dis3 protein was reduced by approximately 70%.

Having established an efficient Dis3 knock down, the effect of Dis3 on Pol II termination of the human β -globin gene was assessed using the RNase Protection Readthrough (RPA-RT) assay described in Chapter 3. Nuclear RNA was isolated from mock treated and Dis3 depleted HeLa cells that were transfected with p β T⁺ and the transactivator Tat (pTat), and hybridised to the HIV-LTR radiolabelled antisense RNA probe. Following RNase digestion, protected products were analysed by PAGE (Figure 5.1B). In the case of efficient transcriptional termination one short band, labelled P, resulting from protection of the probe complementary to the 5' end of β -globin mRNA is detected at the foot of the gel. In the case of inefficient termination two bands are detected, the aforementioned short product P and a larger readthrough product, labelled RT (Figure 3.3b, right-hand side panel). The higher the ratio of RT to P, the higher the termination defect. Band signal strength was quantified by PhosphoImager and the ratio RT/P calculated, normalised to the positive control, mock treated cells, and displayed in the graph below the data panel (Figure 5.1B). The results of this experiment indicate that there is no difference in the ratio of RT over P from the mock treated and Dis3 depleted cells, suggesting that depletion of Dis3 does

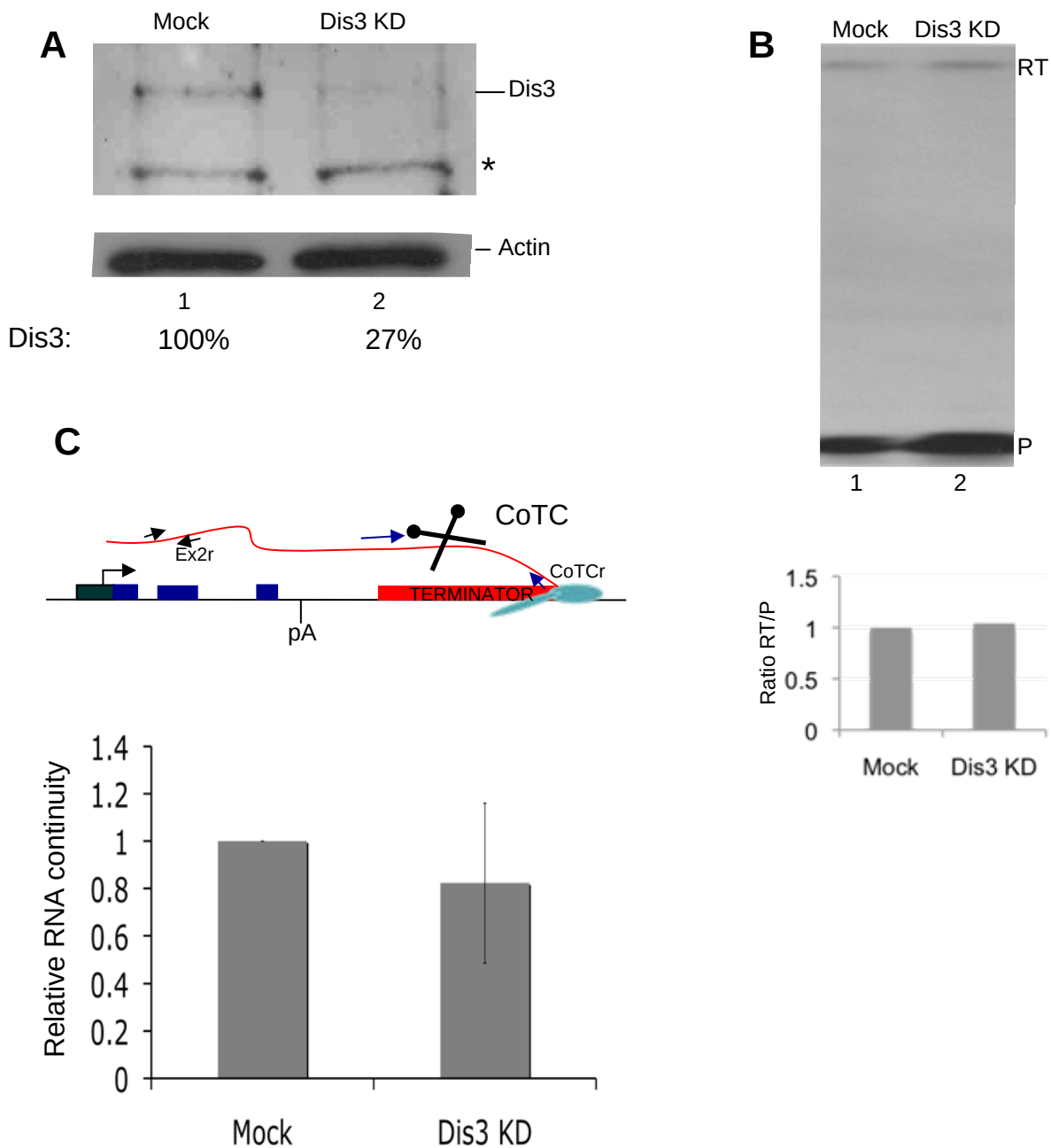


Figure 5.1: Analysis of the role of Dis3 on Pol II termination of the human β -globin Terminator. **A)** Western Blot analysis of the efficiency of Dis3 RNAi mediated knockdown. The actin loading control is shown in the bottom panel. An unidentified non-specific band (asterisk) is detected by the Dis3 antibody. Band signal strength was quantitated by ImageJ, and Dis3 level is shown below the data panel. **B)** RPA analysis of nuclear RNA isolated from mock treated and Dis3 depleted cells transfected with p β T+ and pTat. Nascent RNA is hybridised to a radiolabelled antisense RNA probe that spans the promoter region of the β -globin construct. Band P results from hybridisation of correctly initiated transcripts to the antisense probe. Band RT results from hybridisation of readthrough transcripts to the the antisense probe and represents the readthrough transcripts. Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control (mock treated cells) which is set at 1 and displayed in the graph below the data panel. **C)** Quantitative RT-PCR analysis of nuclear RNA isolated from mock treated and Dis3 depleted cells transfected with p β T+ and pTat. The diagram in the upper panel indicates the positions of the PCR primers. The bar graph in the lower panel shows the ratio of Terminator to intron1/exon2 RNA in Dis3 depleted cells relative to that of mock treated cells. Error bars indicating standard deviation over three biological replicates are shown.

not affect transcriptional termination on the human β -globin gene in this transient transfection system.

The lack of a termination defect upon depletion of Dis3 does not exclude the possibility that Dis3 could mediate CoTC activity, in a manner which is somehow uncoupled from Pol II termination. To test this possibility, qRT-PCR analysis of Terminator transcript continuity was carried out on nuclear RNA isolated from mock treated and Dis3 depleted HeLa cells that were transfected with p β T⁺ and the pTat. cDNAs to detect transcripts of the Terminator region and transcripts within the coding region of the gene were synthesised using reverse primers CoTCr and Ex2r, which are complementary to Terminator and exon 2 transcripts, respectively (see diagram Figure 5.1C). The resulting CoTCr cDNA was amplified using primer pair CoTCf/CoTCr, to detect the Terminator transcript. The Ex2r cDNA was amplified using primer pair Int1f/Ex2r, to detect an intron 1/exon 2 transcript. In this assay, the level of PCR product obtained over the Terminator region gives an indication of the extent of transcript cleavage. To control for transcriptional activity in each construct, the amount of PCR product over the Terminator was normalised to the level of nascent transcription in the body of the gene represented by the intron 1/exon 2 PCR amplicon. The ratio of Terminator to nascent transcript in mock treated cells was set to 1 and the same ratio in Dis3 depleted cells was normalised to that of mock treated cells. The results of this analysis are shown in the bar graph in Figure 5.1C. From the graph it can be seen that in Dis3 depleted cells there appears to be no significant change in Terminator transcript continuity when compared to mock treated cells. These results give a strong indication that Dis3 endonuclease activity does not mediate CoTC of nascent Terminator transcripts.

5.3 Dicer and its involvement in termination of the human β -globin gene

Rnt1, a double stranded RNA-specific ribonuclease III (RNase III) family member, has been linked to co-transcriptional cleavage activity in *S. cerevisiae* (Ghazal et al. 2009). Another member of the RNase III family is the mammalian protein Dicer, which is involved in the RNAi pathway. I therefore decided to investigate the possible role that Dicer may play in CoTC activity. Dicer was depleted from HeLa cells by transfection of a heteroduplex targeting Dicer mRNA. Mock transfections were also carried out in parallel and the efficiency of the knock down was assessed by Western Blot as shown in Figure 5.2A. In lane 1 (mock treated cells), a strong band representing Dicer is detected. In lane 2 (Dicer depleted cells), a much weaker band is detected indicating that Dicer is significantly lower. Band signal strength was quantitated using ImageJ. In this representative experiment it can be seen that Dicer has been reduced by approximately 80% using this approach. Actin was also detected and used as a loading control (Figure 5.2A, lower panel).

To examine if Dicer depletion had any effect on Terminator transcript continuity, I employed the qRT-PCR approach used previously for analysis of Dis3 activity (section 5.2). Total RNA isolated from mock treated and Dicer depleted cells transfected with p β T⁺ and pTat, was used to synthesise cDNA using primers CoTCr and Ex2r (see diagram Figure 5.2B). The resulting cDNAs were amplified using the primer pairs CoTCf/CoTCr and Int1f/Ex2r respectively. The ratio of the Terminator amplicon to the intron 1/exon 2 amplicon in mock treated cells was set to 1 and the same ratio in Dicer depleted cells was normalised to that of mock treated cells. The results of this experiment are presented in the bar graph in Figure 5.2B. From this

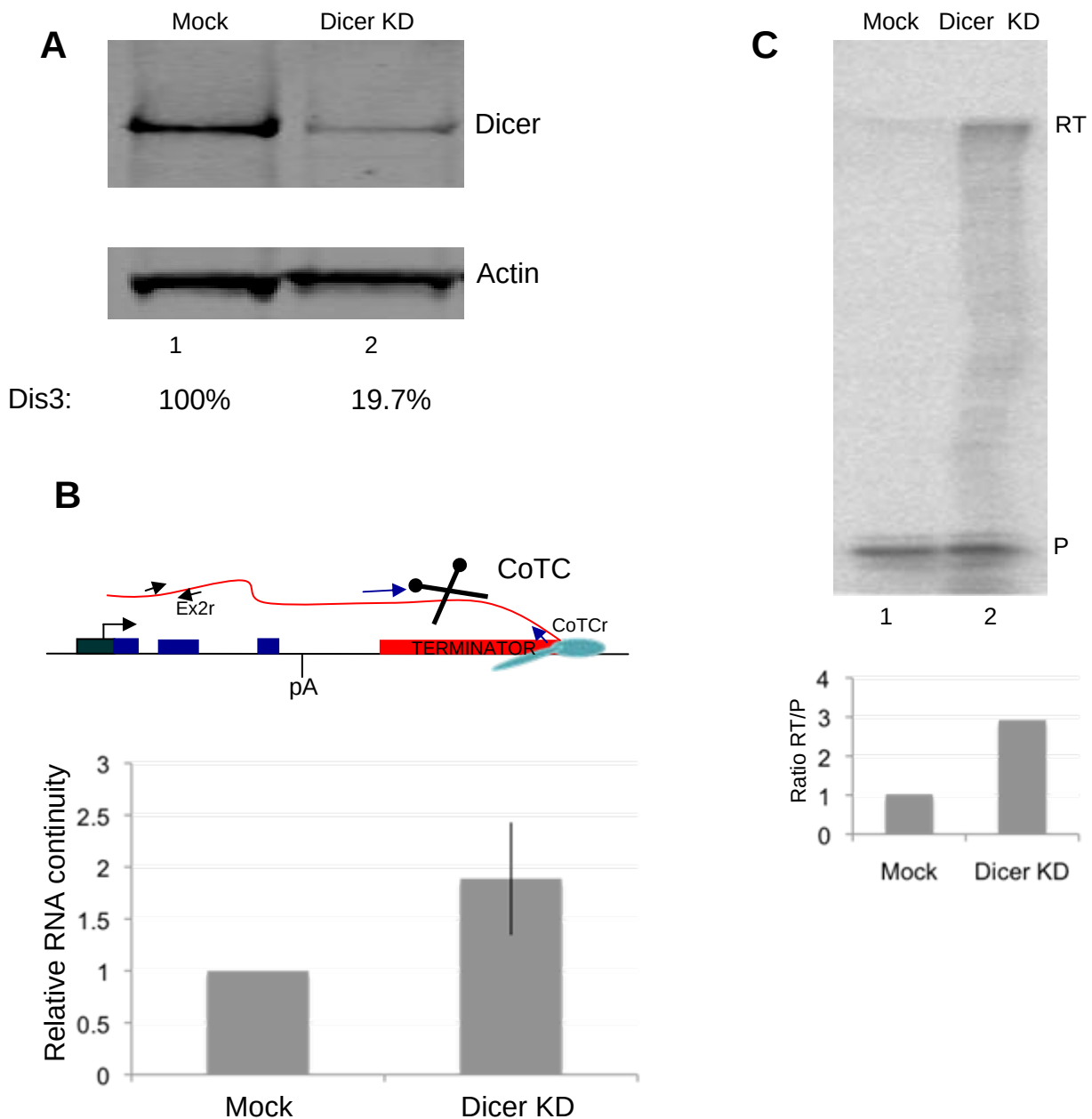


Figure 5.2: Analysis of the role of Dicer on Pol II termination of the human β -globin Terminator. **A)** Western Blot analysis of the efficiency of Dicer RNAi mediated knockdown. The actin loading control is shown in the bottom panel. Band signal strength was quantitated by ImageJ, and Dicer level is shown below the data panel. **B)** Quantitative RT-PCR analysis of nuclear RNA isolated from mock treated and Dicer depleted cells transfected with p β T+ and pTat. The diagram in the upper panel indicates the positions of the PCR primers. The ratio of Terminator to intron1/exon2 RNA in Dicer depleted cells is normalised to that of mock treated cells. Error bars indicating standard deviation over five biological replicates are shown. **C)** RPA analysis of nuclear RNA isolated from mock treated and Dicer depleted cells transfected with p β T+ and pTat. Nascent RNA is hybridised to a radiolabelled antisense RNA probe that spans the promoter region of the β -globin construct. Band P results from hybridisation of correctly initiated transcripts. Band RT results from hybridisation of readthrough transcripts to the the antisense probe and represents the readthrough transcripts. The signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control (mock treated cells) which is set at 1 and displayed in the graph below the data panel.

analysis it can be seen that depletion of Dicer results in approximately two fold increase in the detection of non-cleaved Terminator transcripts (Figure 5.2B), suggesting a reduction in the level of CoTC activity.

The results from the qRT-PCR analysis indicate that CoTC activity may be affected by depletion of Dicer. As mentioned earlier, the transcriptional termination capacity of fragments of the β -globin terminator corresponds to their ability to cleave the transcript (Dye and Proudfoot, 2001). Therefore, if Dicer depletion affects CoTC activity, this should result in a Pol II termination defect. I decided to test this possibility using the RPA-RT assay described above (section 5.2). Nuclear RNA was isolated from mock treated and Dicer depleted HeLa cells that were transfected with p β T+ and pTat, then hybridised to the HIV-LTR radiolabelled antisense RNA probe. Following RNase digestion, protected products were analysed by PAGE (Figure 5.2C). In lane 1 (mock treated cells), there is a band for P and a very weak band for RT indicating efficient termination seen in the presence of the Terminator element. In lane 2 (Dicer depleted cells), there is a band for P, of approximately the same intensity as in lane 1, and a strong band for RT. Band signal strength was quantified by PhosphoImager and the ratio RT/P calculated, normalised to the positive control, mock treated cells, and displayed in the graph below the data panel (Figure 5.2C). The results of this analysis show that the ratio of RT to P in Dicer depleted cells is approximately three times higher than mock treated cells.

In my analysis of *cis* mutations (Chapter 3), accumulation of readthrough transcripts in the RPA-RT assay correlates with a transcription termination defect. Therefore, to directly examine whether there is indeed a transcriptional termination defect upon

depletion of Dicer, I carried out NRO analyses of mock and Dicer depleted cells. In the NRO assay, nascent transcripts were radiolabelled by incubation of isolated nuclei with $\alpha^{32}\text{P}$ -UTP. Transcripts were partially hydrolysed then hybridised to single stranded DNA probes fixed to the hybridisation filters shown in Figure 5.3B. Active Pol II density was analysed over the promoter (P), exon 3 (B3), downstream of the poly(A) site (B4), the vector backbone (A) and the U3 region of the HIV-LTR (Figure 5.3A). In mock treated cells, strong hybridisation signals are detected over P, B3 and B4, and low signals over A and U3 indicating efficient transcriptional termination. This result was expected, as mock treated cells are virtually identical to p β T+ cells used in Figure 3.10B (slightly higher signals are detected over probes A and U3 here, possibly due to the multiple transfections involved in the RNAi procedure). In Dicer depleted cells, a very similar profile is seen with strong hybridisation signals over P, B3 and B4 and low signals over A and U3. NRO signals in mock and Dicer depleted cells were quantified by PhosphoImager, normalised to the number of U residues in each probe and displayed in the graphs besides the data panels (Figure 5.3B, right hand panels). NRO analyses showed that transcriptional termination occurred efficiently in both mock treated and Dicer depleted cells, evident by the low signals over A and U3 (Figure 5.3B). This result indicates that Dicer activity has no significant role in CoTC dependent Pol II termination. However, the data from the RPA-RT analysis and the qRT-PCR analysis indicates that Dicer may be involved in CoTC activity. Nevertheless, it is also possible that Dicer depletion affects the stability of those readthrough transcripts that are detected even in the presence of the Terminator element (Figure 5.2, lane 1), and not CoTC activity itself. Therefore, further experiments will be required to determine whether the affect of Dicer

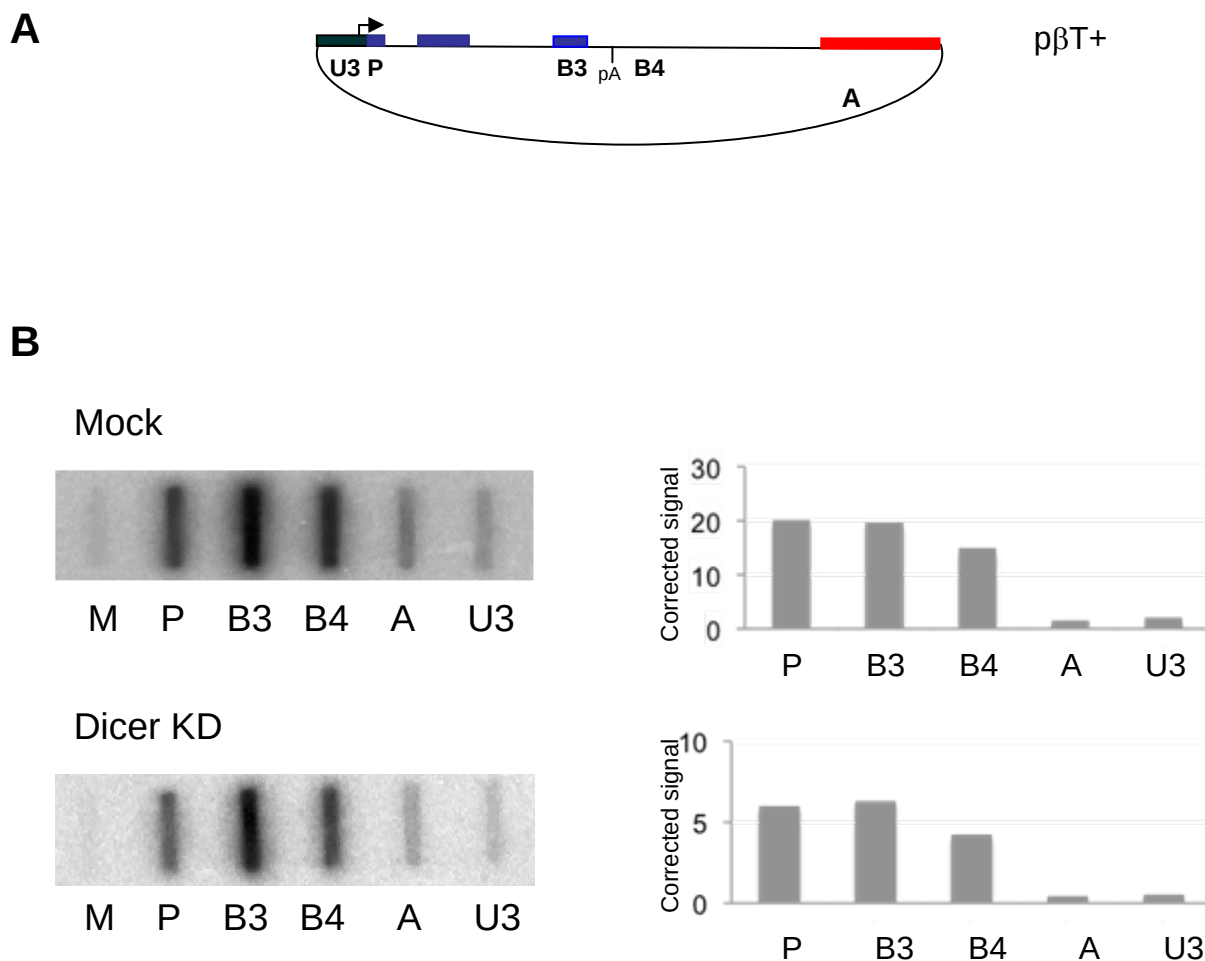
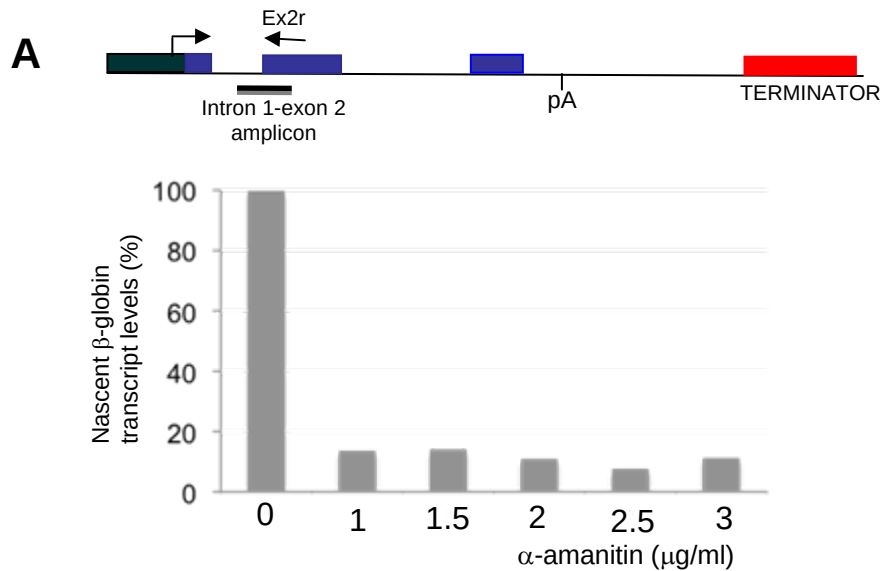


Figure 5.3: Nuclear Run On analysis of the effect of Dicer depletion on termination of the human β -globin Terminator. A) Diagram of the β -globin plasmid. The start site of transcription (arrow) and regions of complementarity of antisense ssDNA M13 probes are indicated (bold). B) NRO analyses of nascent transcripts of mock treated and Dicer depleted cells transfected with p β T+ and pTat. NRO probes are indicated below their respective positions on the filters. M represents M13 vector and is a background control. Graphs on the right of the NRO data panels show strength of hybridisation signals after normalisation for the number of U residues present in each transcript.

depletion on accumulation of RT transcripts is due to a decrease in CoTC activity or an effect on RNA stability.

5.4 Pol II CTD and its involvement in termination of the human β -globin gene

As the CTD of the Rpb1 subunit of Pol II has been shown to be involved in pre-mRNA processing (McCracken *et al.* 1997; Hirose and Manley, 1998), I decided to investigate whether the CTD is required for CoTC activity. To do this, I used a transient transfection system in which cells are transfected with constructs encoding α -amanitin resistant Pol II Rpb1 subunits bearing either a full length CTD of 52 heptapeptide repeats (p52repCTD) or a truncated CTD of 5 heptapeptide repeats (p5repCTD) (Gerber *et al.* 1995). Transcriptional activity of endogenous Pol II Rpb1 subunits in transfected cells is inhibited by addition of α -amanitin. Transcription therefore is derived only from Pol II associated with α -amanitin resistant Rpb1. HEK293 cells were used for this part of the study, as they are more responsive to treatment with α -amanitin and easier to transfect. Initially, I decided to titrate the amount of α -amanitin used to treat cells in order to determine the optimal α -amanitin concentration to inhibit transcription from endogenous Pol II Rpb1 subunit so that it would not interfere with analysis of α -amanitin resistant Pol II Rpb1 transcription. HEK293 cells were transiently transfected with p β T+ and pTat and were either left untreated or were treated with increasing amounts of α -amanitin (1 to 3 μ g/ml) 24 hours after transfection. Total RNA was isolated 48 hours after the α -amanitin treatment and used as a template for cDNA synthesis using primer Ex2r, located in exon 2 of the β -globin gene (see diagram Figure 5.4A). This cDNA was then amplified with forward primer Int1f, located in intron 1 of the β -globin gene, and



B

Transfection

p β T+, pTat & p52repCTD
or p5repCTD or mock

Add α -amanitin

Collect RNA

T=-24hr

T=0hr

T=48hr

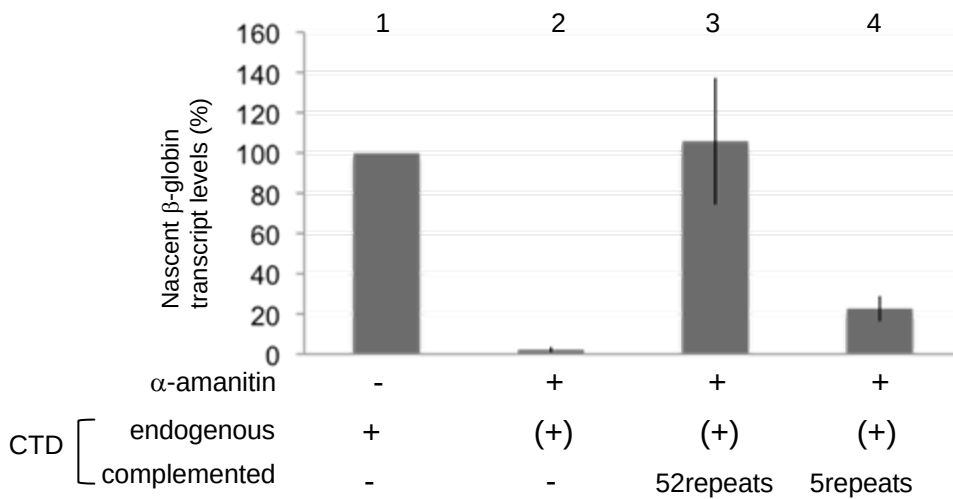


Figure 5.4: The effect of α -amanitin treatment on Pol II transcription. **A)** Quantitative RT-PCR analysis of total RNA isolated from cells transiently transfected with p β T+, pTat and treated with increasing amounts of α -amanitin. In the top diagram of the β -globin gene (upper panel), the position of the PCR amplicon and RT primer is shown. The level of β -globin PCR product (intron 1/exon 2) was normalised to the level of 28S rRNA. The level of β -globin transcripts detected in the absence of α -amanitin is set to 100% and everything else is normalised to that. **B)** The top panel displays a timeline for the CTD complementation experiments. Quantitative RT-PCR analysis of nuclear RNA isolated from cells transiently transfected with p β T+, pTat and/or an α -amanitin resistant Rpb1 with either 52 (p52repCTD) or 5 (p5repCTD) heptapeptide repeats following the scheme shown above. Where the endogenous CTD is inactivated by α -amanitin, it is represented by (+). β -globin RT-PCR products were normalised to the level of 28S rRNA. The level of transcripts detected for samples in the presence of α -amanitin is normalised to that in the absence of α -amanitin which is set to 100%. Error bars indicating standard deviation over three biological replicates are shown.

Ex2r. The amount of this PCR product represents the level of nascent unspliced β -globin transcripts, which I used as an indicator of transcriptional activity in cells exposed to the varying concentrations of α -amanitin. The level of the β -globin PCR product was normalised to the levels of 28S rRNA detected by qRT-PCR, which is unaffected by α -amanitin treatment. The amount of β -globin qRT-PCR product detected in the absence of α -amanitin treatment is set to 100% and qRT-PCR products for other samples were normalised to that. The results of this analysis are shown in the graph in Figure 5.4A (lower panel). Here it can be seen that the level of nascent transcripts drops dramatically (to 14%) when cells are treated with 1 μ g/ml of α -amanitin. This level falls to 7% when cells are treated with 2.5 μ g/ml of α -amanitin. I decided that treating HEK293 cells with 2.5 μ g/ml of α -amanitin decreased transcription by the endogenous Pol II Rpb1 subunit sufficiently for this analysis and so subsequently used this amount.

I next decided on a scheme for analysing transcription from α -amanitin resistant Pol II, as shown in Figure 5.4B. HEK293 cells were transiently transfected with p β T+, pTat and either the p52repCTD, p5repCTD or mock transfected (two samples were processed for this as positive and negative controls). Cells were subsequently treated with 2.5 μ g/ml α -amanitin 24 hours post transfection, with the exception of a positive control (minus α -amanitin), and RNA was isolated 48 hours later. The level of β -globin nascent transcripts in treated and untreated cells is displayed in the bar graph in Figure 5.4B. In lane 1, where no α -amanitin is added and there is endogenous full length CTD, the level of β -globin transcripts is set to 100%. The sharp decrease in the level of β -globin nascent transcripts in the presence of 2.5 μ g/ml α -amanitin and the

absence of an α -amanitin resistant CTD (lane 2), shows that the treatment is effective and has a negative effect on transcription. In lane 3, the level of transcription is restored, owing to complementation with the α -amanitin resistant full length CTD construct (p52repCTD). In lane 4, the truncated α -amanitin resistant CTD (p5repCTD) is only able to partially restore transcription indicating a requirement for the full length CTD for efficient transcription. The data in Figure 5.4B show that α -amanitin treatment effectively abolishes transcription by endogenous Pol II, and that complementation with an α -amanitin resistant Rpb1 subunit is able to restore Pol II transcription. Furthermore, p52repCTD is much more efficient at restoring transcription than p5repCTD.

Using this system, I next analysed any requirement of Pol II CTD in transcription termination. Nuclear RNA isolated from HEK293 cells transiently transfected with p β T+, pTat, complemented with either p52repCTD or p5repCTD, and treated with 2.5 μ g/ml α -amanitin, was hybridised to the HIV-LTR radiolabelled antisense probe. Following RNase digestion, protected products were analysed by PAGE (Figure 5.5A) and compared to protected products from HEK293 cells transfected with p β T+ and pTat in the absence of α -amanitin. In lane 1 (minus α -amanitin), only a very weak RT band is detected and the ratio of RT/P is very small, indicating efficient transcriptional termination. In lane 2, a similar ratio of RT/P bands is observed when the cells were treated with α -amanitin but complemented with p52repCTD. However, in lane 3, when the cells were complemented with p5repCTD, the level of RT increases and the level of P decreases significantly, indicating a termination defect. Band signal strength was quantified by PhosphoImager and the ratio RT/P calculated

for each sample, normalised to the positive control, minus α -amanitin, and displayed in the graph below the data panel (Figure 5.5A). These data show that Pol II with a

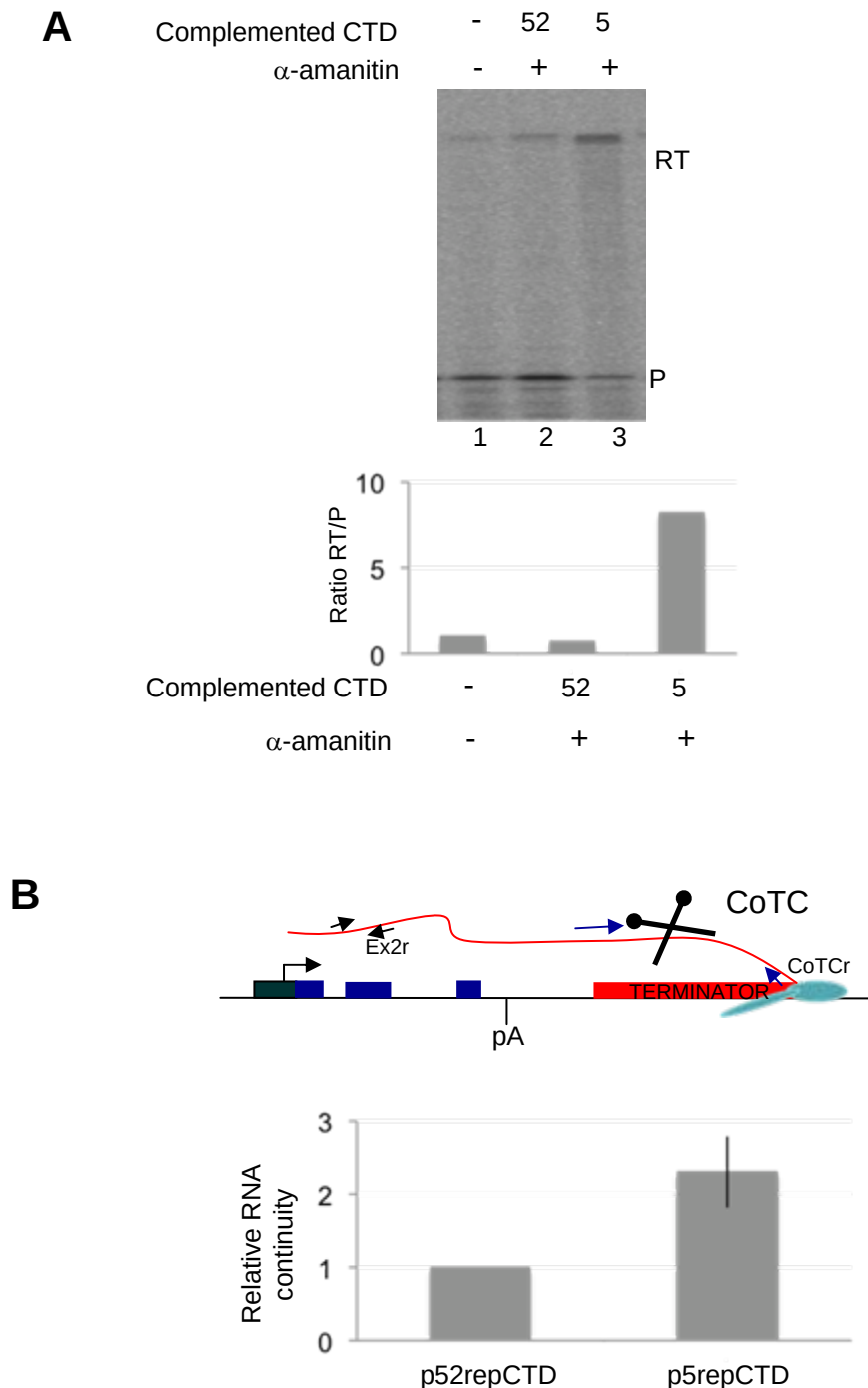


Figure 5.5: Analysis of connections between CTD truncation and CoTC activity of the human β -globin Terminator. **A)** RPA-RT analysis of nuclear RNA isolated from HEK293 cells transiently transfected with p β T+, α -amanitin resistant Rpb1 with either 52 or 5 heptapeptide repeats on the CTD followed by α -amanitin treatment. β -globin transcripts are hybridised to a radiolabelled antisense RNA probe that spans the promoter region. Band P results from hybridisation of correctly initiated transcripts to the antisense probe. Band RT results from hybridisation of readthrough transcripts to the antisense probe and represents readthrough transcripts. Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control (α -amanitin -) and displayed in the graph below the data panel. **B)** Quantitative RT-PCR analysis of total RNA isolated from HEK293 cells transiently transfected with p β T+ and α -amanitin resistant Rpb1 with either 52 or 5 heptapeptide repeats followed by α -amanitin treatment. The diagram indicates the positions of the PCR primers and amplicons. PCR product detected over the Terminator region is indicative of the level of uncleaved transcripts. The ratio of Terminator to intron 1/exon 2 RNA in cells complemented with the α -amanitin resistant Rpb1 composed of 5 heptapeptide repeats (p5repCTD) is normalised to that of the cells complemented with the α -amanitin resistant Rpb1 composed of 52 heptapeptide repeats (p52repCTD). Error bars indicating standard deviation over three biological replicates are shown.

truncated CTD is less capable of directing efficient transcriptional termination of the human β -globin gene. Thus there appears to be a requirement for full length CTD in Pol II termination.

Having shown that CTD truncation causes readthrough transcription on the β -globin gene, I next examined the effect of CTD truncation on CoTC activity using the qRT-PCR assay described in section 5.2. Total RNA isolated from HEK293 cells transfected with p β T+, pTat and complemented with either the p52repCTD or p5repCTD, and treated with 2.5 μ g/ml α -amanitin, was used as a template for cDNA synthesis using primers CoTCr and Ex2r shown in the diagram in Figure 5.5B. Quantitative PCR was then used to measure the abundance of amplicons of the Terminator region and the intron 1/exon 2 junction transcripts. The level of the Terminator region amplicon is representative of the level of transcript continuity while the latter is a measure of the level of transcription. The ratio of the Terminator amplicon over the intron 1/exon 2 amplicon from cells complemented with p52repCTD is set to 1, and the same ratio from cells complemented with p5repCTD is normalised to it. The results of this experiment are displayed in the graph in Figure 5.5B. Here it can be seen that CTD truncation leads to a 2-3 fold increase in the level of uncleaved Terminator RNA, which is indicative of a decrease in the level of CoTC activity. This result could indicate that the CTD mediates CoTC activity by acting as a platform for factors that catalyse CoTC much in the same way as it is thought to be involved in RNA processing events (McCracken et al. 1997). However, as it is well established that the CTD is required for efficient pre-mRNA processing, and that pre-mRNA processing is required for efficient Pol II termination (Whitelaw and Proudfoot, 1986; Dye and Proudfoot, 1999), it is also possible that the effect of CTD

truncation on CoTC activity noted is caused by defects in pre-mRNA processing. I therefore next addressed the possible role of pre-mRNA processing in the CoTC activity.

5.5 Polyadenylation and splicing are required for CoTC activity of the human β -globin gene Terminator

To analyse a role of pre-mRNA processing in CoTC activity, I made two constructs, p β T+ Δ pA and p β T+ Δ SA. Both constructs contain the β -globin gene, 240bp of post poly(A) site sequence and the Terminator element. However, for pT+ Δ pA, the canonical AATAAA poly(A) signal sequence has been inactivated by mutation to AATTCC. In p β T+ Δ SA, the intron 2 splice acceptor signal is inactivated by mutation of the polypyrimidine tract and the AG dinucleotide (Figure 5.6A). I then tested these constructs for effects on termination, using the RPA-RT assay. Nuclear RNA isolated from HeLa cells transiently transfected with pTat and, p β T+, p β T+ Δ pA or p β T+ Δ SA was hybridised to the HIV-LTR radiolabelled antisense RNA probe. Following RNase digestion, protected products were analysed by PAGE (Figure 5.6B). Band signal strength was quantified by PhosphoImager and the ratio RT/P calculated for each construct, normalised to the positive control, p β T+, and displayed in the graph next to the data panel (Figure 5.6B). As expected, in lane 1 (p β T+), there is a band for P, but not for RT, indicating efficient termination. In lane 2 (p β T+ Δ pA), the level of P decreases and the level of RT increases dramatically, as is the case in lane 3 (p β T+ Δ SA), confirming that there is a termination defect in both constructs and that polyadenylation and terminal intron splicing signals are required for termination of the human β -globin gene (Dye and Proudfoot, 1999). I next examined the effect of

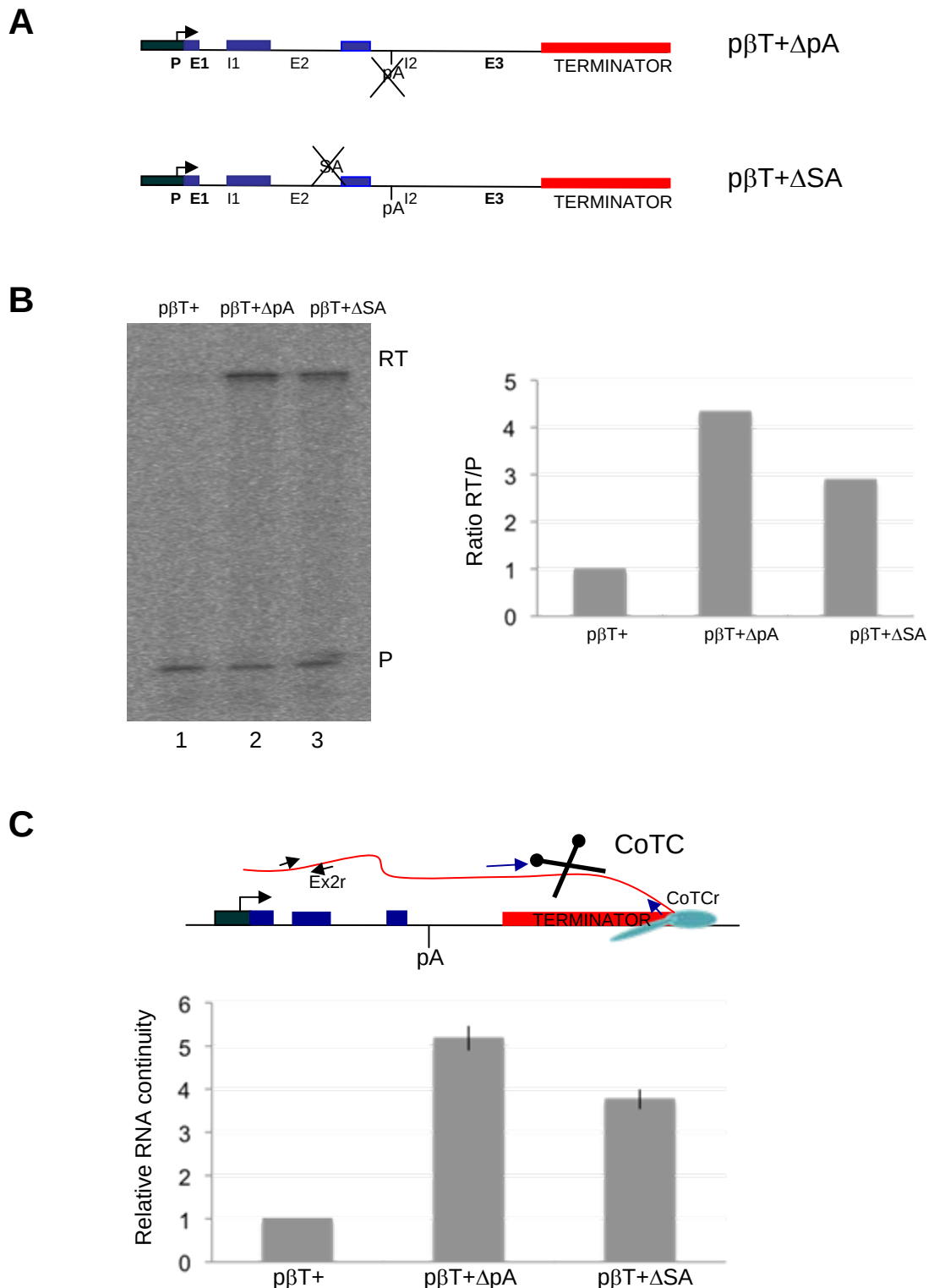


Figure 5.6: Analysis of connections between pre-mRNA processing and CoTC activity of the human β -globin Terminator. **A)** Diagram of β -globin constructs with mutations in the poly(A) site ($p\beta T+\Delta pA$) or the terminal splice acceptor site ($p\beta T+\Delta SA$). **B)** RPA analysis of nuclear RNA isolated from HeLa cells transiently transfected with $p\beta T+$, $p\beta T+\Delta pA$ or $p\beta T+\Delta SA$. Nuclear RNA is hybridised to a radiolabelled antisense RNA probe that spans the promoter region of the β -globin construct. Band P results from hybridisation of correctly initiated transcripts to the antisense probe and band RT results from hybridisation of readthrough transcripts to the antisense probe and represents readthrough transcription. Radioactive signals were quantitated using PhosphorImager, the RT/P ratio calculated and corrected to the positive control, $p\beta T+$, which is set at 1 and displayed in the graph to the right of the data panel. **C)** Quantitative RT-PCR analysis of total RNA isolated from cells transiently transfected with $p\beta T+$, $p\beta T+\Delta pA$ or $p\beta T+\Delta SA$. The diagram shows the positions of the PCR primers and amplicons. The level of PCR product detected over the Terminator region is indicative of the level of uncleaved transcripts. This is then compared to the level of transcription indicated by the intron 1/exon 2 amplicon. The ratio of Terminator to intron 1/exon 2 RNA in $p\beta T+\Delta pA$ and $p\beta T+\Delta SA$ is normalised to that of $p\beta T+$. Error bars indicating standard deviation over three biological replicates are shown.

these mutations on CoTC activity by measuring Terminator transcript continuity using the qRT-PCR assay described in section 5.2. Total RNA was reverse transcribed with primers CoTCr and Ex2r, then the cDNAs were PCR amplified using the CoTCf/CoTCr and Int1f/Ex2r primer pairs, respectively. The level of PCR products detected over the Terminator with primer pair CoTCr/CoTCf was normalised to the level of PCR product detected over the intron 1/exon 2 junction with primer pair Ex2r/Int1f. The ratio of Terminator over nascent transcripts from p β T+ Δ pA and p β T+ Δ SA is normalised to the ratio obtained from the positive control, p β T+. The results shown in Figure 5.6C indicate an almost 5 fold increase in RNA continuity for p β T+ Δ pA and 4 fold for p β T+ Δ SA, suggesting that mutation of these pre-mRNA processing signals reduces CoTC activity.

5.6 Discussion

In this Chapter, I analysed the effect of depleting cells of the exosome subunit Dis3 on CoTC activity. RPA-RT analysis and qRT-PCR based Terminator transcript continuity assay showed that Dis3 depletion affected neither Pol II termination or CoTC indicating that the endonuclease activity of Dis3 does not mediate CoTC of the human β -globin Terminator transcript (Figure 5.1).

I conducted a similar analysis on the RNase III family protein Dicer. RPA-RT analysis shows accumulation of readthrough transcripts in Dicer depleted cells and qRT-PCR analysis indicates approximately a two fold increase in RNA continuity over the Terminator region when Dicer is depleted. However, NRO analysis of nascent transcripts shows that there is no discernible termination defect in Dicer

depleted cells (Figure 5.3). The discrepancy between these results suggests that perhaps the effects observed by qRT-PCR and RPA-RT analyses reflect a change in the stability of the RNA when Dicer is depleted. However, my results do not rule out the possibility that Dicer may be necessary for CoTC activity and/or turnover of readthrough transcripts, but not for termination itself. Furthermore, it is possible that a number of factors are required for Terminator function and loss of one does not abolish termination. It is equally plausible that the Terminator has a dual role; one to pause or slow down Pol II and the other is the CoTC activity. In this scenario, upon depletion of Dicer, the CoTC activity is disrupted but not the pausing/slowing down which could be sufficient for termination to occur. Cleavage at the poly(A) signal would allow Xrn2 entry leading to termination (West et al. 2004). Further investigation into the role of Dicer in CoTC activity is currently underway.

As well as analysing single protein candidates that might mediate CoTC activity, I also obtained some preliminary data for the role of Pol II CTD and pre-mRNA processing in the CoTC activity. In my analysis of the requirement of full length CTD for termination of the human β -globin gene, I recapitulated the results of McCracken et al. (1997), showing a significant termination defect in cells depleted of full length CTD and confirmed the effect they had observed with hybrid genes. I also examined the effect of CTD truncation on CoTC and detected a 2-3 fold increase in Terminator transcript continuity, which could indicate a requirement of full length CTD for CoTC activity.

Finally, I examined whether pre-mRNA processing has a role in CoTC activity. I confirmed that efficient pre-mRNA processing is required for Pol II termination and

showed a significant (4-5 fold) increase in the level of Terminator transcript continuity, when pre-mRNA processing is blocked by *cis* mutations. However, it has previously been published that a functional poly(A) signal is not required for CoTC activity (Dye and Proudfoot, 2001). The reason for this discrepancy could be that the more sensitive qRT-PCR technique is able to detect a reduction in CoTC activity which is undetected by a hybrid selection NRO assay. This is a very interesting result as it may indicate a direct link between pre-mRNA processing and the CoTC activity. However, the possibility that RNA stability of uncleaved CoTC transcripts is increased when transcript processing is defective cannot be discounted. Further work is underway to examine these findings.

CHAPTER 6

GENERAL DISCUSSION

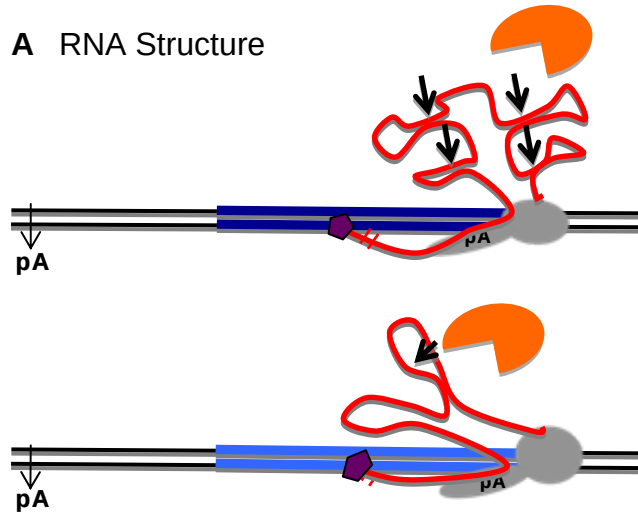
Pol II transcriptional termination is a complex and potentially critical step in gene expression, which depends on the coordinated function of *cis*-acting nucleotide sequences and *trans*-acting factors. Two different types of *cis*-acting sequences are required for Pol II termination – RNA processing signals that mediate mRNA 3' end formation, and specific terminator elements located downstream of the poly(A) site. Work described in this thesis focuses on understanding the molecular basis of one of the most efficient terminator elements described so far, namely the Co-Transcriptional Cleavage (CoTC) element of the human β -globin gene (Figure 1.3B).

It has been shown that CoTC-mediated termination results in enhanced mRNA and protein levels (West and Proudfoot, 2009), a finding which indicates that CoTC terminators are important regulators of gene expression. It is possible that this means of enhancing gene expression is employed for highly expressed genes and/or regulated genes whose products need to be made available quickly. The limited number of examples of CoTC elements described so far has precluded bioinformatics-based identification of further examples. However, work is currently underway in our laboratory to identify more examples of genes employing CoTC-mediated termination by a novel approach using CLIP-based selection of fractionated pre-mRNAs. This approach is based on the mechanism of CoTC-mediated termination whereby cleavage of a CoTC element transcript leads to release of pre-mRNA and Pol II from the DNA template before cleavage at the poly(A) site (West et al. 2008). Released pre-mRNA, which has not been cleaved at the poly(A) signal, may potentially employ a CoTC-mediated termination mechanism. Approximately 20 potential candidates have already been identified, and CoTC termination is being verified using a variety of techniques.

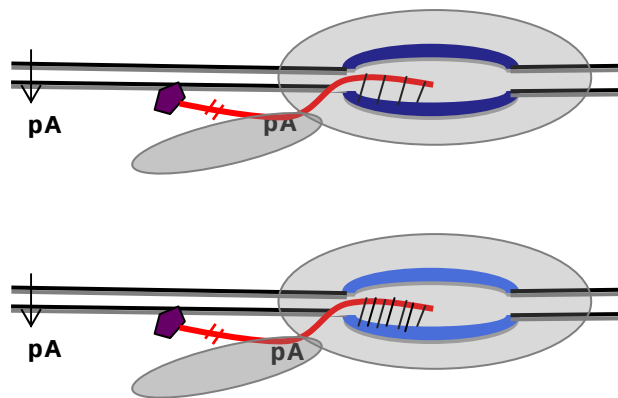
The CoTC terminator element of the human β -globin gene has been divided into three sub-sections, each approximately 300bp, that are independently able to direct CoTC-mediated transcription termination (Dye and Proudfoot, 2001). Sequence comparison between these sub-sections does not reveal any obvious homology. However, all three terminator regions are approximately 75% AT rich. In Chapter 3, I describe the role of homopolymer A tracts located in the 3' end of the β -globin Terminator in Pol II termination. I show that the integrity of the homopolymer tracts is required for co-transcriptional RNA cleavage and subsequent transcription termination.

There are three possible explanations for my observation that homopolymer A tract integrity is required for efficient Pol II termination. Firstly, the A tracts could be required at the RNA level for the establishment of a specific RNA secondary structure. This structure could be required for recognition by a *trans*-acting factor that mediates CoTC (Figure 6.1A). Preliminary structural analysis of the Terminator RNA based on the Mfold RNA structure prediction programs indicates that the structure of the CoTC transcript (Figure 6.2) is significantly altered by A tract disruption (Figure 6.3). However the significance of this finding has not yet been determined. Secondly, the A tracts could be required at the DNA level for the formation of a weak DNA:RNA hybrid in transcribing Pol II (Figure 6.1B). Work in yeast and mammalian systems has shown the importance of weak A:T hybrids, in combination with *trans*-acting factors, in Pol I and Pol III termination. Moreover, the role of A:T hybrids in prokaryotic intrinsic terminators is well established. Thirdly, the A tracts could be required at the DNA level for the formation of a particular chromatin structure. It has been reported that A tracts in genomic DNA have the lowest average nucleosome occupancy (Kaplan et al. 2009). It is therefore possible that disruption of A tracts in

A RNA Structure



B DNA:RNA Hybrid



C Chromatin structure

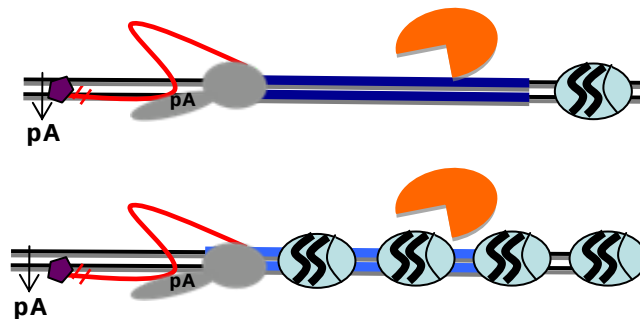


Figure 6.1: Models of CoTC activity and possible effects of Terminator oligo A mutations. Elongating Pol II (grey oval), DNA (black lines), wild type and mutant Terminator (navy blue and blue lines respectively), nascent RNA transcript (red line), cap (purple pentagon) and poly(A) signal (pA) are indicated. **A)** RNA structure. Possible recognition sites (black arrows) and a possible transacting factor (orange packman) are indicated. The secondary structure of the Terminator RNA may be important for CoTC activity possibly by being recognised by a *trans*-acting factor. The mutations can change the RNA structure and lead to a reduction of secondary elements capable of being recognised. **B)** DNA:RNA hybrid. The melting temperature of the DNA duplex is increased due to the mutations and the strength of the DNA:RNA hybrid would also be stronger which could in turn affect CoTC activity. **C)** Chromatin structure. Nucleosomes (light blue ovals with black lines) are indicated. Nucleosome occupancy of the Terminator DNA may be altered due to the mutations. This could in turn affect CoTC activity by preventing a transacting factor to recognise the DNA sequence. A combination of these models could be acting together to affect CoTC activity. Diagrams not drawn to scale.

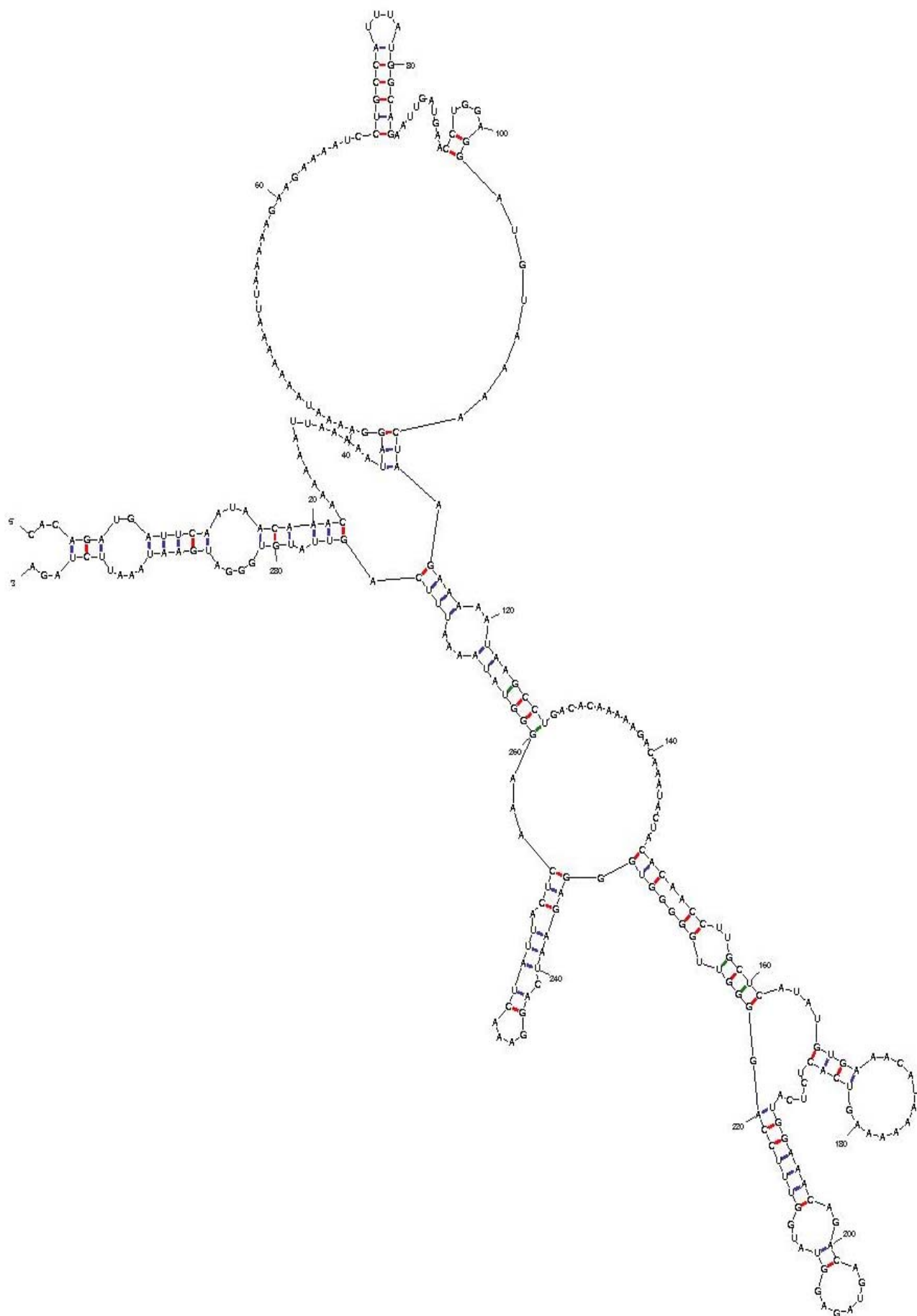


Figure 6.2: Secondary structure prediction of the wild type Terminator RNA as defined by Mfold structure prediction algorithm.

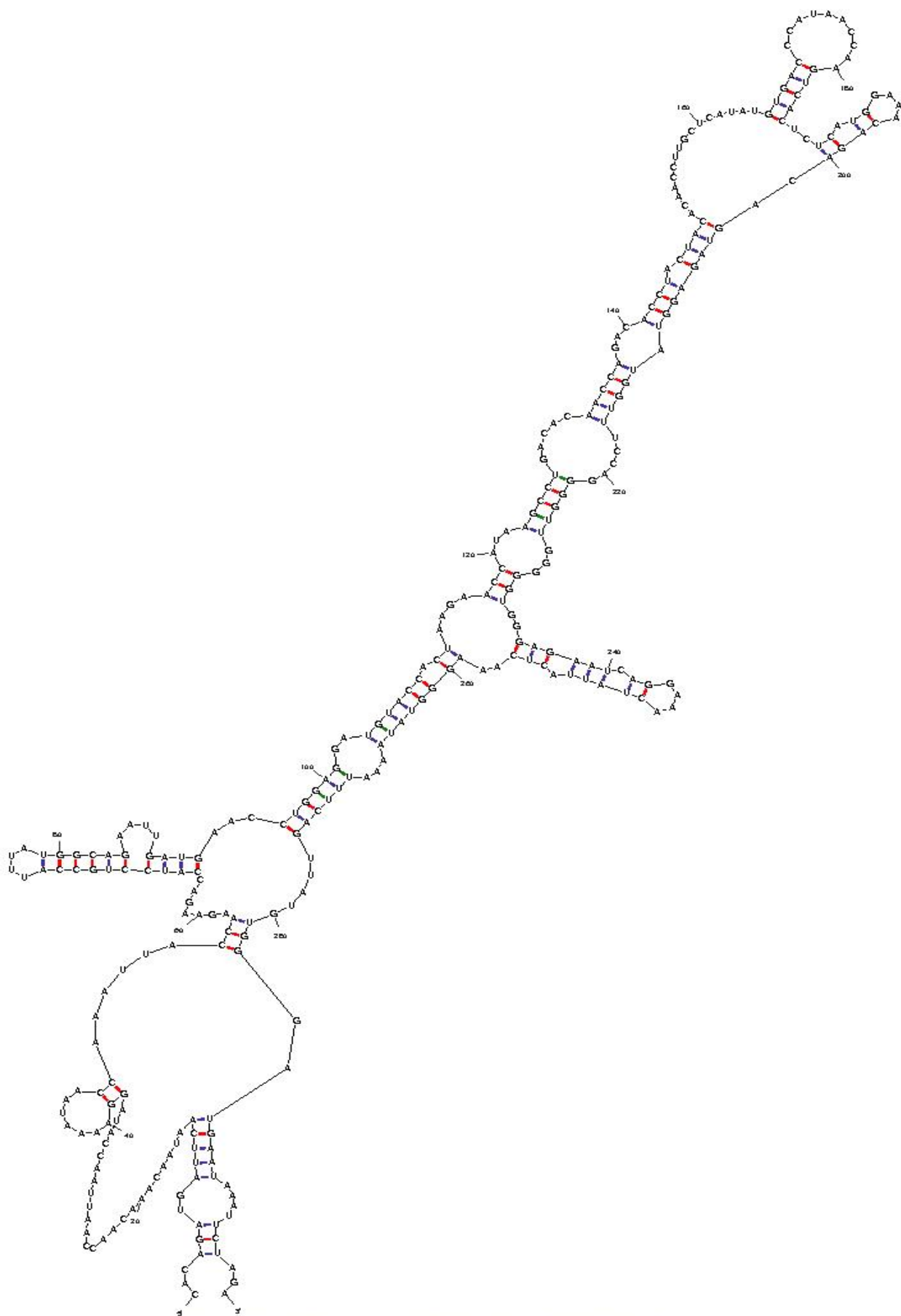


Figure 6.3: Secondary structure prediction of the mutant Terminator RNA as defined by Mfold structure prediction algorithm.

the mutant Terminator leads to an increase in nucleosome density, which could in turn prevent recognition of the CoTC element at the DNA level by necessary *trans*-acting factor(s) (Figure 6.1C). In my analysis of the Terminator activity in *S. pombe*, presented in Chapter 4, I show that the human β -globin poly(A) signal in conjunction with the wild type Terminator, but not the mutant Terminator, is capable of promoting antisense transcription. This observation gives a strong indication that DNA template accessibility for Pol II transcription initiation is dramatically affected by A tract disruption. Finally, it is possible that all three of the factors listed above could operate in concert to mediate efficient Pol II transcriptional termination.

To further understand CoTC-mediated termination, it is clearly critical to identify the activity responsible for the CoTC process. There are three potential mediators of the cleavage event (Figure 6.1B). Firstly, Pol II itself could be responsible, as it has an intrinsic ability to cleave nascent RNA transcripts (Cheung & Cramer, 2011). Data in Chapter 4, however, shows that the CoTC element does not function efficiently in *S. pombe*, which argues against this possibility. The second scenario involves the RNA itself functioning as a ribozyme. This would fit in with a significant role for the secondary structure of this element. However, no significant self-cleaving efficiency is observed *in vitro*, making this an unlikely explanation (Dye et al. 2006). Lastly, there is the possibility that a *trans*-acting factor is involved. This factor would possess endonucleolytic activity, which could act independently or in concert with other factors. Chapter 5 describes an investigation into possible roles of two likely candidates - the exosome subunit Dis3, and the RNase III enzyme Dicer - in CoTC activity. It appears that Dis3 is not involved in the CoTC process. However, the data concerning Dicer remains unresolved and further investigation is required. If Dicer is

the activity responsible, it is surprising that the CoTC element does not function in *S. pombe*, considering that a Dicer homologue is present in these cells. In Chapter 4, the attempt to utilise *S. pombe* to develop a genetic screen for identifying factor(s) involved in CoTC activity was only partially successful. The weak CoTC activity observed in *S. pombe* indicates specific factor(s) in mammalian cells are required to mediate efficient CoTC activity.

In vitro transcribed CoTC element RNA incubated in HeLa nuclear extract is not cleaved (Dye and Proudfoot, 2001) and the CoTC element appears to only mediate RNA cleavage *in vitro* when coupled to transcription (personal communication, M. Dye). These findings indicate that transcription associated factor(s) and/or Pol II itself, are required for the cleavage event. Chapter 5 describes a preliminary investigation into the role of the CTD and pre-mRNA processing events in CoTC activity. The results so far obtained indicate that a full length CTD is required for CoTC activity. However, it is not clear whether CTD acts in a direct manner or indirectly because it is required for pre-mRNA processing reactions. A potential requirement of CTD in CoTC activity could explain the very low efficiency observed for the CoTC element in *S. pombe* as described in Chapter 4. Full length CTD in *S. pombe* is composed of only 29 heptapeptide repeats, which is approximately 50% smaller than the mammalian Pol II CTD. Overall, my experiments imply that the human CTD is required for CoTC activity, suggesting that the CTD might act as a platform or mediator of *trans*-acting factor(s).

In my analysis of CoTC mediated termination presented in this thesis, I have made a series of new observations relating to CoTC activity. Firstly I have shown that the A

tracts in the human β -globin Terminator are important for mediating CoTC. I have also shown that the CoTC element has very low efficiency in *S. pombe*, which suggests that possible specific CoTC recognition and/or cleavage factor(s) are required in mammalian cells. I have also made a number of observations concerning factors that could mediate CoTC cleavage. Although the exact mechanism of CoTC remains unresolved, we are now closer to understanding it.

CHAPTER 7

REFERENCES

- Abou Elela, S. and Ares, M., Jr. 1998. Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *The EMBO journal* **17**(13): 3738-3746.
- Adamson, T.E. and Price, D.H. 2003. Cotranscriptional processing of *Drosophila* histone mRNAs. *Molecular and cellular biology* **23**(12): 4046-4055.
- Agarwal, K., Baek, K.H., Jeon, C.J., Miyamoto, K., Ueno, A., and Yoon, H.S. 1991. Stimulation of transcript elongation requires both the zinc finger and RNA polymerase II binding domains of human TFIIS. *Biochemistry* **30**(31): 7842-7851.
- Ahn, S.H., Kim, M., and Buratowski, S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular cell* **13**(1): 67-76.
- Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Molecular cell* **34**(3): 387-393.
- Alen, C., Kent, N.A., Jones, H.S., O'Sullivan, J., Aranda, A., and Proudfoot, N.J. 2002. A role for chromatin remodeling in transcriptional termination by RNA polymerase II. *Molecular cell* **10**(6): 1441-1452.
- Allison, D.S. and Hall, B.D. 1985. Effects of alterations in the 3' flanking sequence on in vivo and in vitro expression of the yeast SUP4-o tRNA^{Tyr} gene. *The EMBO journal* **4**(10): 2657-2664.
- Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. 1999. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO journal* **18**(19): 5399-5410.
- Anderson, J.S. and Parker, R.P. 1998. The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *The EMBO journal* **17**(5): 1497-1506.
- Ansari, A. and Hampsey, M. 2005. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes & development* **19**(24): 2969-2978.
- Aranda, A. and Proudfoot, N. 2001. Transcriptional termination factors for RNA polymerase II in yeast. *Molecular cell* **7**(5): 1003-1011.
- Aranda, A. and Proudfoot, N.J. 1999. Definition of transcriptional pause elements in fission yeast. *Molecular and cellular biology* **19**(2): 1251-1261.
- Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. 2006. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular cell* **23**(6): 841-851.
- Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. 1997. Intergenic transcription and transinduction of the human beta-globin locus. *Genes & development* **11**(19): 2494-2509.
- Ashfield, R., Enriquez-Harris, P., and Proudfoot, N.J. 1991. Transcriptional termination between the closely linked human complement genes C2 and factor B: common termination factor for C2 and c-myc? *The EMBO journal* **10**(13): 4197-4207.
- Ashfield, R., Patel, A.J., Bossone, S.A., Brown, H., Campbell, R.D., Marcu, K.B., and Proudfoot, N.J. 1994. MAZ-dependent termination between

- closely spaced human complement genes. *The EMBO journal* **13**(23): 5656-5667.
- Azzouz, T.N., Gruber, A., and Schumperli, D. 2005. U7 snRNP-specific Lsm11 protein: dual binding contacts with the 100 kDa zinc finger processing factor (ZFP100) and a ZFP100-independent function in histone RNA 3' end processing. *Nucleic acids research* **33**(7): 2106-2117.
- Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhata, R. 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**(2): 265-276.
- Ballarino, M., Morlando, M., Pagano, F., Fatica, A., and Bozzoni, I. 2005. The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in *Saccharomyces cerevisiae*. *Molecular and cellular biology* **25**(13): 5396-5403.
- Ballarino, M., Pagano, F., Girardi, E., Morlando, M., Cacchiarelli, D., Marchioni, M., Proudfoot, N.J., and Bozzoni, I. 2009. Coupled RNA processing and transcription of intergenic primary microRNAs. *Molecular and cellular biology* **29**(20): 5632-5638.
- Bannister, A.J. and Kouzarides, T. 2011. Regulation of chromatin by histone modifications. *Cell research* **21**(3): 381-395.
- Barilla, D., Lee, B.A., and Proudfoot, N.J. 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **98**(2): 445-450.
- Barnes, W.M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proceedings of the National Academy of Sciences of the United States of America* **91**(6): 2216-2220.
- Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2): 215-233.
- Bartsch, I., Schoneberg, C., and Grummt, I. 1987. Evolutionary changes of sequences and factors that direct transcription termination of human and mouse ribosomal genes. *Molecular and cellular biology* **7**(7): 2521-2529.
- . 1988. Purification and characterization of TTFI, a factor that mediates termination of mouse ribosomal DNA transcription. *Molecular and cellular biology* **8**(9): 3891-3897.
- Baskaran, R., Dahmus, M.E., and Wang, J.Y. 1993. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America* **90**(23): 11167-11171.
- Bentley, D.L. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Current opinion in cell biology* **17**(3): 251-256.
- Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. 2004. Global nucleosome occupancy in yeast. *Genome biology* **5**(9): R62.
- Bernstein, J., Patterson, D.N., Wilson, G.M., and Toth, E.A. 2008. Characterization of the essential activities of *Saccharomyces cerevisiae* Mtr4p, a 3'->5' helicase partner of the nuclear exosome. *The Journal of biological chemistry* **283**(8): 4930-4942.
- Bernstein, P. and Ross, J. 1989. Poly(A), poly(A) binding protein and the regulation of mRNA stability. *Trends in biochemical sciences* **14**(9): 373-377.

- Betz, J.L., Chang, M., Washburn, T.M., Porter, S.E., Mueller, C.L., and Jaehning, J.A. 2002. Phenotypic analysis of Paf1/RNA polymerase II complex mutations reveals connections to cell cycle regulation, protein synthesis, and lipid and nucleic acid metabolism. *Mol Genet Genomics* **268**(2): 272-285.
- Beyer, K., Dandekar, T., and Keller, W. 1997. RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. *The Journal of biological chemistry* **272**(42): 26769-26779.
- Birchmeier, C., Grosschedl, R., and Birnstiel, M.L. 1982. Generation of authentic 3' termini of an H2A mRNA in vivo is dependent on a short inverted DNA repeat and on spacer sequences. *Cell* **28**(4): 739-745.
- Bird, G., Zorio, D.A., and Bentley, D.L. 2004. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Molecular and cellular biology* **24**(20): 8963-8969.
- Birnboim, H.C. and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic acids research* **7**(6): 1513-1523.
- Birse, C.E., Lee, B.A., Hansen, K., and Proudfoot, N.J. 1997. Transcriptional termination signals for RNA polymerase II in fission yeast. *The EMBO journal* **16**(12): 3633-3643.
- Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. 1998. Coupling termination of transcription to messenger RNA maturation in yeast. *Science (New York, NY)* **280**(5361): 298-301.
- Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. 2010. RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. *The Journal of biological chemistry* **285**(1): 188-196.
- Bogenhagen, D.F. and Brown, D.D. 1981. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. *Cell* **24**(1): 261-270.
- Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. 2000. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**(6): 765-775.
- Braglia, P., Heindl, K., Schleiffer, A., Martinez, J., and Proudfoot, N.J. 2010. Role of the RNA/DNA kinase Grc3 in transcription termination by RNA polymerase I. *EMBO reports* **11**(10): 758-764.
- Braglia, P., Kawauchi, J., and Proudfoot, N.J. 2011. Co-transcriptional RNA cleavage provides a failsafe termination mechanism for yeast RNA polymerase I. *Nucleic acids research* **39**(4): 1439-1448.
- Braglia, P., Percudani, R., and Dieci, G. 2005. Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III. *The Journal of biological chemistry* **280**(20): 19551-19562.
- Briggs, D., Jackson, D., Whitelaw, E., and Proudfoot, N.J. 1989. Direct demonstration of termination signals for RNA polymerase II from the sea urchin H2A histone gene. *Nucleic acids research* **17**(20): 8061-8071.
- Briggs, M.W., Burkard, K.T., and Butler, J.S. 1998. Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S

- rRNA 3' end formation. *The Journal of biological chemistry* **273**(21): 13255-13263.
- Brown, K.M. and Gilmartin, G.M. 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Molecular cell* **12**(6): 1467-1476.
- Bucheli, M.E. and Buratowski, S. 2005. Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. *The EMBO journal* **24**(12): 2150-2160.
- Bucheli, M.E., He, X., Kaplan, C.D., Moore, C.L., and Buratowski, S. 2007. Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. *RNA (New York, NY)* **13**(10): 1756-1764.
- Buratowski, S. 2003. The CTD code. *Nature structural biology* **10**(9): 679-680.
- . 2005. Connections between mRNA 3' end processing and transcription termination. *Current opinion in cell biology* **17**(3): 257-261.
- . 2009. Progression through the RNA polymerase II CTD cycle. *Molecular cell* **36**(4): 541-546.
- Cairns, B.R. 2009. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**(7261): 193-198.
- Callebaut, I., Moshous, D., Mornon, J.P., and de Villartay, J.P. 2002. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic acids research* **30**(16): 3592-3601.
- Calvo, O. and Manley, J.L. 2001. Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Molecular cell* **7**(5): 1013-1023.
- Camblong, J., Iglesias, N., Fickentscher, C., Dieppo, G., and Stutz, F. 2007. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* **131**(4): 706-717.
- Campbell, F.E., Jr. and Setzer, D.R. 1992. Transcription termination by RNA polymerase III: uncoupling of polymerase release from termination signal recognition. *Molecular and cellular biology* **12**(5): 2260-2272.
- Carrington, J.C. and Ambros, V. 2003. Role of microRNAs in plant and animal development. *Science (New York, NY)* **301**(5631): 336-338.
- Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. 2007. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA (New York, NY)* **13**(3): 361-373.
- Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D., and Corden, J.L. 2004. Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Molecular and cellular biology* **24**(14): 6241-6252.
- Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A., and Proudfoot, N. 2004. Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Molecular and cellular biology* **24**(10): 4174-4183.
- Chanfreau, G., Elela, S.A., Ares, M., Jr., and Guthrie, C. 1997. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes & development* **11**(20): 2741-2751.
- Chanfreau, G., Legrain, P., and Jacquier, A. 1998. Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *Journal of molecular biology* **284**(4): 975-988.

- Chao, L.C., Jamil, A., Kim, S.J., Huang, L., and Martinson, H.G. 1999. Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. *Molecular and cellular biology* **19**(8): 5588-5600.
- Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. 2007. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science (New York, NY)* **318**(5857): 1780-1782.
- Chekanova, J.A., Shaw, R.J., Wills, M.A., and Belostotsky, D.A. 2000. Poly(A) tail-dependent exonuclease AtRrp41p from *Arabidopsis thaliana* rescues 5.8 S rRNA processing and mRNA decay defects of the yeast ski6 mutant and is found in an exosome-sized complex in plant and yeast cells. *The Journal of biological chemistry* **275**(42): 33158-33166.
- Chen, C.Y., Gherzi, R., Ong, S.E., Chan, E.L., Raijmakers, R., Pruijn, G.J., Stoecklin, G., Moroni, C., Mann, M., and Karin, M. 2001. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* **107**(4): 451-464.
- Cheng, H., He, X., and Moore, C. 2004. The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. *Molecular and cellular biology* **24**(7): 2932-2943.
- Cheung, A.C. and Cramer, P. 2011. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**(7337): 249-253.
- Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes & development* **15**(24): 3319-3329.
- Cho, H., Kim, T.K., Mancebo, H., Lane, W.S., Flores, O., and Reinberg, D. 1999. A protein phosphatase functions to recycle RNA polymerase II. *Genes & development* **13**(12): 1540-1552.
- Chodchoy, N., Pandey, N.B., and Marzluff, W.F. 1991. An intact histone 3'-processing site is required for transcription termination in a mouse histone H2a gene. *Molecular and cellular biology* **11**(1): 497-509.
- Christofori, G. and Keller, W. 1988. 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP. *Cell* **54**(6): 875-889.
- Ciampi, M.S. 2006. Rho-dependent terminators and transcription termination. *Microbiology (Reading, England)* **152**(Pt 9): 2515-2528.
- Colgan, D.F. and Manley, J.L. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes & development* **11**(21): 2755-2766.
- Connelly, S. and Manley, J.L. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & development* **2**(4): 440-452.
- . 1989a. A CCAAT box sequence in the adenovirus major late promoter functions as part of an RNA polymerase II termination signal. *Cell* **57**(4): 561-571.
- . 1989b. RNA polymerase II transcription termination is mediated specifically by protein binding to a CCAAT box sequence. *Molecular and cellular biology* **9**(11): 5254-5259.

- Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. 2000. A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **154**(2): 557-571.
- Costa, P.J. and Arndt, K.M. 2000. Synthetic lethal interactions suggest a role for the *Saccharomyces cerevisiae* Rtf1 protein in transcription elongation. *Genetics* **156**(2): 535-547.
- Cozzarelli, N.R., Gerrard, S.P., Schlissel, M., Brown, D.D., and Bogenhagen, D.F. 1983. Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. *Cell* **34**(3): 829-835.
- Cramer, P. 2002. Multisubunit RNA polymerases. *Current opinion in structural biology* **12**(1): 89-97.
- Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J. et al. 2008. Structure of eukaryotic RNA polymerases. *Annual review of biophysics* **37**: 337-352.
- Cramer, P., Srebrow, A., Kadener, S., Werbajh, S., de la Mata, M., Melen, G., Nogues, G., and Kornblihtt, A.R. 2001. Coordination between transcription and pre-mRNA processing. *FEBS letters* **498**(2-3): 179-182.
- Cuello, P., Boyd, D.C., Dye, M.J., Proudfoot, N.J., and Murphy, S. 1999. Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *The EMBO journal* **18**(10): 2867-2877.
- Cui, M., Allen, M.A., Larsen, A., Macmorris, M., Han, M., and Blumenthal, T. 2008. Genes involved in pre-mRNA 3'-end formation and transcription termination revealed by a *lin-15* operon Muv suppressor screen. *Proceedings of the National Academy of Sciences of the United States of America* **105**(43): 16665-16670.
- Cullen, B.R. 1993. Does HIV-1 Tat induce a change in viral initiation rights? *Cell* **73**(3): 417-420.
- de Vegvar, H.E., Lund, E., and Dahlberg, J.E. 1986. 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell* **47**(2): 259-266.
- de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., and Keller, W. 2000. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *The EMBO journal* **19**(21): 5895-5904.
- d'Aubenton Carafa, Y., Brody, E., and Thermes, C. 1990. Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *Journal of molecular biology* **216**(4): 835-858.
- Dermody, J.L., Dreyfuss, J.M., Villen, J., Ogundipe, B., Gygi, S.P., Park, P.J., Ponticelli, A.S., Moore, C.L., Buratowski, S., and Bucheli, M.E. 2008. Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation. *PloS one* **3**(9): e3273.
- Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W., and Barabino, S.M. 2004. Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *The Journal of biological chemistry* **279**(34): 35788-35797.
- Dheur, S., Voile, T.A., Voisin-Hakil, F., Minet, M., Schmitter, J.M., Lacroute, F., Wyers, F., and Minvielle-Sebastia, L. 2003. Pti1p and Ref2p found in

- association with the mRNA 3' end formation complex direct snoRNA maturation. *The EMBO journal* **22**(11): 2831-2840.
- Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S., and Keller, W. 2002. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *The EMBO journal* **21**(15): 4125-4135.
- Dieci, G. and Sentenac, A. 1996. Facilitated recycling pathway for RNA polymerase III. *Cell* **84**(2): 245-252.
- Dieci, G., Fiorino, G., Castelnovo, M., Teichmann, M., and Pagano, A. 2007. The expanding RNA polymerase III transcriptome. *Trends Genet* **23**(12): 614-622.
- Doma, M.K. and Parker, R. 2007. RNA quality control in eukaryotes. *Cell* **131**(4): 660-668.
- Dominski, Z., Erkmann, J.A., Yang, X., Sanchez, R., and Marzluff, W.F. 2002. A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing. *Genes & development* **16**(1): 58-71.
- Dominski, Z. and Marzluff, W.F. 1999. Formation of the 3' end of histone mRNA. *Gene* **239**(1): 1-14.
- . 2007. Formation of the 3' end of histone mRNA: getting closer to the end. *Gene* **396**(2): 373-390.
- Dominski, Z., Yang, X.C., and Marzluff, W.F. 2005. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123**(1): 37-48.
- Dye, M. 1999. Studies in transcriptional termination of Pol II genes. In. University of Oxford.
- Dye, M.J., Gromak, N., and Proudfoot, N.J. 2006. Exon tethering in transcription by RNA polymerase II. *Molecular cell* **21**(6): 849-859.
- Dye, M.J. and Proudfoot, N.J. 1999. Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Molecular cell* **3**(3): 371-378.
- . 2001. Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* **105**(5): 669-681.
- Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. 2007. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature structural & molecular biology* **14**(1): 15-22.
- Egecioglu, D.E., Henras, A.K., and Chanfreau, G.F. 2006. Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. *RNA (New York, NY)* **12**(1): 26-32.
- Egloff, S., Al-Rawaf, H., O'Reilly, D., and Murphy, S. 2009. Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. *Molecular and cellular biology* **29**(14): 4002-4013.
- Egloff, S. and Murphy, S. 2008. Cracking the RNA polymerase II CTD code. *Trends Genet* **24**(6): 280-288.
- Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. 2007. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science (New York, NY)* **318**(5857): 1777-1779.

- Egloff, S., O'Reilly, D., and Murphy, S. 2008. Expression of human snRNA genes from beginning to end. *Biochemical Society transactions* **36**(Pt 4): 590-594.
- Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and Murphy, S. 2010. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *The Journal of biological chemistry* **285**(27): 20564-20569.
- Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., and Murphy, S. 2011. Ser7 Phosphorylation of the CTD Recruits the RPAP2 Ser5 Phosphatase to snRNA Genes. *Molecular cell*.
- El Hage, A., Koper, M., Kufel, J., and Tollervey, D. 2008. Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes & development* **22**(8): 1069-1081.
- Elela, S.A., Igel, H., and Ares, M., Jr. 1996. RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell* **85**(1): 115-124.
- Enriquez-Harris, P., Levitt, N., Briggs, D., and Proudfoot, N.J. 1991. A pause site for RNA polymerase II is associated with termination of transcription. *The EMBO journal* **10**(7): 1833-1842.
- Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S., and Nudler, E. 2007. An allosteric path to transcription termination. *Molecular cell* **28**(6): 991-1001.
- Epshtein, V., Dutta, D., Wade, J., and Nudler, E. 2010. An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463**(7278): 245-249.
- Evers, R. and Grummt, I. 1995. Molecular coevolution of mammalian ribosomal gene terminator sequences and the transcription termination factor TTF-I. *Proceedings of the National Academy of Sciences of the United States of America* **92**(13): 5827-5831.
- Evers, R., Smid, A., Rudloff, U., Lottspeich, F., and Grummt, I. 1995. Different domains of the murine RNA polymerase I-specific termination factor mTTF-I serve distinct functions in transcription termination. *The EMBO journal* **14**(6): 1248-1256.
- Fairley, J.A., Kantidakis, T., Kenneth, N.S., Intine, R.V., Maraia, R.J., and White, R.J. 2005. Human La is found at RNA polymerase III-transcribed genes in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **102**(51): 18350-18355.
- Fatica, A., Morlando, M., and Bozzoni, I. 2000. Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *The EMBO journal* **19**(22): 6218-6229.
- Ferrari, R., Rivetti, C., Acker, J., and Dieci, G. 2004. Distinct roles of transcription factors TFIIB and TFIIC in RNA polymerase III transcription reinitiation. *Proceedings of the National Academy of Sciences of the United States of America* **101**(37): 13442-13447.
- Flaherty, S.M., Fortes, P., Izaurralde, E., Mattaj, I.W., and Gilmartin, G.M. 1997. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proceedings of the National Academy of Sciences of the United States of America* **94**(22): 11893-11898.
- Frazao, C., McVey, C.E., Amblar, M., Barbas, A., Vonnrhein, C., Arraiano, C.M., and Carrondo, M.A. 2006. Unravelling the dynamics of RNA

- degradation by ribonuclease II and its RNA-bound complex. *Nature* **443**(7107): 110-114.
- French, S.L., Osheim, Y.N., Schneider, D.A., Sikes, M.L., Fernandez, C.F., Copela, L.A., Misra, V.A., Nomura, M., Wolin, S.L., and Beyer, A.L. 2008. Visual analysis of the yeast 5S rRNA gene transcriptome: regulation and role of La protein. *Molecular and cellular biology* **28**(14): 4576-4587.
- Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C., and Faye, G. 2003. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *The EMBO journal* **22**(7): 1588-1598.
- Garas, M., Dichtl, B., and Keller, W. 2008. The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *RNA (New York, NY)* **14**(12): 2671-2684.
- Garcia-Rubio, M., Chavez, S., Huertas, P., Tous, C., Jimeno, S., Luna, R., and Aguilera, A. 2008. Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. *Mol Genet Genomics* **279**(2): 123-132.
- Gerber, H.P., Hagmann, M., Seipel, K., Georgiev, O., West, M.A., Litingtung, Y., Schaffner, W., and Corden, J.L. 1995. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature* **374**(6523): 660-662.
- Ghazal, G., Gagnon, J., Jacques, P.E., Landry, J.R., Robert, F., and Elela, S.A. 2009. Yeast RNase III triggers polyadenylation-independent transcription termination. *Molecular cell* **36**(1): 99-109.
- Gick, O., Kramer, A., Vasserot, A., and Birnstiel, M.L. 1987. Heat-labile regulatory factor is required for 3' processing of histone precursor mRNAs. *Proceedings of the National Academy of Sciences of the United States of America* **84**(24): 8937-8940.
- Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature structural & molecular biology* **15**(1): 71-78.
- Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. 2009. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Molecular and cellular biology* **29**(20): 5455-5464.
- Gottlieb, E. and Steitz, J.A. 1989. Function of the mammalian La protein: evidence for its action in transcription termination by RNA polymerase III. *The EMBO journal* **8**(3): 851-861.
- Greger, I.H., Demarchi, F., Giacca, M., and Proudfoot, N.J. 1998. Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic acids research* **26**(5): 1294-1301.
- Gregory, R.I., Chendrimada, T.P., and Shiekhattar, R. 2006. MicroRNA biogenesis: isolation and characterization of the microprocessor complex. *Methods in molecular biology (Clifton, NJ)* **342**: 33-47.
- Gromak, N., West, S., and Proudfoot, N.J. 2006. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Molecular and cellular biology* **26**(10): 3986-3996.

- Grummt, I., Maier, U., Ohrlein, A., Hassouna, N., and Bachellerie, J.P. 1985. Transcription of mouse rDNA terminates downstream of the 3' end of 28S RNA and involves interaction of factors with repeated sequences in the 3' spacer. *Cell* **43**(3 Pt 2): 801-810.
- Grzechnik, P. and Kufel, J. 2008. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Molecular cell* **32**(2): 247-258.
- Gu, X. and Marzluff, W.F. 1996. 3' Processing and termination of mouse histone transcripts synthesized in vitro by RNA polymerase II. *Nucleic acids research* **24**(19): 3797-3805.
- Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. 2008. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nature structural & molecular biology* **15**(8): 786-794.
- Gullerova, M. and Proudfoot, N.J. 2008. Cohesin complex promotes transcriptional termination between convergent genes in *S. pombe*. *Cell* **132**(6): 983-995.
- Gunnery, S., Ma, Y., and Mathews, M.B. 1999. Termination sequence requirements vary among genes transcribed by RNA polymerase III. *Journal of molecular biology* **286**(3): 745-757.
- Guo, H. and Price, D.H. 1993. Mechanism of DmS-II-mediated pause suppression by Drosophila RNA polymerase II. *The Journal of biological chemistry* **268**(25): 18762-18770.
- Gusarov, I. and Nudler, E. 1999. The mechanism of intrinsic transcription termination. *Molecular cell* **3**(4): 495-504.
- Ha, K.S., Touloukhonov, I., Vassilyev, D.G., and Landick, R. 2010. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *Journal of molecular biology* **401**(5): 708-725.
- Hallegger, M., Llorian, M., and Smith, C.W. 2010. Alternative splicing: global insights. *The FEBS journal* **277**(4): 856-866.
- Hamada, M., Sakulich, A.L., Koduru, S.B., and Maraia, R.J. 2000. Transcription termination by RNA polymerase III in fission yeast. A genetic and biochemically tractable model system. *The Journal of biological chemistry* **275**(37): 29076-29081.
- Hammond, S.M. 2005. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS letters* **579**(26): 5822-5829.
- Haracska, L., Johnson, R.E., Prakash, L., and Prakash, S. 2005. Trf4 and Trf5 proteins of *Saccharomyces cerevisiae* exhibit poly(A) RNA polymerase activity but no DNA polymerase activity. *Molecular and cellular biology* **25**(22): 10183-10189.
- Hausmann, S. and Shuman, S. 2002. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *The Journal of biological chemistry* **277**(24): 21213-21220.
- Hawryluk, P.J., Ujvari, A., and Luse, D.S. 2004. Characterization of a novel RNA polymerase II arrest site which lacks a weak 3' RNA-DNA hybrid. *Nucleic acids research* **32**(6): 1904-1916.
- Henras, A.K., Bertrand, E., and Chanfreau, G. 2004. A cotranscriptional model for 3'-end processing of the *Saccharomyces cerevisiae* pre-ribosomal RNA precursor. *RNA (New York, NY)* **10**(10): 1572-1585.

- Hernandez, N. 1985. Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *The EMBO journal* **4**(7): 1827-1837.
- Hilleren, P., McCarthy, T., Rosbash, M., Parker, R., and Jensen, T.H. 2001. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* **413**(6855): 538-542.
- Hirose, Y. and Manley, J.L. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**(6697): 93-96.
- Hofmann, I., Schnolzer, M., Kaufmann, I., and Franke, W.W. 2002. Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in *Xenopus laevis* oocytes. *Molecular biology of the cell* **13**(5): 1665-1676.
- Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., and Tollervey, D. 2006. Microarray detection of novel nuclear RNA substrates for the exosome. *Yeast (Chichester, England)* **23**(6): 439-454.
- Houseley, J., Kotovic, K., El Hage, A., and Tollervey, D. 2007. Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. *The EMBO journal* **26**(24): 4996-5006.
- Houseley, J., LaCava, J., and Tollervey, D. 2006. RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* **7**(7): 529-539.
- Houseley, J. and Tollervey, D. 2006. Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO reports* **7**(2): 205-211.
- Hsin, J.P., Sheth, A., and Manley, J.L. 2011. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science (New York, NY)* **334**(6056): 683-686.
- Hu, J., Lutz, C.S., Wilusz, J., and Tian, B. 2005. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA (New York, NY)* **11**(10): 1485-1493.
- Huang, Y. and Carmichael, G.G. 1996. Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Molecular and cellular biology* **16**(4): 1534-1542.
- Ioshikhes, I.P., Albert, I., Zanton, S.J., and Pugh, B.F. 2006. Nucleosome positions predicted through comparative genomics. *Nature genetics* **38**(10): 1210-1215.
- Izban, M.G. and Luse, D.S. 1993. The increment of SII-facilitated transcript cleavage varies dramatically between elongation competent and incompetent RNA polymerase II ternary complexes. *The Journal of biological chemistry* **268**(17): 12874-12885.
- Jacobs, E.Y., Ogiwara, I., and Weiner, A.M. 2004. Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. *Molecular and cellular biology* **24**(2): 846-855.
- Jaeger, S., Martin, F., Rudinger-Thirion, J., Giege, R., and Eriani, G. 2006. Binding of human SLBP on the 3'-UTR of histone precursor H4-12 mRNA induces structural rearrangements that enable U7 snRNA anchoring. *Nucleic acids research* **34**(17): 4987-4995.
- Jansa, P. and Grummt, I. 1999. Mechanism of transcription termination: PTRF interacts with the largest subunit of RNA polymerase I and dissociates paused transcription complexes from yeast and mouse. *Mol Gen Genet* **262**(3): 508-514.

- Jansa, P., Mason, S.W., Hoffmann-Rohrer, U., and Grummt, I. 1998. Cloning and functional characterization of PTRF, a novel protein which induces dissociation of paused ternary transcription complexes. *The EMBO journal* **17**(10): 2855-2864.
- Jimeno-Gonzalez, S., Haaning, L.L., Malagon, F., and Jensen, T.H. 2010. The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II. *Molecular cell* **37**(4): 580-587.
- Jinek, M. and Doudna, J.A. 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**(7228): 405-412.
- Johnson, M.R., Norman, C., Reeve, M.A., Scully, J., and Proudfoot, N.J. 1986. Tripartite sequences within and 3' to the sea urchin H2A histone gene display properties associated with a transcriptional termination process. *Molecular and cellular biology* **6**(11): 4008-4018.
- Ju, Q.D., Morrow, B.E., and Warner, J.R. 1990. REB1, a yeast DNA-binding protein with many targets, is essential for growth and bears some resemblance to the oncogene myb. *Molecular and cellular biology* **10**(10): 5226-5234.
- Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. 2007. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. *Genes & development* **21**(14): 1779-1789.
- Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J. et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**(7236): 362-366.
- Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., and Gingeras, T.R. 2005. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome research* **15**(7): 987-997.
- Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *The EMBO journal* **23**(3): 616-626.
- Kawauchi, J., Mischo, H., Braglia, P., Rondon, A., and Proudfoot, N.J. 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes & development* **22**(8): 1082-1092.
- Keller, W., Bienroth, S., Lang, K.M., and Christofori, G. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *The EMBO journal* **10**(13): 4241-4249.
- Kelly, W.G., Dahmus, M.E., and Hart, G.W. 1993. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *The Journal of biological chemistry* **268**(14): 10416-10424.
- Kerwitz, Y., Kuhn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E., and Wahle, E. 2003. Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *The EMBO journal* **22**(14): 3705-3714.
- Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature structural & molecular biology* **17**(10): 1279-1286.

- Kim, H.D., Choe, J., and Seo, Y.S. 1999. The *sen1(+)* gene of *Schizosaccharomyces pombe*, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. *Biochemistry* **38**(44): 14697-14710.
- Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. 2004a. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *The EMBO journal* **23**(2): 354-364.
- Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedeia, E., Greenblatt, J.F., and Buratowski, S. 2004b. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**(7016): 517-522.
- Kim, M., Suh, H., Cho, E.J., and Buratowski, S. 2009. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *The Journal of biological chemistry* **284**(39): 26421-26426.
- Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. 2006. Distinct pathways for snoRNA and mRNA termination. *Molecular cell* **24**(5): 723-734.
- Kim, Y.K. and Kim, V.N. 2007. Processing of intronic microRNAs. *The EMBO journal* **26**(3): 775-783.
- Kireeva, M.L., Komissarova, N., and Kashlev, M. 2000. Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *Journal of molecular biology* **299**(2): 325-335.
- Kiss, T. 2002. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**(2): 145-148.
- Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and cellular biology* **25**(8): 3305-3316.
- Kolev, N.G. and Steitz, J.A. 2005. Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. *Genes & development* **19**(21): 2583-2592.
- Kolev, N.G., Yario, T.A., Benson, E., and Steitz, J.A. 2008. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. *EMBO reports* **9**(10): 1013-1018.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & development* **14**(19): 2452-2460.
- Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. 2002. Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Molecular cell* **10**(5): 1151-1162.
- Kong, S.E., Kobor, M.S., Krogan, N.J., Somesh, B.P., Sogaard, T.M., Greenblatt, J.F., and Svejstrup, J.Q. 2005. Interaction of Fcp1 phosphatase with elongating RNA polymerase II holoenzyme, enzymatic mechanism of action, and genetic interaction with elongator. *The Journal of biological chemistry* **280**(6): 4299-4306.
- Kornberg, R.D. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science (New York, NY)* **184**(139): 868-871.

- Kornberg, R.D. and Lorch, Y. 1999. Chromatin-modifying and -remodeling complexes. *Current opinion in genetics & development* **9**(2): 148-151.
- Krishnamurthy, S., Ghazy, M.A., Moore, C., and Hampsey, M. 2009. Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. *Molecular and cellular biology* **29**(11): 2925-2934.
- Kuehner, J.N., Pearson, E.L., and Moore, C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* **12**(5): 283-294.
- Kufel, J., Dichtl, B., and Tollervy, D. 1999. Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA (New York, NY)* **5**(7): 909-917.
- Kuhn, A. and Grummt, I. 1989. 3'-end formation of mouse pre-rRNA involves both transcription termination and a specific processing reaction. *Genes & development* **3**(2): 224-231.
- Kuhn, A., Normann, A., Bartsch, I., and Grummt, I. 1988. The mouse ribosomal gene terminator consists of three functionally separable sequence elements. *The EMBO journal* **7**(5): 1497-1502.
- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. 2007. Functional architecture of RNA polymerase I. *Cell* **131**(7): 1260-1272.
- Kuhn, U., Gundel, M., Knoth, A., Kerwitz, Y., Rudel, S., and Wahle, E. 2009. Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *The Journal of biological chemistry* **284**(34): 22803-22814.
- Kyburz, A., Friedlein, A., Langen, H., and Keller, W. 2006. Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Molecular cell* **23**(2): 195-205.
- Kyburz, A., Sadowski, M., Dichtl, B., and Keller, W. 2003. The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. *Nucleic acids research* **31**(14): 3936-3945.
- Labhart, P. and Reeder, R.H. 1986. Characterization of three sites of RNA 3' end formation in the *Xenopus* ribosomal gene spacer. *Cell* **45**(3): 431-443.
- . 1987. A 12-base-pair sequence is an essential element of the ribosomal gene terminator in *Xenopus laevis*. *Molecular and cellular biology* **7**(5): 1900-1905.
- LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervy, D. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**(5): 713-724.
- Landick, R. and Yanofsky, C. 1987. Isolation and structural analysis of the *Escherichia coli* trp leader paused transcription complex. *Journal of molecular biology* **196**(2): 363-377.
- Lang, W.H., Morrow, B.E., Ju, Q., Warner, J.R., and Reeder, R.H. 1994. A model for transcription termination by RNA polymerase I. *Cell* **79**(3): 527-534.
- Lang, W.H. and Reeder, R.H. 1995. Transcription termination of RNA polymerase I due to a T-rich element interacting with Reb1p. *Proceedings of the National Academy of Sciences of the United States of America* **92**(21): 9781-9785.

- Lang, W.H., Platt, T., and Reeder, R.H. 1998. Escherichia coli rho factor induces release of yeast RNA polymerase II but not polymerase I or III. *Proceedings of the National Academy of Sciences of the United States of America* **95**(9): 4900-4905.
- Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science (New York, NY)* **294**(5543): 858-862.
- Lebreton, A. and Seraphin, B. 2008. Exosome-mediated quality control: substrate recruitment and molecular activity. *Biochimica et biophysica acta* **1779**(9): 558-565.
- Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics* **39**(10): 1235-1244.
- Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. 2004. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* **23**(20): 4051-4060.
- Lejeune, F., Li, X., and Maquat, L.E. 2003. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Molecular cell* **12**(3): 675-687.
- Lian, Z., Karpikov, A., Lian, J., Mahajan, M.C., Hartman, S., Gerstein, M., Snyder, M., and Weissman, S.M. 2008. A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation. *Genome research* **18**(8): 1224-1237.
- Liang, S. and Lutz, C.S. 2006. p54nrb is a component of the snRNP-free U1A (SF-A) complex that promotes pre-mRNA cleavage during polyadenylation. *RNA (New York, NY)* **12**(1): 111-121.
- Libri, D., Dower, K., Boulay, J., Thomsen, R., Rosbash, M., and Jensen, T.H. 2002. Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Molecular and cellular biology* **22**(23): 8254-8266.
- Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Molecular cell* **9**(5): 1101-1111.
- Lin, P.S., Marshall, N.F., and Dahmus, M.E. 2002. CTD phosphatase: role in RNA polymerase II cycling and the regulation of transcript elongation. *Progress in nucleic acid research and molecular biology* **72**: 333-365.
- Lin-Marq, N. and Clarkson, S.G. 1998. Efficient synthesis, termination and release of RNA polymerase III transcripts in *Xenopus* extracts depleted of La protein. *The EMBO journal* **17**(7): 2033-2041.
- Liu, Q., Greimann, J.C., and Lima, C.D. 2006. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**(6): 1223-1237.
- Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S. 2009. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Molecular and cellular biology* **29**(17): 4852-4863.
- Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr., and Shenk, T. 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse

- beta maj-globin gene. *Proceedings of the National Academy of Sciences of the United States of America* **84**(23): 8306-8310.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**(6648): 251-260.
- Lund, E. and Dahlberg, J.E. 2006. Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. *Cold Spring Harbor symposia on quantitative biology* **71**: 59-66.
- Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhardt, A., and Varani, G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nature structural & molecular biology* **17**(10): 1195-1201.
- Luo, W., Johnson, A.W., and Bentley, D.L. 2006. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes & development* **20**(8): 954-965.
- Maciolek, N.L. and McNally, M.T. 2007. Serine/arginine-rich proteins contribute to negative regulator of splicing element-stimulated polyadenylation in rous sarcoma virus. *Journal of virology* **81**(20): 11208-11217.
- MacRae, I.J. and Doudna, J.A. 2007. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Current opinion in structural biology* **17**(1): 138-145.
- Mandel, C.R., Bai, Y., and Tong, L. 2008. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* **65**(7-8): 1099-1122.
- Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**(7121): 953-956.
- Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E., and Jensen, T.H. 2010. Crosstalk between mRNA 3' end processing and transcription initiation. *Molecular cell* **40**(3): 410-422.
- Maraia, R.J. and Intine, R.V. 2001. Recognition of nascent RNA by the human La antigen: conserved and divergent features of structure and function. *Molecular and cellular biology* **21**(2): 367-379.
- Maraia, R.J., Kenan, D.J., and Keene, J.D. 1994. Eukaryotic transcription termination factor La mediates transcript release and facilitates reinitiation by RNA polymerase III. *Molecular and cellular biology* **14**(3): 2147-2158.
- Marzluff, W.F., Wagner, E.J., and Duronio, R.J. 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature reviews* **9**(11): 843-854.
- Mason, S.W., Sander, E.E., and Grummt, I. 1997. Identification of a transcript release activity acting on ternary transcription complexes containing murine RNA polymerase I. *The EMBO journal* **16**(1): 163-172.
- Matlin, A.J. and Moore, M.J. 2007. Spliceosome assembly and composition. *Advances in experimental medicine and biology* **623**: 14-35.
- Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C. et al. 2008. Nucleosome organization in the Drosophila genome. *Nature* **453**(7193): 358-362.

- Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nature structural & molecular biology* **17**(10): 1272-1278.
- Mazabraud, A., Scherly, D., Muller, F., Rungger, D., and Clarkson, S.G. 1987. Structure and transcription termination of a lysine tRNA gene from *Xenopus laevis*. *Journal of molecular biology* **195**(4): 835-845.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**(6614): 357-361.
- Medlin, J., Scurry, A., Taylor, A., Zhang, F., Peterlin, B.M., and Murphy, S. 2005. P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *The EMBO journal* **24**(23): 4154-4165.
- Medlin, J.E., Uguen, P., Taylor, A., Bentley, D.L., and Murphy, S. 2003. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *The EMBO journal* **22**(4): 925-934.
- Meinhart, A. and Cramer, P. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**(6996): 223-226.
- Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S., and Cramer, P. 2005. A structural perspective of CTD function. *Genes & development* **19**(12): 1401-1415.
- Minvielle-Sebastia, L., Preker, P.J., Wiederkehr, T., Strahm, Y., and Keller, W. 1997. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. *Proceedings of the National Academy of Sciences of the United States of America* **94**(15): 7897-7902.
- Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular cell* **41**(1): 21-32.
- Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. 1997. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell* **91**(4): 457-466.
- Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L., and Proudfoot, N.J. 1998. The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes & development* **12**(16): 2522-2534.
- Morillon, A., O'Sullivan, J., Azad, A., Proudfoot, N., and Mellor, J. 2003. Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. *Science (New York, NY)* **300**(5618): 492-495.
- Morlando, M., Ballarino, M., Greco, P., Caffarelli, E., Dichtl, B., and Bozzoni, I. 2004. Coupling between snoRNP assembly and 3' processing controls box C/D snoRNA biosynthesis in yeast. *The EMBO journal* **23**(12): 2392-2401.
- Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. 2008. Primary microRNA transcripts are processed co-transcriptionally. *Nature structural & molecular biology* **15**(9): 902-909.
- Morlando, M., Greco, P., Dichtl, B., Fatica, A., Keller, W., and Bozzoni, I. 2002. Functional analysis of yeast snoRNA and snRNA 3'-end formation

- mediated by uncoupling of cleavage and polyadenylation. *Molecular and cellular biology* **22**(5): 1379-1389.
- Mukherjee, D., Gao, M., O'Connor, J.P., Raijmakers, R., Pruijn, G., Lutz, C.S., and Wilusz, J. 2002. The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *The EMBO journal* **21**(1-2): 165-174.
- Murthy, K.G. and Manley, J.L. 1992. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *The Journal of biological chemistry* **267**(21): 14804-14811.
- . 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes & development* **9**(21): 2672-2683.
- Nabavi, S. and Nazar, R.N. 2010a. Cleavage-induced termination in U2 snRNA gene expression. *Biochemical and biophysical research communications* **393**(3): 461-465.
- . 2010b. Pac1 endonuclease and Dhp1p 5'→3' exonuclease are required for U3 snoRNA termination in *Schizosaccharomyces pombe*. *FEBS letters* **584**(15): 3436-3441.
- Nag, A., Narsinh, K., Kazerouninia, A., and Martinson, H.G. 2006. The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. *RNA (New York, NY)* **12**(8): 1534-1544.
- Nag, A., Narsinh, K., and Martinson, H.G. 2007. The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nature structural & molecular biology* **14**(7): 662-669.
- Narita, T., Yung, T.M., Yamamoto, J., Tsuboi, Y., Tanabe, H., Tanaka, K., Yamaguchi, Y., and Handa, H. 2007. NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Molecular cell* **26**(3): 349-365.
- Narlikar, G.J., Fan, H.Y., and Kingston, R.E. 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**(4): 475-487.
- Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L., and Greenblatt, J. 2003. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *The Journal of biological chemistry* **278**(35): 33000-33010.
- Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. 2008. The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Molecular cell* **29**(5): 577-587.
- Nelson, H.C., Finch, J.T., Luisi, B.F., and Klug, A. 1987. The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **330**(6145): 221-226.
- Ng, H.H., Robert, F., Young, R.A., and Struhl, K. 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* **11**(3): 709-719.

- Ni, Z., Schwartz, B.E., Werner, J., Suarez, J.R., and Lis, J.T. 2004. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Molecular cell* **13**(1): 55-65.
- Niwa, M. and Berget, S.M. 1991. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes & development* **5**(11): 2086-2095.
- Niwa, M., MacDonald, C.C., and Berget, S.M. 1992. Are vertebrate exons scanned during splice-site selection? *Nature* **360**(6401): 277-280.
- Orban, T.I., Izaurralde, E. 2005. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA (New York, NY)*.
- Orozco, I.J., Kim, S.J., and Martinson, H.G. 2002. The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. *The Journal of biological chemistry* **277**(45): 42899-42911.
- Paule, M.R. and White, R.J. 2000. Survey and summary: transcription by RNA polymerases I and III. *Nucleic acids research* **28**(6): 1283-1298.
- Peterlin, B.M. and Price, D.H. 2006. Controlling the elongation phase of transcription with P-TEFb. *Molecular cell* **23**(3): 297-305.
- Pillai, R.S., Grimmier, M., Meister, G., Will, C.L., Luhrmann, R., Fischer, U., and Schumperli, D. 2003. Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes & development* **17**(18): 2321-2333.
- Plant, K.E., Dye, M.J., Lafaille, C., and Proudfoot, N.J. 2005. Strong polyadenylation and weak pausing combine to cause efficient termination of transcription in the human Ggamma-globin gene. *Molecular and cellular biology* **25**(8): 3276-3285.
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science (New York, NY)* **322**(5909): 1851-1854.
- Prescott, E.M., Osheim, Y.N., Jones, H.S., Alen, C.M., Roan, J.G., Reeder, R.H., Beyer, A.L., and Proudfoot, N.J. 2004. Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proceedings of the National Academy of Sciences of the United States of America* **101**(16): 6068-6073.
- Proudfoot, N. 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Current opinion in cell biology* **16**(3): 272-278.
- Proudfoot, N.J. and Brownlee, G.G. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**(5574): 211-214.
- Proudfoot, N.J., Furger, A., and Dye, M.J. 2002. Integrating mRNA processing with transcription. *Cell* **108**(4): 501-512.
- Qiu, H., Hu, C., and Hinnebusch, A.G. 2009. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Molecular cell* **33**(6): 752-762.
- Racki, L.R. and Narlikar, G.J. 2008. ATP-dependent chromatin remodeling enzymes: two heads are not better, just different. *Current opinion in genetics & development* **18**(2): 137-144.

- Rasmussen, T.P. and Culbertson, M.R. 1998. The putative nucleic acid helicase Sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nucleolar RNAs in *Saccharomyces cerevisiae*. *Molecular and cellular biology* **18**(12): 6885-6896.
- Reeder, R.H. 1999. Regulation of RNA polymerase I transcription in yeast and vertebrates. *Progress in nucleic acid research and molecular biology* **62**: 293-327.
- Richard, P. and Kiss, T. 2006. Integrating snoRNP assembly with mRNA biogenesis. *EMBO reports* **7**(6): 590-592.
- Richard, P. and Manley, J.L. 2009. Transcription termination by nuclear RNA polymerases. *Genes & development* **23**(11): 1247-1269.
- Robertson, H.D., Webster, R.E., and Zinder, N.D. 1968. Purification and properties of ribonuclease III from *Escherichia coli*. *The Journal of biological chemistry* **243**(1): 82-91.
- Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., and Bradley, A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome research* **14**(10A): 1902-1910.
- Rondon, A.G., Gallardo, M., Garcia-Rubio, M., and Aguilera, A. 2004. Molecular evidence indicating that the yeast PAF complex is required for transcription elongation. *EMBO reports* **5**(1): 47-53.
- Rondon, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. 2009. Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Molecular cell* **36**(1): 88-98.
- Rosonina, E., Ip, J.Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J., and Blencowe, B.J. 2005. Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. *Molecular and cellular biology* **25**(15): 6734-6746.
- Rougemaille, M., Gudipati, R.K., Olesen, J.R., Thomsen, R., Seraphin, B., Libri, D., and Jensen, T.H. 2007. Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. *The EMBO journal* **26**(9): 2317-2326.
- Ruegsegger, U., Beyer, K., and Keller, W. 1996. Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *The Journal of biological chemistry* **271**(11): 6107-6113.
- Ryan, K., Calvo, O., and Manley, J.L. 2004. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA (New York, NY)* **10**(4): 565-573.
- Sachs, A.B., Sarnow, P., and Hentze, M.W. 1997. Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell* **89**(6): 831-838.
- Sadowski, M., Dichtl, B., Hubner, W., and Keller, W. 2003. Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *The EMBO journal* **22**(9): 2167-2177.
- Saguez, C., Schmid, M., Olesen, J.R., Ghazy, M.A., Qu, X., Poulsen, M.B., Nasser, T., Moore, C., and Jensen, T.H. 2008. Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. *Molecular cell* **31**(1): 91-103.
- Saini, H.K., Griffiths-Jones, S., and Enright, A.J. 2007. Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America* **104**(45): 17719-17724.

- Sambrook, J., Fritsch, E.F., Maniatis, T. 1989. *Molecular cloning - a laboratory manual*. Cold Spring Harbour Laboratory Press, Cold Spring Harbour. N.Y.
- Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M., and van Hoof, A. 2009. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nature structural & molecular biology* **16**(1): 56-62.
- Schilders, G., Raijmakers, R., Raats, J.M., and Pruijn, G.J. 2005. MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation. *Nucleic acids research* **33**(21): 6795-6804.
- Schneider, C., Leung, E., Brown, J., and Tollervey, D. 2009. The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic acids research* **37**(4): 1127-1140.
- Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes & development* **14**(19): 2435-2440.
- Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. 2006. A genomic code for nucleosome positioning. *Nature* **442**(7104): 772-778.
- Shankar, S., Hatoum, A., and Roberts, J.W. 2007. A transcription antiterminator constructs a NusA-dependent shield to the emerging transcript. *Molecular cell* **27**(6): 914-927.
- Sheets, M.D., Ogg, S.C., and Wickens, M.P. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic acids research* **18**(19): 5799-5805.
- Sheldon, K.E., Mauger, D.M., and Arndt, K.M. 2005. A Requirement for the *Saccharomyces cerevisiae* Paf1 complex in snoRNA 3' end formation. *Molecular cell* **20**(2): 225-236.
- Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Molecular cell* **33**(3): 365-376.
- Shuman, S. 2001. Structure, mechanism, and evolution of the mRNA capping apparatus. *Progress in nucleic acid research and molecular biology* **66**: 1-40.
- Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell* **42**(6): 794-805.
- Sopta, M., Carthew, R.W., and Greenblatt, J. 1985. Isolation of three proteins that bind to mammalian RNA polymerase II. *The Journal of biological chemistry* **260**(18): 10353-10360.
- Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. 2009. Biased chromatin signatures around polyadenylation sites and exons. *Molecular cell* **36**(2): 245-254.
- Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M., and Hartzog, G.A. 2002. The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *The EMBO journal* **21**(7): 1764-1774.

- Staals, R.H., Bronkhorst, A.W., Schilders, G., Slomovic, S., Schuster, G., Heck, A.J., Raijmakers, R., and Pruijn, G.J. 2010. Dis3-like 1: a novel exoribonuclease associated with the human exosome. *The EMBO journal* **29**(14): 2358-2367.
- Steinmetz, E.J. and Brow, D.A. 1996. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Molecular and cellular biology* **16**(12): 6993-7003.
- . 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. *Proceedings of the National Academy of Sciences of the United States of America* **95**(12): 6699-6704.
- . 2003. Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Molecular and cellular biology* **23**(18): 6339-6349.
- Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413**(6853): 327-331.
- Steinmetz, E.J., Ng, S.B., Cloute, J.P., and Brow, D.A. 2006. cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. *Molecular and cellular biology* **26**(7): 2688-2696.
- Stevens, A. and Poole, T.L. 1995. 5'-exonuclease-2 of *Saccharomyces cerevisiae*. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1. *The Journal of biological chemistry* **270**(27): 16063-16069.
- Suraweera, A., Becherel, O.J., Chen, P., Rundle, N., Woods, R., Nakamura, J., Gatei, M., Criscuolo, C., Filla, A., Chessa, L. et al. 2007. Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. *The Journal of cell biology* **177**(6): 969-979.
- Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., and Lavin, M.F. 2009. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Human molecular genetics* **18**(18): 3384-3396.
- Swinburne, I.A., Meyer, C.A., Liu, X.S., Silver, P.A., and Brodsky, A.S. 2006. Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. *Genome research* **16**(7): 912-921.
- Takagaki, Y. and Manley, J.L. 1997. RNA recognition by the human polyadenylation factor CstF. *Molecular and cellular biology* **17**(7): 3907-3914.
- . 2000. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Molecular and cellular biology* **20**(5): 1515-1525.
- Takagaki, Y., Ryner, L.C., and Manley, J.L. 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes & development* **3**(11): 1711-1724.
- Tantravahi, J., Alvira, M., and Falck-Pedersen, E. 1993. Characterization of the mouse beta maj globin transcription termination region: a spacing sequence is required between the poly(A) signal sequence and multiple downstream termination elements. *Molecular and cellular biology* **13**(1): 578-587.

- Tarun, S.Z., Jr. and Sachs, A.B. 1996. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *The EMBO journal* **15**(24): 7168-7177.
- Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N.J., and Akoulitchiev, A. 2004. Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* **432**(7016): 526-530.
- Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. 2006. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Molecular cell* **23**(6): 853-864.
- Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S. et al. 2010. Chemical-genomic dissection of the CTD code. *Nature structural & molecular biology* **17**(9): 1154-1161.
- Tomecki, R., Kristiansen, M.S., Lykke-Andersen, S., Chlebowski, A., Larsen, K.M., Szczesny, R.J., Drazkowska, K., Pastula, A., Andersen, J.S., Stepień, P.P. et al. 2010. The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *The EMBO journal* **29**(14): 2342-2357.
- Toulokhonov, I., Artsimovitch, I., and Landick, R. 2001. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science (New York, NY)* **292**(5517): 730-733.
- Toulokhonov, I., Zhang, J., Palangat, M., and Landick, R. 2007. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Molecular cell* **27**(3): 406-419.
- Uguen, P. and Murphy, S. 2003. The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *The EMBO journal* **22**(17): 4544-4554.
- Ursic, D., Chinchilla, K., Finkel, J.S., and Culbertson, M.R. 2004. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic acids research* **32**(8): 2441-2452.
- Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F., and Culbertson, M.R. 1997. The yeast SEN1 gene is required for the processing of diverse RNA classes. *Nucleic acids research* **25**(23): 4778-4785.
- Vagner, S., Vagner, C., and Mattaj, I.W. 2000. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes & development* **14**(4): 403-413.
- van Hoof, A., Lennertz, P., and Parker, R. 2000. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Molecular and cellular biology* **20**(2): 441-452.
- Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. 2005. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS biology* **3**(6): e189.
- Vasiljeva, L. and Buratowski, S. 2006. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Molecular cell* **21**(2): 239-248.

- Vasiljeva, L., Kim, M., Terzi, N., Soares, L.M., and Buratowski, S. 2008. Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. *Molecular cell* **29**(3): 313-323.
- Venkataraman, K., Brown, K.M., and Gilmartin, G.M. 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes & development* **19**(11): 1315-1327.
- Vignali, M., Hassan, A.H., Neely, K.E., and Workman, J.L. 2000. ATP-dependent chromatin-remodeling complexes. *Molecular and cellular biology* **20**(6): 1899-1910.
- Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B., and Mattaj, I.W. 1996. A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. *The Journal of cell biology* **133**(1): 5-14.
- Vodala, S., Abruzzi, K.C., and Rosbash, M. 2008. The nuclear exosome and adenylation regulate posttranscriptional tethering of yeast GAL genes to the nuclear periphery. *Molecular cell* **31**(1): 104-113.
- Volloch, V., Schweitzer, B., and Rits, S. 1996. Antisense globin RNA in mouse erythroid tissues: structure, origin, and possible function. *Proceedings of the National Academy of Sciences of the United States of America* **93**(6): 2476-2481.
- Wagner, E.J. and Marzluff, W.F. 2006. ZFP100, a component of the active U7 snRNP limiting for histone pre-mRNA processing, is required for entry into S phase. *Molecular and cellular biology* **26**(17): 6702-6712.
- Walther, T.N., Wittop Koning, T.H., Schumperli, D., and Muller, B. 1998. A 5'-3' exonuclease activity involved in forming the 3' products of histone pre-mRNA processing in vitro. *RNA (New York, NY)* **4**(9): 1034-1046.
- Wang, Z., Bai, L., Hsieh, Y.J., and Roeder, R.G. 2000. Nuclear factor 1 (NF1) affects accurate termination and multiple-round transcription by human RNA polymerase III. *The EMBO journal* **19**(24): 6823-6832.
- Wang, Z. and Roeder, R.G. 1996. TFIIC1 acts through a downstream region to stabilize TFIIC2 binding to RNA polymerase III promoters. *Molecular and cellular biology* **16**(12): 6841-6850.
- . 1998. DNA topoisomerase I and PC4 can interact with human TFIIC to promote both accurate termination and transcription reinitiation by RNA polymerase III. *Molecular cell* **1**(5): 749-757.
- Weitzer, S. and Martinez, J. 2007. hClp1: a novel kinase revitalizes RNA metabolism. *Cell cycle (Georgetown, Tex)* **6**(17): 2133-2137.
- West, S., Gromak, N., Norbury, C.J., and Proudfoot, N.J. 2006a. Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Molecular cell* **21**(3): 437-443.
- West, S., Gromak, N., and Proudfoot, N.J. 2004. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**(7016): 522-525.
- West, S. and Proudfoot, N.J. 2008. Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic acids research* **36**(3): 905-914.
- . 2009. Transcriptional termination enhances protein expression in human cells. *Molecular cell* **33**(3): 354-364.

- West, S., Proudfoot, N.J., and Dye, M.J. 2008. Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Molecular cell* **29**(5): 600-610.
- West, S., Zaret, K., and Proudfoot, N.J. 2006b. Transcriptional termination sequences in the mouse serum albumin gene. *RNA (New York, NY)* **12**(4): 655-665.
- Whitehall, S.K., Bardeleben, C., and Kassavetis, G.A. 1994. Hydrolytic cleavage of nascent RNA in RNA polymerase III ternary transcription complexes. *The Journal of biological chemistry* **269**(3): 2299-2306.
- Whitelaw, E. and Proudfoot, N. 1986. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *The EMBO journal* **5**(11): 2915-2922.
- Wood, V. Gwilliam, R. Rajandream, M.A. Lyne, M. Lyne, R. Stewart, A. Sgouros, J. Peat, N. Hayles, J. Baker, S. Basham, D. Bowman, S. Brooks, K. Brown, D. Brown, S. Chillingworth, T. Churcher, C. Collins, M. Connor, R. Cronin, A. Davis, P. Feltwell, T. Fraser, A. Gentles, S. Goble, A. Hamlin, N. Harris, D. Hidalgo, J. Hodgson, G. Holroyd, S. Hornsby, T. Howarth, S. Huckle, E.J. Hunt, S. Jagels, K. James, K. Jones, L. Jones, M. Leather, S. McDonald, S. McLean, J. Mooney, P. Moule, S. Mungall, K. Murphy, L. Niblett, D. Odell, C. Oliver, K. O'Neil, S. Pearson, D. Quail, M.A. Rabinowitsch, E. Rutherford, K. Rutter, S. Saunders, D. Seeger, K. Sharp, S. Skelton, J. Simmonds, M. Squares, R. Squares, S. Stevens, K. Taylor, K. Taylor, R.G. Tivey, A. Walsh, S. Warren, T. Whitehead, S. Woodward, J. Volckaert, G. Aert, R. Robben, J. Grymonprez, B. Weltjens, I. Vanstreels, E. Rieger, M. Schafer, M. Muller-Auer, S. Gabel, C. Fuchs, M. Dusterhoft, A. Fritzc, C. Holzer, E. Moestl, D. Hilbert, H. Borzym, K. Langer, I. Beck, A. Lehrach, H. Reinhardt, R. Pohl, T.M. Eger, P. Zimmermann, W. Wedler, H. Wambutt, R. Purnelle, B. Goffeau, A. Cadieu, E. Dreano, S. Gloux, S. Lelaure, V. Mottier, S. Galibert, F. Aves, S.J. Xiang, Z. Hunt, C. Moore, K. Hurst, S.M. Lucas, M. Rochet, M. Gaillardin, C. Tallada, V.A. Garzon, A. Thode, G. Daga, R.R. Cruzado, L. Jimenez, J. Sanchez, M. del Rey, F. Benito, J. Dominguez, A. Revuelta, J.L. Moreno, S. Armstrong, J. Forsburg, S.L. Cerutti, L. Lowe, T. McCombie, W.R. Paulsen, I. Potashkin, J. Shpakovski, G.V. Ussery, D. Barrell, B.G. and Nurse, P. 2002. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**(6874): 871-880.
- Wu, X., Wilcox, C.B., Devasahayam, G., Hackett, R.L., Arevalo-Rodriguez, M., Cardenas, M.E., Heitman, J., and Hanes, S.D. 2000. The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *The EMBO journal* **19**(14): 3727-3738.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B. et al. 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**(5): 725-737.
- Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M.M., Manley, J.L., and Tong, L. 2010. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**(7316): 729-733.

- Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. 2009. Structure and function of the 5'→3' exoribonuclease Rat1 and its activating partner Rai1. *Nature* **458**(7239): 784-788.
- Xu, Y.X., Hirose, Y., Zhou, X.Z., Lu, K.P., and Manley, J.L. 2003. Pin1 modulates the structure and function of human RNA polymerase II. *Genes & development* **17**(22): 2765-2776.
- Yang, P.K., Hoareau, C., Froment, C., Monsarrat, B., Henry, Y., and Chanfreau, G. 2005. Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. *Molecular and cellular biology* **25**(8): 3295-3304.
- Yang, X.C., Sullivan, K.D., Marzluff, W.F., and Dominski, Z. 2009. Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Molecular and cellular biology* **29**(1): 31-42.
- Yang, X.C., Xu, B., Sabath, I., Kunduru, L., Burch, B.D., Marzluff, W.F., and Dominski, Z. 2011. FLASH is required for the endonucleolytic cleavage of histone pre-mRNAs but is dispensable for the 5' exonucleolytic degradation of the downstream cleavage product. *Molecular and cellular biology* **31**(7): 1492-1502.
- Yonaha, M. and Proudfoot, N.J. 1999. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Molecular cell* **3**(5): 593-600.
- Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., and Corden, J.L. 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proceedings of the National Academy of Sciences of the United States of America* **93**(14): 6975-6980.
- Zhang, J. and Corden, J.L. 1991. Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *The Journal of biological chemistry* **266**(4): 2290-2296.
- Zhang, Z., Fu, J., and Gilmour, D.S. 2005. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes & development* **19**(13): 1572-1580.
- Zhang, Z. and Gilmour, D.S. 2006. Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Molecular cell* **21**(1): 65-74.
- Zhang, Z., Klatt, A., Henderson, A.J., and Gilmour, D.S. 2007. Transcription termination factor Pcf11 limits the processivity of Pol II on an HIV provirus to repress gene expression. *Genes & development* **21**(13): 1609-1614.
- Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. 2011. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science (New York, NY)* **332**(6032): 977-980.
- Zhang, Z., Wu, C.H., and Gilmour, D.S. 2004. Analysis of polymerase II elongation complexes by native gel electrophoresis. Evidence for a novel carboxyl-terminal domain-mediated termination mechanism. *The Journal of biological chemistry* **279**(22): 23223-23228.
- Zhao, J., Hyman, L., and Moore, C. 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63**(2): 405-445.

- Zhou, D., Frendewey, D., and Lobo Ruppert, S.M. 1999. Pac1p, an RNase III homolog, is required for formation of the 3' end of U2 snRNA in *Schizosaccharomyces pombe*. *RNA (New York, NY)* **5**(8): 1083-1098.
- Zhou, K., Kuo, W.H., Fillingham, J., and Greenblatt, J.F. 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proceedings of the National Academy of Sciences of the United States of America* **106**(17): 6956-6961.
- Zhu, B., Mandal, S.S., Pham, A.D., Zheng, Y., Erdjument-Bromage, H., Batra, S.K., Tempst, P., and Reinberg, D. 2005. The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes & development* **19**(14): 1668-1673.
- Zuo, Y., Vincent, H.A., Zhang, J., Wang, Y., Deutscher, M.P., and Malhotra, A. 2006. Structural basis for processivity and single-strand specificity of RNase II. *Molecular cell* **24**(1): 149-156.