RNA POLYMERASE II COLLISION & ITS ROLE IN TRANSCRIPT ELONGATION

David James Hobson

University College London

and

Cancer Research UK London Research Institute

PhD Supervisor: Jesper Q. Svejstrup

A thesis submitted for the degree of

Doctor of Philosophy

University College London

August 2012

Declaration

I, David James Hobson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Antisense non-coding transcripts, genes-within-genes, and convergent gene pairs are prevalent among eukaryotes. The existence of such transcription units raises the question of what happens when RNA polymerase II (RNAPII) molecules collide head-to-head. In this study a combination of biochemical and genetic approaches in yeast are used to show that polymerases transcribing opposite DNA strands cannot bypass each other. RNAPII stops, but does not dissociate upon head-to-head collision *in vitro*, suggesting that opposing polymerases represent insurmountable obstacles for each other. Head-to-head collision *in vivo* also results in RNAPII stopping, and removal of collided RNAPII from the DNA template can be achieved via ubiquitylation-directed proteolysis. Indeed, in cells lacking efficient RNAPII poly-ubiquitylation, the half-life of collided polymerases increases, so that they can be detected between convergent genes. These results provide new insight into fundamental mechanisms of gene traffic control, and point to an unexplored effect of antisense transcription on gene regulation via polymerase collision.

Acknowledgements

I would like to take this opportunity to offer my sincere gratitude to my PhD supervisor Jesper Svejstrup. Firstly, for providing me with an opportunity to perform this work in his laboratory and secondly for all the guidance he has provided through the many useful and interesting discussions we have had over the years.

Also, I thank my thesis committee: Vincenzo Costanzo and Mark Petronczki, for their aid and guidance.

I wish to thank both Wu Wei and Lars Steinmetz for being excellent collaborators and performing the bioinformatic analysis of our genome-wide studies.

I also want to thank all the members of the Svejstrup laboratory (past and present) for enlightening and lively scientific discussion, and for making Clare Hall such a pleasant place to work.

Furthermore, I would like to thank all the members of the public who have kindly and generously donated to Cancer Research UK, without whom our work would not be made possible.

Also, thank you to *Boehringer Ingelheim Fonds* for funding part of this PhD project.

Lastly, but certainly not least, a very big thank you to my wife, parents and family for their support, kind encouragement and constant belief.

Table of Contents

Table of Figures

List of tables

Abbreviations

bp – Base pair(s)

ChIP – Chromatin immuno-precipitation

ChIP-Seq – Chromatin immuno-precipitation followed by advanced sequencing

CID – CTD-interacting domain

CPD – Cyclobutane pyrimidine dimer

CTD – Carboxy-terminal domain (of Rpb1)

CUT – Cryptic unstable transcript

DMSO – Dimethyl sulfoxide

DRB - 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole

DTT – Dithiothreitol

DUB – De-ubiquitylation enzyme

E. coli – Escherichia coli

EC – Elongation complex

EDTA – Ethylenediaminetetraacetic acid

FL – Full-length (generally referring to RNA size)

Gal – Galactose

GG-NER – General genome nucleotide excision repair

Glc – Glucose

GTF – General transcription factor

HAT – Histone acetyltransferase

HDAC – Histone deacetylase

HMT – Histone methyltransferase

HRP – Horseradish peroxidase

IPTG - Isopropyl β-D-1-thiogalactopyranoside

kb – Kilo-bases

kDa – Kilo-Dalton

ncRNA – Non-coding RNA

NFR – Nucleosome-free region

NP-40 – Nonylphenol Polyethylene Glycol

nt – Nucleotide(s)

NTP – Nucleoside triphosphate

NTS – Non-transcribed strand

ORF – Open reading frame

PAGE – Polyacrylamide gel electrophoresis

PAS – Poly(A) signal

PCR – Polymerase chain reaction

PEG – Polyethylene glycol

PIC – Pre-initiation complex

PMSF – Phenylmethanesulfonylfluoride

qPCR – Quantitative polymerase chain reaction

RNAPII – RNA polymerase II

RRM – RNA recognition motif

S. cerevisiae – Saccharomyces cerevisiae

SDS – Sodium dodecyl-sulphate

snoRNA – Small nucleolar RNA

SUT – Stable un-annotated transcript

TBP – TATA-binding protein

TC-NER – Transcription-coupled nucleotide excision repair

TS – Transcribed strand

TSS – Transcription start site

UAS – Upstream activating sequence

WT – Wild-type (W303)

YPD – Yeast extract peptone dextrose

Chapter 1. Introduction

1.1 RNA Polymerase II transcription

The central dogma of molecular biology states that the genetic code stored in DNA is transcribed to mRNA, which provides a template for translation and protein synthesis. In eukaryotes, the protein responsible for transcribing DNA into mRNA is the multi-subunit complex RNA polymerase II (RNAPII). Transcription is a highly regulated and fundamental biological process, being an end-point of virtually all cell-signalling pathways. Additionally, deregulation of transcription is responsible for numerous diseases, including cancer.

Transcription can be divided into three main stages: initiation, elongation and termination. Initiation involves the binding of transcriptional activators to enhancers, recruitment of general transcription factors and loading of RNAPII onto promoter DNA. Following promoter escape, RNAPII enters into the elongation phase, during which RNA is produced, prior to termination and dissociation of polymerase from the DNA.

Extensive regulation occurs at all stages of RNAPII transcription and will be covered in more detail in the following sections. The main focus of this introduction will be on topics directly relevant to this thesis: RNAPII structure and function, transcript elongation and its associated factors, genome organisation and transcriptomics. Other aspects of transcription are discussed, albeit in less depth.

1.1.1 RNAPII structure and function

Knowledge of the structure and function of RNAPII is a pre-requisite for a comprehensive understanding of transcription, and particularly, transcript elongation.

RNAPII is a 12-subunit heteromeric complex, consisting of Rpb1-12, with a total mass in excess of 500 kDa. RNAPII is highly conserved between even distantly related eukaryotes, with a primary protein sequence identity of 53% and in excess of 80% overall homology (similarity) between *Saccharomyces cerevisiae* and *Homo sapiens* (Cramer et al., 2001). Remarkably, in complementation studies the expression of mammalian RNAPII subunits, Rpb5, 6, 7, 8, 9 and 10 were able to restore viability to yeast carrying genetic deletions of the counterpart genes (Shpakovski et al., 1995, McKune et al., 1995, Khazak et al., 1995).

RNAPII is a DNA-dependent RNA polymerase and so uses DNA as a template for the production of complementary RNA. Incoming nucleoside triphosphates (NTPs) base-pair to the DNA template strand and via an RNAPII-catalysed condensation reaction, the chain is elongated with a concomitant release of inorganic pyrophosphate and forward movement of RNAPII (Herbert et al., 2008). Details of this mechanism will be discussed later.

Although complete RNAPII contains 12 subunits, the population of cellular RNAPII is not homogeneous; it also exists as a 10-subunit complex lacking Rpb4 and Rpb7, which form an autonomous sub-complex. The association between Rpb4 and Rpb7 was confirmed by genetic manipulation in yeast, as *rpb4*Δ cells yielded only the 10-subunit RNAPII, lacking Rpb7 (Woychik and Young, 1989, Woychik et al., 1991). Dissociation of the Rpb4/7 complex is probably regulatory and may play a role in the stress-response. Indeed, *in vitro* assays have shown that the 10-subunit RNAPII is unable to initiate transcription at a promoter, but is catalytically active in a non-specific transcription assay, suggesting this is a mechanism by which RNAPII can be made initiation-incompetent (Edwards et al., 1991).

Early attempts at RNAPII crystallisation were marred by the heterogeneity of the complex, but upon preparation of highly purified RNAPII from *rpb4*Δ strains, EM structural topography was revealed at 16Å resolution (Darst et al., 1991), followed by a 5 Å X-ray structure (Fu et al., 1999). Improvements were made in the crystallisation process leading to increased resolution (Cramer et al., 2000) and ultimately to a 2.8 Å resolution model of free RNAPII (Cramer et al., 2001), as well as a model of an RNAPII elongation complex (EC) at 3.3 Å (Gnatt et al., 2001). The studies revealed the position of Rpb1 and Rpb2, which form a positively charged active centre cleft spanned by the bridge helix, with the smaller subunits distributed around the periphery (Figure 1.1A). The EC model showed that double-stranded DNA passes into the cleft between the upper jaw (composed of Rpb1 and Rpb9) and the lower jaw (Rpb5), with the transcription bubble and DNA-RNA hybrid positioned at the active site (Gnatt et al., 2001). DNA is bent to an angle of approximately 90° as it passes through RNAPII due to the presence of the Rpb2 protein wall blocking the back of the cleft.

During transcription, it is necessary that the transcription bubble of approximately 18 bp, containing an 8-9 nt RNA hybrid, is maintained (Saeki and Svejstrup, 2009, Wang et al., 2006, Kireeva et al., 2000). The structure implicates two features called “fork loops” in this process (Figure 1.1B). Fork loop 2 of the Rpb2 subunit is responsible for separating the DNA prior to it reaching the bridge helix, whilst fork loop 1 of Rpb1 interacts with, and stabilises, the DNA-RNA hybrid. Upstream of the hybrid the DNA rewinds, but the energetically favourable DNA-RNA hybrid has to be disrupted to allow this. This is achieved by the rudder and lid of RNAPII, with the rudder contacting and stabilising the DNA whilst a phenylalanine

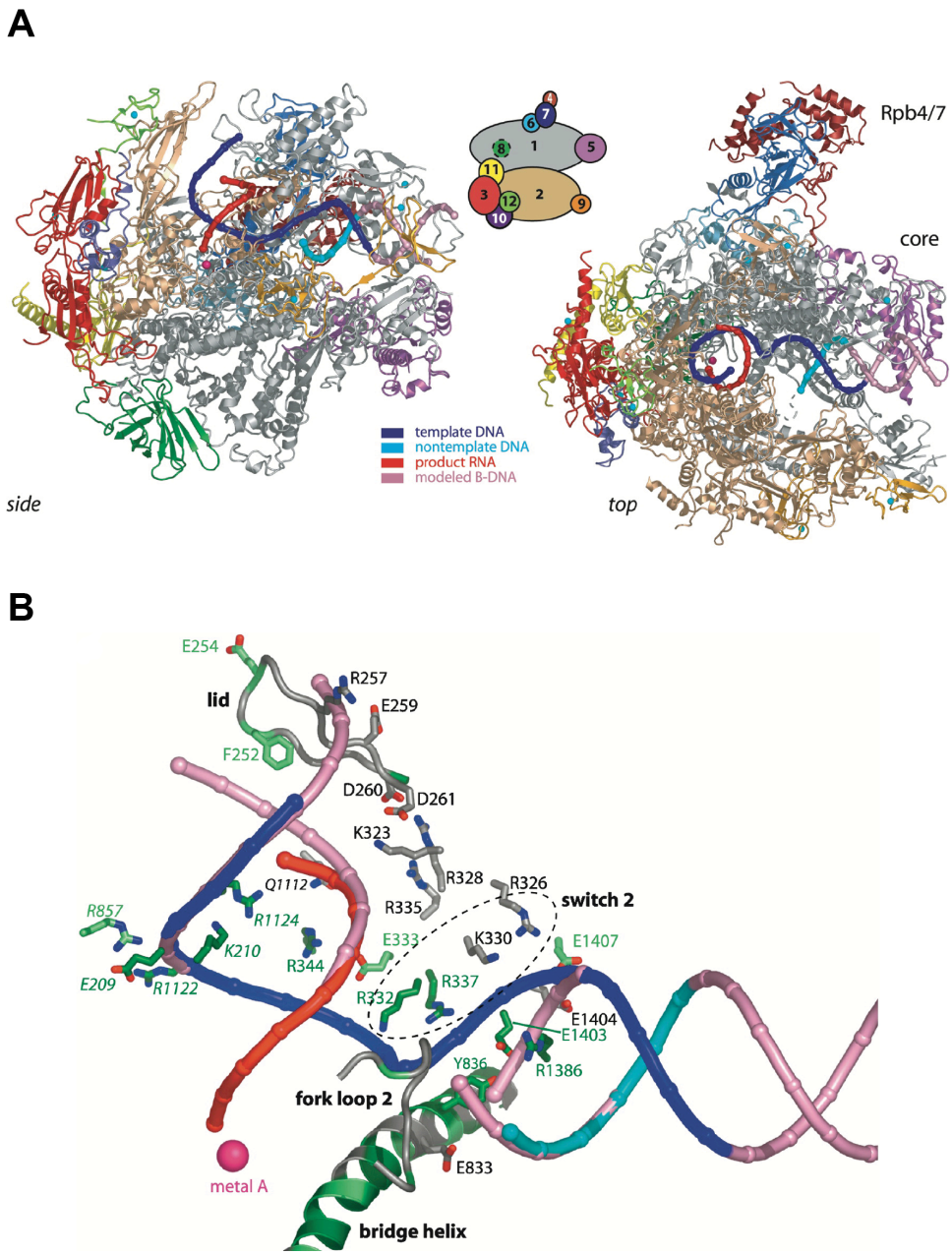

Figure 1.1 RNAPII elongation complex structure.

A. (Left) Side-view of RNAPII-EC. (Right) Top-view of RNAPII-EC. DNA TS, NTS and RNA are coloured as indicated. RNAPII subunits are coloured as in schematic representation (top). **B.** Detailed structure of RNAPII-nucleic acid interactions, with key structural features labelled. See text for details. Reproduced with permission (Kettenberger et al., 2004).

residue of the lid acts as a wedge to separate the DNA-RNA hybrid (Westover et al., 2004b).

At the floor of the RNAPII cleft is the active site where NTP addition occurs through a two metal-ion mechanism common to all polymerases (Steitz, 1998). A Mg^{2+} cation (metal A) is stably bound by Rpb1 aspartates 481, 483 and 485 and binds the RNA 3' end (Cramer et al., 2001). Metal B is not always stably bound and arrives with the NTP, interacting with aspartate 481, 837 and glutamate 836 of Rpb2 (Westover et al., 2004a, Cramer et al., 2001). Switch regions 1 and 2 of Rpb1 bind to the DNA template "active base" and flip it approximately 90° in a left-handed direction towards the pore, a channel present at the bottom of the cleft. This is assisted by the bridge helix, over which the template strand runs, which separates the active base from the downstream base.

Two important sites have been defined in the active centre of RNAPII: the entry site (E-site), which NTPs first encounter upon entry through the pore, and the addition site (A-site), where catalysis occurs and only NTPs correctly matched to DNA are observed (Westover et al., 2004a). An NTP in the A-site causes the trigger loop (Rpb1 residues between ~1070 and 1100 (Kettenberger et al., 2003)) to swing into position, leading to changes in the bridge helix conformation. The interactions between NTPs and the bridge helix and trigger loop are extensive. These occur between amino acids and the β-phosphate, 2'-OH and 3'-OH, allowing RNAPII to recognise the correct NTP for catalysis (Wang et al., 2006). Following catalysis, PP_i release leads to a disruption of the trigger loop interaction, altered conformation of the bridge helix and a "relaxing" of the active site, allowing translocation of the DNA-RNA hybrid. Entry of RNAPII into a

post-translocated state allows the cycle of NTP addition to be repeated (Wang et al., 2006).

Another key element of RNAPII is the Rpb1 clamp, a mobile domain shown to occupy two different open states (Cramer et al., 2001), but which enters into a closed state when in a complex with DNA and RNA (Gnatt et al., 2001). Closure of the clamp involves movement of some RNAPII residues by up to 30 Å and is implicated in RNAPII processivity, as the DNA-RNA hybrid is secured (Gnatt et al., 2001). Additionally, it has been suggested that this is the cause of EC stability in high salt (Gnatt, 2002, Levin et al., 1987). As mentioned earlier, the switch regions interact with the DNA-RNA hybrid, but also with each other. The presence of a DNA-RNA hybrid results in switch 3 becoming ordered and movement of switch 2 to prevent steric collision; the end-point of these molecular re-arrangements is clamp closure (Gnatt et al., 2001).

A note-worthy, non-catalytic feature of RNAPII is the Rpb1 carboxy-terminal domain (CTD) consisting of 52 heptapeptide repeats of a motif with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ in humans, and 26 repeats in yeast (Corden, 1990). The CTD is linked flexibly to the RNAPII core close to the RNA exit site (Cramer et al., 2001). The CTD can be phosphorylated by various kinases on tyrosine 1 (Tyr1), serine 2 (Ser2), threonine 4 (Thr4), serine 5 (Ser5) and serine 7 (Ser7) and acts as a platform for the recruitment of elongation-associated factors involved in 5’ or 3’ RNA processing events (Hirose and Manley, 2000, Proudfoot et al., 2002, Mayer et al., 2012), the details of which will be discussed later.

In RNAPII structural studies, no protein density was observed for the CTD of Rpb1, possibly due to disorder as a result of its flexibility (Cramer et al., 2001). CTD

peptides were used to determine the crystal structure of unphosphorylated, Ser5 phosphorylated and Ser2 phosphorylated CTD (Meinhart and Cramer, 2004). The CTD was shown to be a left-handed β-spiral, most compact when unphosphorylated, stabilised by Ser2 phosphorylation and opened following Ser5 phosphorylation, thus providing a structural basis for the temporal recruitment of elongation-associated factors.

1.2 Initiation

The binding of RNAPII to the promoter region of DNA and its assembly into a pre-initiation complex (PIC) was for many years believed to be the principal step at which transcription was controlled. Indeed, in many of the early systems used to study initiation (such as at *GAL4* in *S. cerevisiae* and the *lac* operon in *E. coli*) RNAPII recruitment is a limiting factor. Further studies have shown that regulation of transcription is a multi-faceted process with myriad other points for regulation, including at promoter clearance and during transcript elongation. Nevertheless, the process of initiation is highly controlled with many proteins playing a role.

1.2.1 Formation of the pre-initiation complex

The PIC is a complex of multiple proteins that direct the recruitment of RNAPII to the promoter and catalyse DNA melting, thus enabling transcription to occur. The PIC is made up of general transcription factors (GTFs) in a complex with RNAPII. The minimal set of GTFs required to direct basal (non-activated) transcription on a nucleosome-free DNA template are: TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Nechaev and Adelman, 2011). Formation of the PIC *in vivo* is highly dynamic and most likely occurs in a stepwise manner, however, there is a suggestion that assembly can also occur via the single-step recruitment of a pre-assembled RNAPII holoenzyme (McNally

et al., 2000, Koh et al., 1998, Wilson et al., 1996, Maldonado et al., 1996, Ossipow et al., 1995).

Sequence elements in core promoters direct PIC assembly, ensuring that transcription occurs at appropriate positions within the genome (Hahn, 2004, Cosma, 2002). There are a number of different sequence elements found in core promoters, including the TATA element, BRE (TFIIB-recognition element), Inr (initiator element) and DPE (downstream promoter element) (Smale and Kadonaga, 2003). These sequences are binding sites for proteins of the PIC and serve to correctly orientate RNAPII, determining the direction of transcription. Although most promoters contain one or more of these elements no particular one is absolutely required to facilitate initiation.

The TATA element is the most ancient and widely used core promoter element in nature (Figure 1.2A). The element has the consensus sequence TATAWAAR with the upstream nucleotide at -31 to -30 relative to the transcription start site (TSS) in mammals (Carninci et al., 2006, Ponjavic et al., 2006). The TATA DNA is bound by the TATA-binding protein (TBP), a saddle-shaped molecule consisting of two imperfect repeats. TBP binds to the widened minor groove of the TATA element, unwinding DNA approximately one third of a helical turn and bending it by approximately 80 Å towards the major groove (Kim et al., 1993b, Kim et al., 1993a). TBP bound to TATA DNA represents the initial step in GTF recruitment at this promoter element (Figure 1.2), with other GTFs being necessary to direct the orientation of TBP binding (Cox et al., 1997).

The BRE element is bound by TFIIB and is also well-conserved, the archael TFB binds a similar sequence, where it is the primary determinant of transcription

orientation (Lagrange et al., 1998, Qureshi and Jackson, 1998, Bell et al., 1999, Littlefield et al., 1999). The other two well-characterised elements are Inr and DPE. The Inr element is able to bind TBP-associated factors TAF1 and TAF2 (Chalkley and Verrijzer, 1999), while the DPE element is in close proximity to TAF6 and TAF9 (Burke and Kadonaga, 1997). Approximately fourteen TAFs and TBP form a complex known as TFIID, which is highly conserved in evolution (Sanders and Weil, 2000). The proper function of DPE-based promoters relies on the presence of the Inr element because they together co-ordinate TFIID binding (Burke and Kadonaga, 1997).

In addition to being a subunit of TFIID, TBP has been found in other transcription-related complexes, an interesting one being SAGA, a histone acetyltransferase (HAT) complex recruited to upstream activating sequences (UAS). TBP directly interacts with the Spt3 subunit of the complex (Laprade et al., 2007, Eisenmann et al., 1992, Mohibullah and Hahn, 2008). Indeed, studies have even shown that there is a high degree of similarity between the composition of SAGA and TFIID, and that these two complexes have a remarkably similar shape (Wu et al., 2004, Leurent et al., 2002). In agreement with the two complexes also having similar functions, studies using yeast mutants showed that approximately 10% of genes were dependent on SAGA rather than TFIID for their transcription, and that highly-inducible genes with clear TATA boxes were SAGA-dependent whilst their housekeeping counterparts with no recognisable TATA box preferentially recruited TFIID (Basehoar et al., 2004, Huisinga and Pugh, 2004).

The relative rarity of TATA-based promoters in many higher eukaryotes such as *Drosophila melanogaster* suggests that the method of promoter core recognition is more complicated than reductionist models would suggest (Ohler et al., 2002). *Drosophila*

and human TATA-less promoters have both the Inr and DPE elements, suggesting that in this subset of genes the TAFs have a greater functional role than TBP for PIC assembly, but TBP probably confers stability owing to protein-protein interactions between the GTFs (Smale and Kadonaga, 2003).

Following the binding of TFIID or SAGA to the core promoter (Figure 1.2B), the next event is likely to be binding of TFIIA or TFIIB, both of which interact specifically with TBP and have affinity for the bent promoter DNA (Figure 1.2B) (Nikolov et al., 1995, Geiger et al., 1996, Tan et al., 1996).

TFIIA is a heterodimer composed of two domains, the C-terminal is responsible for the TBP-DNA interaction whilst the N-terminal domain points away from TBP. TFIIA has two main functions: it stabilises TBP-DNA binding and promotes TFIID binding to DNA by competing with the N-terminal domain of TAF1, which masks the TBP DNA-binding surface in free TFIID (Weideman et al., 1997, Kokubo et al., 1998, Liu et al., 1998, Sanders et al., 2002).

TFIIB has two domains both of which are highly conserved; the N-terminal is a zinc-ribbon domain connected via a flexible linker to the C-terminal core domain, which binds TBP-DNA. TFIIB has a role in RNAPII recruitment as both domains co-operatively bind RNAPII with the functional surface of the ribbon domain absolutely required for this function (Chen and Hahn, 2003, Pardee et al., 1998, Hahn and Roberts, 2000). Part of the flexible linker known as the B-finger is positioned in the active centre of RNAPII and both biochemical and structural studies have implicated it in determination of the TSS (Bushnell et al., 2004, Kostrewa et al., 2009, Liu et al., 2010).

The next step in PIC assembly following formation of the TFIID/TBP-TFIIA-TFIIB sub-complex is likely to be binding of TFIIF, which probably has a role in the

correct positioning of RNAPII (Figure 1.2B). TFIIF has been shown to reduce the non-specific DNA affinity of RNAPII *in vitro*, consistent with this assertion (Finkelstein et al., 1992). The final two GTFs to bind the PIC are TFIIE and TFIIH, which interact strongly, and bind the DNA both upstream and downstream of the TSS (Figure 1.2C) (Bushnell et al., 1996). RNAPII associates with the GTFs and promoter DNA where the conserved heterodimer TFIIE binds to the polymerase jaw and probably has a role in regulating jaw conformation and the transition from the closed to the open complex (Leuther et al., 1996). Also, TFIIE has the ability to stimulate both the kinase and helicase activities of TFIIH (Figure 1.2D) (Lu et al., 1992, Ohkuma and Roeder, 1994).

Initiation at most promoters is ATP-dependent and this is due to the final stage, which requires TFIIH helicase activity. Low resolution EM of TFIIH has shown it to be a ring-shaped complex with two helicase subunits located either side of a prominent protrusion (Schultz et al., 2000, Chang and Kornberg, 2000). Studies have shown the ATPase activity of the helicase-related TFIIH subunit XPB (Ssl2 in yeast) to be essential for DNA strand separation and transition to the open complex (Figure 1.2D) (Coin et al., 1999). Recent work in the Hahn laboratory has shown that Ssl2 interacts directly with TFIIE and acts as a double-stranded DNA translocase. The fixed position of Ssl2, along with the anchoring of upstream DNA by the other GTFs (namely TBP, TFIIA and TFIIB), and combined with Ssl2 DNA translocation, leads to right-handed rotation of the DNA helix, unwinding and feeding of 15 bp of single-stranded DNA into the RNAPII cleft (Figure 1.2D and E) (Kim et al., 2000, Grunberg et al., 2012). Additionally, the TFIIH kinase Cdk7 (Kin28 in yeast) is responsible for Ser5 phosphorylation of the Rpb1-CTD, a key step, which destabilises the interaction between RNAPII, Mediator and the PIC, allowing the transition to the next step in

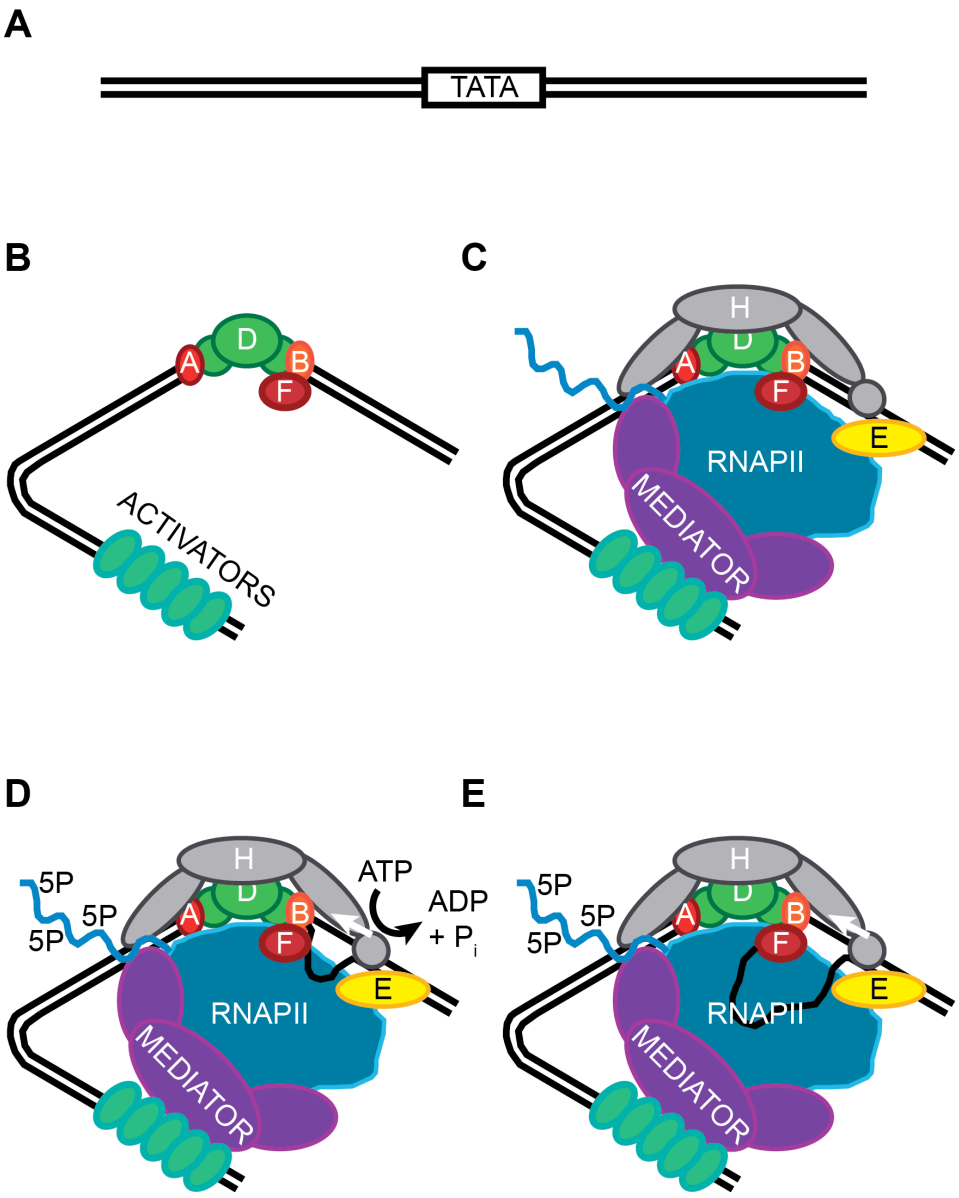

Figure 1.2 Pre-initiation complex formation and initiation

A. Promoter DNA with a TATA box indicated. **B.** TFIID/TBP (green) binds TATA DNA, bending it. Next TFIIB (orange), TFIIA (light red) and TFIIF (dark red) bind. **C.** TFIIE (yellow) and TFIIH (grey) bind DNA. Mediator (purple) associated with UAS-bound activators (light green) associates with the PIC and RNAPII (blue). **D.** Mediator stimulates kinase activity of TFIIH leading to CTD Ser5 phosphorylation (5P), whilst the Ssl2 translocase opens DNA in an ATP-dependent manner (white arrow). **E.** Single-stranded DNA is fed into the RNAPII active site and RNA polymerisation begins.

transcription (Figure 1.2D and E) (Lu et al., 1992, Liu et al., 2004, Svejstrup et al., 1997, Max et al., 2007).

1.2.2 The role of additional factors in initiation

There are numerous other factors that have a role in initiation, including myriad activators and repressors, which bind at promoter distal elements, namely UASs and enhancers. However, this section will only focus on a few factors that also play a role in general transcription initiation, the first being Mediator.

Mediator is a 24-subunit complex first isolated in yeast and shown to be essential for activated transcription, whilst also increasing both the efficiency of basal transcription and of TFIIH-catalysed CTD phosphorylation (Kim et al., 1994). Mediator interacts with un-phosphorylated RNAPII; approximately 40% of it is in a stable complex with polymerase (Rani et al., 2004). Mediator has three domains: the head, middle and tail, which bind to RNAPII via Rpb3 and Rpb11 (Davis et al., 2002). Genome-wide studies have shown Mediator to be present upstream of almost all active, and some inactive, genes. This is presumably a result of interactions between Mediator and activators bound to upstream regulatory sequences (Andrau et al., 2006, Zhu et al., 2006). Consistent with this assertion, crosslinking studies have shown that Mediator can be recruited to promoters prior to binding of the basal transcription machinery, probably due to activator-based recruitment, which could lead to Mediator being transferred to the PIC (Bhoite et al., 2001, Bryant and Ptashne, 2003, Cosma et al., 2001). The formation of the PIC and order of GTF binding seems to be a highly dynamic process; evidence suggests that this is also the case for Mediator, with some *in vitro* data showing that TFIIB, RNAPII and Mediator can be recruited to the promoter co-operatively (Ranish et al., 1999). Nevertheless, binding of Mediator is a key stage in

PIC formation and its stimulation of TFIIH kinase activity could represent a checkpoint during initiation, with the resulting CTD phosphorylation leading to a disruption of the Mediator-RNAPII interaction (Svejstrup et al., 1997, Max et al., 2007). Mediator subsequently stabilises a sub-complex of basal factors that remain behind at the promoter following movement of RNAPII into the early elongation phase, and likely accelerates the rate of subsequent re-initiation events (Reeves and Hahn, 2003, Esnault et al., 2008, Yudkovsky et al., 2000).

The transcription factor TFIIS (*see section 1.3.2.1.1*) has a well-characterised role as an elongation factor stimulating transcript cleavage by RNAPII (Fish and Kane, 2002). However, it can also be found associated with TFIIB, TFIIE and RNAPII, in addition to promoter-associated SAGA via its Spt8 subunit, and the Mediator subunit Med13 (Pan et al., 1997, Hirst et al., 1999, Wery et al., 2004). This strongly suggests that TFIIS has an additional role, in transcription initiation, and supporting evidence shows that TFIIS lacking the ability to stimulate transcript cleavage but able to interact with RNAPII is able to stimulate PIC formation (Kim et al., 2007). There is, therefore, a clear distinction between the elongation and initiation-dependent roles of TFIIS, and TFIIS may help promote PIC formation through its affinity for the basal factors and RNAPII. Recruitment of TFIIS to the *GAL1* promoter is dependent on Mediator and SAGA but independent of RNAPII, but TFIIS loss also has an affect on the recruitment of TBP and RNAPII to promoters *in vivo* (Prather et al., 2005). Altogether, this suggests that TFIIS plays an important role in PIC recruitment and association during initiation.

1.3 Elongation

Elongation is a highly regulated process with many auxiliary factors controlling and facilitating RNAPII progression across the gene. The process of elongation can broadly be divided into two stages: early elongation and processive elongation.

1.3.1 Early elongation/promoter escape

Early elongation is a loosely defined term to describe the phase in which promoter escape and formation of the early transcribing complex occurs. Following open promoter complex formation, RNAPII begins synthesis of RNA complimentary to the DNA template. During this period, the EC is characteristically unstable with a tendency to slip backwards on the DNA and release its short nascent transcript, a process termed abortive initiation (Pal and Luse, 2002). Footprinting studies of prokaryotic RNAP indicate that polymerase is unable to leave the promoter during abortive initiation, and that a commitment to stable elongation occurs only after production of a 10 nt RNA (Krummel and Chamberlin, 1989, Carpousis and Gralla, 1985), approximately consistent with the length of the stable DNA-RNA hybrid observed in both structural (Gnatt et al., 2001) and biochemical studies (Kireeva et al., 2000). At a mechanistic level, abortive initiation results from the unproductive release of energy stored in a stressed intermediate (Straney and Crothers, 1987). Single-molecule studies have revealed that RNAP remains stationary at the promoter and pulls downstream DNA into itself; an additional turn of DNA unwinding occurs and serves as an intermediate to store the energy that facilitates promoter escape (Revyakin et al., 2006, Kapanidis et al., 2006). Abortive transcription is likely also exacerbated by steric collision between RNA and the B-reader loop of TFIIB, which reaches through the

RNA exit site towards the active centre of RNAPII. Although this factor stabilises the production of short RNAs, it must dissociate to allow the production of longer RNAs. This process occurs when the transcript length exceeds 7 nt (Bushnell et al., 2004, Kostrewa et al., 2009, Liu et al., 2010). Concomitant with the production of a more stable DNA-RNA hybrid, DNA upstream of the bubble collapses, stabilising the EC and possibly providing additional energy for promoter escape (Pal et al., 2005).

Formation of a 23 nt RNA transcript is associated with increased EC stability (Pal and Luse, 2003). Stable ECs have been shown to pause approximately 50 bp downstream of the TSS in a process termed promoter-proximal pausing (Marshall and Price, 1992, Rahl et al., 2010). The proteins involved in regulating this, and the mechanism, are discussed in more detail below.

1.3.1.1 DSIF, NELF, P-TEFb and promoter-proximal pausing

DRB (5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole)-sensitivity-inducing factor (DSIF) is an elongation factor initially identified as causing DRB-dependent RNAPII pausing in a nuclear-extract-driven transcription system, along with stimulating elongation in the absence of DRB (Wada et al., 1998). More recent studies suggest that the DRB-dependent transcriptional inhibition actually resulted from contamination of the transcription reactions with NELF (discussed in more detail below) (Yamaguchi et al., 1999). DSIF is a heterodimeric complex formed of a 160 kDa subunit (p160) and 14 kDa (p14) subunit, shown to be homologous to yeast Spt5 and Spt4, respectively (Wada et al., 1998). Spt4 is a positive elongation factor, with *spt4*Δ mutants displaying reduced mRNA levels, whilst *spt4*Δ extracts lack processivity when used in *in vitro* transcription assays (Rondon et al., 2003). Spt5 also stimulates elongation, by preventing premature termination and pausing during late elongation (Bourgeois et al., 2002).

Negative elongation factor (NELF), another factor involved in promoter-proximal pausing is a five subunit complex (NELF-A, -B, -C, -D and –E) with no known homologue in yeast. NELF induces RNAPII pausing in the presence of DSIF through interaction with a DSIF-RNAPII complex (Yamaguchi et al., 1999).

NELF and DSIF regulate transcript elongation in concert with a third factor, P-TEFb (positive transcription elongation factor b), a heterodimeric complex consisting of the kinase CDK9 with cyclin T1, T2a, or T2b (Peng et al., 1998a, Peng et al., 1998b). P-TEFb was identified as a CTD kinase capable of stimulating elongation *in vitro* (Marshall and Price, 1992, Marshall et al., 1996). Later studies in *C. elegans* provided evidence that P-TEFb specifically phosphorylates Ser2 and not Ser5 of the Rpb1 CTD (Shim et al., 2002). Yeast does not have a direct homologue of P-TEFb, but the bulk of Ser2 phosphorylation is performed by CTDK-I a complex consisting of Ctk1 (a CDK homologue), Ctk2 (a cyclin homologue) and Ctk3 (Buratowski, 2009). A recent study showed that Ctk1 is homologous to the human cyclin-dependent kinases CDK12 and CDK13, which also have CTD-kinase activity (Bartkowiak et al., 2010). Additionally, the Bur1/2 kinase (CDK9 in metazoans) is also able to phosphorylate the CTD at Ser2, as well as acting upon Spt5 (Qiu et al., 2009, Liu et al., 2009).

DSIF/NELF mediates RNAPII arrest during early elongation, with NELF competing with the stimulatory factor TFIIF for RNAPII binding (Renner et al., 2001, Cheng and Price, 2007). Importantly, promoter-proximal pausing serves as a checkpoint for correct pre-mRNA capping, which occurs as a result of the sequential activity of 5'-triphosphatase, guanylyltransferase and methyltransferase. Capping enzyme is recruited and stimulated by the Ser5 phosphorylated Rpb1 CTD (Yue et al., 1997, Ho and Shuman, 1999, Moteki and Price, 2002). Additionally, capping enzyme is able to

interact directly with both the Spt5 subunit of DSIF and the Rpb1-CTD phosphorylated at Ser5 (Wen and Shatkin, 1999, Cho et al., 1997). Early arrest is relieved by P-TEFb-mediated phosphorylation of both the Rpb1-CTD and the Spt5 subunit of DSIF, which causes NELF to dissociate (Yamada et al., 2006, Chen et al., 2009). Capping enzyme was also shown to relieve NELF-mediated pausing, possibly providing a feedback between correct RNA capping and checkpoint release (Mandal et al., 2004).

The loss of NELF from an elongation complex allows movement into the next transcription phase, processive elongation.

1.3.2 Processive elongation

It is in this phase of elongation that the bulk of RNA is produced, with RNAPII able to synthesise RNA in excess of 2 million nucleotides long without dissociating from DNA (Tennyson et al., 1995). Elongation is characterised by high rates of transcription, the average reported at between 1.1-1.4 kb per minute (Femino et al., 1998) and even as high as 50 kb per minute (Maiuri et al., 2011). However, it is important to note that these differences in transcription rate are quite distinct from RNAPII catalytic rate, which does not change. The source of rate variability is due to the fraction of time RNAPII spends in an on-pathway, rather than an off-pathway, state. Indeed, *in vivo* studies have shown several kinetically distinct populations of RNAPII interacting with a single specific gene (Darzacq et al., 2007).

On-pathway events are those that result in productive catalytic activity, namely forward translocation of RNAPII and elongation of the nascent transcript. The extraordinary stability of ECs has been mentioned earlier (*see section 1.1.1*) and in agreement with this, single-molecule studies have shown that the EC remains stable even when loads of up to 30 pN are applied to it (Dalal et al., 2006, Neuman et al., 2003,

Wang et al., 1998). During elongation, RNAPII functions like a Brownian ratchet; there is no ATP-dependent 'power-stroke' during translocation. RNA polymerases are able to move backwards as well as forwards on the DNA template, with forward movement favoured through NTP binding and hydrolysis (Kireeva et al., 2000, Bar-Nahum et al., 2005).

Movement of RNAPII into off-pathway states frequently interrupts elongation but may be important for regulation and recruitment of appropriate factors such as those responsible for mRNA capping, splicing, poly-adenylation and transcription termination (de la Mata et al., 2003, Yonaha and Proudfoot, 1999). RNAPII in an off-pathway state can be either paused or arrested. Pausing is a transient and common feature during elongation, which predominates when forward motion is impeded, for example by DNA damage or nucleotide mis-incorporation (Donahue et al., 1994, Jeon and Agarwal, 1996). The transition of an EC from a paused to an arrested state is a function of the dwell time of RNAPII at a pause site (Hawley et al., 1993, Gu and Reines, 1995). Arrested ECs are defined as those unable to resume transcription, whilst still being catalytically active (Arndt and Chamberlin, 1990, Krummel and Chamberlin, 1992). Arrested RNAPII is thus in a backtracked state in which the RNAPII active site is out of register with the 3'-OH group of the nascent transcript (Chamberlin, 1992). The incorrectly positioned RNA is then extruded and fills the pore of RNAPII, forming protein interactions and stabilising the backtracked state (Wang et al., 2009). Factors facilitating the re-entry of RNAPII into on-pathway elongation are discussed in the following section.

1.3.2.1 Factors assisting elongation

Factors have evolved to control all aspects of transcript elongation *in vivo* and maintain RNAPII processivity in the presence of blocks to elongation. A selection of relevant factors is discussed below.

1.3.2.1.1 TFIIS

A major factor aiding RNAPII in overcoming transcription barriers is TFIIS, a protein identified by its ability to stimulate transcript elongation *in vitro* (Natori et al., 1973). TFIIS was shown to act after transcription initiation (Sekimizu et al., 1976, Reinberg and Roeder, 1987, Sawadogo et al., 1981) by stimulating the extent and efficiency of RNA production, but not the rate (Sluder et al., 1989, Reinberg and Roeder, 1987, Rappaport et al., 1987, Bengal et al., 1991, Izban and Luse, 1992a). TFIIS was also shown to stimulate read-through of A-T-rich DNA sequences, which are inhibitory to transcription (Kerppola and Kane, 1990, Hawley et al., 1993) as well as stimulate transcription past sequence-specific DNA-binding proteins (Izban and Luse, 1992a, Reines and Mote, 1993). The mechanism of this was found to result from the ability of TFIIS to stimulate transcript cleavage. The transcript cleavage activity is a conserved and integral part of RNAP itself, supported by studies showing it to be α-amanitin-sensitive and independent of accessory factors (Johnson and Chamberlin, 1994). Intrinsic RNAP-mediated transcript cleavage, however, occurs at a very low rate, but is greatly enhanced by accessory factors: the Gre factors in prokaryotes and TFIIS in eukaryotes (Orlova et al., 1995, Wang and Hawley, 1993, Rudd et al., 1994). A recent study performed in the Svejstrup laboratory isolated a TFIIS mutant able to inhibit the intrinsic cleavage activity of RNAPII, which when expressed *in vivo* resulted

in a loss of viability and transcriptional arrest (Sigurdsson et al., 2010). This shows that in the absence of functional TFIIS the intrinsic cleavage activity of RNAPII is absolutely required, as polymerase arrest must occur frequently *in vivo*.

TFIIS can act on backtracked and arrested ECs, the cleavage activity causing up to 17 nt fragments of RNA to be released, whilst realigning the 3'-OH of the RNA in the active site to restart transcription (Izban and Luse, 1992b, Izban and Luse, 1993a, Izban and Luse, 1993b). Importantly, cleavage factors do not affect the fraction of RNAP able to transcribe past an elongation block but instead enable RNAP to have multiple attempts at bypass by restarting transcription (Reines, 1992).

TFIIS is a single-subunit protein consisting of three domains: I, II and III with a 19 amino acid linker connecting domains II and III. The N-terminal domain I is a four-helix bundle dispensable for cleavage activity (Booth et al., 2000), the central domain is a three-helix bundle, which forms an additional three helices upon RNAPII interaction (Morin et al., 1996, Olmsted et al., 1998, Kettenberger et al., 2003), and the C-terminal domain III is a zinc-ribbon fold with a β-hairpin, containing the activating acidic loop (Olmsted et al., 1998, Qian et al., 1993). The linker residues between domains II and III are necessary for TFIIS activity and have been shown to confer species-specificity (Shimasaki and Kane, 2000, Awrey et al., 1998).

Structural studies have helped elucidate the mechanism of TFIIS activity (Kettenberger et al., 2003, Kettenberger et al., 2004). Domains II, III and the linker are required for RNAPII binding by TFIIS. Domain II docks to the Rpb1 jaw domain, the linker passing through and opening a crevice on RNAPII causing structural rearrangements, and domain III inserts into the RNAPII pore entering the active site (Kettenberger et al., 2003). The β-hairpin of domain III becomes highly ordered through

extensive contacts with RNAPII, including interactions with the trigger loop and bridge helix, and causes a 4 Å shift in the position of the RNA (Kettenberger et al., 2004). Two highly conserved acidic residues in the β-hairpin are required for TFIIS activity (Jeon et al., 1994), which occurs through the co-ordination of two catalytic metal ions by the acidic residues, in a manner similar to DNA cleavage by the Klenow DNA polymerase (Beese and Steitz, 1991, Joyce and Steitz, 1994, Kettenberger et al., 2004). TFIIS aspartate 290 in the acidic residue patch directly contacts and activates a nucleophilic water molecule for an S_N2 attack on the RNA phosphodiester bond (Kettenberger et al., 2004, Wang et al., 2009). This model is in agreement with a requirement for divalent metal ions for RNA cleavage by RNAPII-TFIIS (Izban and Luse, 1992b, Reines, 1992, Wang and Hawley, 1993, Weilbaecher et al., 2003).

Control of TFIIS can be achieved through phosphorylation of serine and threonine residues in domain I (Sekimizu et al., 1981) resulting in the inability of TFIIS to stimulate elongation in a reconstituted system. However, phosphorylated TFIIS is still active when added to purified RNAPII (Hirai et al., 1988).

1.3.2.1.2 TFIIF

Yeast TFIIF is a complex consisting of three proteins, Tfg1, Tfg2 and Tfg3, the first two being homologous to human RAP74 and RAP30, but the latter having no known mammalian counterpart. TFIIF has roles in both initiation and elongation, where it has been shown to stabilise TFIIB in the PIC (Cabart et al., 2011), suppress abortive initiation and stimulate promoter escape, whilst functioning in tandem with TFIIH to prevent arrest of early elongation intermediates (Yan et al., 1999). Although a recent genome-wide study showed that TFIIF is most commonly localised to promoter regions

(Mayer et al., 2010), it has also been found at the 3' ends of human genes (Cojocaru et al., 2008). During elongation, TFIIF stimulates elongation by re-associating with stalled ECs to convert them into an elongation-competent form (Zawel et al., 1995) and co-operates with TFIIS to regulate RNAPII processivity (Zhang et al., 2003).

1.3.2.1.3 Elongins

The Elongin complex (SIII) is a three-subunit complex consisting of Elongin A, the active subunit and regulatory subunits Elongins B and C (Bradsher et al., 1993a, Aso et al., 1995, Aso et al., 1996, Garrett et al., 1995, Takagi et al., 1996). Studies have shown that like TFIIF, the Elongin complex has a positive role in transcript elongation, achieved through suppression of transient EC pausing, but unlike TFIIF is not required for PIC formation and has no other roles during initiation (Bradsher et al., 1993b). Data suggests that the effect of Elongin and TFIIF on ECs are mutually exclusive, as TFIIF must be lost in order to confer Elongin-sensitivity (Moreland et al., 1998). Very little is known about the precise molecular mechanism of Elongin action but it has been suggested that it is able to realign the 3'-OH end of the nascent transcript in the RNAPII active site (Takagi et al., 1995). Other Elongin A-related proteins have been discovered (Elongins A2 and A3), which are able to stimulate elongation but are not regulated by Elongin B or Elongin C.

In yeast, the proteins Elc1 and Ela1 are homologues of Elongins C and A, however, no Elongin B homologue has been identified (Aso and Conrad, 1997). Unlike the mammalian proteins, Ela1 and Elc1 are unable to stimulate transcript elongation (Koth et al., 2000), but have been shown to have a role in RNAPII poly-ubiquitylation and degradation (Ribar et al., 2006, Ribar et al., 2007, Harreman et al., 2009), which

will be discussed in more detail later. Not surprisingly, mammalian Elongins also have a role in ubiquitylation. Elongins B and C form a stable complex with the von Hippel-Lindau tumour suppressor protein (pVHL), Cul2 and Rbx1. This complex is remarkably similar to the yeast complex of Ela1, Elc1, Cul3 and Roc1 and in support of them having a similar role, mammalian Elongin C is homologous to yeast Skp1, a subunit of the SCF complex, which targets proteins for poly-ubiquitylation and degradation (Ribar et al., 2007, Harreman et al., 2009, Ivan and Kaelin, 2001). Furthermore, the ElonginBC-pVHL complex has been shown to target hypoxia-inducible factor (HIF) for proteasome-mediated degradation; however, further studies are necessary to fully elucidate the role of the Elongins and their various complexes in transcript elongation (Ivan and Kaelin, 2001, Conaway and Conaway, 2002).

1.3.2.1.4 ELL family proteins

ELL is a product of the eleven-nineteen lysine-rich leukaemia gene and was found to be functionally analogous to TFIIF and Elongin, with the ability to aid RNAPII processivity by suppressing transcriptional pausing, thus limiting the amount of time spent by RNAPII in an off-pathway state (Shilatifard et al., 1996). There are three ELLs in humans designated ELL1, ELL2 and the testis-specific ELL3, all of which have stimulatory roles during elongation (Shilatifard et al., 1997, Miller et al., 2000).

Drosophila has a single homologue of ELL (dELL) which co-localises with RNAPII on polytene chromosomes and is essential for normal development, suggesting that it plays a non-redundant role in elongation, distinct from that of Elongin and TFIIF (Gerber et al., 2001, Eissenberg et al., 2002). Additionally, mutations in the gene encoding dELL result in transcription defects, the effect of which is mainly confined to

longer genes (Eissenberg et al., 2002, Shilatifard, 2004). This is consistent with ELL having an important role in transcript elongation.

In addition to interacting with RNAPII, ELL proteins have been shown to exist in other complexes. The first complex is with ELL-associated proteins EAP20, EAP30 and EAP45, which has positive elongation activity and lacks the uncharacterised negative activity that ELL alone can have on elongation *in vitro* (Shilatifard, 1998). ELL has also been shown to be part of the super elongation complex (SEC) a multi-protein complex with several different sub-forms, consisting of: P-TEFb, AFF1, AFF4 (the main scaffold protein), ELL1, ELL2, ENL and AF9 (Sobhian et al., 2010, He et al., 2010, Lin et al., 2010, Yokoyama et al., 2010, Mueller et al., 2009). The SEC is able to interact with the PAF complex (a multi-functional complex which has a role in elongation, co-transcriptional histone modification and mRNA biogenesis (Zhou et al., 2012)) and is required for efficient induction of some rapidly-induced genes in response to differentiation signals in metazoans (Sobhian et al., 2010, Lin et al., 2011). These data are consistent with those obtained in *Drosophila* and implicate ELL in development and differentiation.

1.3.3 Blocks to transcription

1.3.3.1 Chromatin

All DNA metabolic events occur in the context of chromatin, the basic unit of which is the nucleosome, an octameric protein complex consisting of two of each of the core histones H2A, H2B, H3 and H4 (Kornberg and Lorch, 1999). Histones H2A and H2B interact to form a dimer as do histones H3 and H4. Nucleosomes are encircled by

approximately 147 bp of DNA and have the ability to regulate protein-DNA interactions, such as those between GTFs and promoter DNA (Luger et al., 1997, Luger, 2006).

The structure of chromatin is important for DNA metabolism, alteration of its state either allows or inhibits the occurrence of such processes. The presence of chromatin has an impact on the efficiency of transcription. Indeed, linear poly-nucleosomal arrays inhibit RNAPII elongation *in vitro*, but RNAPII is able to transcribe past single nucleosomes (Lorch et al., 1987, Izban and Luse, 1991, Orphanides et al., 1998). Biochemical studies performed by the Kashlev laboratory have shown that RNAPII has the ability to cause dissociation of the H2A/H2B dimer from a nucleosome when transcribing through it (Kireeva et al., 2002). Further work has shown that a proportion of RNAPII backtracks following pausing at a nucleosome and that TFIIS, or the presence of a trailing RNAPII molecule, helps reduce backtracking of the leading polymerase, aiding nucleosome bypass (Kireeva et al., 2005, Jin et al., 2010). Such bypass is likely to be the result of elongating RNAPII stripping DNA off the surface of the nucleosome, and this is supported by the finding that RNAPII does not change the position of the nucleosome on DNA whilst transcribing past it (Kireeva et al., 2005, Kireeva et al., 2002).

In addition to the intrinsic, albeit inefficient, ability of RNAPII to transcribe through nucleosomes, factors have evolved to change the state of chromatin, facilitating the progression of RNAPII and alleviating the role of chromatin as a transcriptional block. These factors fall into three broad groups and some key examples of each are discussed below.

The first group are ATP-dependent chromatin remodellers, in which four families have been identified: SWI-SNF, ISWI, CHD and INO80/SR. This group of

remodellers is able to use the energy stored in ATP to alter the position of nucleosomes on DNA, as well as destabilise or eject them; activities essential for correct promoter function and allowing PIC proteins access to their respective binding sequences (Clapier and Cairns, 2009, Schnitzler, 2008). In addition to promoter-based remodelling, this group of factors has been shown to influence chromatin on coding regions. For example, RSC was shown to increase mono-nucleosome bypass efficiency *in vitro* through its SWI-SNF remodelling activity (Carey et al., 2006).

The second group are histone chaperones, proteins involved in the dynamic assembly and disassembly cycle of nucleosomes. FACT (facilitates chromatin transcription) was the first chaperone found to have a role in transcript elongation and does so by destabilising the nucleosome and aiding in the removal of an H2A/H2B dimer upon RNAPII passage, probably aiding stripping of DNA off the surface of the nucleosome by the transcribing polymerase (Belotserkovskaya et al., 2003). The NAP-like family of histone chaperones also have transcription-associated roles: Vps75 binds histones and has been implicated in RNAPII progression during transcription (Selth and Svejstrup, 2007, Selth et al., 2009). Another key example of a histone chaperone is Spt6, an H3/H4 binding protein, which has a role in the maintenance of chromatin structure in yeast and specifically the restoration of chromatin structure following RNAPII progression across a gene (Bortvin and Winston, 1996). In support of this, studies have shown that Spt6 is required to prevent cryptic transcription occurring at the 3' end of genes (Kaplan et al., 2003). Therefore, this group of proteins can be thought of as nucleosome 'recyclers'.

The final group are covalent histone modifying proteins, enzymes with the ability to transfer a chemical moiety or protein (e.g. ubiquitin) to histones. The

protruding, unstructured amino-terminal 'tails' of histones are the targets for a diverse array of post-translational modifications, which serve to alter the packaging and structure of chromatin, as well as provide a binding surface for histone-associated proteins (Luger et al., 1997). The three modifications that will be discussed are: acetylation, methylation and ubiquitylation.

1.3.3.1.1 Acetylation of histones

Histone acetylation occurs through the activity of histone acetyltransferases (HATs), which use acetyl-CoA as a substrate for the reaction. This activity is antagonised by histone deacetylases (HDACs), which remove the modification. Both HATs and HDACs are found in the coding region of genes, suggesting substantial cycling of acetylation on active genes (Carrozza et al., 2005, Wang et al., 2002, Gilbert et al., 2004, Close et al., 2006, Keogh et al., 2005, Govind et al., 2007).

In support of a role for acetylation in elongation, mutants of the Elongator HAT complex, a complex isolated through its interaction with elongating RNAPII (Otero et al., 1999), display hypoacetylation of histones on the coding region of genes and transcriptional impairment (Kristjuhan et al., 2002). Additionally, Elongator is able to stimulate transcription on a chromatin template *in vitro*, suggesting that acetylation occurs co-transcriptionally (Otero et al., 1999, Wittschieben et al., 1999, Svejstrup, 2007b, Kim et al., 2002). It seems that this is not a feature unique to Elongator, as the p300/CBP-associated factor (PCAF) also binds hyper-phosphorylated RNAPII and is required for efficient elongation on a chromatin template *in vitro* (Cho et al., 1998, Obrdlik et al., 2008, Guermah et al., 2006). The SAGA subunit Gcn5 has also been

implicated in co-transcriptional acetylation, as H3 acetylation levels are reduced in the coding regions of *gcn5*Δ mutants, which show elongation and nucleosome eviction defects (Kristjuhan and Svejstrup, 2004, Govind et al., 2007, Kristjuhan et al., 2002, Govind et al., 2005). This highlights the importance of this co-transcriptional event in allowing processive elongation to occur.

In addition to the role of acetylation in elongation, hyperacetylation is a hallmark of active promoters and probably serves to relax chromatin structure in addition to recruiting ATP-dependent remodellers such as SWI-SNF (Workman and Kingston, 1998, Pokholok et al., 2005, Hassan et al., 2001). Therefore, this chromatin modification is necessary for the correct function of transcription at multiple different stages.

1.3.3.1.2 Histone methylation and ubiquitylation

The role of histone methylation is less defined than that of acetylation, as the addition of this small chemical group is unlikely to cause significant structural changes in chromatin. It is much more likely that this modification acts as a platform for recruitment of effector proteins containing chromodomains, PHD and tudor domains (Daniel et al., 2005). The addition of methyl groups to histones occurs via the action of histone methyltransferases (HMTs), which are able to mono-, di-, or tri-methylate histone H3, whilst demethylases serve to remove this modification.

There are several methylation marks associated with transcription, of which a few shall be discussed in further detail in this section. Co-transcriptional methylation occurs at H3K36 in coding regions as a result of the association between the HMT Set2

and RNAPII, via the Ser2 phosphorylated form of the CTD (Krogan et al., 2003, Li et al., 2003, Schaft et al., 2003, Xiao et al., 2003). The accumulation of di- and tri-methylated H3K36 in coding regions is, in turn, responsible for recruitment of the Rpd3S HDAC complex (Li et al., 2009). The methyl mark serves as a platform for binding of the Eaf3 and Rco1 subunits of the HDAC, via their respective chromodomain and PHD domain, ultimately leading to co-transcriptional histone de-acetylation (Carrozza et al., 2005, Keogh et al., 2005, Joshi and Struhl, 2005, Li et al., 2009, Li et al., 2007). These findings suggest that H3K36 methylation has a regulatory role in the transcription-associated acetylation-deacetylation cycle.

A further modification is histone ubiquitylation. In addition to its role in targeting proteins for degradation, generally associated with the addition of ubiquitin chains (poly-ubiquitylation), addition of one ubiquitin moiety (mono-ubiquitylation) serves as a signalling mark, resulting in a variety of downstream effects (Pickart, 2001). Mono-ubiquitylation of histone H2B, which occurs at K123 (K120 in mammals) is no exception. This histone mark occurs as a consequence of transcriptional activation and is found at promoters and coding regions (Henry et al., 2003, Kao et al., 2004, Xiao et al., 2005). H2B mono-ubiquitylation requires active transcription and is dependent on the TFIIH kinase, which phosphorylates Ser5 of the Rpb1-CTD, marking the transition from initiation to elongation (Xiao et al., 2005). The mechanism of H2B ubiquitylation is a complex one, requiring multiple factors. Transcribing RNAPII associates with the elongation factor Spt5, which can be phosphorylated by the Bur1/2 kinase. Spt5 phosphorylation leads to the recruitment of the Paf complex (Liu et al., 2009, Zhou et al., 2009), which associates with the Rad6/Bre1/Lge1 complex and mono-ubiquitylates histone H2B (Weake and Workman, 2008).

Histone H2B mono-ubiquitylation is important for RNAPII movement through the coding region of genes and is associated with histone H3 methylation. Methylation of H3K4 (performed by Set1 in yeast and MLL in humans, part of the COMPASS complex) and H3K79 (performed by Dot1) are linked to H2B ubiquitylation (Gerber and Shilatifard, 2003). H3K4 mono-methylation is able to occur in the absence of ubiquitylated H2B, but the Cps35 subunit of COMPASS, required for the subsequent di- and tri-methylation is absent (Lee et al., 2007). Furthermore, the requirement of ubiquitylated H2B in H3K79 methylation is likely mediated by Cps35, which is able to interact with Dot1 and is required for efficient H3K79 methylation (Selth et al., 2010). Although the exact mechanism of this is unclear, it highlights the interplay between different histone modifications and their ultimate effect on transcript elongation.

In addition to the indirect effects of histone H2B ubiquitylation on transcript elongation, there are also more direct effects. This histone mark has been shown to enhance RNAPII transcription on nucleosomal templates *in vitro*, probably as a result of FACT-dependent histone H2A/H2B dissociation and has been shown to stimulate FACT activity (Pavri et al., 2006, Fleming et al., 2008). Furthermore, ubiquitylated H2B is involved in nucleosome reassembly in the wake of RNAPII transcription. Mutant H2B that could not be ubiquitylated showed a reduction in re-assembly of nucleosomes at highly expressed genes (Batta et al., 2011), a phenomenon dependent on the Chd1 chromatin remodeller (Lee et al., 2012).

In conclusion, chromatin forms a barrier to transcription at various stages of the transcription cycle. The presence of chromatin prevents cryptic initiation events, as well as unscheduled initiation, adding another level of control to gene expression. The

presence of accessory factors allows control of transcription at the level of elongation and can facilitate the passage of RNAPII through this strong transcription inhibitor.

1.3.3.2 DNA damage and RNAPII ubiquitylation

DNA lesions arising from endogenous and exogenous damaging agents form a block to elongation and thus interrupt gene traffic. RNAPII can transcribe past helix-distorting lesions such as cyclobutane pyrimidine dimers (CPDs) *in vitro*, but bypass is extremely slow and inefficient (Brueckner et al., 2007, Walmacq et al., 2012), and may not occur often *in vivo*. Other forms of DNA damage are also a problem: O^6-Methylguanine lesions impair elongation, but to a much lesser extent than the bulky lesions (Dimitri et al., 2008), whilst transcription factors have been shown to aid in the bypass of non-bulky thymine glycol (TFIIF, Elongin, CSB) and 8-oxoguanine lesions (Elongin, CSB, TFIIS), presumably by suppressing RNAPII pausing and arrest (Charlet-Berguerand et al., 2006). Upon encountering a damage site, especially one containing a bulky lesion, RNAPII pauses and eventually arrests. This has two results: blockage of the gene preventing further transcription, and initiation of transcription-coupled nucleotide excision repair (TC-NER) (Saeki and Svejstrup, 2009, Donahue et al., 1994, Svejstrup, 2007a, Svejstrup, 2003). Arrested RNAPII is also inhibitory to other metabolic events occurring on DNA; therefore, it can be necessary to remove immobile RNAPII from the gene.

Removal of arrested RNAPII from a gene occurs via ubiquitylation and degradation of the Rpb1 subunit. This represents a last resort pathway, occurring as a result of irreversibly arrested RNAPII, rather than as a DNA damage response *per se* (Woudstra et al., 2002, Anindya et al., 2007, Somesh et al., 2005, Sigurdsson et al., 2010). Clearance of RNAPII from a coding region results in the continuation of

transcription by trailing polymerases, or, if damage is still present (if TC-NER did not occur), access of proteins responsible for its repair (general genome nucleotide excision repair, GG-NER).

Ubiquitylation of yeast RNAPII occurs at two key lysine residues of Rpb1: K330 and K695 (Somesh et al., 2007). As in all ubiquitylation pathways, a triple-enzyme cascade is responsible for Rpb1 ubiquitylation, consisting of an: E1 (activating enzyme), E2 (conjugating enzyme) and E3 (ligating enzyme). Both *in vitro* and *in vivo* studies have shown Rsp5 to be the E3 involved in Rpb1 ubiquitylation post-DNA damage, where it associates with the Rpb1-CTD (Somesh et al., 2005, Harreman et al., 2009, Beaudenon et al., 1999). Furthermore, this reaction can be reconstituted *in vitro* using the highly purified factors: Uba1, Ubc5 and Rsp5 (as the E1, E2 and E3, respectively) (Harreman et al., 2009, Somesh et al., 2005). Surprisingly, utilising Rsp5 *in vitro* results in the formation of lysine-63-linked (K63) poly-ubiquitin chains, which are not detected on RNAPII *in vivo*. However, the Rsp5-associated de-ubiquitylation enzyme (DUB) Ubp2 has been shown to recognise K63-linked chains and trim them back to a single mono-ubiquitin moiety on Rpb1, probably antagonising the formation of unphysiological ubiquitin chains *in vivo* (Kee et al., 2005). Importantly, these data suggested that additional factors are necessary for productive poly-ubiquitylation.

Subsequent studies discovered a role for the Elongin-Cullin complex (Elc1, Ela1, Cul3 and Roc1 (Rbx1)) in Rpb1 ubiquitylation and degradation *in vivo*, suggesting it could be the main E3 ubiquitin ligase (Ribar et al., 2006, Ribar et al., 2007). Purification of the Elongin-Cullin complex and its use in biochemical assays has revealed that it is the second E3 ubiquitin ligase, acting in a two-step mechanism with Rsp5 (Harreman et al., 2009). A more complete model has now been elucidated in

which Rsp5 ubiquitylates Rpb1 following irreversible arrest, whilst Ubp2 trims back any K63-linked poly-ubiquitin chains to a single mono-ubiquitin moiety on Rpb1 (Figure 1.3, steps 1 and 2). The mono-ubiquitylated intermediate is then recognised by the second E3 ubiquitin ligase, the Elongin-Cullin complex, which mediates K48-linked poly-ubiquitylation, resulting in proteasome-mediated Rpb1 degradation (Figure 1.3, steps 3 and 4) (Harreman et al., 2009). This multi-step process likely helps ensure that RNAPII is only degraded as a last resort when transcription problems are encountered.

Another factor, Def1 (RNAPII degradation factor 1), is also essential for RNAPII poly-ubiquitylation, and like cells lacking subunits of the Elongin-Cullin E3 ligase, *def1*Δ mutants still contain mono-ubiquitylated Rpb1, showing they are not required for this Rsp5-mediated initial step (Woudstra et al., 2002, Somesh et al., 2005). The precise mechanism of Def1 function is currently unknown. However, it associates with RNAPII after DNA damage and can accelerate Rpb1 ubiquitylation *in vitro* (Somesh et al., 2005, Reid and Svejstrup, 2004).

Initiation of the ubiquitylation pathway does not necessarily result in degradation of RNAPII and there is in fact a proofreading element allowing reversal of poly-ubiquitylation. Another DUB, Ubp3, is capable of removing K48-linked chains and can even hydrolyse the bond between K330/K695 and the mono-ubiquitin moiety (Kvint et al., 2008). Therefore, Ubp3 can be thought of as a ubiquitylation proofreading protein and represents a checkpoint for ubiquitylation and degradation, allowing its reversal.

To summarise, bypass of DNA damage is inefficient and effectively requires the repair of the damaged DNA to allow the resumption of transcription. However, cells

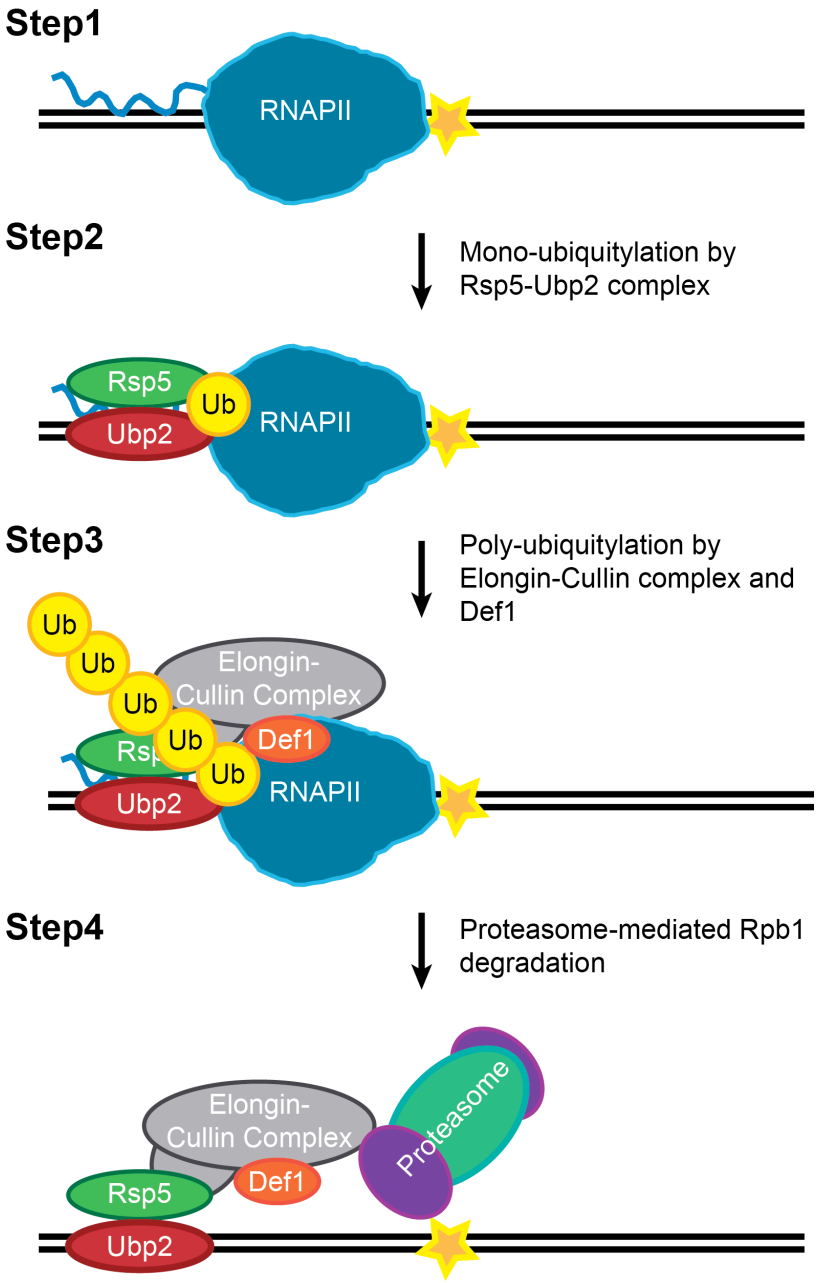

Figure 1.3 Ubiquitylation and degradation of Rpb1

RNAPII stalled at a damage lesion on DNA (yellow star) (**Step 1**) is mono-ubiquitylated by Rsp5, which is associated with Ubp2, a DUB able to trim back any K63-linked poly-ubiquitin chains (**Step 2**). This is followed by poly-ubiquitylation by the Elongin-Cullin complex, facilitated in an unknown manner by Def1 (**Step 3**). The K48-linked poly-ubiquitin chains target the Rbp1 subunit of RNAPII for proteasome-mediated degradation (**Step 4**).

have evolved a pathway to deal with arrested RNAPII, allowing it to be removed to allow more time for access of the repair machinery to the lesion.

1.4 Termination

Termination is the final stage of the transcription cycle and is essential for successful gene expression. It is in this stage that the nascent RNA is released and the ternary complex disrupted, allowing RNAPII to be recycled back to the promoter (of the same or other genes) for further rounds of transcription. Termination can be broadly divided into two sections, termination of poly(A) mRNA and termination of non poly(A) transcripts.

1.4.1 Termination of protein-coding genes

At the 3' end of a gene, RNAPII encounters regulatory elements required for processing of the nascent transcript and the subsequent events associated with termination. One such well-characterised sequence element is the poly(A) signal (PAS), which is degenerate in yeast but has the consensus sequence AAUAAA in humans (Proudfoot, 2011). The PAS has a major role in termination, as it is responsible for the initiation of 3' nascent mRNA processing. Upon transcribing this element the transcript is processed by endonucleolytic cleavage and stabilised via the addition of a poly(A) tail (of 70 nt in yeast and 200 nt in humans) (Bienroth et al., 1993). It is essential that during 3' processing, RNA cleavage and the subsequent polyadenylation occur at the correct positions in the nascent transcript. This is reliant upon other less conserved sequence recognition elements present at the PAS. At most genes they consist of three elements: an adenine-rich efficiency element with the consensus sequence TAYRTA, a pyrimidine-rich positioning element with the consensus AAWAAA (located

approximately 30 nt upstream of the cleavage site) and a uridine-rich element, which spans the cleavage position and the poly(A) addition site (Guo and Sherman, 1995, Bardwell et al., 1991, Dichtl and Keller, 2001, Valentini et al., 1999).

Two multimeric protein complexes are responsible for catalysing the cleavage and poly-adenylation of pre-mRNA and have subunits that interact with the PAS and its associated elements. These complexes, CPF (cleavage and polyadenylation factor) and CF (cleavage factor I A and B), are transferred from the Rpb1-CTD onto their substrate, the nascent mRNA, where cleavage of the RNA is catalysed by the Ysh1 subunit and the subsequent poly-adenylation catalysed by the poly(A) polymerase Pap1, which interacts with CPF (Keller and Minvielle-Sebastia, 1997, Ryan et al., 2004, Mandel et al., 2006, Zhelkovsky et al., 2006, Bienroth et al., 1993, Wahle, 1991, Kuhn et al., 2009, Ezeokonkwo et al., 2012). The association of these complexes with RNAPII is dependent on the Pcf11 subunit of CFI and its binding is highly co-ordinated. Interestingly, recent work has shed light on this regulation, showing that Pcf11 binds Ser2 phosphorylated Rpb1-CTD, but only in the absence of Tyr1 and Ser5 phosphorylation (Mayer et al., 2012). The presence of Tyr1 phosphorylation across the vast majority of the gene may therefore act as a safeguard to ensure against premature termination should RNAPII transcribe a sequence resembling the PAS.

Upon completion of 3' processing events, the pre-mRNA is converted into mature mRNA ready for export and translation. However, RNAPII is still associated with the DNA template and, surprisingly, actually continues producing transcript a long way past the PAS. Regulation and transcription of downstream genes can be affected by continued transcription of RNAPII into downstream promoters, so it must be removed from the DNA (Greger et al., 2000). Removal of RNAPII from the gene after transcript

cleavage is poorly understood, but must occur via EC destabilisation, with two main models having been proposed for how this could occur: the allosteric model and the torpedo model.

The allosteric model suggests that upon transcription of the PAS there is a change in the factors associated with RNAPII, specifically the loss of factors promoting processive elongation and binding of negative elongation factors (Logan et al., 1987). This model has been supported by studies in mammalian cells, which indicate that termination can be separated into two distinct stages: EC pausing and RNAPII dissociation (Park et al., 2004).

The torpedo model suggests that following cleavage of pre-mRNA and its poly-adenylation, a new 5' end is generated on the RNA still connected to the RNAPII active site (Connelly and Manley, 1988). As this transcript is un-capped, it could act as an entry point for either a 5'-3' exonuclease or helicase activity, which could move along the nascent RNA, catch up with the polymerase, and cause EC destabilisation. In support of this model, a 5'-3' exonuclease, Rat1, was found to promote termination in yeast by degrading the nascent transcript downstream of the PAS (Kim et al., 2004). Furthermore, inactivation of both Rat1 and its human homologue Xrn2 show termination defects, resulting in read-through of normal termination sites (Kim et al., 2004, West et al., 2004). For the dissociation factor to remove polymerase it is presumably necessary for the exonuclease/helicase to catch-up with it, which could be made possible by RNAPII pausing downstream from the transcript cleavage site. Thus, it seems likely that the actual mechanism of termination utilises a mixture of the allosteric and torpedo models. RNAPII likely pauses at and after the PAS, the transcript is cleaved and poly-adenylated, prior to Rat1 being loaded onto the nascent transcript,

which it degrades. Upon contact with RNAPII, Rat1 could cause the release of the remaining transcript, subsequently leading to EC destabilisation, in a manner similar to that of bacterial Rho factor (Epshtein et al., 2010). Alternatively, and arguably more closely related to the bacterial mechanism, the Sen1 RNA helicase, which has a role in non-coding RNA termination (at snoRNA genes), could also act as the EC destabilisation factor (Kawauchi et al., 2008, Rondon et al., 2009, Steinmetz et al., 2001). Further studies are necessary to fully comprehend this mechanism, but given the lack of dramatic termination phenotypes associated with mutation of any individual gene encoding a 'torpedo'- or EC-destabilising factor, it is probable that the interplay between multiple factors is required for efficient termination.

1.4.2 Termination of non-coding RNAs

RNAPII not only transcribes all protein-coding genes, but is also responsible for the expression of some stable non-coding RNAs. Processing of the latter transcripts differs from that of mRNA. Non-coding RNAs generally lack poly(A) tails and are subject to a different form of termination. Termination of small nucleolar RNAs (snoRNAs) is by far the most extensively characterised of the non-coding RNA termination processes.

SnoRNA termination occurs through the action of the trimeric NRD complex consisting of: Nrd1, Nab3 and Sen1 (Rasmussen and Culbertson, 1998, Steinmetz et al., 2001). Both Nrd1 and Nab3 have a role in RNA binding and contain RNA recognition motifs (RRMs) (Steinmetz and Brow, 1996, Hobor et al., 2011). Additionally, Nrd1 is responsible for the association of the NRD complex with RNAPII, via its CTD-interacting domain (CID), which can associate with the Rpb1-CTD (Vasiljeva et al.,

2008). The third subunit, Sen1, is an RNA helicase, as mentioned previously (*see section 1.4.1*).

Sites of snoRNA termination are defined by arrays of two motifs: the first, UCUU, forms a binding site for Nab3, whilst the second, GUA(A/G), binds Nrd1 (Carroll et al., 2004). Following interaction of the NRD complex with the nascent RNA, it has been suggested that the RNA is processed by the NRD-associated exosome complex (Carroll et al., 2007, Carroll et al., 2004, Vasiljeva and Buratowski, 2006). Finally, in a similar manner to that proposed for the torpedo model, Sen1 acts as a Rho-like factor and is able to disrupt the EC, leading to RNAPII dissociation (Chinchilla et al., 2012).

Although the proteins involved in ncRNA termination are different from those involved in termination of poly-adenylated transcripts, they most certainly share a common theme. It is also very likely that there is redundancy between the two pathways, probably providing a failsafe mechanism for termination (Kawauchi et al., 2008, Rondon et al., 2009).

1.5 The transcriptional landscape

The transcriptome is made up of two main classes of RNA: mRNA, which forms the template for protein synthesis, and non-coding RNA (ncRNA), a large heterogeneous group of RNA, much of which has no known function. This section will focus on the organisation of the genome into distinct gene units, before discussing pervasive transcription, and the manner in which the two classes of transcripts and the RNAPII molecules producing them may interact.

1.5.1 Orientation of protein-coding genes

Gene units arranged on DNA can be divided into pairs and the arrangement classified by the spatial orientation of the genes within the pair. Gene pairs can be arranged tandemly, divergently, or convergently (Figure 1.4). Tandem pairs are arranged in a head-to-tail configuration, with the terminator of one gene being spatially close to the promoter of the next (Figure 1.4A). In this configuration the transcribed (sense) strands of the two genes are the same. Divergent pairs are arranged head-to-head so that the promoters are spatially close, each being transcribed away from the other on the opposite DNA strand (Figure 1.4B). Finally, convergent genes are arranged in a tail-to-tail configuration, with RNAPII molecules from each gene transcribing towards the other, on opposite strands of DNA (Figure 1.4C).

The orientation of genes has implications for their transcription and regulation. For example, many expression profile studies have shown that neighbouring genes are co-regulated. Divergent and tandem gene pairs show strong positive expression correlation (i.e. the pair are transcribed simultaneously) (Kraakman et al., 1989, Nakao et al., 1986, Osley et al., 1986, West et al., 1984, Cohen et al., 2000), whilst convergent pairs often negatively affect each other's expression (Xu et al., 2011, Hongay et al., 2006). Furthermore, as a result of the compact nature of the yeast genome, many of the genes that are arranged convergently overlap at their 3' ends. Presumably, simultaneous expression of genes would result in RNAPII interaction and possibly collision, likely affecting the transcripts produced. A final gene class, observed mainly in higher eukaryotes, is nested genes (genes-within-genes), which are embedded in the body of another gene in either a sense or antisense orientation (Figure 1.4D, an antisense embedded gene). An example of an antisense nested gene is *F8A* (Factor VIII-

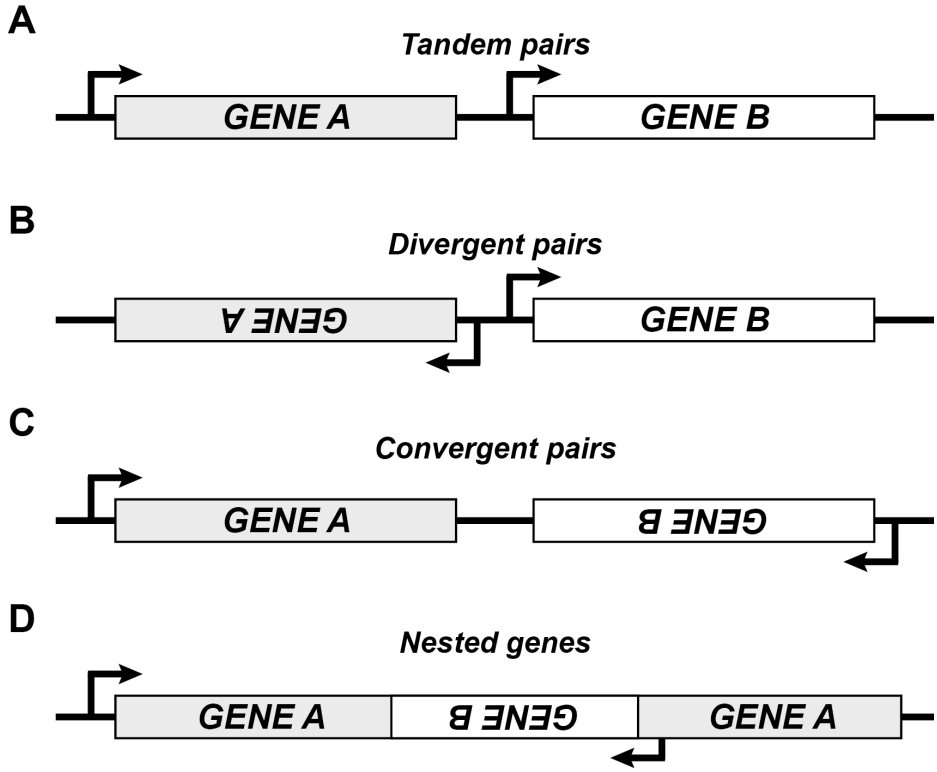

Figure 1.4 Orientation of gene pairs.

A. Schematic showing *GENE A* (grey) and *GENE B* (white) arranged as a tandem gene pair. Arrow indicates promoter position and direction of transcription. **B.** *GENE A* and *GENE B* arranged divergently. **C.** *GENE A* and *GENE B* arranged convergently. **D.** A nested gene pair, showing *GENE B* within *GENE A*, in a convergent orientation.

associated gene A), which encodes HAP40, a huntingtin-associated protein. *F8A* is found in intron 22 of the *Factor VIII* gene, and transcribed convergently to it (Levinson et al., 1990, Peters and Ross, 2001).

1.5.2 Non-coding RNA production

The second class of transcripts is the ncRNAs. These include RNAPI and RNAPIII-synthesised transcripts, such as ribosomal RNAs and tRNAs, but the focus here will be on RNAPII-synthesised ncRNAs, which includes snoRNAs, but also multiple other transcripts, often of unknown function. With the advent of micro-array analysis and subsequently next-generation sequencing, came the discovery of a large number of un-annotated transcripts, which had previously been disregarded as artefacts. Indeed, the genomes of diverse organisms, such as yeast (David et al., 2006, Dutrow et al., 2008, Nagalakshmi et al., 2008, Wilhelm et al., 2008), plants (Li et al., 2006), *Drosophila* (Stolc et al., 2004) and mammals (Bertone et al., 2004, Carninci et al., 2005, He et al., 2008) show extensive transcription, not resulting in the production of a stable transcript. This transcription is disproportionate to the number of protein-coding genes present, in *S. cerevisiae* which contains 5,654 open reading frames (ORFs), up to 85% of the entire genome is transcribed (David et al., 2006). These additional transcripts were found to arise from transcription of intergenic regions, as well as from overlapping transcription within genes (David et al., 2006, Xu et al., 2009). ncRNAs can overlap extensively with protein-coding genes and these can be transcribed in a sense (tandem) or antisense (convergent) orientation in relation to the gene. Genome-wide studies have highlighted the prevalence of this, one showing 50,111 overlapping transcript pairs in the human genome and another that up to 88% of protein-coding ORFs display

concurrent antisense transcription (Katayama et al., 2005, Vallon-Christersson et al., 2007).

The ncRNAs produced from this pervasive transcription can be divided into two main classes: short ncRNAs, or long ncRNAs. Short ncRNAs include short-interfering RNAs and micro RNAs, which often have *trans* regulatory roles, related to RNA interference in higher eukaryotes (Jacquier, 2009). In yeast the long ncRNAs are comprised of two groups: the stable un-annotated transcripts (SUTs) and the cryptic unstable transcripts (CUTs). These two classes of ncRNA differ from one another in several respects. Firstly, SUTs (as their name would imply) are stable, whereas CUTs are rapidly degraded and only detectable in exosome mutants, such as *rrp6*Δ (Xu et al., 2009). Secondly, there is a clear difference in length between the two groups: CUTs are typically 200-600 nts long and have a heterogeneous 3' end, whilst SUTs are much longer with a median size of 760 nts (Xu et al., 2009). Frequently, CUTs and SUTs are found to initiate from nucleosome-free regions (NFRs) associated with promoters and 3' regions of genes (Xu et al., 2009, Neil et al., 2009). Strikingly, this was found to be a result of the inherent bi-directionality, characteristic of eukaryotic promoters (Xu et al., 2009, Neil et al., 2009). Therefore, many initiation events result in RNAPII firing in both directions away from the promoter, transcribing opposite strands of DNA. Additional ncRNA production from 3'NFRs is likely to result from altered chromatin structure following gene transcription. In support of this, cells lacking the gene for the chromatin remodelling protein Isw2 display increased SUT and CUT production, owing to enhanced 3' nucleosome depletion (Whitehouse et al., 2007).

The function of many ncRNAs is, however, unknown, but some studies have shown that it is the act of transcribing the ncRNA itself that has a function, in many

cases serving to regulate neighbouring genes (Martens et al., 2004, Hongay et al., 2006, Xu et al., 2011, Martens et al., 2005). A gene involved in the serine biosynthesis pathway, *SER3*, is subject to this form of *cis* regulation by a ncRNA, *SRG1* (Martens et al., 2004). *SRG1* is transcribed from upstream of the *SER3* promoter, in a tandem (sense) orientation to the gene. The act of transcribing *SRG1* was shown to occlude the *SER3* promoter, preventing the binding of activators/initiation factors, and ultimately, repressing *SER3* expression (Martens et al., 2004).

Similarly, the *IME4* locus in yeast is also regulated by the production of a ncRNA. Expression of the *IME4* gene is necessary to commit a cell to meiosis. Therefore it is, necessarily, tightly repressed in haploid cells and temporally regulated in diploids. Repression of *IME4* in haploids was found to result from antisense transcription across the gene (Hongay et al., 2006). The diploid-specific a1/ α2 protein was shown to be able to bind and inhibit production of the antisense transcript, when activation of *IME4* is required. The authors suggest that the mechanism of repression by the ncRNA could be due to transcriptional interference, either through RNAPII interactions or promoter occlusion. In this case, transcription of the antisense ncRNA could be providing a mechanism for fine-tuning expression of *IME4*, ensuring that the gene is tightly repressed when necessary. There is likely to be an evolutionary advantage to this, as untimely expression of a gene, particularly one committing a cell to a developmental fate, could ultimately be fatal.

In addition to the repressive role of ncRNA transcription, it has also been shown to have an activating function. This is certainly the case at the phosphate biosynthesis pathway gene, *PHO5*. *PHO5* was found to have an associated antisense ncRNA, transcribed when the gene is in a repressed state (Uhler et al., 2007). This study found

that histone eviction at the *PHO5* promoter is associated with antisense RNAPII transcription, which enhances chromatin plasticity. In support of this mechanism, abrogating transcription of the ncRNA led to a delay in both chromatin remodelling and RNAPII recruitment to the *PHO5* promoter. Therefore, this suggests a role for ncRNA transcription at genes that are required to respond rapidly to external cues. Indeed, this was found to be the case in a recent genome-wide study from the Steinmetz laboratory (Xu et al., 2011). Antisense expression was found at genes with a large expression range, i.e. they show high levels of expression when active, and very low levels of expression when repressed. Moreover, stress-response and environment-specific genes, which are required to respond in a switch-like manner, are characterised by high levels of antisense transcription. Altogether, this suggests that ncRNA, and especially antisense ncRNA, can serve to regulate genes by a variety of methods.

Altogether, numerous recent studies indicate that transcription is surprisingly pervasive and that the transcriptional landscape is highly dynamic and inter-regulated. The orientation and spatial arrangement of genes has an impact on their expression, whilst the abundance of ncRNAs, and specifically the act of transcribing them, suggests that they are important for gene regulation. Furthermore, it highlights a gap in our knowledge of ncRNA production, at the level of transcript elongation.

1.6 Aims and scope of this thesis

1.6.1 Relevant background

There has been very little research into gene traffic and progression of RNAPII on busy coding regions, where multiple DNA metabolic events are occurring concurrently. The usual reductionist methods tend to treat transcription in isolation, but it is

affected by and affects other events occurring on DNA, possibly the best-studied being the interactions between transcription and replication machineries (Pomerantz and O'Donnell, 2008, Pomerantz and O'Donnell, 2010), or between transcription and DNA repair (Selth et al., 2010). Studying transcript elongation in the context of other DNA events is important to gain a more comprehensive understanding of the process *in vivo*.

Recent research in the Svejstrup laboratory has contributed to the study of gene traffic by analysing interactions between tandemly transcribing RNAPII molecules, mimicking events occurring on highly transcribed genes (Saeki and Svejstrup, 2009). The findings of the study were numerous and far-reaching, yielding mechanistic and conceptual insights, with implications for the processivity of transcription *in vivo*. The experiments were all performed with a highly purified RNAPII scaffold system allowing reconstitution of transcription with two polymerases transcribing in the same direction using pure RNAPII, DNA- and RNA oligonculeotides. The results obtained showed that RNAPII is conformationally flexible so that when a leading EC was stalled on DNA (mimicking arrest at a damage site or other elongation impediment) the trailing RNAPII substantially invaded the space of the leading polymerase, ultimately resulting in significant backtracking of the trailing RNAPII molecule upon collision. Most interestingly, when the leading RNAPII was paused at a poly-A/T tract, the trailing RNAPII was able to push the former through the transcription block. This finding has important implications for our understanding of elongation *in vivo* and suggests that tandemly transcribing RNAPIIs are able to confer increased processivity on each other in the face of transcriptional pauses. The latter findings are also consistent with those observed for prokaryotic RNAP (Epshtein and Nudler, 2003).

While the studies described above focused on head-to-tail transcription, little is known about convergent transcription (head-to-head collision) and its mechanistic basis. The finding that RNAPII produces numerous antisense, non-coding RNAs, raises important questions when considering their production. The transcription of antisense RNAs often occurs from 3' ends of active genes and most yeast promoters are bi-directional (Neil et al., 2009). This would undoubtedly lead to interactions between promoter-driven ECs and ECs originating from 3' cryptic initiation events. Likewise, interactions would occur between RNAPII terminating at the 3' end of tandem genes and bi-directional transcripts originating from downstream promoters. However, undoubtedly the largest source of RNAPII interactions would be between convergent genes, which in many cases have no terminator (and/or only a very short distance) between them.

Convergent transcription was the subject of a recent biochemical study using T3 and T7 bacteriophage polymerases (Ma and McAllister, 2009). The main finding was that a T7 polymerase can bypass a T3 polymerase stalled on the DNA template and that following this, the T3 polymerase is able to continue transcription. The authors suggest that there is a temporary release of the NTS from the colliding polymerases, allowing transcriptional bypass. The findings of this study have important implications for convergent transcription in eukaryotes, in which the core of polymerase and its nucleic acid interactions may be similar (Durniak et al., 2008, Cramer et al., 2001, Gnatt et al., 2001). From a nucleic acid-centric perspective there should thus be no impediment to bypass, which in theory requires only a widening of the transcription bubble (Figure 1.5A). However, this is more difficult to envisage when looking at a structure of the elongation complex, with RNAPII being of significant size and structural complexity

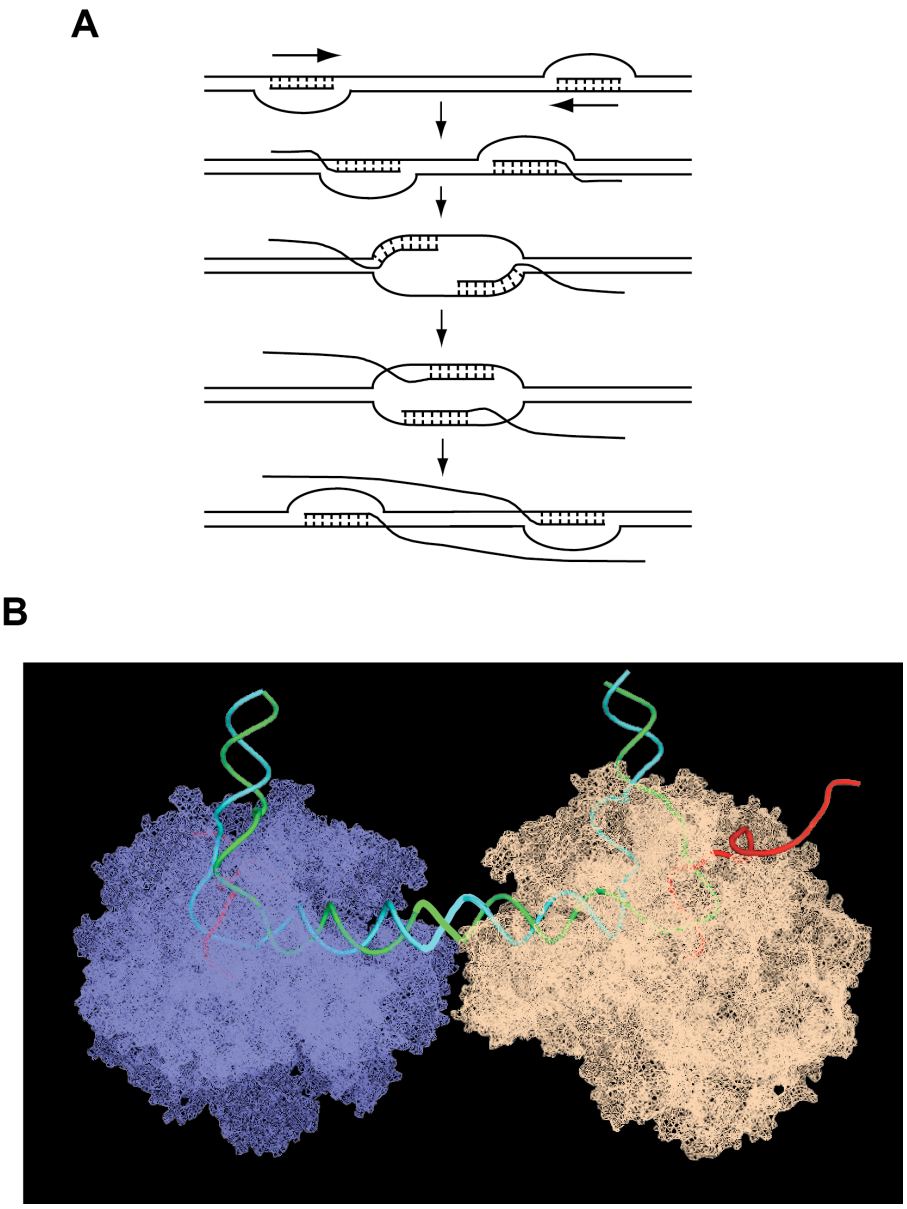

Figure 1.5 Representations of convergent transcription.

A. Schematic showing nucleic-acid centric view of transcription. Bypass could occur by widening of the transcription bubble. **B.** Protein-centric view of elongation, showing a crystallographic model of convergently transcribing RNAPII molecules (Gnatt et al., 2001). DNA (blue and green) and RNA (red) are represented.

(Figure 1.5B). This is in stark contrast to the single-subunit bacteriophage polymerases, which are much smaller and share only the conserved core, lacking the outer protein density (Durniak et al., 2008). Interestingly, this may be an indication that interactions between convergent polymerases could be very different in the two systems.

In support of a difference between bacteriophage RNAP and eukaryotic RNAPII, an *in vivo* study from the Proudfoot laboratory found that *GAL* genes orientated convergently without terminators between them produced only truncated transcripts and showed a severe reduction in steady-state RNA levels, presumably at least partly as a result of RNAPII collision (Prescott and Proudfoot, 2002). Additionally, control constructs arranged in a tandem orientation did not result in reduced RNA levels, arguing against RNA-interference effects. However, a limitation of this study was the inability to distinguish between promoter occlusion and elongation effects arising from convergent transcription. For a better understanding of convergent transcription, it needs to be studied mechanistically, both *in vivo* and *in vitro*, at the level of transcript elongation.

1.6.2 Objectives of this work

There is a distinct lack of understanding concerning the fundamental nature of convergent transcription. This shortfall is highlighted in the literature (*see section 1.5*). Numerous studies discuss the production of antisense ncRNA, which is undoubtedly, in some cases, transcribed concurrently with sense gene mRNA (Vallon-Christersson et al., 2007). Additionally, the presence of overlapping convergent transcription units and nested genes are acknowledged (*see section 1.5.1*). However, no studies have addressed the mechanistic basis of this convergent transcription at the level of RNAPII elongation and surprisingly, none have attempted to address the absolutely fundamental question:

can convergently transcribing RNAPII molecules transcribe past one another? Either answer to this question will necessarily yield exciting new insights into transcript elongation on busy genomic regions. If RNAPII cannot bypass then what happens to the collided polymerases? If it can bypass, then by what mechanism? What effect does this event have on sense-antisense transcription, or transcription at convergent gene pairs?

This thesis aims to address these issues: firstly by utilising an *in vitro* biochemical assay to determine the outcome of RNAPII convergent transcription and then building upon these findings *in vivo* with a convergent gene construct and later genome-wide studies. The data acquired in each system fully complements one another and has provided insight into the outcome of RNAPII convergent transcription, with important implications for the understanding of sense-antisense transcription *in vivo*.

Chapter 2. Materials & Methods

2.1 Buffers, media and solutions

2.1.1 Yeast media

2.1.1.1 YPD

- 1% Yeast Extract (DIFCO)
- 1% Peptone (DIFCO)
- 2% Glucose

Supplemented with adenine to a final concentration of 40 μg/ml.

2.1.1.2 Synthetic Complete Medium

- 2% Glucose or raffinose
- 6.7 mg/ml Yeast Nitrogen Base (DIFCO)
- 1.4 mg/ml Yeast Synthetic Drop-out Medium Supplement without Histidine, Leucine, Tryptophan and Uracil (Sigma)
- 40 μg/ml Tryptophan
- 40 μg/ml Histidine
- 80 μg/ml Leucine
- 40 μg/ml Uracil
- 24 μg/ml Adenine

To make media selective for yeast carrying a marker gene required for its synthesis, that particular amino acid was left out of the mixture. Selective plates were made in the same way, except that water was replaced with 1.6% agar.

2.1.2 Bacterial media

2.1.2.1 LB (rich medium)

- 1% Bacto Tryptone (DIFCO)
- 0.5% Yeast Extract (DIFCO)
- 1% NaCl

pH adjusted to 7.

- -/+ 100 μg/ml Ampicillin (Melford Laboratories)

2.1.2.2 SOC (heat-shock recovery medium)

- 2 % Bacto Tryptone (DIFCO)
- 0.5% Yeast Extract (DIFCO)
- 10 mM NaCl
- 2.5 mM KCl
- 10 mM $MgCl_2$
- 10 mM $MgSO_4$
- 20 mM Glucose

pH adjusted to 7.

2.1.3 General solutions

2.1.3.1 PBS (phosphate-buffered saline)

- 130 mM NaCl
- 7 mM Na_2HPO_4
- 2 mM NaH_2PO_4

Adjusted to pH 7.5.

2.1.3.2 TBS (Tris-buffered saline)

- 50 mM Tris-Cl (pH 7.6)
- 150 mM NaCl

2.1.3.3 TE (Tris-EDTA)

- 10 mM Tris-Cl pH 7.5
- 1 mM EDTA pH 8.0

This was also made up as a 10 x solution.

2.1.3.4 TE low EDTA

- 10 mM Tris-Cl pH 7.5
- 0.1 mM EDTA pH 8.0

2.1.3.5 TBE (Tris-Borate-EDTA)

- 89 mM Tris Base
- 89 mM Boric Acid
- 2 mM EDTA

2.1.3.6 TE/LiOAc

- 1 x TE pH 7.5
- 100 mM Lithium acetate

2.1.3.7 PEG/TE/LiOAc

- 1 x TE pH 7.5
- 100 mM Lithium acetate
- 50% PEG (3350)

2.1.3.8 10 x Agarose gel DNA loading Buffer

- 20 mM EDTA pH 8.0
- 50% Glycerol
- 0.05% Bromophenol blue

2.1.3.9 5x SDS-PAGE loading buffer

- 250 mM Tris-Cl pH 6.8
- 500 mM DTT
- 10% SDS
- 0.5% Bromophenol blue
- 50% Glycerol

2.1.3.10 Denaturing PAGE loading buffer

- 95% De-ionised formamide (Sigma)
- 5 mM EDTA pH 8.0
- 0.004% Bromophenol blue
- 0.004% Xylene cyanol

2.1.3.11 100 x Protease inhibitor cocktail

- 28.4 µg/ml Leupeptin
- 137 µg/ml Pepstatin A
- 17 µg/ml PMSF
- 33 µg/ml Benzamidine

All components dissolved in 100% ethanol.

2.1.4 Buffers for RNAPII purification

2.1.4.1 Yeast Lysis Buffer

- 50 mM Tris-Cl (pH 8.0)
- 150 mM KCl
- 10% Glycerol
- 10 µM $ZnCl_2$
- 1 mM EDTA
- 10 mM DTT
- 1 x Protease inhibitors

2.1.4.2 HSB150 (Heparin sulphate chromatography)

- 50 mM Tris-Cl (pH 8.0)
- 150 mM KCl
- 1 mM EDTA
- 10% Glycerol
- 10 mM DTT

- 10 µM $ZnCl_2$
- 1 x Protease inhibitors

2.1.4.3 HSB600

- 50 mM Tris-Cl (pH 8.0)
- 600 mM KCl
- 1 mM EDTA
- 10% Glycerol
- 10 mM DTT
- 10 µM $ZnCl_2$
- 1 x Protease inhibitors

2.1.4.4 TEZ_0

- 50 mM Tris-Cl pH 7.5
- 1 mM EDTA
- 10 µM $ZnCl_2$
- 10 mM DTT
- 1 x Protease inhibitors

2.1.4.5 TEZ_{500}

- 50 mM Tris-Cl pH 7.5
- 500 mM Ammonium sulphate
- 1 mM EDTA
- 10 µM $ZnCl_2$

- 10mM DTT
- 1 x Protease inhibitors

2.1.4.6 $TEZ_{Elution}$

- 50 mM Tris-Cl pH 7.5
- 500 mM Ammonium sulphate
- 50% Glycerol
- 1 mM EDTA
- 10 µM $ZnCl_2$
- 10 mM DTT
- 1 x Protease inhibitors

2.1.4.7 MonoQ Dilution Buffer

- 50 mM Tris-Cl pH 7.5
- 10% Glycerol
- 0.5 mM EDTA
- 10 µM $ZnCl_2$
- 5 mM DTT
- 1 x Protease inhibitors

2.1.4.8 MonoQ200

- 50 mM Tris-Cl pH 7.5
- 200 mM KCl
- 10% Glycerol

- 0.5 mM EDTA
- 10 μM $ZnCl_2$
- 5 mM DTT
- 1 x Protease inhibitors

2.1.4.9 MonoQ1000

- 50 mM Tris-Cl pH 7.5
- 1 M KCl
- 10% Glycerol
- 0.5 mM EDTA
- 10 μM $ZnCl_2$
- 5 mM DTT
- 1 x Protease inhibitors

2.1.4.10 RNAPII Buffer

- 20 mM HEPES-KOH pH7.6
- 100 mM KOAc
- 0.1 mM EDTA
- 10% Glycerol
- 2 mM DTT
- 0.1 mM PMSF

2.1.5 Buffers for reconstitution and analysis of elongation *in vitro*

2.1.5.1 Reconstitution Buffer

- 20 mM Tris-Cl (pH 7.9)
- 40 mM KCl
- 0.2 mM $MgCl_2$
- 20 μM $ZnCl_2$
- 5 mM DTT

2.1.5.2 Transcription Buffer

- 20 mM Tris-Cl (pH 7.9)
- 40 mM KCl
- 7 mM $MgCl_2$
- 20 μM $ZnCl_2$
- 5 mM DTT
- 0.75 mg/ml BSA (NEB)

2.1.5.3 EC Wash Buffer

- 50 mM Tris-Cl pH 7.5
- 500 mM NaCl
- 0.5 mM EDTA
- 10 μM $ZnCl_2$
- 0.05% NP-40
- 10% Glycerol

2.1.5.4 FLAG-EB

The same recipe as transcription buffer but with 300 µg/ml FLAG peptide added.

2.1.5.5 Stop Buffer

- 20 mM Tris pH7.9
- 40 mM KCl
- 20 µM $ZnCl_2$
- 5 mM DTT
- 20 mM EDTA
- -/+ 1 mg/ml Proteinase K

2.1.5.6 Native agarose loading buffer

- 3% Glycerol
- 0.75 mg/ml BSA (NEB)
- 2.5 mM 2-mercaptoethanol

2.1.5.7 Native agarose gel

- 0.7% NuSieve GTG Agarose (Lonza)
- 22.25 mM Tris base
- 22.25 mM Boric acid
- 5 mM 2-mercaptoethanol
- 0.1 mM $MgCl_2$
- 10 µM $ZnCl_2$

2.1.5.8 Urea denaturing PAGE

- 1 x TBE
- 8.3 M Urea
- 0.07% TEMED
- 0.05% APS

The correct percentage gel was prepared by adding 40% Acrylamide/Bis Solution (19:1) (BioRad).

2.1.6 Buffers for analysis of transcription *in vivo*

2.1.6.1 Buffer Y1

- 1 M Sorbitol
- 100 mM EDTA pH 7.4

2.1.6.2 Spheroplasting buffer

- Buffer Y1
- 0.1% 2-mercaptoethanol
- 100 U Zymolase-20T (MP Biomedicals)

2.1.7 Buffers for ChIP

2.1.7.1 FA-Lysis

- 50 mM Tris-Cl (pH 7.5)
- 140 mM NaCl
- 1 mM EDTA

- 1% Triton X-100
- 0.1% Na-deoxycholate

2.1.7.2 FA-500

- 50 mM Tris-Cl (pH 7.5)
- 500 mM NaCl
- 1 mM EDTA
- 1% Triton X-100
- 0.1% Na-deoxycholate

2.1.7.3 ChIP-WB

- 10 mM Tris-Cl (pH 8.0)
- 250 mM LiCl
- 0.5% NP-40
- 1 mM EDTA
- 0.5% Na-deoxycholate

2.1.7.4 TES

- 10 mM Tris-Cl (pH 7.5)
- 1mM EDTA
- 100 mM NaCl

2.1.7.5 ChIP-Elution

- 50 mM Tris-Cl (pH 8.0)

- 10 mM EDTA
- 1% SDS

2.2 Bacterial techniques

2.2.1 Transformation of chemically-competent *E. coli*

TOP10 cells (Invitrogen) were used for transformations with plasmid and newly-ligated DNA. Typically 200 ng of plasmid DNA or 2 µl of ligation mixture were added to bacteria and incubated for 5 minutes on ice prior to heat-shocking for 30 seconds at 42 °C. Cells were placed on ice to recover and 250 µl of SOC medium (*see section 2.1.2.2*) added. Cells were incubated at 37 °C with shaking and plated onto pre-warmed LB-Amp plates.

2.2.2 Plasmid mini-prep

Typically 2 ml bacterial cultures were grown in LB-Amp (*see section 2.1.2.1*) liquid culture overnight at 37 °C and plasmid DNA extracted using QIAprep Miniprep Kit (QIAGEN) following standard procedures.

2.2.3 Expression of recombinant proteins

Bacteria carrying a plasmid for protein expression were grown in LB-Amp to an OD_{600} density of 0.7, 1mM IPTG added and transferred to a 30 °C incubator for 3 hours to induce protein expression. Bacteria were harvested by centrifugation and lysed by sonication.

2.3 DNA techniques

2.3.1 Polymerase Chain Reaction (PCR)

PCRs were performed using the KOD Hot Start DNA Polymerase Kit (Novagen) following standard cycling procedures. A typical PCR reaction was set up as follows:

- 1x KOD reaction buffer
- 1.5 mM $MgSO_4$
- 0.2 mM dNTPs
- 0.3 µM Sense primer
- 0.3 µM Antisense primer
- 0.22 ng/µl Template DNA
- 0.02 U/µl KOD Hot Start DNA Polymerase

2.3.2 Quantitative PCR (qPCR)

qPCR reactions were prepared in a 20 µl reaction volume as follows:

- 1 x iQ Custom SYBR Green SuperMix (BioRad)
- 0.2 µM Sense primer
- 0.2 µM Antisense primer
- µl DNA (from ChIP)

The reactions were then placed in a CFX-96 Real-Time System (BioRad) and standard cycling conditions used.

2.3.3 Restriction digestion

Restriction digests were performed with enzymes obtained from NEB and generally the reaction was allowed to continue overnight at 37 °C. The following reaction mixture is typical:

- 1 x NEB reaction buffer (1-4 dependent on enzyme used)
- 1 µg DNA
- 1 µl Restriction enzyme
- 1 µl Restriction enzyme 2 (optional)
- dH_2O added up to 50 µl

2.3.4 DNA ligations

DNA ligations were usually set up in a volume of 10 µl with 100 ng of vector, a 3 times molar excess of insert, 1 x reaction buffer (Roche) and 1 µl T4 DNA ligase (Roche). The reaction was allowed to proceed for 1 hour at 25 °C prior to transformation.

2.3.5 DNA purification

Typically DNA was purified following PCR, restriction digest and ChIP using QIAquick PCR Purification Kit (QIAGEN) following the manufacturers recommendations. In the case of ChIP DNA purification, standard procedures were followed but DNA was eluted from the column with 100 µl of nuclease-free dH_2O.

2.3.6 Sequencing of plasmid DNA

Plasmid DNA was sequenced at the Cancer Research UK Sequencing Facility using standard methods. 20 µl sequencing reactions were set up as follows:

- 3 µl BigDye Terminator 5x Sequncing Buffer (Applied Biosystems)
- 2 µl BigDye Terminator (Applied Biosystems)
- 0.64 µl Primer (3.2 pmol final concentration)
- 200 ng Plasmid DNA

The reaction took place in a thermal cycler, using the conditions indicated (Table 2.1).

Table 2.1 Cycling conditions used during the sequencing reaction.

STEP NUMBER	TEMPEARATURE (°C)	TIME
1	96	1 minute
2	96	10 seconds
3	50	5 seconds
4	60	4 minutes
GO TO STEP 2 (x 25)		
5	4	Forever

2.3.7 De-proteinisation of DNA

DNA solution was mixed with an equal volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) saturated with 10 mM Tris-Cl pH 8.0, 1mM EDTA (Sigma) and isolated using MaXtract High Density phase-lock tubes (QIAGEN) following the manufacturer's recommendations.

2.3.8 Ethanol precipitation of DNA

DNA was ethanol-precipitated by adding 300 mM NaOAc (pH 5.2), -/+ 5 µg glycogen and 2 times volume of ice-cold absolute ethanol. The mixture was placed at

minus 20 °C overnight prior to centrifugation and removal of supernatant. Pellets were washed with 70% ethanol and resuspended in TE (*see section 2.1.3.3*).

2.3.9 De-salting of DNA

DNA desalting was performed using MicroSpin G-25 columns (GE Healthcare) following the manufacturer's recommendations.

2.3.10 Standard agarose gel electrophoresis

Horizontal agarose gel electrophoresis was used to separate DNA of different sizes. The agarose content was dependent on the size of fragments being resolved, with 0.7-2% commonly used. Agarose was added to 1 x TBE (*see section 2.1.3.5*) and heated until dissolved. Once cool, ethidium bromide (Sigma) was added to a final concentration of 0.5 μg/ml and the gel poured. DNA in 1 x Agarose gel DNA loading buffer (*see section 2.1.3.8*) was loaded into the wells. The buffer for electrophoresis was 1 x TBE and the voltage used typically 100 V.

2.3.11 Purification of DNA from agarose gels

DNA was visualised on a UV transilluminator and a scalpel used to isolate the appropriate DNA band. QIAquick Gel Extraction Kit (QIAGEN) and standard protocols were used to extract the DNA from the agarose.

2.3.12 ^{32}P-end labelling of DNA oligonucleotides

DNA oligonucleotides were 5' end-labelled using ^{32}P-γATP in the following reaction mixture:

- 20 pmol DNA

- 1 x T4 polynucleotide kinase buffer (NEB)
- 63 μCi ^{32}P-γATP (Perkin-Elmer)
- 10 U T4 polynucleotide kinase (NEB)

The reaction was allowed to proceed at 37 °C for 30 minutes, then another 10 U of T4 polynucleotide kinase added for another 30 minutes at 37 °C. The reaction was stopped by the addition of EDTA to a final concentration of 20 mM and deproteinised by phenol-chloroform extraction (*see section 2.3.7*). Unincorporated ^{32}P-γATP was removed by G-25 de-salting column (*see section 2.3.9*) followed by ethanol precipitation (*see section 2.3.8*), this process was repeated and ^{32}P-end labelled DNA resuspended in TE low EDTA (*see section 2.1.3.4*).

2.3.13 Purification of DNA oligonucleotides via denaturing PAGE

Full-length DNA oligonucleotides were purified from shorter intermediates of incomplete synthesis using denaturing PAGE. A 5.2% urea polyacrylamide gel was placed in the Sequi-Gen GT Sequencing Cell (BioRad) and used to resolve DNA, which was visualised by UV-shadowing and isolated with a scalpel. The gel slice was macerated and eluted using TE at 25 °C overnight, followed by a second elution at 37 °C for 3 hours. The two eluates were pooled and passed twice through 0.2 μm NANOSEP MF filtration columns (Pall). The DNA solution was then concentrated by butanol extraction (using standard techniques) followed by phenol-chloroform extraction (*see section 2.3.7*) and ethanol precipitation (- glycogen) (*see section 2.3.8*). DNA was finally de-salted (*see section 2.3.9*) and subjected to another round of ethanol precipitation and resuspension in TE low EDTA (*see section 2.1.3.4*).

To test the purity of the isolated DNA oligonucleotide, it was ^{32}P-end labelled (*see section 2.3.12*) and resolved by 5.2% denaturing PAGE.

2.4 RNA techniques

2.4.1 Extraction of total RNA from yeast cells

Yeast cells were grown to mid-log phase and where necessary, galactose or glucose added. Cells were harvested and washed in ice-cold TBS prior to resuspension in spheroplasting solution (*see section 2.1.6.2*) for 10 minutes at 30 °C with agitation. Spheroplasts were lysed and RNA extracted using the RNeasy Kit (QIAGEN) following the manufacturer's recommendations.

2.4.2 De-proteinisation of RNA

De-proteinisation of RNA was performed in the same manner as for DNA (*see section 2.3.7*).

2.4.3 Ethanol precipitation of RNA

RNA was precipitated by the addition of 3 M NaOAc pH 5.2 to a final concentration of 300 mM, -/+ 5 µg glycogen and 3 times volume of ice-cold absolute ethanol. Samples were incubated at -20 °C overnight prior to centrifugation and washing twice with 80% ethanol. Finally, pellets were resuspended in TE low EDTA (*see section 2.1.3.4*) or for direct analysism in formamide gel loading buffer (*see section 2.1.3.10*).

2.4.4 ^{32}P-end labelling of RNA oligonucleotides

RNA oligonucleotides were ^{32}P-end labelled in the same manner as DNA oligonucleotides (*see section 2.3.12*), however, the ethanol precipitation was carried out as outlined for RNA (*see section 2.4.3*).

2.4.5 Purification of RNA oligonucleotides via denaturing PAGE

RNA oligonucleotides were purified in a manner similar to DNA oligonucleotides (*see section 2.3.13*), except that electrophoresis was on a 20% denaturing gel and ethanol precipitation was performed as outlined for RNA (*see section 2.4.3*).

2.5 Yeast techniques

2.5.1 Yeast strains used in this study

All *in vivo* experiments were performed with *S. cerevisiae* strains congenic to W303 (Thomas and Rothstein, 1989) and include *elc1*Δ*::URA3* and *def1*Δ*::URA3* (Harreman et al., 2009, Woudstra et al., 2002), full details of the strains are listed below (Table 2.2).

Table 2.2 Yeast strains used in this study.

STRAIN	GENOTYPE
W303-1A	*MAT***a** *leu2-3,112 his3-11,15 ade2-1 ura3-1 trp1-1 can1-100*
W303-1B	*MAT*α *leu2-3,112 his3-11,15 ade2-1 ura3-1 trp1-1 can1-100*
*elc1*Δ	*MAT***a** *leu2-3,112 his3-11,15 ade2-1 ura3-1 trp1-1 can1-100 elc1*Δ*::URA3*
*def1*Δ	*MAT***a** *leu2-3,112 his3-11,15 ade2-1 ura3-1 trp1-1 can1-100 def1*Δ*::URA3*

2.5.2 Conditions for culturing yeast

Starter cultures of yeast were set up in 5 ml of YPD (*see section 2.1.1.1*) and grown overnight at 30 °C to stationary phase. Cells were typically diluted in the morning to 2 x 10^6 cells/ml.

In the case of yeast harbouring the pRS314-*GAL10-GAL7* or pRS314-*GAL10-GAL7*Δ*TATA* plasmids, cells were grown for 24 hours in synthetic complete media (*see section 2.1.1.2*) lacking tryptophan and with raffinose as the only sugar source. Cells were then diluted to approximately 2 x 10^6 cells/ml and grown to mid-log phase where 2% glucose or galactose was added for the indicated time prior to harvesting or crosslinking.

2.5.3 Lithium-acetate transformations

Yeast cultures were set up in 5 ml of YPD (*see section 2.1.1.1*) and grown overnight at 30 °C to stationary phase. Cells were diluted in YPD to a concentration of 2 x 10^6 cells/ml in a total volume of 50 ml and grown to mid-log phase, before being centrifuged and washed twice in distilled water. The yeast pellets were washed in 1 ml of TE/LiOAc (*see section 2.1.3.6*) prior to resuspension in 500 µl of TE/LiOAc. 100 µl of yeast suspension (1 x 10^8 cells/ml) was used per transformation, to which was added 100 ng of plasmid DNA, 50 µg of salmon-sperm DNA (which was boiled for 3 minutes and placed on ice before adding) and 600 µl of PEG/TE/LiOAc (*see section 2.1.3.7*). A control transformation was always performed, where TE rather than plasmid DNA was added. Transformation mixtures were placed in a 30 °C incubator for 30 minutes and once removed 70 µl of DMSO was added and heat-shock performed at 42 °C for 15 minutes. Lastly, samples were centrifuged and the supernatant discarded prior to

washing in distilled water and plating onto the appropriate selective media (*see section 2.1.1.2*).

2.5.4 Preparing yeast extracts

Yeast cultures were pelleted and washed twice in TBS (*see section 2.1.3.2*) and washed in 1 ml of FA-lysis buffer (*see section 2.1.7.1*) prior to resuspension in 700 µl of FA-lysis buffer and addition of 0.5 mm Zirconia/Silica Beads (BioSpec Products, Inc.). Lysis of cells was performed in a FastPrep-24 Tissue and Cell Homogenizer (MP Biomedicals) with the following conditions: level 5.5 for 30 seconds, with cooling on ice for 1 minute, repeated 6 times.

2.6 Protein techniques

2.6.1 SDS-PAGE

Precast 4-12% Bis-Tris gradient gels (Biorad) were routinely used to separate proteins. The running buffer was XT MES buffer (Biorad) and electrophoresis was carried out in Criterion chambers (Biorad). The Precision Plus protein marker from Biorad was used to determine the size of the detected protein. For analysis of RNAPII 4-12 % gradient gels were usually run at 180 V for 45 minutes and protein detected by staining with Instant*Blue* (Expedeon), according to the manufacturer's instructions.

2.6.2 Western-blotting

Following electrophoresis of proteins by SDS-PAGE, gels were briefly washed in water before soaking in transfer buffer (25 mM Tris, 60 mM Glycine, 20% Methanol). Hybond C-Extra membrane (Amersham Biosicences) was placed on top of the gel and

sandwiched between two pieces of Whatman 3MM paper. This was then placed in a Biorad Criterion Blotter and transferred (in transfer buffer) at 4 °C for 90 minutes at 500 mA. The membrane was then incubated with PBS (*see section 2.1.3.1*) +2% milk powder (Marvel) for 30 minutes at 25 °C.

2.6.3 Detection of blotted proteins

Blotted proteins were detected using a primary antibody against the protein of interest, or against an epitope tag on the protein. Typically 1:10,000 dilution of antibody (for detecting Rpb1 of RNAPII) in PBS +2% milk was incubated with the blot for 2 hours at 25 °C. The blot was then washed 3 times 10 minutes in PBS. Next a 1:10,000 dilution of secondary antibody (with conjugated HRP) in PBS +2 % milk was added for 1 hour at 25 °C, again followed by 3 times 10 minute washes in PBS. The binding of primary and secondary antibody was detected using a SuperSignal West Pico Chemiluminescent Substrate kit (Thermo Scientific) and exposure of the membrane to Amersham Hyperfilm ECL (GE Healthcare).

2.6.4 RNA polymerase II purification

RNAPII purification was carried out based on a previously developed protocol (Cramer et al., 2001). 100 L of yeast cells were grown to a high density and harvested. Cells were lysed by ball-milling and the extract cleared by centrifugation. Extract was loaded at 5 ml/minute onto a 350 ml heparin-sulphate column, equilibrated with HSB150 (*see section 2.1.4.2*). 3 column volumes (CV) of HSB150 were pumped through the column to remove non-specifically bound proteins. Protein was eluted by addition of HSB600 (*see section 2.1.4.3*). Collected fractions were pooled and placed in

a container with a magnetic stirrer. Ammonium sulphate powder was added in 20 g amounts with 15 minutes between each addition, until a saturation of 50 % had been achieved. The solution was centrifuged in a GSA rotor at 12000 rpm for 20 minutes and the supernatant discarded. A second spin was performed at 10,000 rpm for 15 minutes to ensure any remaining supernatant was removed.

The pellet obtained from ammonium sulphate precipitation was resuspended in TEZ_0 (*see section 2.1.4.4*) in a volume of 20 ml and the conductivity checked against TEZ_{500} (*see section 2.1.4.5*) to determine if more TEZ_0 should be added. Once the conductivity was similar to that of TEZ_{500}, the sample was mixed with 4 ml of 8WG16-agarose resin and bound in batch for 2 hours at 4 °C. The protein-resin slurry was poured into a column and washed under gravity with 10 CV of TEZ_{500}. Another 10 CV TEZ_{500} wash was performed at 25 °C prior to elution with $TEZ_{Elution}$ (*see section 2.1.4.6*) at 25 °C and collection of 1.5 ml fractions. Samples were taken from each fraction and SDS-PAGE loading buffer (*see section 2.1.3.9*) added prior to SDS-PAGE of the samples (*see section 2.6.1*). Peak fractions were pooled and diluted in 3 times volume of MonoQ Dilution Buffer (*see section 2.1.4.7*) and loaded onto a 1ml MonoQ column at a rate of 0.5 ml/minute. A gradient was set up from 200 mM KCl in MonoQ200 (*see section 2.1.4.8*) to 1M KCl in MonoQ1000 (*see section 2.1.4.9*) and 1 ml fractions collected. Fractions were analysed by SDS-PAGE to determine the purity of isolated RNAPII. Fractions containing pure RNAPII were dialysed into RNAPII Buffer (*see section 2.1.4.10*)

2.6.5 Purification of recombinant TFIIS

E. coli BL21-CodonPlus (DE3) RIL strain (Agilent) was transformed with pET15b-DST1 (*see section 2.2.1*), grown in LB containing ampicillin and chloramphenicol to an OD_{600} of 0.7 and protein expression induced by the addition of 1 mM IPTG at 37 °C for 3.5 hours. Harvested cells were resuspended in CBB (50 mM KH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM Imidazole, 1 mM 2-mercaptoethanol and 1 x PIs) and sonicated. Binding of His-tagged protein to Ni-NTA Agarose (QIAGEN) was performed in batch for 2 hours at 4 °C, prior to 30 CV wash with Ni300 (50 mM KH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol, 1 x PIs) and 5 CV wash with Ni150 (50 mM KH_2PO_4 pH 8.0, 150 mM NaCl, 30 mM imidazole, 1 mM 2-mercaptoethanol, 1 x PIs). Elution buffer (50 mM KH_2PO_4 pH 8.0, 150 mM NaCl, 250 mM imidazole, 1 mM 2-mercaptoethanol, 1 x PIs) was added and 1.5 ml fractions collected. Fractions containing recombinant TFIIS were dialysed into Transcription Buffer (*see section 2.1.5.2*).

2.7 Reconstitution of elongation complexes and transcription *in vitro*

2.7.1 Reconstitution of elongation complexes

Elongation complexes were reconstituted using 150 nucleotide DNA oligonucleotides from DNA technology and 9 nt RNA oligonucleotides purchased from Dharmacon, the sequences of which are shown (Table 2.3). The DNA oligonucleotides were received HPLC purified, but subjected to further purification by 5.2% denaturing PAGE (*see section 2.3.13*), whilst the RNA oligonucleotides were purified by 20% denaturing PAGE (*see section 2.4.5*).

Table 2.3 Oligonucleotides used for elongation complex reconstitution

OLIGONUCLEOTIDE	SEQUENCE
TS	GCGTGGTATAGGAGTAGAGTATGGAGAGGTGTT GTTGTTGTGAGTTGGTTATGGTAGGTGAGTGTGT GATTGTGTGTTAGTAGTTGGTTAGTTGGGTGTAT GTTTGATGTGTATCCTGGTAAGTGATGTGTGGTT GGAATGTGGTTGCGG
NTS	CCGCAACCACATTCCAACCACACATCACTTACCA GGATACACATCAAACATACACCCAACTAACCAA CTACTAACACACAATCACACACTCACCTACCATA ACCAACTCACAACAACAACACCTCTCCATACTCT ACTCCTATACCACGC
RNA1	CCAGGAUAC
RNA2	AUGGAGAGG

Mono-ECs were assembled in Reconstitution Buffer (*see section 2.1.5.1*) by incubating 1 pmol of TS with 2 pmol of RNA1 (including 400 counts per second (CPS) of ^{32}P end-labelled RNA1) at 65 °C for 5 minutes followed by step-wise cooling to 25 °C over a 40-minute period. 3 pmol of FLAG-tagged RNAPII was added for 25 minutes at 25 °C, followed by the addition of 6 pmol of NTS (which was placed at 65 °C for 5 minutes then stored on ice before addition) at 37 °C for 10 minutes.

Di-ECs were formed using the same hybridisation temperatures and times as mono-ECs by incubation of 1 pmol of TS with 2 pmol of RNA1 (including 400 CPS ^{32}P end-labeled RNA1) followed by addition of 3 pmol of FLAG-tagged RNAPII (designated ^{32}P-RNA1-EC) whilst in a separate tube 6 pmol of NTS was incubated with 12 pmol of unlabelled RNA2 and then 18 pmol of HA-tagged RNAPII (RNA2-EC). ^{32}P-RNA1-EC and RNA2-EC were mixed and incubated at 37 °C for 10 minutes to allow hybridisation to occur.

2.7.2 Purification of elongation complexes

Mono- and di-ECs were incubated with Anti-FLAG M2 Affinity Gel (Sigma) for 1 hour at 4°C with shaking, prior to washing with EC Wash Buffer (*see section 2.1.5.3*) and 20 mM Tris-Cl, pH 7.5, followed by elution with FLAG-EB (*see section 2.1.5.4*). Purified ECs were then incubated with Anti-HA Affinity Matrix (Roche) for 1 hour at 4 °C with shaking to isolate di-ECs, followed by repeated washes with EC Wash Buffer and 20mM Tris-Cl, pH 7.5.

2.7.3 Ubiquitylation of elongation complexes

Unpurified ECs were incubated with yeast Uba1, Ubc5 and Rsp5 -/+ ubiquitin (with all lysine residues mutated) for 90 minutes at 30 °C prior to transcription. Ubiquitylation efficiency was assayed by Western blot with 4H8 Rpb1-CTD antibody.

2.7.4 Transcription and analysis

Elongation complexes in Transcription Buffer (*see section 2.1.5.2*) were incubated with 600 μM NTPs (either AUC or AUCG) for 5 minutes at 25 °C (or as specified). Transcription was stopped with Stop Buffer (*see section 2.1.5.5*), and, if proteinase K was added, incubated at 37 °C for 30 minutes. Samples were either phenol-chloroform extracted (*see section 2.4.2*) and RNA ethanol precipitated (*see section 2.4.3*) for analysis of transcripts by 8.3 M urea denaturing PAGE (6% acrylamide) (*see section 2.1.5.8*), or elongation complexes analysed by 0.7% agarose gel electrophoresis (*see section 2.1.5.7*) in native agarose loading buffer (*see section 2.1.5.6*).

Purification of nascent RNA from the agarose gel was performed by staining in

a 1:10,000 dilution of Sybr Gold, allowing visualization and excision of the band corresponding to the mono- or di-EC, followed by maceration, phenol-chloroform extraction (*see section 2.4.2*), butanol extraction and ethanol precipitation (*see section 2.4.3*). Transcripts were then analysed by denaturing PAGE (*see section 2.1.5.8*) and exposure to phosphor imager screens or Kodak BioMax MR film.

2.7.5 Assaying TFIIS activity

2.7.5.1 Oligonuceotides used

The oligonucleotides used for reconstitution of mono-ECs to test TFIIS activity are presented (Table 2.4).

Table 2.4 Oligonucleotides used for TFIIS activity assays.

OLIGONUCLEOTIDE	SEQUENCE
Poly-A/T TS 3' biotinylated	CCTTTCCTACCTACATACACCACACACCACACCG AGAAAAAAAAATTACCCCTTCACCTTTACCCTTA CCCCTCTCCATACCACACCACCTTACCTACCACC CACCTTCCCTTACCCTTC-**B**
Poly-A/T NTS	GAAGGGTAAGGGAAGGTGGGTGGTAGGTAGGGT GGTGTGGTATGGAGAGGGGTAAGGGTAAAGGTG AAGGGGTAATTTTTTTTTCTCGGTGTGGTGTGTG GTGTATGTAGGTAGGAAAGG
G-Stop TS 3' biotinylated	CCTTTCCTACCTACATACACCACACACCACACCG AGCCCAACCACTTACCCCTTCACCTTTACCCTTA CCCCTCTCCATACCACACCACCCTACCTACCACC CACCTTCCCTTACCCTTC-**B**
G-Stop NTS	GAAGGGTAAGGGAAGGTGGGTGGTAGGTAGGGT GGTGTGGTATGGAGAGGGGTAAGGGTAAAGGTG AAGGGGTAAGTGGTTGGGCTCGGTGTGGTGTGT GGTGTATGTAGGTAGGAAAGG
RNA	AUGGAGAGG

2.7.5.2 Poly A/T tract assay

Mono-ECs were reconstituted in a similar manner as before with 1pmol of TS, 2 pmol of ^{32}P-end labelled RNA, 3 pmol of RNAPII and 6 pmol NTS (*see section 2.7.1*)

but in this case the oligonucleotides used contained a poly-A/T tract (Table 2.4). Following reconstitution, ECs were bound to Streptavidin MagneSphere Paramagnetic Particles (Promega) and washed with Transcription Buffer (*see section 2.1.5.2*) three times, prior to aliquoting ECs into 4 tubes and resuspending in 45 µl of Transcription Buffer. 5 µl of NTPs (final concentration 500 µM) were added for 1 minute at 25 °C, before adding Transcription Buffer, or increasing amounts of recombinant TFIIS (0.1:1, 0.2:1 or 1:1 molar ratio of TFIIS to RNAPII) for 4 minutes. The reaction was halted by the addition of Stop Buffer (*see section 2.1.5.5*). RNA was isolated and analysed by denaturing PAGE (10 % polyacrylamide) in the usual manner (*see section 2.7.4*).

2.7.5.3 Backtracking assay

Mono-ECs were reconstituted using the G-Stop DNA oligonucleotides (Table 2.4), purified on Streptavidin beads and aliquoted as before (*see section 2.7.5.2*) but this time transcription (NTPs, no CTP) was allowed to proceed for 5 minutes, prior to washing 3 times in Transcription Buffer to remove NTPs. Transcription Buffer or increasing amounts of recombinant TFIIS (0.1:1, 0.2:1 or 1:1 molar ratio of TFIIS to RNAPII) were added to the ECs and incubated at 25 °C for 5 minutes before the reaction was stopped with Stop Buffer. Transcripts were isolated and analysed as before (*see section 2.7.4*).

2.8 Analysis of transcription *in vivo*

2.8.1 Construction of collision and control plasmids

The plasmid *GAL10-GAL7* used throughout this study was constructed by obtaining a *GAL10* PCR product using primers targeted 343 bp upstream of the ORF to

include all promoter elements (5'AGAGAAAGCTCGAG CTTTATTGTTCGGAGCAGTGCGG3') and 47 bp upstream of the stop codon, to ensure no terminator elements were present (5'AGAGAGAGATCGAT TCAAGGTTACACAATCTTTCCAGTTCTC3'). This product was cloned between the *Xho*I and *Cla*I sites of pRS314. A *GAL7* PCR product (or *GAL7*Δ*TATA* PCR product for the control) was obtained with primers hybridizing 446 bp upstream of the ORF (5'AGAGAGAGGAGCTC ATATCACTCACAACTATTGCGAAGCG3') and 38bp upstream of the *GAL7* stop codon (5'AGAGAGAGACTAG TTCTTAGTTTTTCAGCAGCTTGTTCCG3'). The fragment was cloned into the *Sac*I and *Spe*I sites in a convergent orientation to the *GAL10* gene. A 100 bp G-less cassette was PCR amplified from pGAL4CG- (Lue et al., 1989) and cloned into an *Eco*RI site at the 5' end of the *GAL10* ORF to form the promoter proximal G-less cassette (forward 5'AGAGAGAGGAATTCACTCACCCAATACTCCCTACTC3'; reverse 5'AGAGAGAGGAATTCGGGAGTGGAATGAGAAATG3'). Finally a 371 bp G-less cassette obtained from the 365 bp G-less cassette of pGAL4CG- (forward 5'AGAGAGAGATCGATCCTCCATACCCTTCCTCC3'; reverse 5'AGAGAGAGACTAGTGGGAGTGGAATGAGAAATG3') was cloned into the *Spe*I and *Cla*I sites between the 3' ends of the *GAL7* (or *GAL7*Δ*TATA*) and *GAL10* ORFs. The plasmids pYC10-7Fus and pYC10-7Fus-Δ7 (Prescott and Proudfoot, 2002) used as templates for the PCR amplifications above were kindly provided by Nick Proudfoot.

2.8.2 RNA extraction and Northern blotting

Cells were grown in synthetic complete minimal media with raffinose (*see section 2.5.2*) and harvested following addition of 2% glucose or galactose for 75

minutes (or after time indicated) at 30 °C and RNA extracted (*see section 2.4.1*). Equal amounts of RNA were treated with 200 U of RNase T1 (Roche) for 1 hour at 37 °C, prior to phenol-chloroform extraction (*see section 2.4.2*). Denaturing-PAGE loading buffer (*see section 2.1.3.10*) was then added to the RNA, which was heated to 65 °C for 10 minutes and separated by 7% denaturing PAGE (*see section 2.1.5.8*). RNA was transferred to a Hybond-N^+ nylon membrane (GE Healthcare) using semi-dry transfer blotting at 400 mA for 1 hour and UV-crosslinked. Northern membranes were incubated with a random-primed ^{32}P-labelled double-stranded DNA probe (corresponding to the long G-less cassette) for 1 hour at 65 °C. This was followed by 4 washes with WB1 (2 x SSC, 0.05% SDS) for 10 minutes at 25 °C and 2 washes with WB2 (0.1 x SSC, 0.1% SDS), each for 15 minutes at 50 °C. Probed membranes were exposed to phosphor imager screens or Kodak BioMax MR film. '% Distal Cassette Transcribed' was calculated using data from the phosphor imager. The proximal cassette signals were equalized for '-/+ Convergent Transcription'. The '+ Convergent Transcription' distal cassette signal was then calculated as a percentage of the '- Convergent Transcription' distal cassette signal (=100%). The mean value and standard error were calculated from 2 biological replicates.

2.8.3 RNA advanced sequencing

WT and *elc1*Δ cells were grown in YPD to mid-log phase prior to total RNA extraction (Qiagen RNeasy Kit). RNA was subjected to standard library preparation techniques (Illumina), including Ribo-Zero hybrid selection (Epicentre Biotechnologies), and Advanced Sequencing on an Illumina GAIIx sequencer.

2.9 Chromatin immuno-precipitation

2.9.1 Chromatin immuno-precipitation

Chromatin-immuno-precipitation (ChIP) was performed in WT or *elc1*Δ strains containing *GAL10-GAL7* or *GAL10-GAL7*Δ*TATA*. For the steady-state ChIP (Figure 4.3B), cells were crosslinked with 1% formaldehyde for 20 minutes at 25 °C. For the kinetic ChIP (Figure 4.7), the 0 minute timepoint was taken after 2 hours in galactose and crosslinked prior to the addition of 2% glucose. Timepoints were then taken every 2 minutes (up to 10 minutes) and crosslinked. Crosslinking was quenched with 200 mM glycine prior to resuspension and cell lysis in FA-Lysis buffer (*see section 2.1.7.1*). The chromatin was sonicated to a fragment length of 200-500 base pairs, and then incubated with 2.2 µg of 4H8 (anti-CTD antibody), or 2.2 µg mouse IgG (where appropriate) for 2 hours prior to incubation with Protein G Agarose (Pierce) for 2 hours. Beads were washed 3 times for 3 minutes at room temperature in FA-Lysis, once in FA-500 (*see section 2.1.7.2*), once in ChIP-WB (*see section 2.1.7.3*) and once in TES (*see section 2.1.7.4*). Finally, 100 µl ChIP-Elution buffer (*see section 2.1.7.5*) was added and the samples incubated at 65 °C for 10 minutes. 35 µg RNase A (Sigma) was added to the eluate for 30 minutes at 37 °C, followed by addition of 20 µg Proteinase K (Roche) for 2 hours at 42 °C. DNA-protein crosslinks were reversed by incubating at 65 °C for 6 hours, and the DNA was purified using a PCR Purification Kit (QIAGEN).

Quantitative PCR was performed as described (*see section 2.3.2*) using primers targeted to the distal cassette of the *GAL10-GAL7* construct (sense 5'GAGGGGATATGGAAAGGGAA3'; antisense 5'CCGGTGATTTCTTGTCTGCT3') as well as control primers directed to the

telomeres of chromosome 6 (sense 5'TAACAAGCGGCTGGACTACTTT3'; antisense 5'GATAACTCTGAACTGTGCATCC3') and primers targeted to endogenous *GAL1* (sense 5'ACGAGTCTCAAGCTTCTTGC3'; antisense 5'TATAGACAGCTGCCCAATGC3').

The steady-state ChIP (Figure 4.3B) values were divided by the input and telomere values and normalised to the signal for *GAL10-GAL7ΔTATA* (= 1). Values obtained for the kinetic ChIP (Figure 4.7) were divided by the input and normalised to the 0 minute timepoint (= 100%). Columns on the graphs represent the mean value and error bars show standard error, calculated from three biological replicates.

2.9.2 Chromatin-immuno-precipitation followed by advanced sequencing (ChIP-Seq)

WT and *elc1*Δ cells were grown in YPD to mid-log phase and crosslinked as indicated earlier; however, cells were resuspended in FA-500 before lysis. Sonicated chromatin was incubated for 2 hours with 2.2 µg 4H8 or mouse IgG prior to incubation with Protein G Agarose (Pierce) for 1 hour. Beads were washed 3 times for 3 minutes with FA-500, twice with FA-Lysis, twice with ChIP-WB and twice with TES. Elution, RNase treatment and reverse crosslinking were performed as described for standard ChIP (*see section 2.9.1*). DNA was purified by two rounds of phenol-chloroform extraction (*see section 2.3.7*) followed by ethanol precipitation (*see section 2.3.8*) before being subjected to standard library preparation techniques (Illumina) and Advanced Sequencing on an Illumina GAIIx DNA sequencer.

2.10 Bioinformatics and computational analyses

Bioinformatic analysis was performed in collaboration with Wu Wei and Lars Steinmetz (Genome Biology Unit, EMBL). Short read sequences from ChIP-Seq and RNA-Seq were aligned to the S288c reference genome (Version 20110326 downloaded from the *Saccharomyce*s Genome Database (SGD) (Cherry et al., 1998)) using the Novoalign (http://www.novocraft.com) software. Reads at each position along the genome were extracted. In ChIP-Seq data analysis, ChIP signals (Log2 values) were divided by the relative IgG control for each genomic position excluding the ones with <5 reads coverage. Quantile normalisation between samples was then performed. Normalised ChIP-Seq values for the regions flanking the middle positions of intergenic regions (+/- 500bp) between convergent or divergent gene pairs were extracted and the median values of the signal in each position in these gene pairs were calculated and plotted.

RNA-Seq coverage at each position along the genome was divided by the total number of reads mapped in each sample. Regions flanking the end positions of every gene (+/- 500bp) were taken, and normalised RNA-Seq signals extracted for every gene. The mean values of RNA-Seq signals at each position in all genes were calculated and plotted.

Chapter 3. A study of RNA Polymerase II convergent transcription *in vitro*

3.1 Introduction

The underlying complexity of the transcriptional landscape has been highlighted in recent studies which show that non-coding transcription (and especially concurrent antisense transcription (Katayama et al., 2005, Vallon-Christersson et al., 2007)) is highly prevalent at genes (Carninci, 2010, Jacquier, 2009, Berretta and Morillon, 2009). Additionally, genes can be organised into sense-antisense pairs, sometimes without terminators between them (Xu et al., 2011). This raises important questions about sense-antisense transcription and the ability of RNAPII molecules to transcribe the same piece of DNA at the same time, as they indeed do. Convergent transcription must result in RNAPII interactions/collision, even if this is a rare occurrence.

It is the aim of this chapter to present results concerning the nature of interactions between convergently transcribing RNAPII molecules. There are several unanswered and important questions concerning convergent transcription, answers to which are required to understand this process *in vivo*:

1. Can RNAPII molecules transcribe past one another?
2. If not, what is the fate of the RNAPII molecules upon collision?
3. Are there auxiliary factors that control/facilitate transcriptional bypass?

In order to address these questions, a convergent transcription template had to be developed that was amenable to biochemical manipulation and capable of providing robust mechanistic data. Interactions between tandemly transcribing RNAPII molecules have been extensively studied using an *in vitro* biochemical scaffold system (Saeki and

Svejstrup, 2009) (*see section 1.6*), modified from a basic assay originating in the Kashlev laboratory (Kireeva et al., 2000). This provided the template for the development of a convergent transcription template and was selected due to its many advantages over initiation factor-based or tailed-template systems. Firstly, elongation can be reconstituted and studied directly without the need for additional factors. Secondly, reconstituted ECs have an identical footprint and transcription bubble size as that of promoter-driven ECs (Saeki and Svejstrup, 2009). Thirdly, extended DNA-RNA hybrids do not form (as in the tailed-template system). Fourthly, all aspects of the system can be controlled, and finally, purified proteins can be added (Sigurdsson et al., 2010). Furthermore, initiation in promoter-based systems is remarkably inefficient, whilst the scaffold system used here allows pre-positioning and purification of appropriate templates before starting transcription.

The first part of this chapter will be concerned with characterising the experimental system to reconstitute convergent transcription *in vitro*. Later sections will detail the use of this system to deduce the outcome of RNAPII interactions, focussing specifically on the nature of collision and its effect on EC stability. Finally, experiments addressing addition of TFIIS and transcription-competent extract will be presented, with a focus on the implications of this for our understanding of RNAPII interactions and gene traffic in general.

3.2 Elongation can be reconstituted on an oligonucleotide-based scaffold system with purified RNAPII

Eukaryotic RNAPII is highly conserved even between organisms as diverse as humans and yeast (*see section 1.1.1*). Additionally, the ease of purifying large amounts of RNAPII from *Saccharomyces cerevisiae* makes it an ideal source of protein for reconstitution assays.

RNAPII was purified to homogeneity from yeast extract using a procedure developed in the Kornberg lab (Cramer et al., 2001) (*see section 2.6.4*) (Figure 3.1A): heparin purification, ammonium sulphate precipitation, 8WG16 immuno-precipitation and finally MonoQ anion exchange chromatography. Significant purification occurred as a result of immuno-precipitation using the CTD-specific antibody 8WG16, followed by stringent high salt washes. A key advantage of the 8WG16 antibody column is its high specificity and the ease with which RNAPII can be eluted by the addition of buffer containing 50% glycerol. SDS-PAGE analysis of eluted fractions shows the almost pure 12-subunit RNAPII and a few co-eluting protein contaminants (Figure 3.1B, * and **). The peak fractions from the 8WG16 column were then pooled and subjected to anion exchange chromatography on a MonoQ column and eluted over a KCl gradient. The peak fractions were analysed by SDS-PAGE (Figure 3.1C). This step resulted in the recovery of highly concentrated and pure RNAPII, free from the starting contaminants (compare to Figure 3.1B).

To reconstitute ECs on the oligonucleotide scaffold, it is imperative that highly purified reagents are used. This aids in reproducibility ensuring that reconstituted ECs are more homogeneous. Therefore, in addition to pure RNAPII, all oligonucleotides were PAGE-purified before use in the assay (*see section 2.3.13 and 2.4.5*). This is

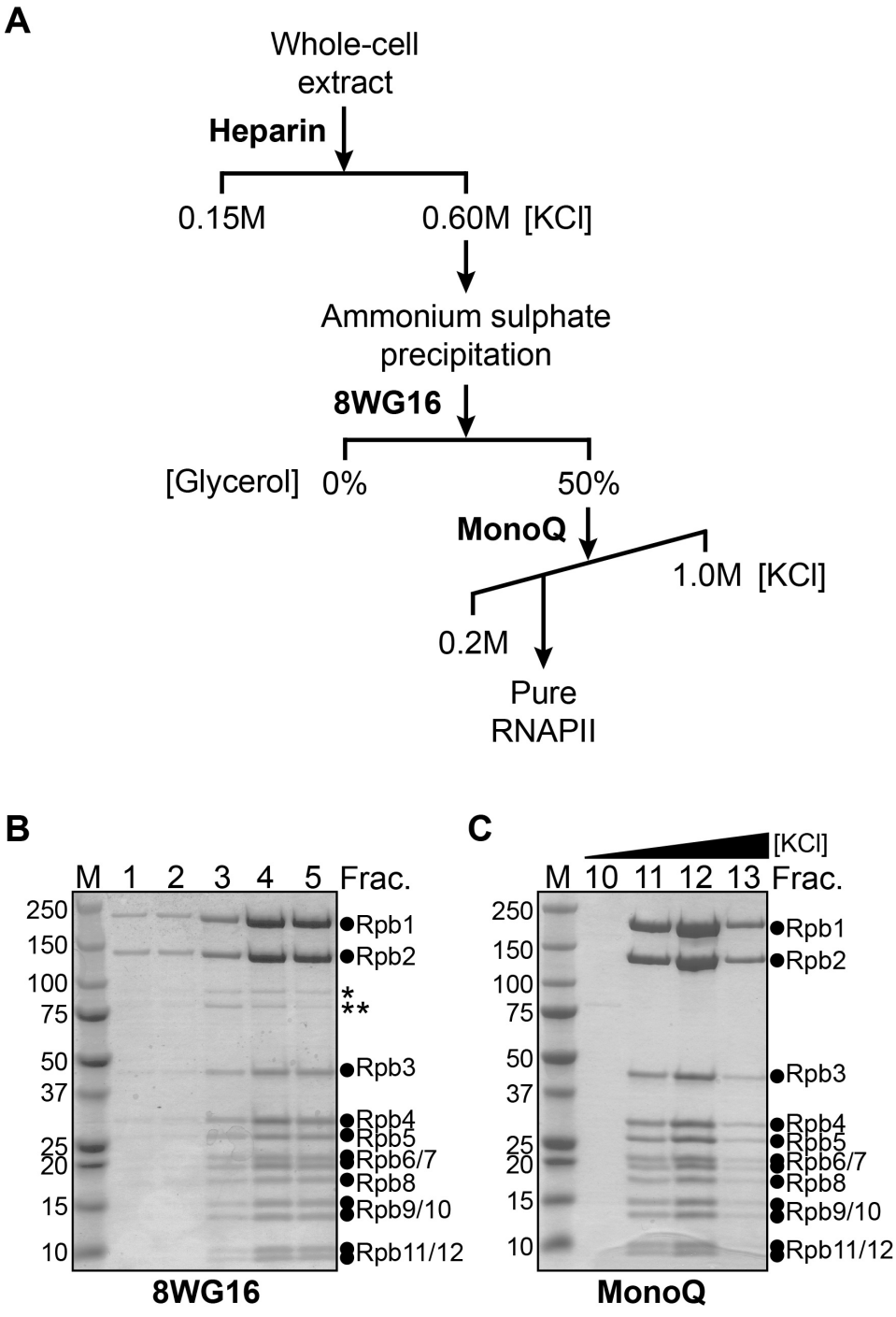

Figure 3.1 Purification of RNAPII from *S. cerevisiae*.

A. Schematic showing the purification schedule employed. **B.** Coomassie-stained SDS-polyacrylamide gel showing the first five fractions eluted from the 8WG16 column. * and ** are co-eluting contaminants. RNAPII subunits indicated, right. **C.** Coomassie-stained SDS-gel of the MonoQ peak fractions.

especially relevant here, as the oligonucleotides are 150 nt long and synthesis of such long DNA is relatively inefficient. Next, it was necessary to determine if the highly purified RNAPII was functionally active and could be used to successfully reconstitute ECs on the scaffold system. Therefore, ECs were reconstituted from oligonucleotides and the purified RNAPII in a stepwise manner (Figure 3.2A) (*see section 2.7.1*). Native agarose electrophoresis was used to assess EC formation, with the presence of a discrete low-mobility band indicating correct formation (Saeki and Svejstrup, 2009). The purified RNAPII did indeed form ECs efficiently giving rise to a discrete band (Mono-EC, Figure 3.2B). Next, it was important to determine if the reconstituted EC was active and able to transcribe the DNA template. NTPs were added to initiate transcription for 5 minutes, the RNA was isolated and analysed by denaturing PAGE. The data show that the RNAPII isolated and ECs formed were indeed highly active and able to extend the 9 nt RNA to the full-length product, with very few ECs arresting on the template or producing truncated RNAs (Figure 3.2C), which would be indicative of low RNAPII processivity or problems during reconstitution.

A system for purification of RNAPII and reconstitution of ECs has thus been implemented. The purity of transcription and the homogeneous nature of ECs, as shown by native agarose electrophoresis, provide a platform for more detailed elongation assays. Due to the efficiency of mono-EC transcription, any inhibition of full length RNA production should be clearly evident in subsequent convergent transcription experiments. This assay was, therefore, suitable as the basis for the following experiments in which a second EC was reconstituted on the opposite strand of DNA.

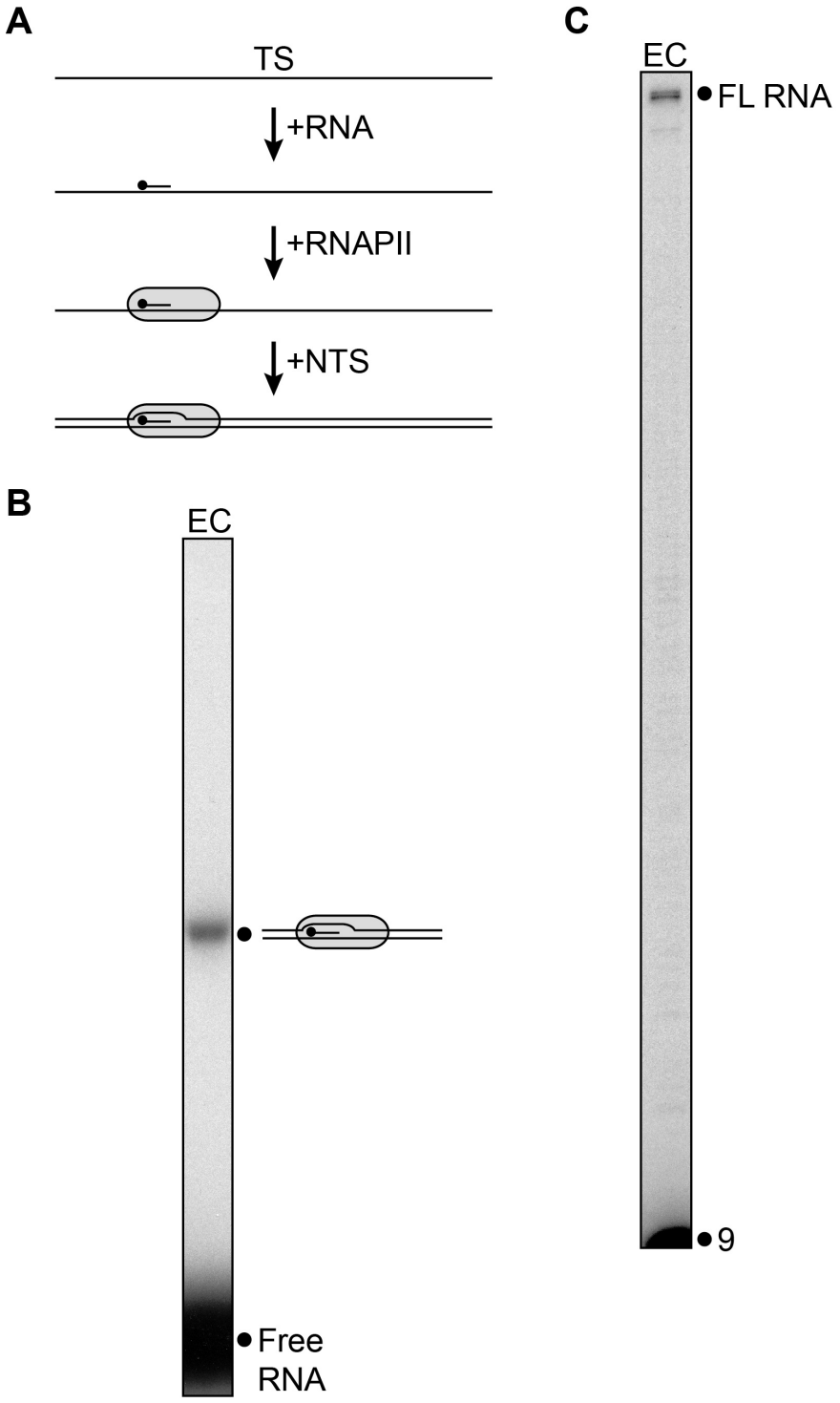

Figure 3.2 Reconstitution of elongation *in vitro* using purified RNAPII.

A. Schematic indicating the method for EC reconstitution, grey oblong is RNAPII and black sphere ^{32}P label. **B.** Autoradiograph of native agarose gel showing a reconstituted EC, which forms a discrete band. **C.** Denaturing PAGE of RNA isolated from EC analysed in **B**. FL RNA is full-length RNA.

3.3 Di-elongation complexes can be reconstituted *in vitro*

In order to study the phenomenon of convergent transcription, modifications were made to the scaffold system. Conditions were found in which two RNAPII molecules could be loaded onto the same piece of double-stranded DNA, orientated to transcribe towards each other. Scaffolds of this nature were obtained by forming separate DNA-RNA-RNAPII intermediates and finding conditions in which they would hybridise to form one transcription template (Figure 3.3A, Di-EC).

During the reconstitution reaction, two pools of partially formed ECs (DNA-RNA-RNAPII) were assembled. One of these ECs was radioactively labelled so that its transcription could be assessed (designated ^{32}P-RNA1-EC). The DNA oligonucleotide used as the NTS for this radioactively labelled EC was incubated with a complimentary RNA primer and RNAPII (referred to as RNA2-EC), and to bias di-EC formation increasing amounts of this complex was mixed with fixed amounts of the complementary TS-RNA-RNAPII complex, carrying the radioactively end-labelled RNA (^{32}P-RNA1-EC) (Figure 3.3A, Di-EC). Upon mixing, the two intermediate ECs should hybridise and form di-ECs (see Figure 3.3A, Di-EC schematic). Reconstituted mono-ECs and putative di-ECs were incubated with NTPs and transcription allowed to proceed for 5 minutes. The RNA was isolated and resolved by denaturing PAGE.

The mono-EC transcribed to the end of the DNA template, producing full-length RNA, as expected (Figure 3.3B, lane 1). Addition of increasing concentrations of RNA2-EC resulted in a decrease in the amount of full-length RNA produced from ^{32}P-RNA1-EC (Figure 3.3B lanes 2-4), along with the formation of novel truncated RNAs, of approximately half the size of full-length RNA. This finding was indicative of an

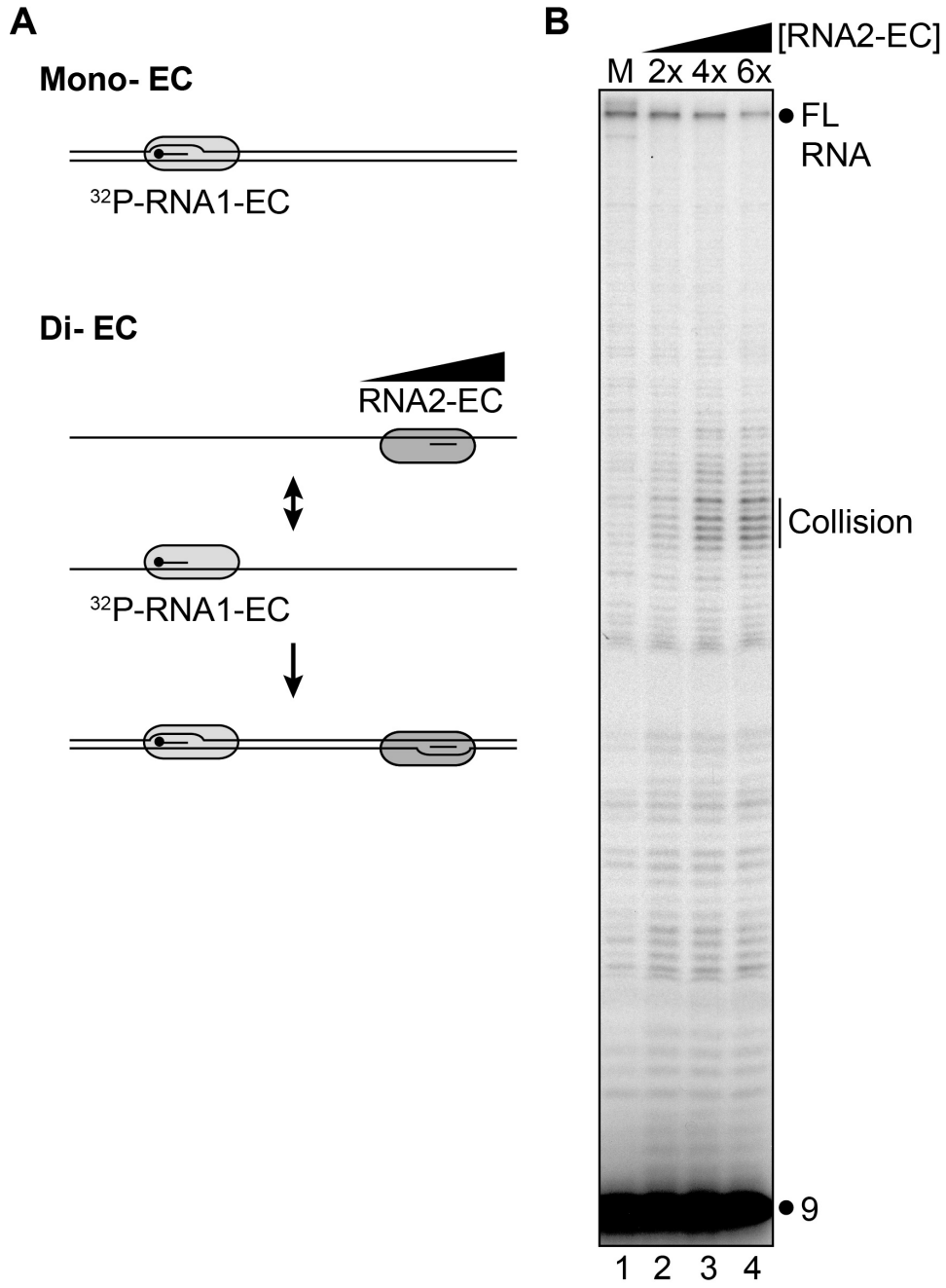

Figure 3.3 Reconstitution of di-ECs *in vitro*.

A. Schematic showing a mono-EC (upper) and the method for reconstituting di-ECs (lower) using increasing amounts of RNA2-EC. **B.** Denaturing PAGE of RNA isolated from mono-ECs (M) and complexes formed by titrating 2, 4, or 6 times molar excess of RNA2-EC. Full-length RNA (FL) is indicated along with 'Collision' products.

impediment to transcription. Interestingly, there was a correlation between the production of truncated RNA and the disappearance of the full-length product and this occurred in an RNA2-EC concentration-dependent manner. A 6x excess of RNA2-EC over ^{32}P-RNA1-EC produced the greatest amount of truncated, and smallest amount of full-length RNA, presumably arising as a result of a transcriptional block imposed by the presence of the second RNAPII molecule (Figure 3.3B, lane 4). In further support of this assertion, the length of truncated products is approximately consistent with expectations after considering the predicted footprint of the second EC on the DNA (Gnatt et al., 2001). If ECs do indeed impede each others progression, then transcription from the opposite direction should also produce RNA of a similar size. To test this, di-ECs were again reconstituted, but this time with RNA2-EC containing the end-labelled RNA and incubated with 6x excess of RNA1-EC (Figure 3.4A). The transcription products isolated show that transcription in the opposite direction also produces truncated RNA products, of a similar size to those seen before and reduced levels of full-length RNA (compare Figure 3.4B lane 2 to Figure 3.3B, lane 4). This suggests that the truncated RNA is indeed produced as a result of collision between the two ECs rather than as the result of an artefact.

In conclusion, the data shown so far indicate that the presence of another RNAPII molecule on the DNA template is inhibitory to elongation, as RNAPII processivity is reduced. This finding has important implications for transcription on highly transcribed DNA. However, the presence of some full-length RNA product indicates that there might be a fraction of RNAPIIs with the ability to bypass one another, and the next section, therefore, focuses on the isolation of purified di-ECs (lacking contamination by mono-ECs) for the purpose of studying this.

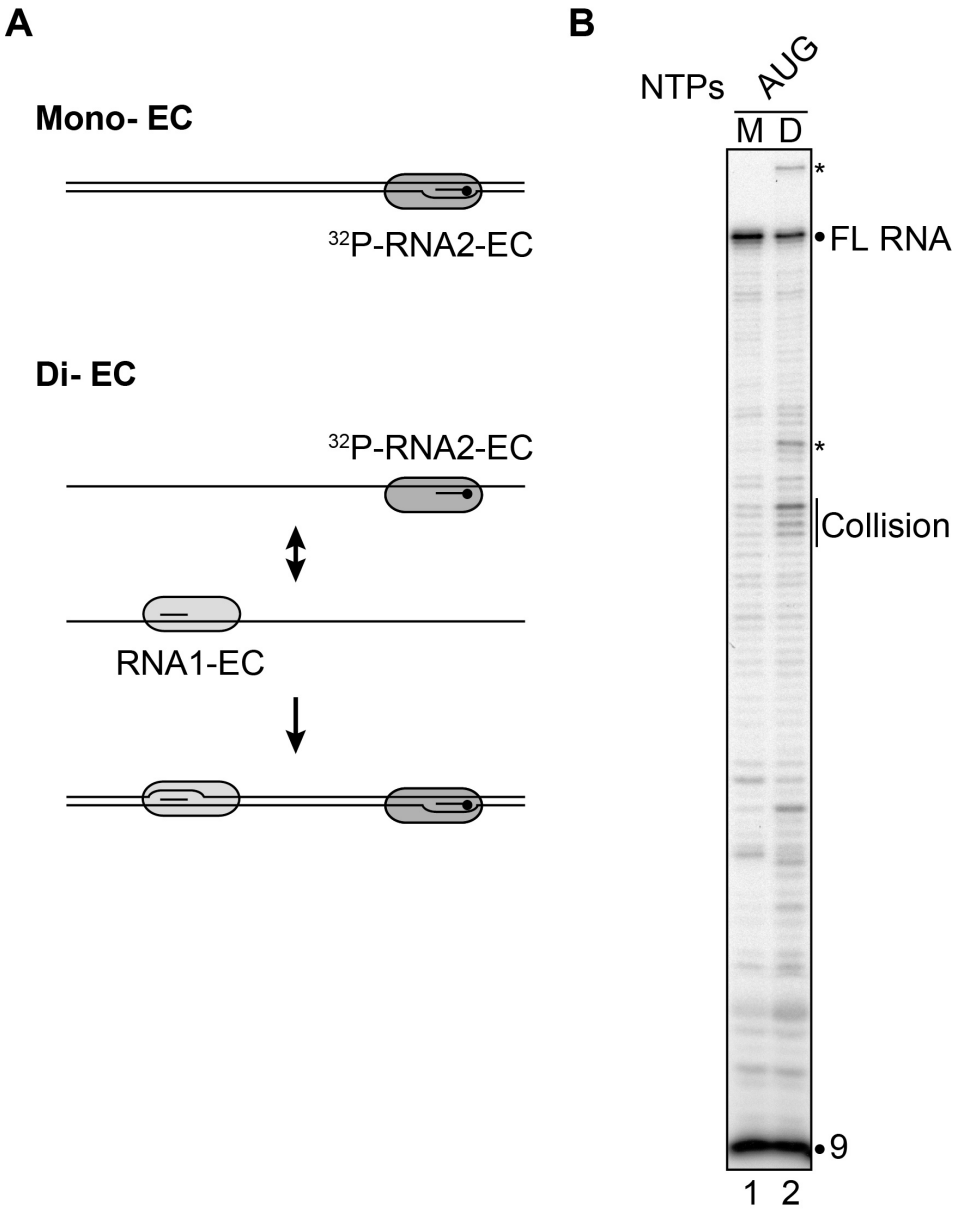

Figure 3.4 Analysis of transcripts from the other EC.

A. Diagram indicating EC position on the DNA template and location of labelled RNA. **B.** Transcription products isolated from ECs. Full-length RNA and Collision bands indicated, in addition to RNA produced by mis-primed ECs *.

3.4 Purified di-ECs show that convergently transcribing RNAPII molecules cannot transcribe past each other

Having found conditions for the formation of di-ECs, it was necessary to determine if the full-length RNA observed in the previous experiment was a result of contaminating mono-ECs or a fraction of RNAPII molecules bypassing one another. Due to the nature of the reconstitution reaction and its inherent inefficiency, a range of different products will undoubtedly be present in the EC mixture and so a method was devised to isolate di-ECs from other observable formation intermediates, which may be giving rise to full-length RNA in the di-EC reactions (Figure 3.5A). The approach used employed the formation of di-ECs with two differently epitope-tagged RNAPII molecules, enabling isolation of di-ECs by two-step affinity-purification (Figure 3.5B).

RNAPII was purified from two yeast strains: one with sequence encoding a C-terminal FLAG-tag inserted into the *RPB3* gene and the other with a sequence encoding a C-terminal HA-tag inserted into the *RPB1* gene. The presence of the FLAG-tag on Rpb3 resulted in a visible mobility shift during SDS-PAGE (Figure 3.6A). Di-ECs were reconstituted with the two different RNAPII forms (using the same method as before) and then both mono- and di-ECs were incubated with Anti-FLAG M2 Agarose to isolate templates containing the labelled EC. Bound complexes were washed stringently with a high-salt buffer followed by several washes in 20 mM Tris pH 7.5 and elution with FLAG-peptide. Di-ECs were then subjected to a second round of purification using Anti-HA Affinity Matrix, to isolate transcription templates that also contained the non-labelled EC, followed by washes and resuspension in Transcription Buffer (*see section 2.7.2*). This method allowed the isolation of true di-ECs.

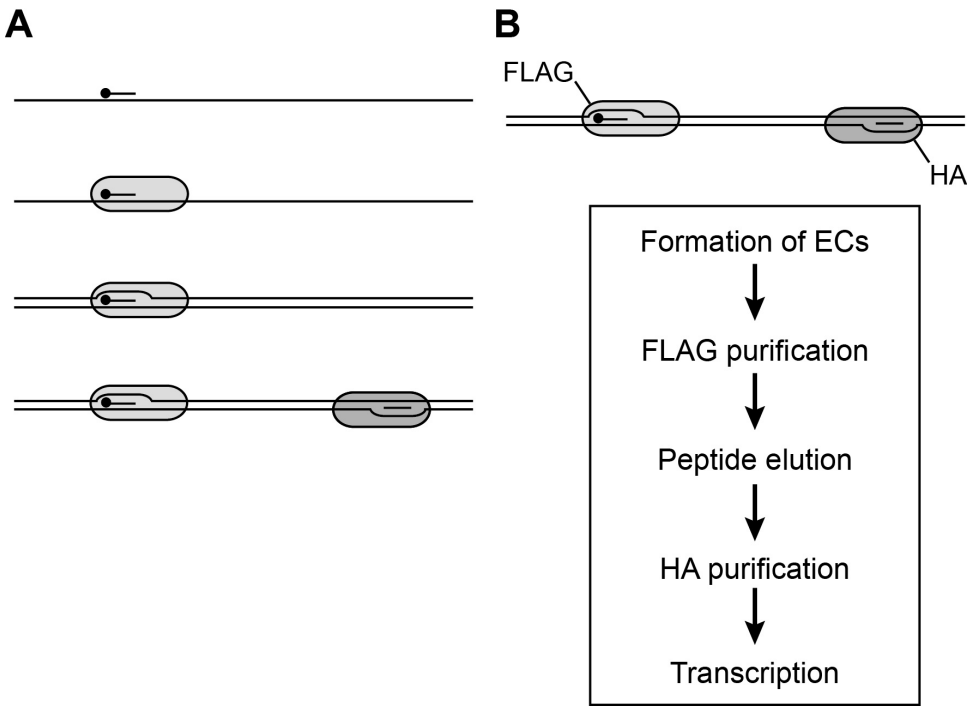

Figure 3.5 Representation of reaction products and di-EC purification schedule.

A. Diagram showing expected observable reaction products. **B.** Di-EC with positioned epitope-tagged RNAPII molecules (upper) and a purification schedule for isolating di-ECs.

The exact starting position of RNAPII on the DNA template is known; therefore, the expected length of RNA products can be calculated, with both mono- and di-ECs having the potential to produce 119 nt RNA when incubated with all NTPs (AUCG) (Figure 3.6B). NTPs were added and isolated RNA products resolved by denaturing PAGE (Figure 3.6C). As expected, the mono-ECs transcribed the full length of the template and produced a run-off product of 119 nt (lane 1). However, the di-ECs did not (compare lanes 1 and 2), instead creating a range of prominent truncated RNAs (indicated, right). The absence of full-length RNA indicates that convergent transcription leads to RNAPII head-to-head collisions, which impede further transcription. These data indicate that RNAPII molecules are unable to bypass one another, but instead impede one another's progress during elongation.

By adding all NTPs in this experiment both RNAPII molecules had the potential to transcribe the length of the template. It is likely that the smaller truncated RNA products (26-40 nt) are a result of collision distributed across the DNA template, with both polymerases moving forward leading to collision at random sites. Interestingly, however, these truncated RNAs appear to correspond to sites of RNAPII pausing in the mono-EC reaction as well (Figure 3.6C, compare lanes 1 and 2). Presumably, these transcripts are also more abundant in the di-EC reactions because the longer residence time at pause sites increases the likelihood of an encounter with the opposing polymerase, preventing continued transcription.

The absence of RNAPII transcriptional bypass was slightly surprising, considering the observations reported by others in the bacteriophage system (Ma and McAllister, 2009). It is also surprising when taking into account all the antisense transcription that has been reported *in vivo* (*see section 1.5.2*). However, in this basic

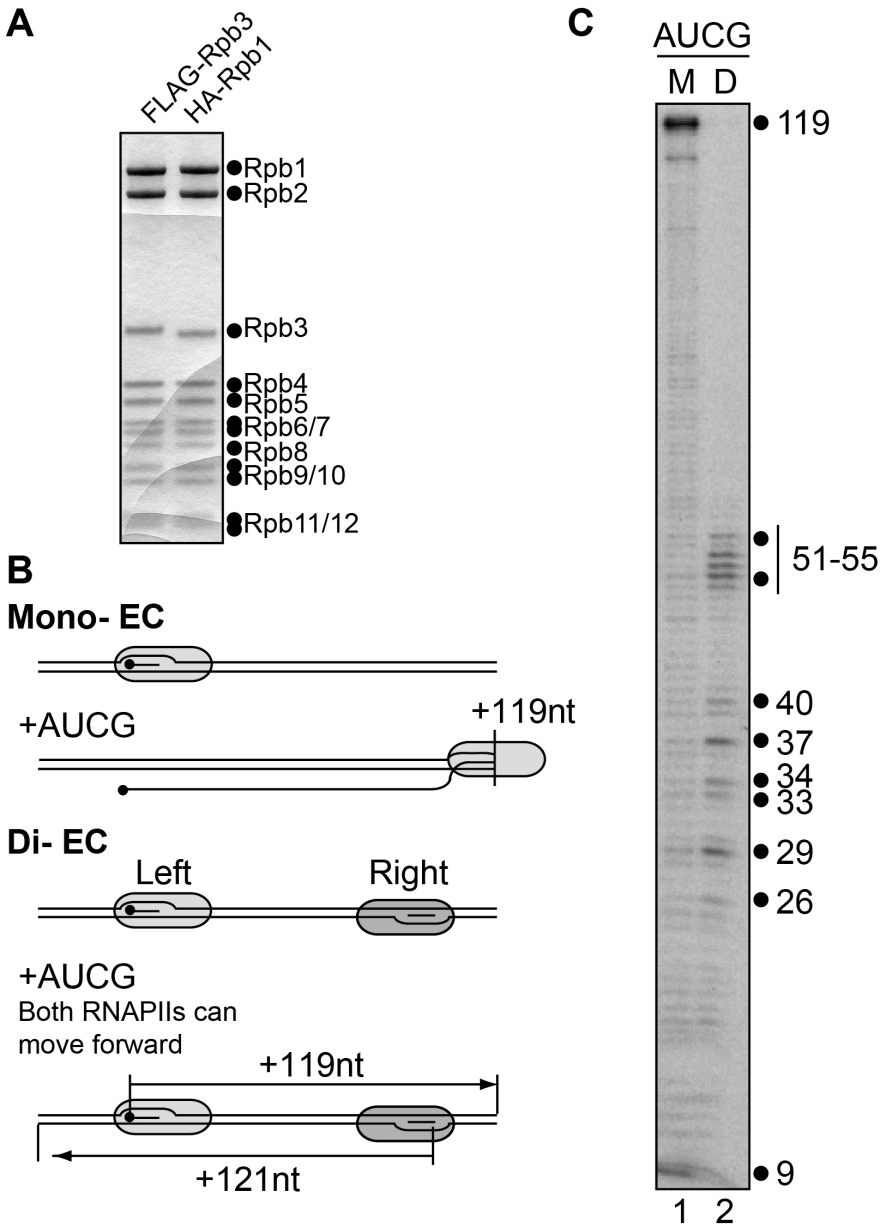

Figure 3.6 Formation and purification of di-ECs with FLAG- and HA-RNAPII.

A. Coomassie-stained SDS-polyacrylamide gel of FLAG-RNAPII and HA-RNAPII. **B.** Schematic representation of mono-ECs (upper) and di-ECs (lower) showing starting positions and predicted transcript lengths. **C.** Denaturing PAGE of RNA products isolated from mono- (M) and di-ECs (D) in the presence of all NTPs (AUCG). Observed transcript length, right.

biochemical system, RNAPII is not associated with any of the myriad elongation factors that facilitate transcription *in vivo* (*see section 1.3.2.1*). Conceivably, some of these could be necessary also to facilitate transcriptional bypass. This theory was tested with experiments presented later in this chapter.

3.5 Spatial invasion does not occur as a result of head-to-head RNAPII collision

The truncated RNAs produced in the presence of all NTPs had to be characterised further to fully elucidate their origin and identity. This was achieved by reconstituting ECs on DNA templates containing a C-less TS downstream of RNA1-EC but a C-rich TS for RNA2-EC. C-residues were positioned such that in the presence of NTPs (no GTP) the EC carrying the labelled RNA could in theory transcribe to the other side of RNA2-EC in the absence of added GTP, to produce a 117 nt RNA product, whereas the unlabelled EC in di-ECs would stop after transcribing just 1 nt (Figure 3.7A). The importance of performing this kind of experiment is two-fold; firstly, the point of RNAPII collision can be precisely defined to test which truncated transcripts are a result of randomly distributed collision and secondly, due to prior knowledge of the DNA region covered by RNAPII in an elongation complex (Gnatt et al., 2001), the ability of RNAPII molecules to invade one another's space on DNA can be studied. This was particularly pertinent, given that head-to-tail collision has been shown to involve dramatic temporary invasion (Saeki and Svejstrup, 2009).

The reconstituted ECs were purified as before, incubated with NTPs (no GTP) and RNA isolated. As expected, the mono-EC transcribed up to the C-stop and produced the full-length 117 nt transcript (Figure 3.7B, lane 1). Also as expected, the

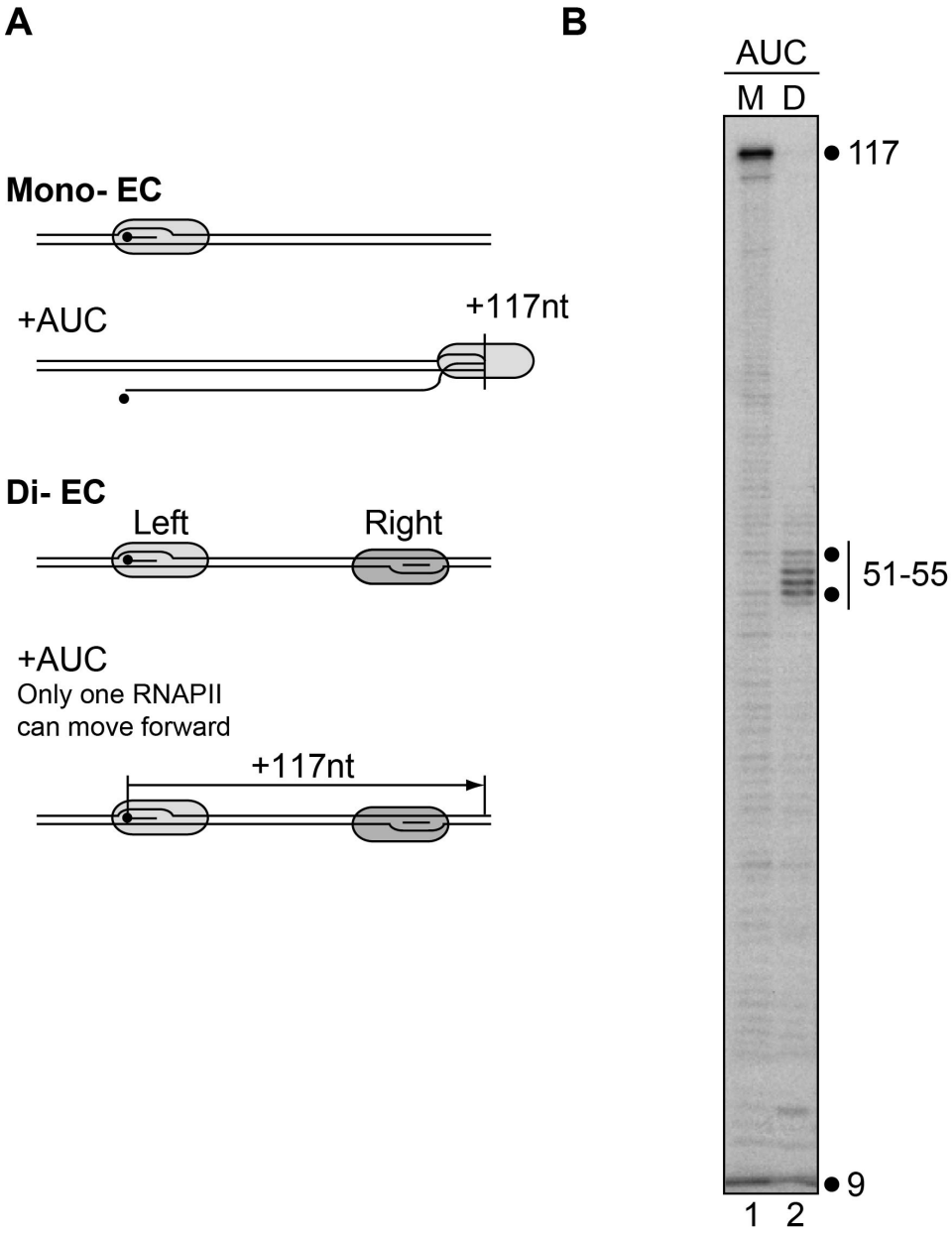

Figure 3.7 Transcription into a stalled RNAPII EC.

A. Diagram showing reconstituted mono- and di-ECs with theoretical RNA products indicated. **B.** Denaturing PAGE of RNA products isolated following transcription (ATP, UTP and CTP). Transcripts indicated, right. Note the disappearance of transcripts between 26 and 40 nt length (inclusive).

di-ECs did not reach the C-stop but instead produced the prominent 51-55 nt RNA products seen previously (Figure 3.7B, lane 2). Interestingly, the shorter transcripts produced in the presence of all NTPs (26-40 nts) (Figure 3.6C, lane 2) were not observed, confirming that these were indeed the result of randomly distributed collision, mainly localising to pause sites.

The nature of the scaffold system and the use of high purified reagents allows the exact positioning of RNAPII through the use of an RNA primer, providing knowledge of the precise starting- and end-point of transcription (Figure 3.8A). The footprint of RNAPII on DNA and the position of the DNA-RNA hybrid are also known from structural studies (Gnatt et al., 2001, Wang et al., 2006, Kettenberger et al., 2004). Therefore, if the EC containing the labelled RNA is moved forward 55 bp (corresponding to the longest RNA produced) then the degree of overlap between RNAPII molecules (and thus invasion) can be determined. The outcome of this analysis is that convergently transcribing RNAPII molecules appear to be virtually unable to invade one another – transcription stops as soon as the leading edges of the polymerases touch (Figure 3.8B). Interestingly, this is in sharp contrast to tandem transcription collisions, where temporal, but extensive, structural intermingling of polymerases was observed (Saeki and Svejstrup, 2009). This helps explain why RNAPII molecules in this highly purified system are unable to bypass one another, with the presence of bulky protein domains likely inhibiting displacement of the non-transcribed strands. Furthermore, this might explain the finding that bacteriophage RNAP molecules can bypass one another. In this simple system, there is a lack of protein density around the polymerase core, which could allow the enzymes to transcribe the same template (Ma and McAllister, 2009, Durniak et al., 2008, Cramer et al., 2001). This suggests that the

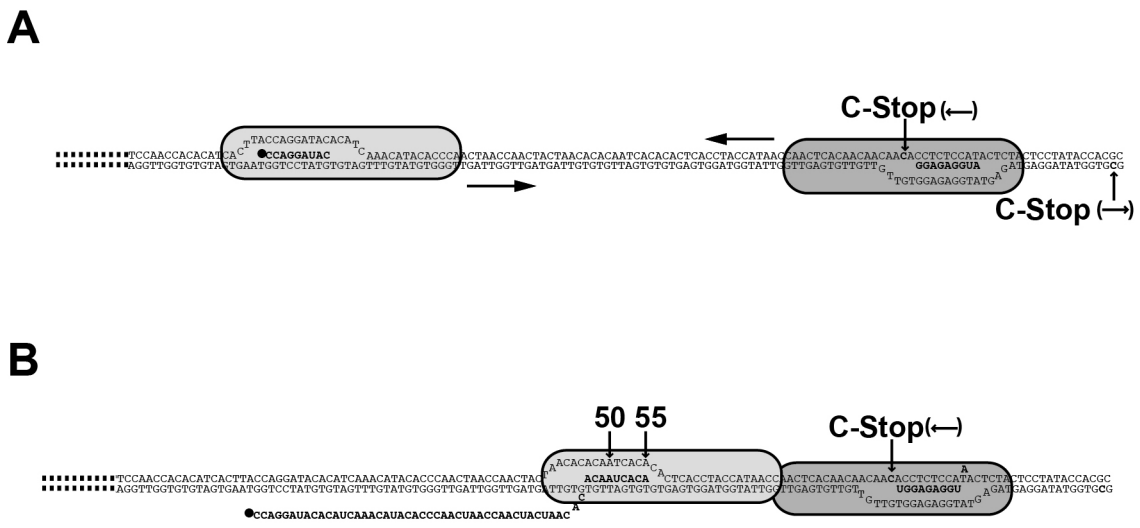

Figure 3.8 Diagram representing ECs before and after collision.

A. Representation of di-ECs before transcription showing the direction of RNAPII movement (arrows) and indicating the position of C-stops on the DNA template. RNAPII (oblong) footprint derived from the crystal structure (Gnatt et al., 2001). **B.** ECs shown after transcription with transcript lengths observed (Fig. 3.7B) indicated.

mechanism of convergent transcription is very different between eukaryote and bacteriophage polymerases.

3.6 Increased transcription time is insufficient to allow RNAPII molecules to bypass each other

In all of the experiments presented above, transcription reactions were allowed to occur for a relatively short time period (5 minutes). It is an interesting possibility that transcription time could have an affect on the ability of polymerases to bypass one another and indeed, as previous studies have shown, some transcription past a DNA damage lesion or pause site can occur with increased transcription time (Walmacq et al., 2012, Izban and Luse, 1991, Dimitri et al., 2008). Therefore, increased transcription times were tested to determine if they would have an effect on transcriptional bypass.

Di-ECs were reconstituted, purified by dual immuno-purification (*see section 3.4*) and then incubated with NTPs for increasing periods of time. Transcription products isolated from these di-ECs (Figure 3.9) displayed a similar size distribution, even after increased transcription time (compare lanes 1 and 4). Neither the intensity of the truncated RNA bands nor the level of 119 nt RNA changed considerably with increased transcription time, ruling out the possibility that significant transcriptional bypass occurs with time.

In conclusion, the data presented so far indicates either that di-ECs collide, impede each others progress, and remain fairly stable upon collision, or that following collision, one of the polymerases falls off the DNA, and the other RNAPII enters into an arrested or backtracked state, requiring re-activation by external factors.

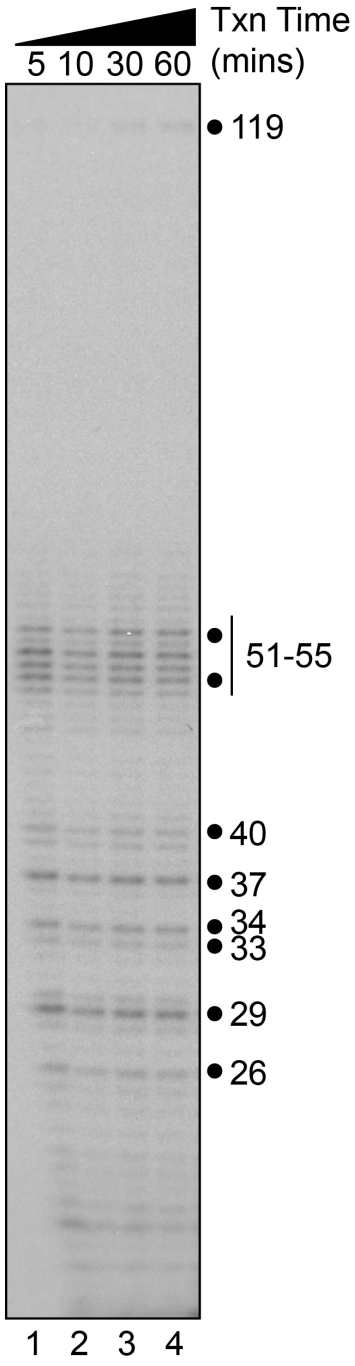

Figure 3.9 Timecourse of di-EC transcription.

Denaturing PAGE of RNA products obtained from purified di-ECs incubated with NTPs for 5, 10, 30, or 60 minutes prior to terminating transcription. RNA length indicated, right.

3.7 Elongation complexes are stable following collision

The observation that RNAPII molecules are unable to transcribe past each other raises questions about the fate of collided ECs. It is possible, for example, that the force of impact between the two molecules causes stress to the structure of the EC and possibly leads to RNAPII dissociation from DNA. On the other hand, ECs are inherently extremely stable (Kireeva et al., 2000), and collisions between tandemly transcribing di-ECs, where transient but significant structural changes occur, lead to little or no EC dissociation (Saeki and Svejstrup, 2009).

To assess EC stability upon collision, native agarose gel electrophoresis of reconstituted mono- and di-ECs was performed either before or after transcription (Figure 3.10A). Mono-ECs (lane 1 and 3) migrated on the agarose gel as seen previously (Figure 3.2B), whereas di-ECs migrated more slowly, owing to their increased molecular weight (lanes 2 and 4). A similarly reduced mobility was seen previously for tandemly arranged di-ECs (Saeki and Svejstrup, 2009). Interestingly, after transcription, little, or no, di-EC dissociation was observed (compare the intensity of di-EC in lane 2 and lane 4) and nascent RNA could be isolated directly from the bands corresponding to mono- and di-ECs on the native agarose gel and resolved by denaturing PAGE (Figure 3.10B). The isolated RNA was of an identical size to that seen previously for mono- and di-ECs (Figure 3.7) and thus confirms that the bands observed on the native agarose gel are indeed mono- and di-ECs.

A similar analysis was performed, adding all NTPs to initiate transcription from both RNAPII molecules. Again, the vast majority of Di-ECs remained intact (Figure 3.11A) and isolated nascent RNA showed the profile observed for immuno-purified di-ECs (compare Figure 3.11B, lanes 1 and 2 to Figure 3.6C lanes 1 and 2). These data

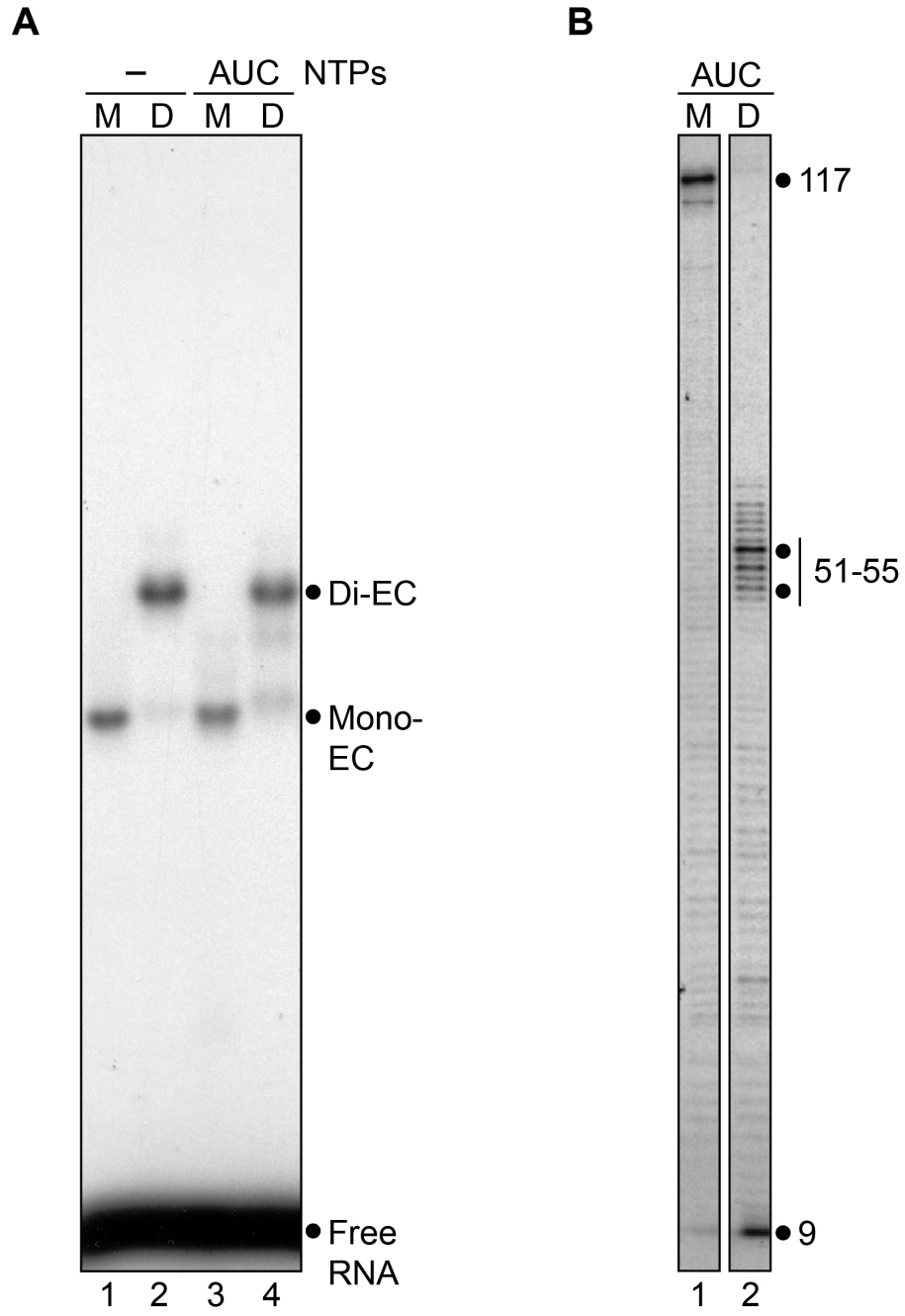

Figure 3.10 Analysis of di-ECs following transcriptional collision.

A. Native agarose gel of unpurified mono- (M) and di-ECs (D) before (lanes 1 and 2) or after (lanes 3 and 4) transcription. Unincorporated RNA primer is indicated (Free RNA). **B.** Denaturing PAGE of RNA isolated from the mono- (M) or di-EC (D) positions on a native gel. RNA sizes indicated, right.

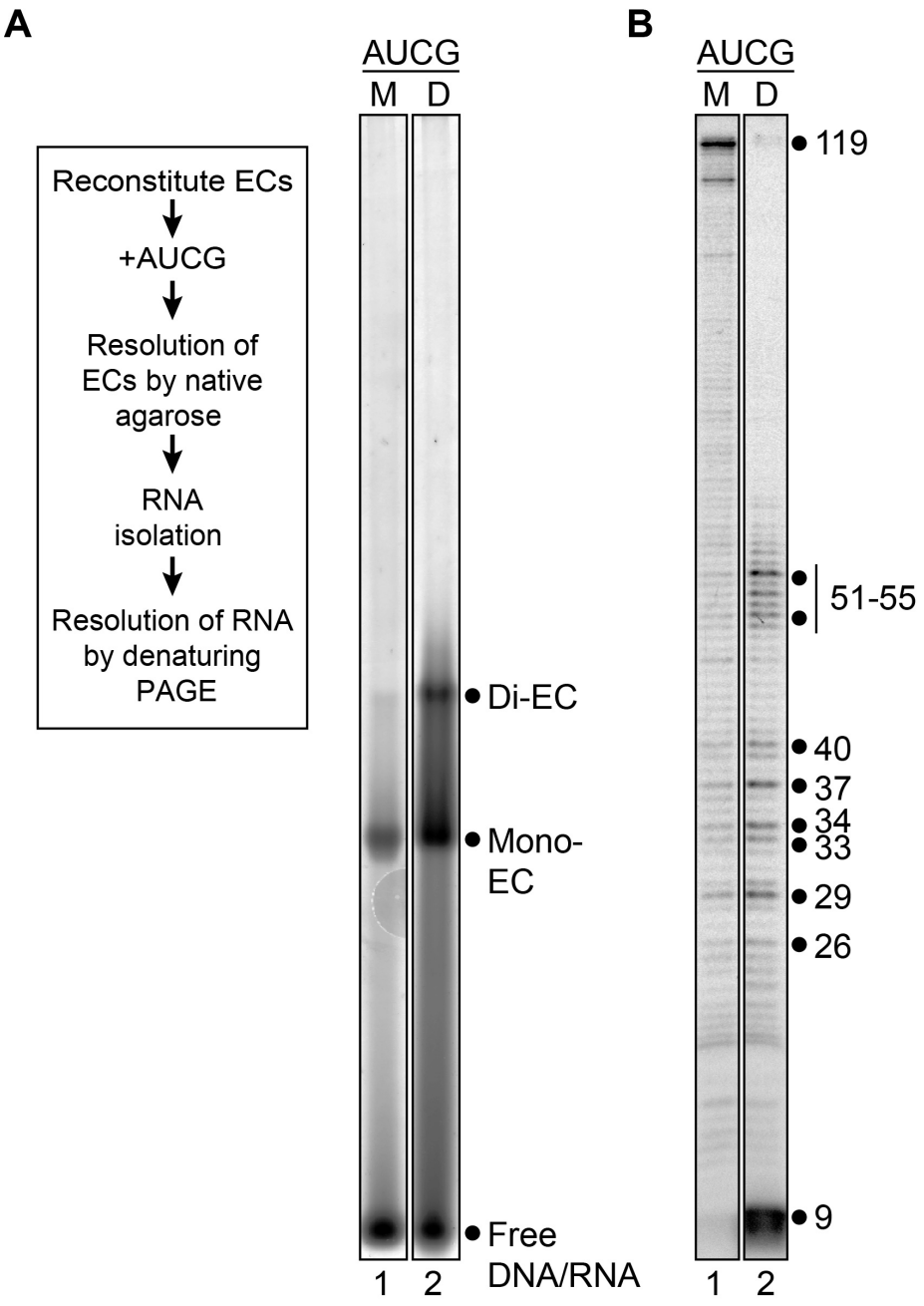

Figure 3.11 Stability of ECs when both polymerases are transcribing.

A. Sybr Gold-stained native agarose gel of mono-(M) and di-ECs (D) after transcription. **B.** Denaturing PAGE analysis of nascent transcripts isolated from 'Mono' and 'Di' positions in **A**.

show that the forces generated during collision are insufficient to push RNAPII off the DNA template, supporting previous data concerning EC stability (Kireeva et al., 2000, Saeki and Svejstrup, 2009). Importantly, this rules out EC dissociation as a major method for RNAPII bypass and suggests that alternative mechanisms and factors may exist *in vivo* to support RNAPII bypass. The following sections will focus on testing this hypothesis.

3.8 RNAPII cannot overcome transcriptional collision through TFIIS activity

The elongation factor TFIIS is able to stimulate RNAPII intrinsic RNA cleavage activity and can be purified from recombinant sources (Sigurdsson et al., 2010). TFIIS has previously been shown to stimulate RNAPII transcription through various elongation blocks (Fish and Kane, 2002). Therefore, TFIIS could play a role in facilitating transcriptional bypass between di-ECs. It is also likely that upon initial transcriptional collision, RNAPII backtracks, as seen for tandem di-ECs (Saeki and Svejstrup, 2009), resulting in arrest. As arrested/backtracked RNAPII requires TFIIS to resume transcription, arrest could provide an explanation for the lack of transcriptional bypass observed even during increased transcription time (*see section 3.6*).

A His-tagged form of yeast TFIIS was expressed in *E.coli*, purified via nickel-affinity chromatography (Figure 3.12A) and tested for activity using two assays. The first assay measured transcription through a pause-inducing poly A/T tract (Saeki and Svejstrup, 2009, Sigurdsson et al., 2010) (Figure 3.12B, schematic). Previously, it has been shown that RNAPII is able to transcribe through a pause site and produce full-length RNA upon TFIIS addition (Sigurdsson et al., 2010). To test the activity of

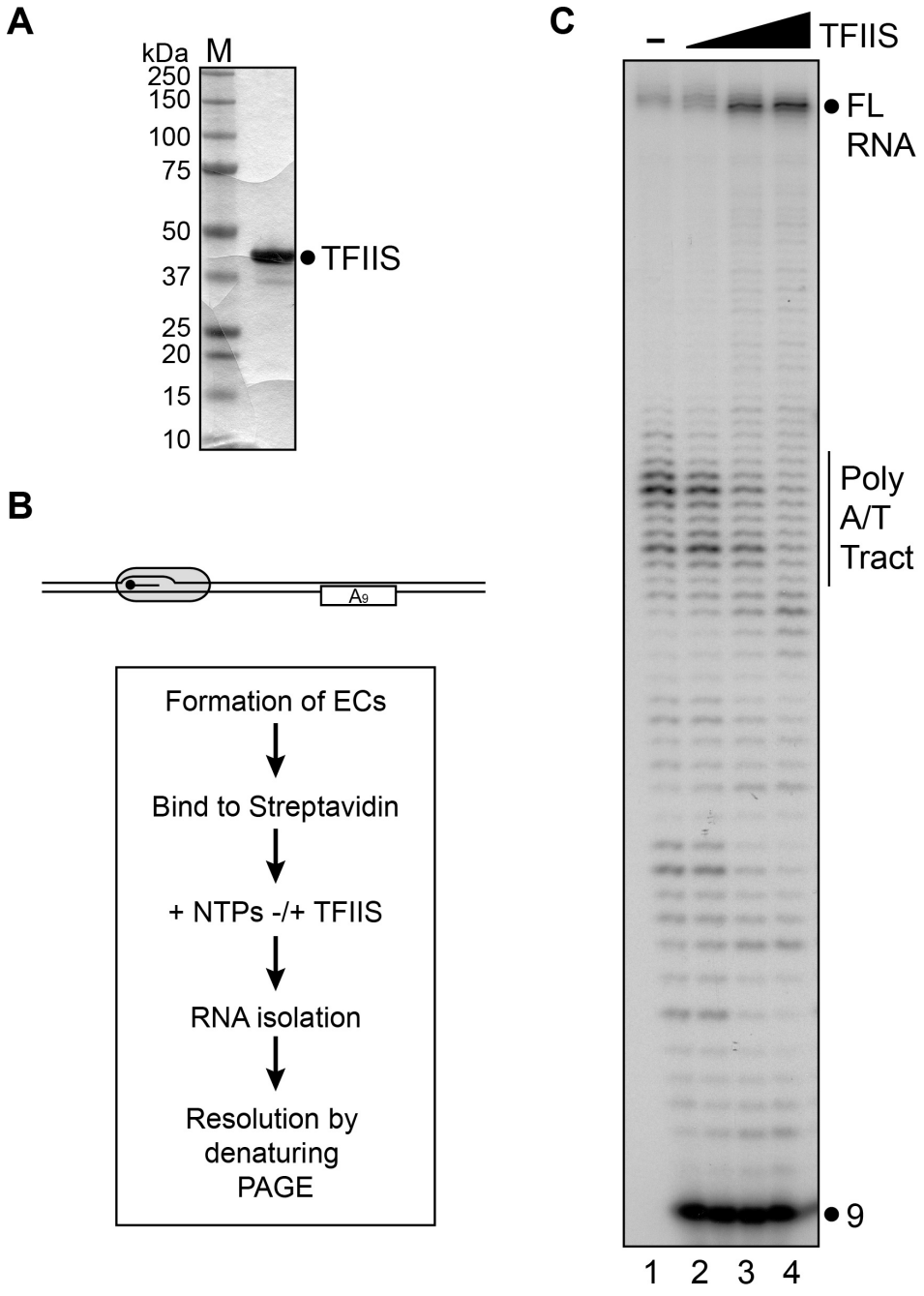

Figure 3.12 Purification of TFIIS and transcription through a pause-site.

A. Coomassie-stained SDS-polyacrylamide gel of Ni-NTA peak fraction of recombinant TFIIS. **B.** Upper, diagram representing mono-EC reconstituted with poly-A/T tract represented (A_9). Lower, schematic showing experimental procedure. **C.** Denaturing PAGE of RNA products isolated from ECs without (-) TFIIS or with increasing amounts of protein (0.1:1 molar ratio of TFIIS to RNAPII (lane 2), 0.2:1 (lane 3) and 1:1 (lane 4)). Presence of the poly-A/T pause site indicated, right.

purified TFIIS, mono-ECs carrying a 3' biotin tag on the TS (Table 2.4) were reconstituted and bound to Streptavidin beads. The ECs were then incubated with NTPs for 1 minute followed by incubation with buffer or increasing amounts of TFIIS for 4 minutes. RNA isolated from these ECs clearly shows that the recombinant TFIIS could facilitate transcription past the pause site and that this occurred in a concentration-dependent manner (Figure 3.12C compare lanes 1 to 2, 3 and 4). When RNAPII and TFIIS were at an equimolar ratio very few polymerases remained paused at the A/T tract, with the majority transcribing the entire template (lane 4). These data show that the recombinant TFIIS could stimulate cleavage of nascent RNA by RNAPII, allowing repeated attempts at transcribing across the pause site.

As a second assay, ECs were formed on a transcription template with a G-less region immediately downstream of the EC, ending with two guanines in a row (a G-stop) (Table 2.4), and bound to Streptavidin beads to measure the stimulation of nascent transcript cleavage by TFIIS. In this instance NTPs (no CTP) were added to the elongation complexes and incubated for 5 minutes to allow transcription to the G-stop, prior to washing away NTPs and the addition of buffer (Figure 3.13B, lane 1), or increasing amounts of TFIIS (lanes 2-4) for 5 minutes. Full-length RNA can be observed for the –TFIIS control. However, its level is dramatically decreased upon the addition of increasing amounts of TFIIS. Indeed, at a 1:1 molar ratio (lane 4), ECs that have not yet begun transcribing are caught and instead, begin to cleave their RNA primer to less than 9 nt in length. These data support the conclusion that RNAPII is indeed backtracking and cleaving its RNA in the presence of recombinant TFIIS.

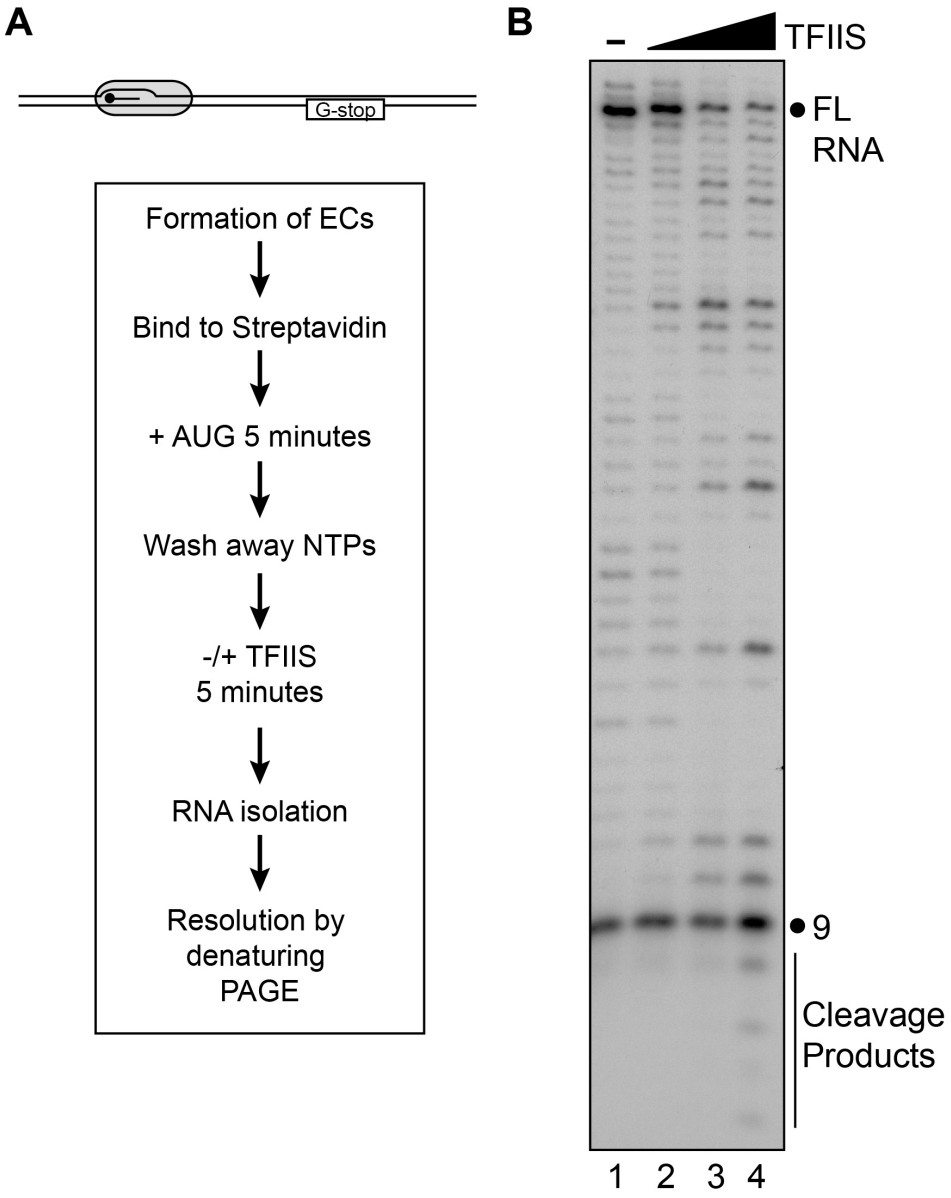

Figure 3.13 Assay of TFIIS ability to stimulate transcript cleavage activity.

A. Schematic of EC showing the presence of G-stop and experimental procedure. **B.** Denaturing PAGE of RNA isolated from ECs without (-) or in the presence of, increasing amounts of TFIIS (0.1:1 molar ratio of TFIIS to RNAPII (lane 2), 0.2:1 (lane 3) and 1:1 (lane 4)).

These experiments showed that the recombinant TFIIS was highly active, and it was therefore now added to di-ECs to determine its ability to stimulate polymerase-polymerase bypass. Di-ECs were reconstituted as before and TFIIS added. Di-ECs were incubated with NTPs and TFIIS for 5 to 60 minutes and RNA analysed by denaturing PAGE. In contrast to the results obtained for the timecourse in the absence of TFIIS (*see section 3.6*), di-ECs incubated with TFIIS were able to produce a small amount of full-length RNA at the 5 minute timepoint (Figure 3.14, lane 1) and this increased in a time-dependent manner (lanes 2-4).

Previous work with tandem, collided di-ECs has shown that after RNAPII collision, polymerases oscillate back and forth between a backtracked and forward translocated state in the presence of NTPs and TFIIS (Saeki and Svejstrup, 2009). At first glance, it seems likely that this is also the case during head-to-head collision and that it helps facilitate transcriptional bypass, but for a number of reasons it is difficult to draw firm conclusions from this experiment. Firstly, 'bypass' is very low, even upon incubation with TFIIS and NTPs for 1 hour; secondly 'bypass' correlates with a significant increase in the level of truncated RNA 9mer (<9 nt length). This opens up the possibility that unlabelled ECs that did not start transcribing prior to TFIIS action are being destabilised by cleaving their short RNA primer, and thus dissociating from the DNA. This would then allow the labelled EC to transcribe the template unimpeded. Following these inconclusive data, it was therefore decided to test transcription-competent extract for a more pronounced activity.

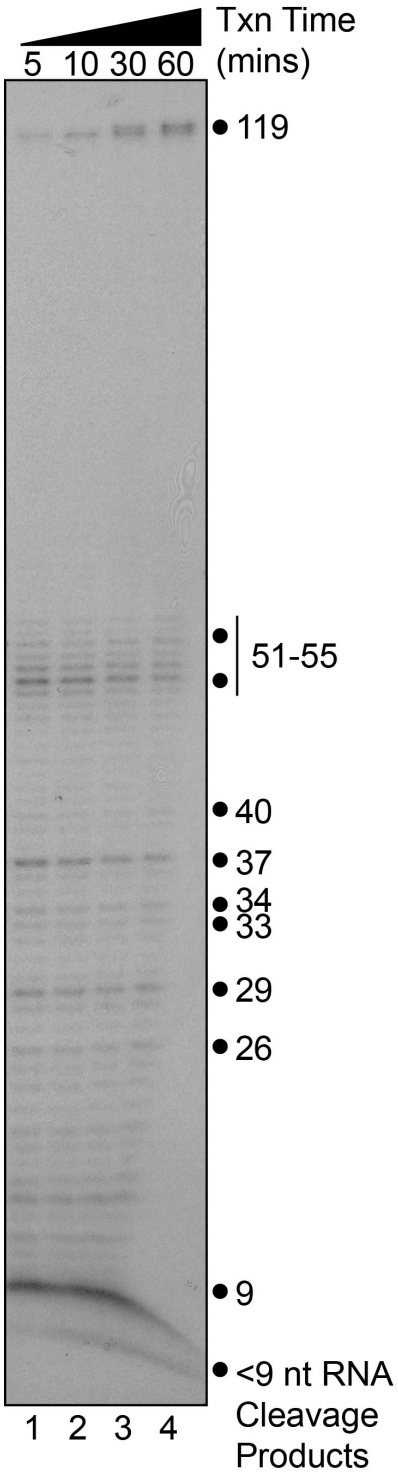

Figure 3.14 A timecourse of transcription by di-ECs in the presence of TFIIS.
Denaturing PAGE of RNA products from a timecourse of NTP and TFIIS incubation. RNA lengths are indicated (right).

3.9 Addition of transcription-competent extract to di-ECs

Following the somewhat ambiguous evidence for a role of TFIIS in assisting RNAPII bypass, transcription-competent extract from *dst1*Δ cells was tested for a bypass activity. *dst1*Δ cells lack the gene encoding TFIIS and should therefore allow detection of any TFIIS-independent bypass activity. Mono- and di-ECs were reconstituted and purified using the dual immuno-purification technique (Figure 3.15A), and incubated with extract for 30 minutes before the RNA products were analysed (Figure 3.15B). Pre-incubation with extract had a dramatic effect on mono-EC transcription efficiency, with higher levels of full-length RNA being produced (compare lanes 1 and 2). This effect was maximal upon the addition of 50 μg of extract, suggesting that saturation had been reached (compare lanes 2 and 3). The stimulation of transcription through the addition of extract is interesting and likely arises as a result of unknown elongation factors acting on ECs to increase RNAPII processivity (*see section 1.3.2.1*). Yeast lacks basic elongation factors such as Elongin and ELL, and TFIIF does not stimulate transcript elongation by yeast RNAPII. It would therefore be of future interest to isolate the responsible activity. Interestingly, pre-incubation of di-ECs with extract also stimulated the production of 119 nt RNA (compare lanes 4-6), suggesting transcriptional bypass. Unfortunately, there was also a concomitant increase in 'collision' RNA products, which might indicate alternative explanations. In these experiments it, thus, cannot be ruled out that the addition of extract stimulates a very minor fraction of contaminating mono-ECs, usually not observable in the absence of stimulation, or that the extract contains a factor that can dissociate elongation complexes, so that the labelled EC in some cases can progress unhindered. The

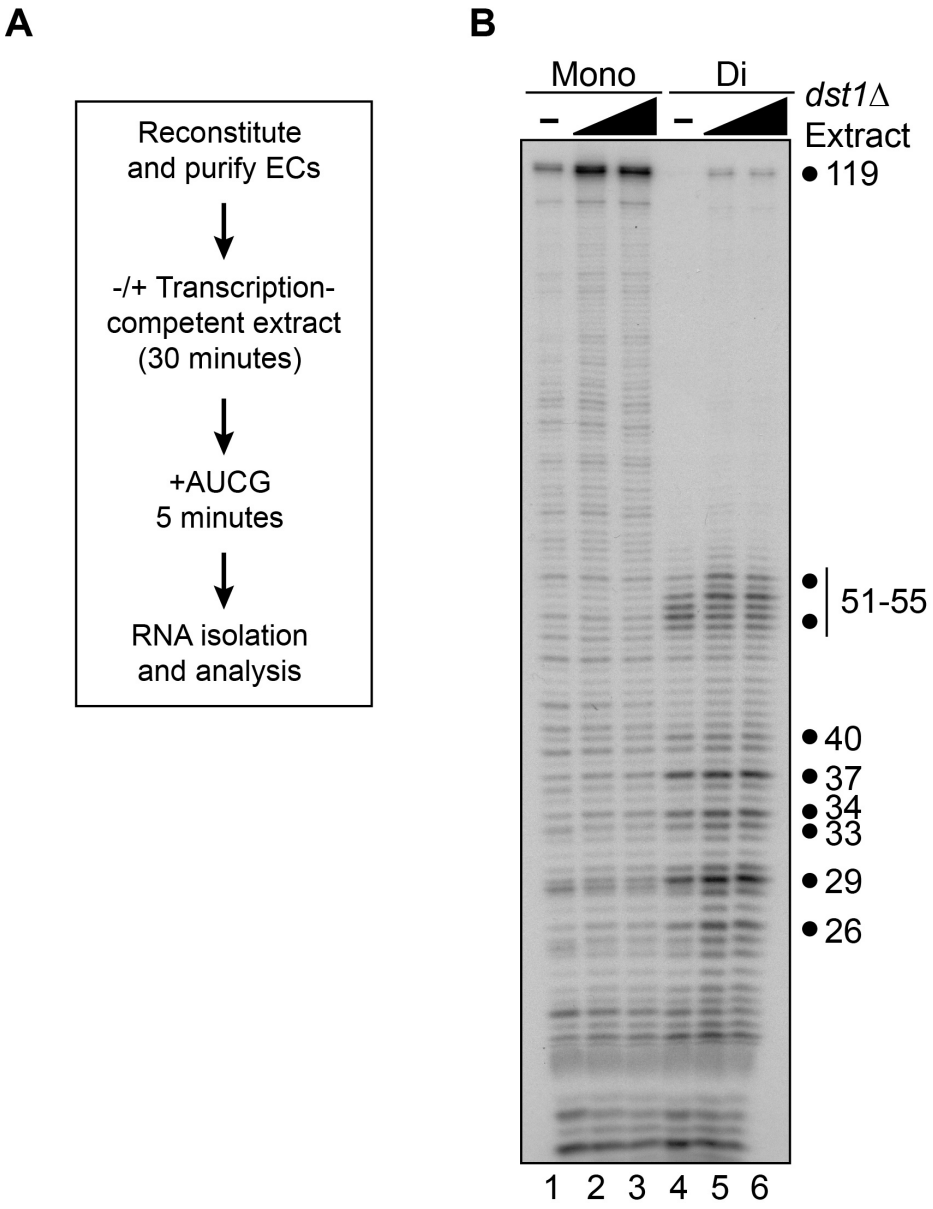

Figure 3.15 ECs pre-incubated with extract.

A. Schematic detailing experimental procedure. **B.** Mono- (lanes 1-3) and di-ECs (lanes 4-6) incubated with 0 µg (-), 50 µg, or 100 µg *dst1*Δ extract. Sizes of transcription products indicated, right.

presence of increased amounts of truncated RNA upon the addition of extract (compare lane 4 to 5 and 6) supports this interpretation. In an ideal situation, increased full-length RNA would be accompanied by a concomitant loss of truncated RNA if transcriptional bypass were occurring.

A different approach was also employed to assay activity present in the extract. To prevent bias towards an EC removal factor, ECs were reconstituted and *dst1*Δ extract added simultaneously with NTPs, followed by 5 minutes incubation at room temperature (Figure 3.16A), prior to native agarose electrophoresis and isolation of di-ECs. Interestingly, RNA from these isolated di-ECs (Figure 3.16B) showed no stimulation of bypass in the presence of extract (compare lane 4 with 5 and 6). Due to the experimental setup, only intact di-ECs were isolated and the levels of contaminating mono-ECs or dissociated complexes were much lower than when the dual-IP method was used.

In conclusion, although tentative evidence for the existence of an RNAPII bypass factor was uncovered using the *in vitro* system, the difficulty of producing and purifying a large amount of di-ECs means that it is not an amenable substrate for use in subsequent biochemical fractionation experiments.

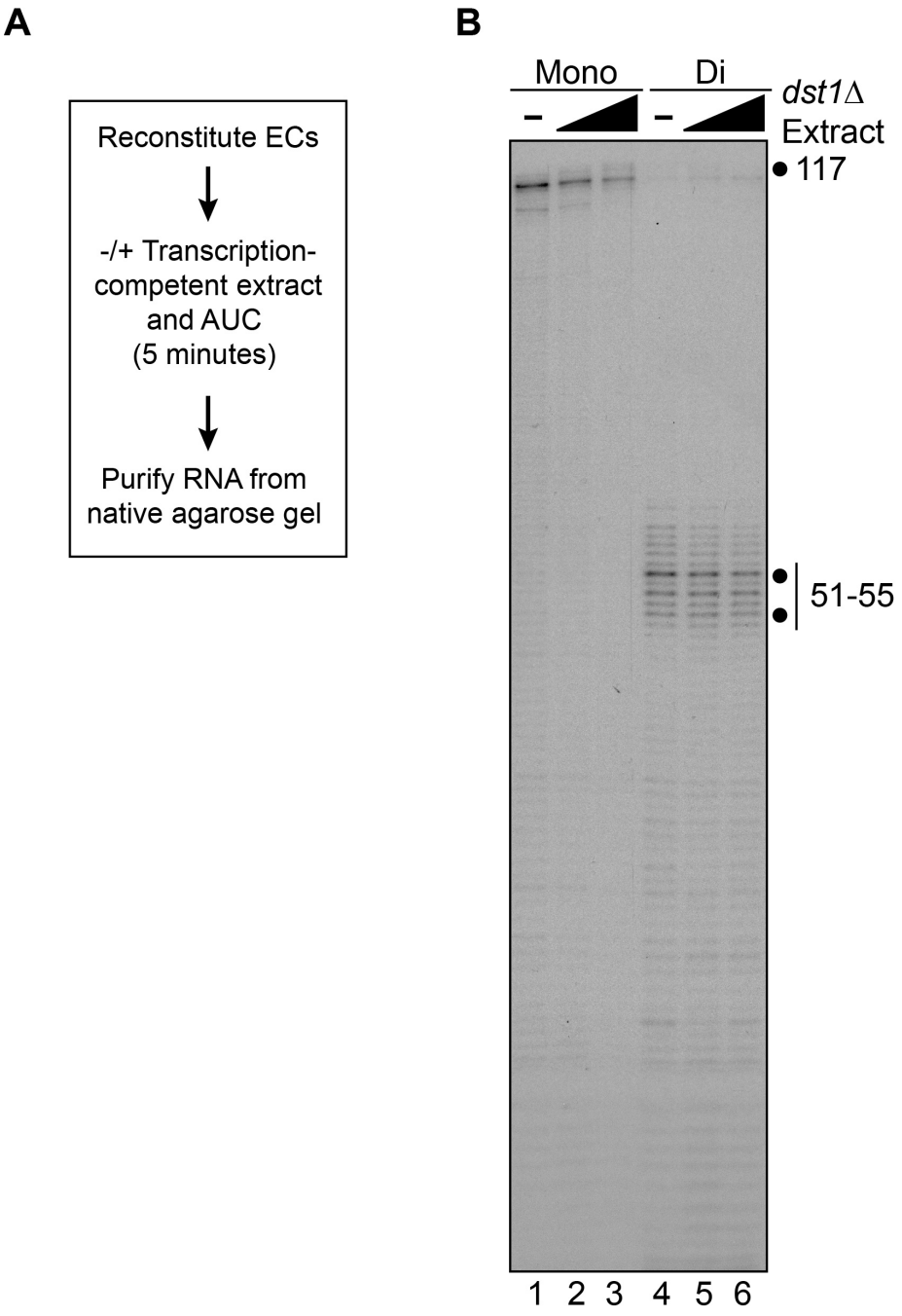

Figure 3.16 ECs incubated with extract and NTPs simultaneously.

A. Schematic detailing experimental procedure. **B.** Transcription analysis arranged as in Figure 3.15B.

3.10 Conclusions

The data presented in this chapter concern the fundamental nature of convergent transcription *in vitro*, providing a biochemical insight into the consequence of collision. Numerous experiments were performed using a robust convergent transcription system reconstituted from highly purified reagents. The presence of a convergently transcribing EC forms a potent block to transcript elongation, resulting in RNAPII arrest as soon as the front edge of the proteins touch. These data indicate that collided RNAPII molecules are highly stable; remaining bound to the DNA template and associated with their nascent RNA product.

Due to the numerous RNAPII-associated factors present *in vivo*, more complex experiments were performed in the presence of TFIIS and cell-free yeast extract. These results show that arrested elongation complexes can be restarted by the addition of recombinant TFIIS, but this seems unable to stimulate significant RNAPII bypass. Likewise, while the addition of transcription-competent extract provides tentative evidence for the presence of an RNAPII bypass factor it first and foremost results in increased transcription efficiency, possibly resulting from the action of Spt4/5 or other unknown factors present in the extract (*see section 1.3.2.1*).

Taken together, the *in vitro* data present convergent transcription as a highly problematic event, which is further exacerbated by the extreme stability of ECs on DNA. *In vivo,* such an event would, at the very least, be expected to result in transcription arrest, preventing further expression of a particular gene and requiring alternative mechanisms to resolve it. It is possible that the collision bypass/resolution mechanism involves proteins inactive in, or co-factors absent from, the extract experiments.

Interestingly, however, it is also an obvious possibility that collision elicits a response similar to that triggered by RNAPII stalled at DNA damage, namely ubiquitylation and proteasome-mediated polymerase degradation (*see section 1.3.3.2*). In order to test these possibilities, it was necessary to move to *in vivo* systems, allowing the study of convergent transcription in its physiological setting.

Chapter 4. RNAPII collision and its resolution *in vivo*

4.1 Introduction

The finding that RNAPII convergent transcription leads to head-to-head collision and polymerase arrest has profound consequences for transcription *in vivo*. Such arrest of transcription in cells is undesirable and can be detrimental, as stalled RNAPII can interfere with other DNA metabolic events, such as replication (Pomerantz and O'Donnell, 2010). Moreover, the presence of arrested RNAPII on a gene is, of course, a potent block to further transcription of that gene by other polymerases, and consequently pathways have evolved to efficiently remove or re-start stalled RNAPII (Svejstrup, 2007a, Somesh et al., 2005, Woudstra et al., 2002, Harreman et al., 2009, Sigurdsson et al., 2010) (*see section 1.3.3.2*). Thus, it can be expected that a pathway(s) exists that resolves collisions following convergent transcription. Additionally, it is necessary to study the nature of collisions *in vivo* because the ECs in the basic, reconstituted system do not fully mimic the multi-factored, dynamic ECs found in cells.

The aim of this chapter is, therefore, to perform an analysis of convergent transcription *in vivo* using the genetically malleable organism, *S. cerevisiae*. The experiments performed aim to answer the following questions:

1. Can transcriptional collision be observed *in vivo*. If so, does transcriptional bypass occur?
2. If not, what is the fate of RNAPII upon collision?
3. How is collision resolved?

The overriding goal is to determine the relevance of the biochemical data (*see Chapter 3*) in a physiological setting. The data presented here are derived from two different

experimental sources: firstly, a convergent gene construct with which *in vivo* biochemistry as well as chromatin immuno-precipitation (ChIP) experiments were performed, and, secondly, endogenous gene pairs for genome-wide ChIP studies. The results obtained outline the consequence of head-to-head collision *in vivo* and support a role for the RNAPII ubiquitylation system in its resolution.

4.2 A convergent transcription construct to study RNAPII collision *in vivo*

The occurrence of RNAPII collision *in vivo* is likely to be stochastically distributed across large areas of DNA making its detection at any specific site difficult. It was therefore surmised that the likelihood of observing it at any specific site is minimal. In order to induce collisions at a specific, observable DNA locus, a plasmid was constructed using the well characterised, inducible, and highly transcribed *GAL* genes. The construct was based on a plasmid previously used in the Proudfoot laboratory (Prescott and Proudfoot, 2002) but modified so that a distinction could be made between promoter occlusion (initiation suppression) and transcriptional collision (elongation suppression). The *GAL10* and *GAL7* genes were cloned opposite each other in a convergent orientation without a terminator between them. Secondly, to allow the measurement of transcription, two cassettes (resulting in G-less *GAL10* RNA) were cloned into the plasmid - one short (105 bp), *GAL10* 'promoter-proximal cassette', and another long (371 bp), 'distal cassette', inserted into the region between the *GAL10* and *GAL7* ORFs, approximately 2 kb downstream of the *GAL10* promoter (Figure 4.1A). Finally, a control construct was created containing a mutated version of the *GAL7*

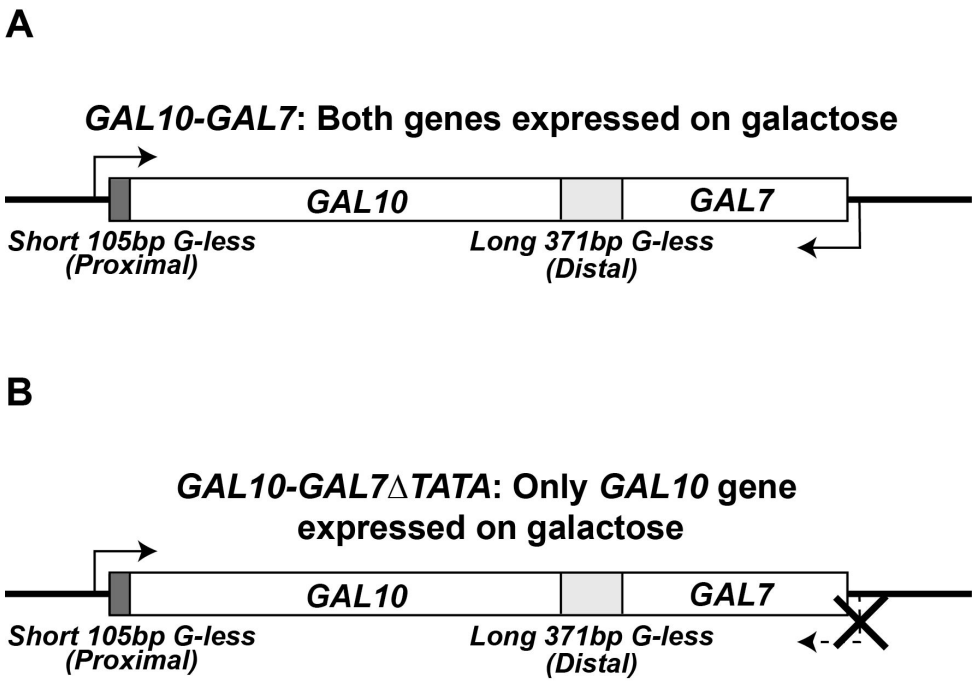

Figure 4.1 Plasmids constructed to study collision *in vivo*.

A. Schematic of the *GAL10-GAL7* construct with short and long G-less cassettes displayed. **B.** *GAL10-GAL7ΔTATA* construct with *GAL7* promoter inactive.

promoter (*GAL10-GAL7ΔTATA*) (Prescott and Proudfoot, 2002) to eliminate any transcription from *GAL7* (Figure 4.1B).

G-less cassettes have been used in the past to study transcript elongation *in vitro* (Lee and Greenleaf, 1997, Sigurdsson et al., 2010) and here the same principles were applied to study RNAPII elongation *in vivo*. Yeast cells were grown in minimal media with raffinose, and then glucose (repressing the *GAL* genes) or galactose (activating the *GAL* genes) was added for 75 minutes prior to RNA extraction and RNase T1 digestion (Figure 4.2A). The RNase T1 specifically cleaves RNA after every G residue, degrading all transcripts but the 105 nt proximal, and 371 nt distal, G-less cassettes of the plasmid-borne *GAL10* RNA. The products were then Northern blotted and detected using a radioactive probe directed against the G-less cassettes (Figure 4.2B). In the presence of glucose, no G-less RNA was produced, showing that the *GAL* genes used were indeed repressed, as expected (Figure 4.2B, lanes 1 and 2). However, following galactose-induced activation, products corresponding to the two G-less cassettes were observed (lanes 3 and 4, proximal and distal). Importantly, the level of distal cassette produced varied greatly between *GAL10-GAL7ΔTATA* (in which there is no collision) and *GAL10-GAL7*, with the collision construct showing a reduction of distal cassette transcription by ~79% (compare lanes 3 and 4). However, the transcription of proximal cassette was largely similar in *GAL10-GAL7ΔTATA* and *GAL10-GAL7*, indicating that changes in the level of distal cassette production were occurring at the level of elongation, not as a result of *GAL10* initiation problems in the collision condition. This allows an important distinction to be made between transcriptional interference as a result of promoter occlusion and head-to-head collisions between actively transcribing

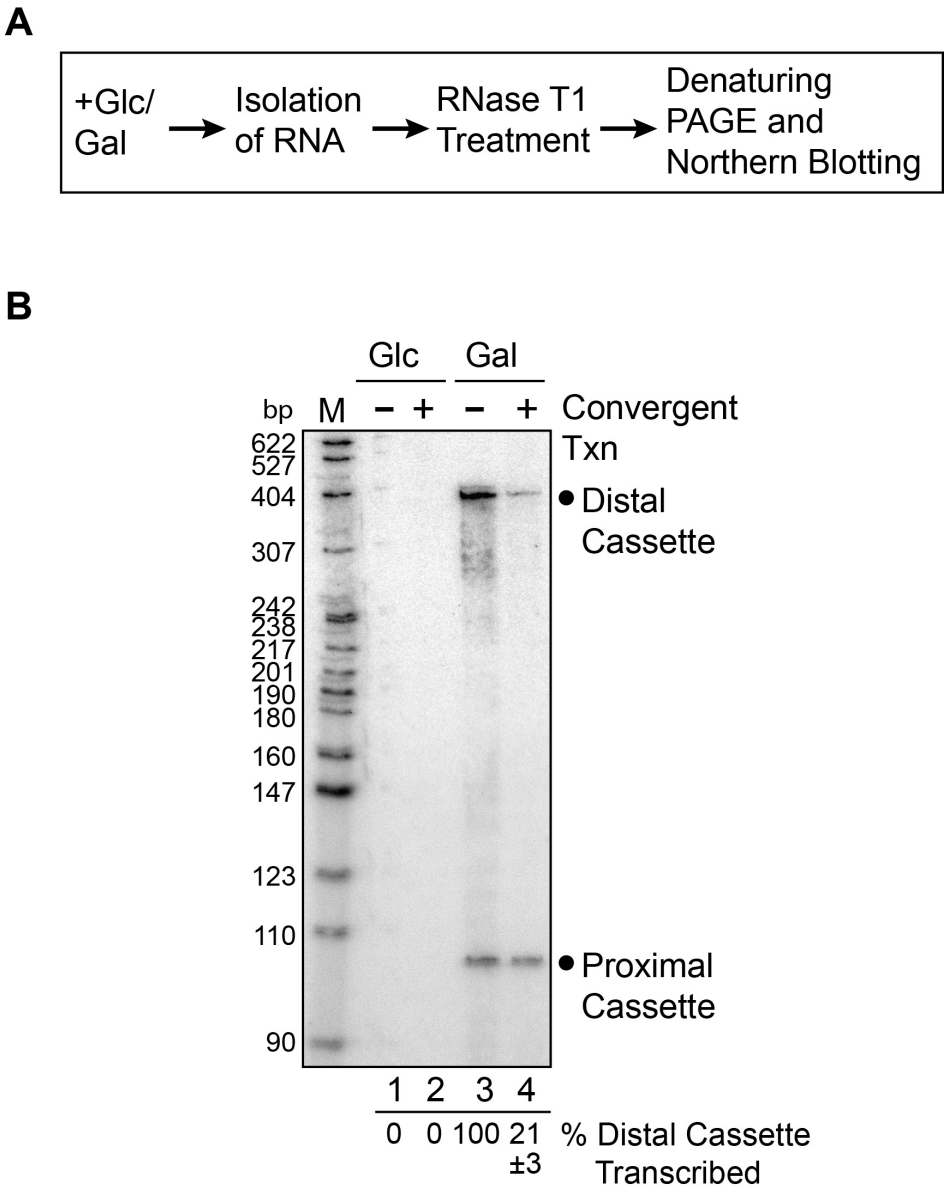

Figure 4.2 Analysis of RNAPII convergent transcription *in vivo*.

A. Schematic representing experimental layout. **B.** Northern blot showing RNA from *GAL10-GAL7ΔTATA* (-) (lanes 1 and 3) and *GAL10-GAL7* (+) (lanes 3 and 4) isolated from strains induced with glucose (lanes 1 and 2) or galactose (lanes 3 and 4). Proximal and distal cassette signals are indicated as well as '% Distal Cassette Transcribed' calculated by normalising to proximal cassette. Values presented are the mean with standard error (calculated from 2 biological replicates) indicated below.

RNAPII ECs. Specifically, these results indicate that collisions are occurring in the ~2 kb region between the G-less cassettes when both promoters are active. These results are in agreement with the biochemical data obtained (*see Chapter 3*) and suggest that RNAPII head-to-head collision is also a potent block to elongation *in vivo*. Additionally, these data confirm and expand upon observations seen in the Proudfoot laboratory (Prescott and Proudfoot, 2002), where run-on assays with *GAL10-GAL7* showed truncated transcripts being produced from both promoters.

The ability to localise RNAPII convergent collisions on a DNA construct formed a platform to study other aspects of this event. Importantly, knowing the location of high-frequency collision, chromatin immuno-precipitation (ChIP) could be performed to assess the fate of the collided ECs.

4.3 Upon collision RNAPII is removed from DNA in a ubiquitylation-dependent manner

The potent block to transcript elongation observed upon convergent transcription *in vivo* suggested that cells do not efficiently deal with RNAPII head-to-head collision, at least on the highly transcribed gene construct. Conceptually, it can be assumed that if two highly active promoters are driving transcription, polymerases would accumulate on the gene and the result would be an RNAPII 'traffic jam'. In order to gain an insight into the events occurring at the interface of collision (specifically the fate of polymerases), RNAPII ChIP was employed.

Yeast containing the collision constructs were grown in liquid culture and transcription activated, prior to formaldehyde crosslinking, chromatin extraction, and sonication. ChIP was performed with the 4H8 antibody, directed against the C-terminal

domain of the largest RNAPII subunit, Rpb1, whilst non-specific ChIP using IgG was performed as the control (Figure 4.3A). Unexpectedly, instead of an increase in RNAPII levels in the region between convergent promoters in the *GAL10-GAL7* collision construct (Figure 4.3B, lower, WT, white bar), the level was much reduced when compared to *GAL10-GAL7ΔTATA* (WT, black bar). These data suggest RNAPII is removed following transcriptional collision.

Previous studies have shown that arrested RNAPII can be removed from DNA in a ubiquitylation-dependent manner. This is catalysed by recognition of arrested RNAPII, mono-ubiquitylation by Rsp5 and poly-ubiquitylation by the Elongin-Cullin complex in a Def1-dependent manner (*see section 1.3.3.2*). It was surmised that the ubiquitin-proteasome system could also play a role in the removal of arrested RNAPII from the convergent gene construct post-collision. In order to test this, ChIP was performed in two key mutants defective for RNAPII poly-ubiquitylation, namely *elc1*Δ and *def1*Δ. In contrast to the decrease observed in WT cells, ChIP in *elc1*Δ cells showed a reproducible accumulation of RNAPII on the *GAL10-GAL7* construct (Figure 4.3B, compare WT and *elc1*Δ). A similar ChIP profile was observed in *def1*Δ cells (Figure 4.3B, compare *elc1*Δ and *def1*Δ). The accumulation of RNAPII in these two RNAPII ubiquitylation mutants argues strongly that this pathway is involved in RNAPII removal from the gene following transcriptional collision, in all likelihood through proteasome-mediated degradation of the RNAPII subunit Rpb1.

Next it was important to determine if the accumulation of RNAPII, observed as a result of the inability to poly-ubiquitylate and degrade polymerases, had an affect on transcription *in vivo*. The G-less cassette constructs were used as a measure of transcript

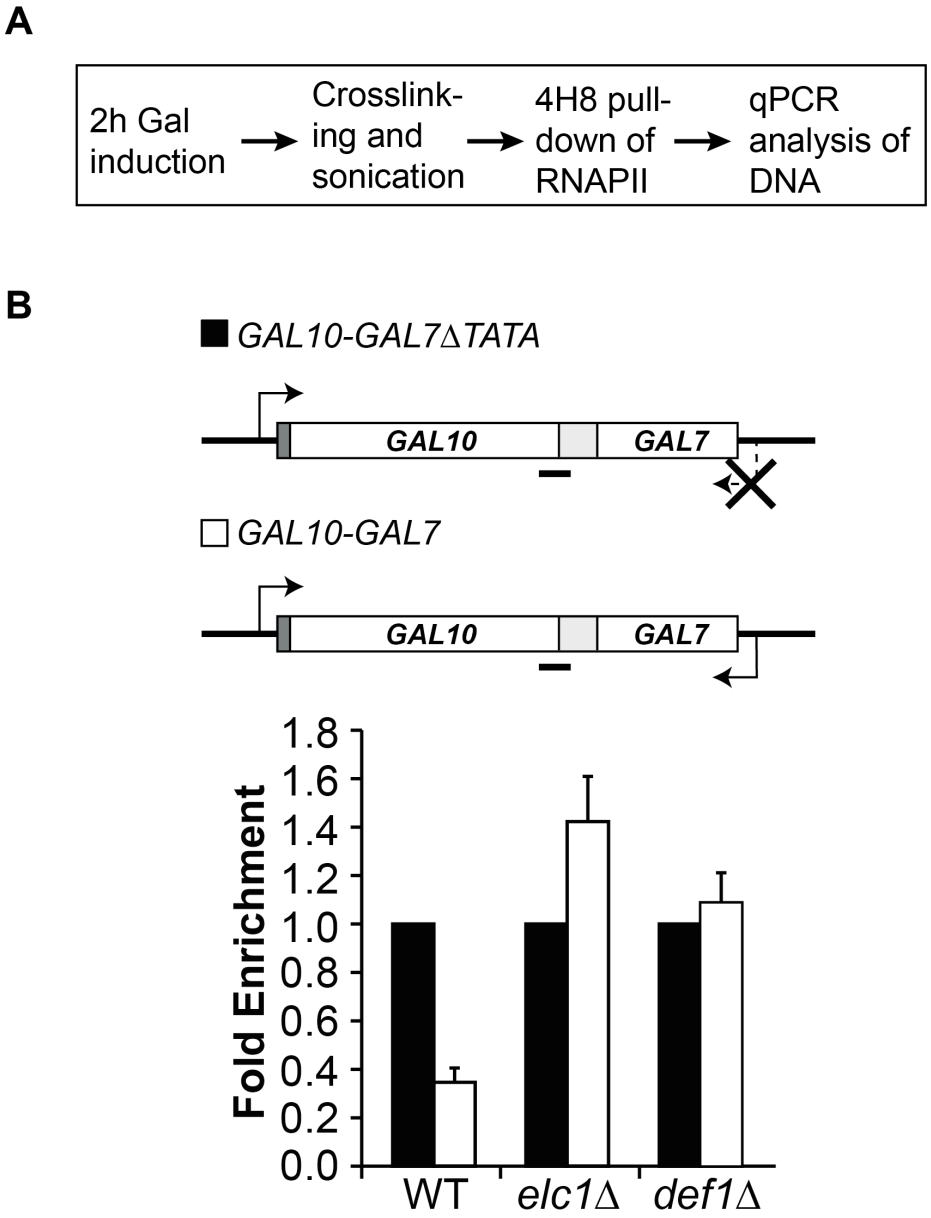

Figure 4.3 ChIP of RNAPII in WT cells and Rpb1 ubiquitylation mutants.

A. Schematic showing procedure for performing ChIP. **B.** Upper, diagram of the *GAL* genes showing the position of qPCR primers (black bar). Lower, qPCR analysis of immunoprecipitated DNA from WT, *elc1*Δ and *def1*Δ cells. Results normalised to *GAL10-GAL7*Δ*TATA* (=1). Error bars show standard error calculated from 3 biological replicates.

elongation in the presence of convergent transcription in WT and *elc1*Δ cells. The experiment was performed as before except that this time transcription was measured at various time points after induction. Surprisingly, there was little effect of deleting *ELC1* on transcript levels, though collision seems to shut down transcription of the distal cassette more rapidly, and may spread with time so that even the proximal cassette is relatively less transcribed in *elc1*Δ compared to WT (Figure 4.4A and B, compare the distal and proximal cassette transcription over time). However, these differences are surprisingly small. It is possible that the removal of collided polymerases by ubiquitylation/degradation in WT cells is not rapid enough to keep up with the high initiation rate and collision on this construct.

The notion that ubiquitylation might have a role in bypass is an interesting one and had actually already been tested *in vitro* using the biochemical scaffold system, prior to the *in vivo* observation. Although this chapter is focussed on *in vivo* work, this seems an appropriate place to present those findings. The two documented ubiquitylation sites of RNAPII are Rpb1 K330 (Figure 4.5A) and K695 (Figure 4.5B), with K330 located surprisingly close to the RNAPII active centre in a position that might potentially alter DNA conformation, or the interaction of Rpb1 with the NTS (Figure 4.5A). To determine if ubiquitylation directly stimulates transcriptional bypass, mono- and di-ECs were reconstituted and ubiquitylated *in vitro* using highly purified E1 (Uba1), E2 (Ubc5), E3 (Rsp5) and lysine-free ubiquitin, to allow only mono-ubiquitylation of Rpb1 (Figure 4.6A). Otherwise, Rsp5 generates (non-physiological) K63-chains on RNAPII *in vitro*, and it was surmised that if RNAPII ubiquitylation played a role in elongation then the presence of a bulky ubiquitin moiety would be sufficient to show an effect. Di-ECs were efficiently ubiquitylated prior to transcription,

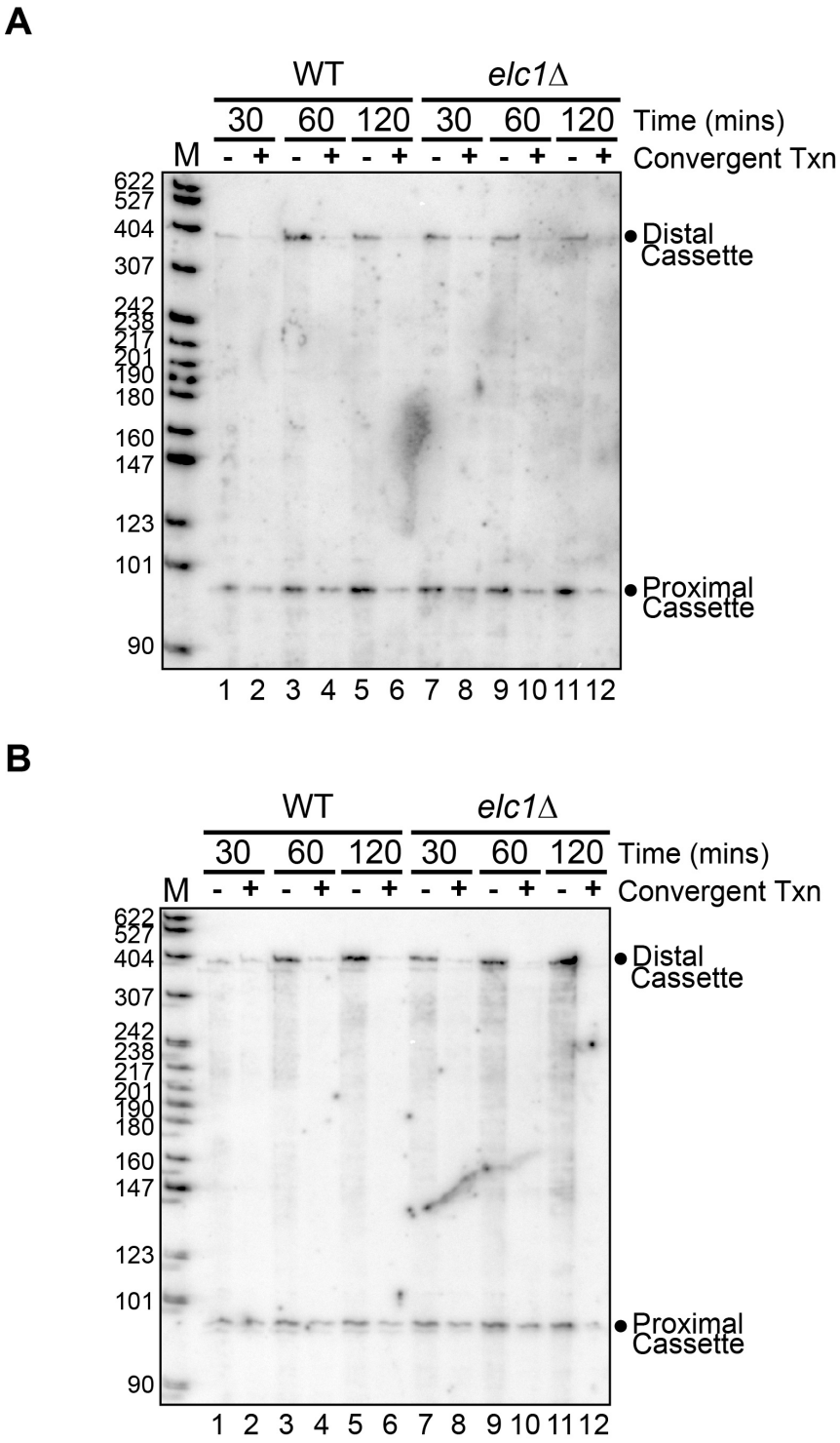

Figure 4.4 An analysis of transcription in WT and *elc1*Δ cells.

A. Northern blot of RNA isolated from WT (lanes 1-6) or *elc1*Δ (lanes 7-12) cells containing *GAL10-GAL7*Δ*TATA* (-) or *GAL10-GAL7* (+) 30, 60, or 120 minutes after galactose induction. Proximal and distal cassette are indicated (right). **B.** A biological replicate of **A**.

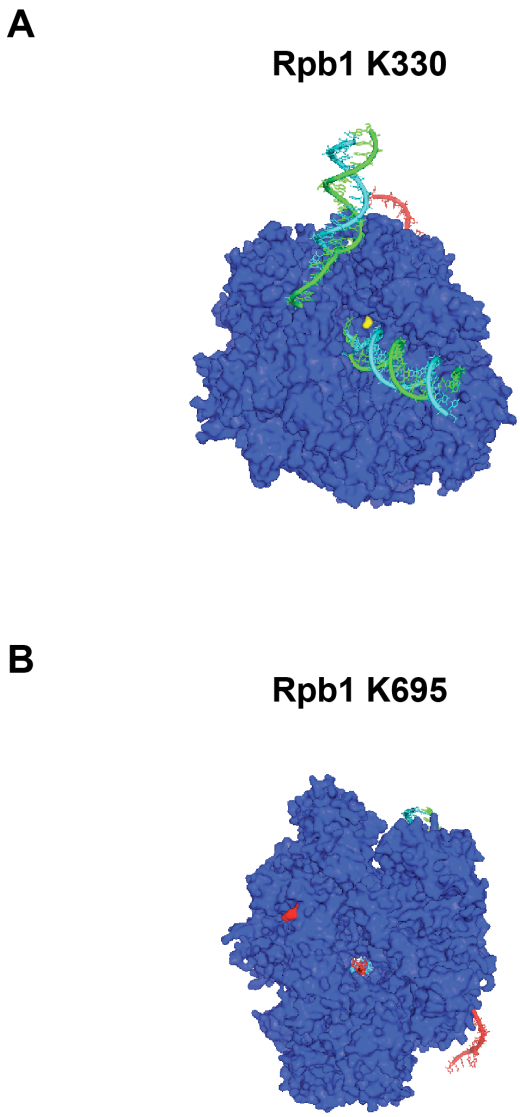

Figure 4.5 RNAPII structure with highlighted ubiquitylation sites.

A. Crystal structure of an RNAPII-EC (from above) with Rpb1 K330 highlighted (yellow). TS (cyan), NTS (green) and RNA (red) are also shown. **B.** RNAPII-EC structure (from underneath) with K695 highlighted (red). The RNAPII pore can be identified (centre) by the DNA and RNA being visible through it.

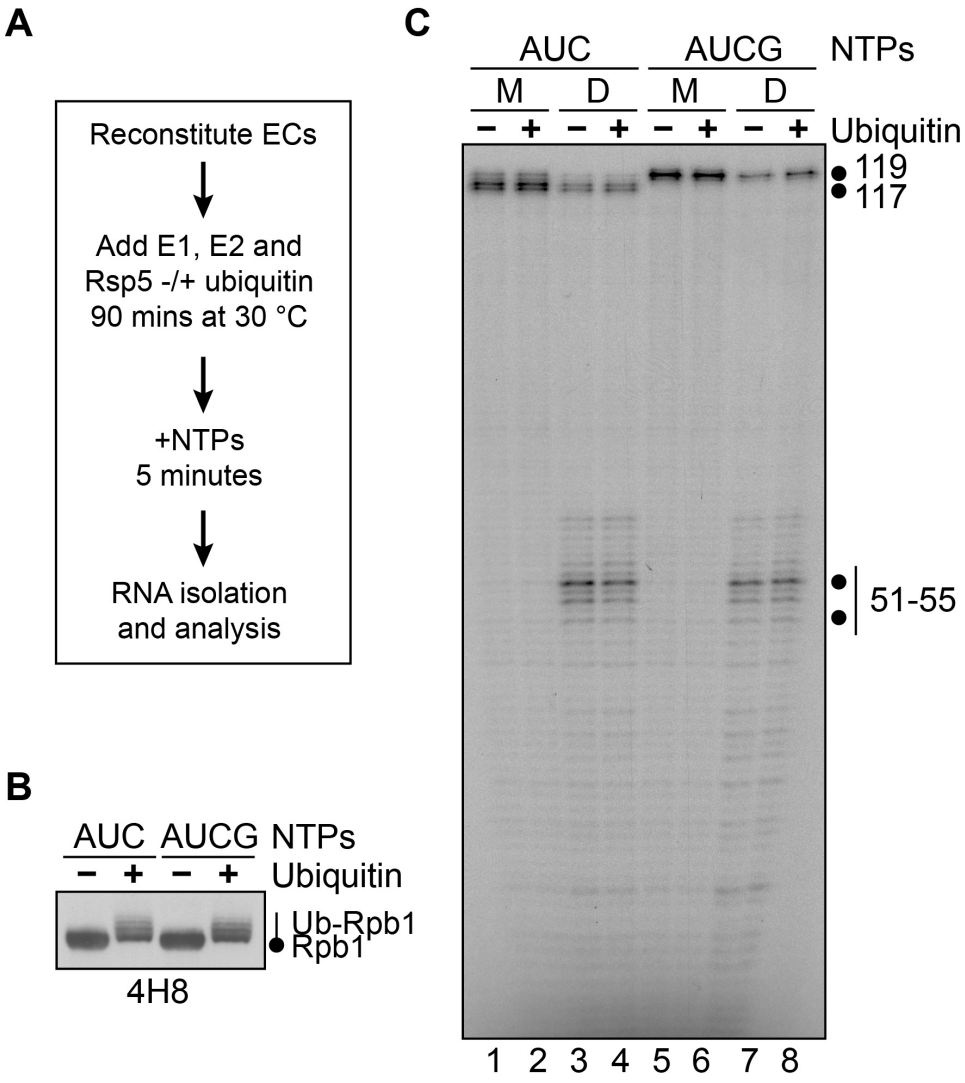

Figure 4.6 Ubiquitylation of ECs.

A. Experimental schematic. **B.** Western blot against Rpb1 (4H8 antibody) of di-ECs used in **C** prior to initiating transcription. Rpb1 and ubiquitylated Rpb1 indicated, right. **C.** Transcription analysis of mono- (M) and di-ECs (D) non-ubiquitylated (-) or ubiquitylated (+) prior to transcription with AUC (lanes 1-4) or AUCG (lanes 5-8).

with the vast majority of Rpb1 being ubiquitylated at one or more sites, as expected (see mobility shifts in Figure 4.6B). Transcription was initiated via the addition of a selection of NTPs (AUC), (Figure 4.6C, lanes 1-4) or all NTPs (AUCG) (Figure 4.6C, lanes 5-8), and RNA isolated. As these elongation complexes were un-purified, a certain percentage of mono-EC contamination was present in the reaction, resulting in some full-length RNA production. However, the level of full-length RNA did not change significantly during transcription with ubiquitylated di-ECs, nor did the level of the different truncated RNAs decrease (compare lanes 3 and 4 and lanes 7 and 8). These data show that, somewhat surprisingly, RNAPII ubiquitylation does not significantly affect transcription. More specifically, it also fails to induce transcriptional bypass. These data add further support to the assertion that ubiquitylation-dependent removal of RNAPII from the *GAL10-GAL7* gene *in vivo* is a result of proteasome-mediated degradation, rather than ubiquitylation-mediated bypass.

4.4 RNAPII collision intermediates are stabilised in an *elc1*Δ mutant

To further characterise the post-collision role of Elc1, an experiment was designed to study the rate of RNAPII removal from DNA. The addition of glucose to cells results in the rapid shutdown of transcription from *GAL* genes and has been extensively characterised (Mason and Struhl, 2005). Therefore, this could be used to determine the rate of RNAPII removal from *GAL10-GAL7* compared to the internal control *GAL1* (*GAL1* is co-regulated with *GAL10* at the endogenous locus and therefore provides an ideal control for normal transcriptional shutdown and RNAPII

disappearance). It was surmised that collided RNAPII should have a longer half-life than regularly transcribing and terminating RNAPII, especially in an *elc1*Δ mutant.

Analysis of RNAPII occupancy on the *GAL1* gene (Figure 4.7A) shows a significant, rapid shutdown of transcription upon glucose addition, with minimal levels of RNAPII remaining on the gene after 6 minutes. At this gene, both WT and *elc1*Δ show similar repression kinetics, with only a slight initial delay in *elc1*Δ. However, even in WT cells, RNAPII occupancy on the *GAL10-GAL7* collision construct did not decrease as rapidly as seen at the *GAL1* control, with ~50% RNAPII remaining even after 6 minutes (Figure 4.7B, black bars). This difference was even more pronounced in *elc1*Δ, which showed an increase in the level of RNAPII arrested at *GAL10-GAL7*, which stabilised at ~80% RNAPII following 6 minutes of repression (Figure 4.7B, grey bars). These data show that, not surprisingly, the rate of RNAPII removal from a gene is similar in WT and *elc1*Δ when collision is absent. However, in the presence of head-to-head collision, arrested ECs require poly-ubiquitylation and degradation in order to be removed from the gene, with collision intermediates stabilised in the ubiquitylation-deficient *elc1*Δ mutant.

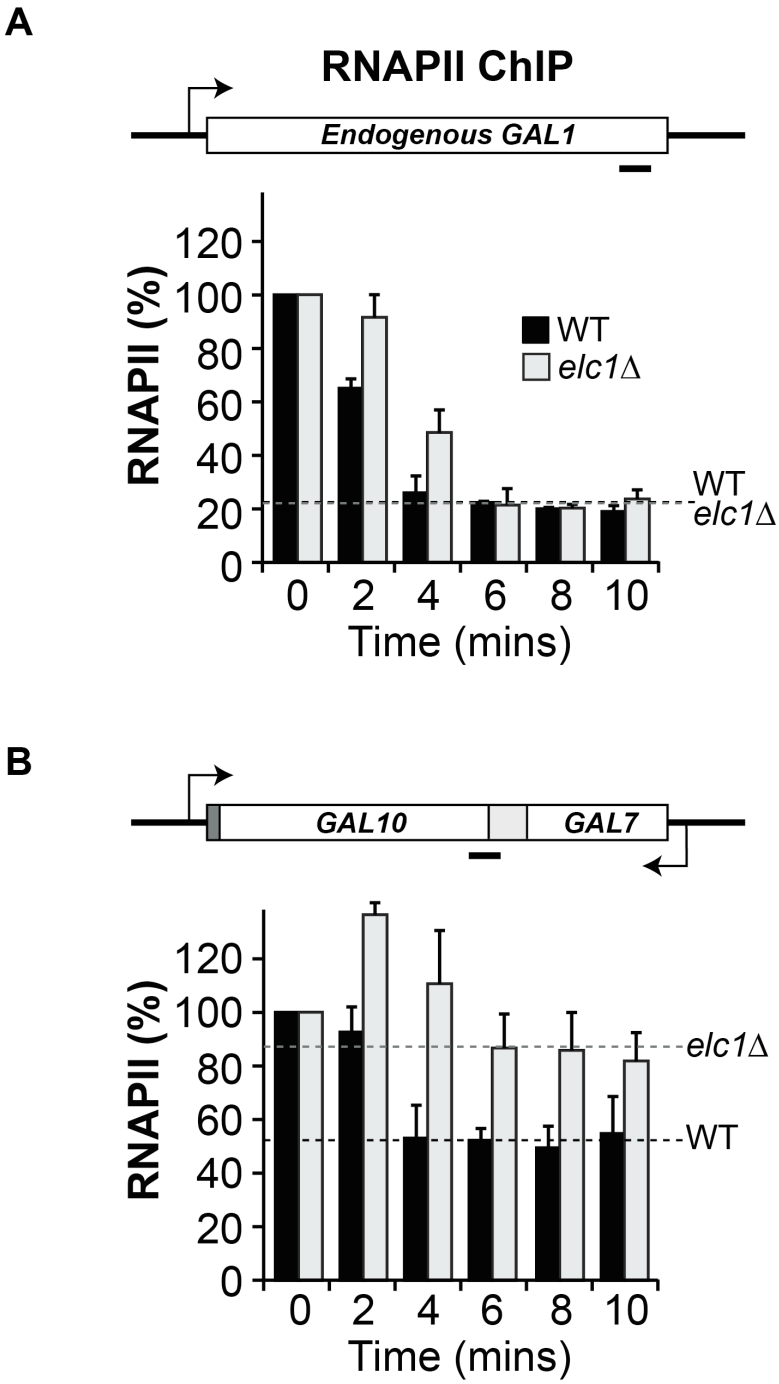

Figure 4.7 Timecourse of RNAPII occupancy on endogenous *GAL1* or *GAL10-GAL7* collision construct.

A. Upper, diagram of endogenous *GAL1* (no-collision control) showing the position of the qPCR product (black bar). Lower, RNAPII occupancy following transcriptional shutdown (time after, indicated) in WT or *elc1*Δ cells. Values normalised to 0 minute timepoint (=100%). Error bars are standard error from 3 biological replicates. **B.** As in **A** but for *GAL10-GAL7* collision construct.

4.5 RNAPII accumulates between natural convergent genes in an *elc1*Δ mutant

In an *elc1*Δ mutant, the inability to poly-ubiquitylate and degrade RNAPII means that the usually short-lived post-collision intermediates are stabilised. As mentioned previously, RNAPII head-to-head collisions are likely to be stochastically distributed under normal circumstances, making their detection difficult. However, the increased stability, and likely accumulation, of collided complexes in an *elc1*Δ strain raised the possibility of observing collision at natural genes. It is important to study this at natural genes expressed at endogenous levels, as arguably, the convergent gene construct used in the previous experiments does not properly represent the true *in vivo* situation, being highly expressed and engineered for the purpose of promoting RNAPII collision. Additionally, detection of collided RNAPII at endogenous loci would provide support for this phenomenon occurring in cells.

A genome-wide approach was employed to study collided RNAPII on endogenous genes in both WT and in an *elc1*Δ mutant. ChIP of RNAPII was performed under highly stringent conditions, again using the CTD-specific antibody 4H8, or IgG as a control. The ChIP was efficient and highly specific for Rpb1 of RNAPII across both cell types and between replicates (Figure 4.8, compare lanes 3 and 5). ChIP specificity is of even greater importance in genome-wide studies, where relatively low levels of DNA need to be detected. The presence of false ChIP signals can thus lead to masking of true occupancy signals and generation of false positives.

The DNA from the immuno-precipitated samples was subjected to library preparation at the CRUK-LRI Advanced Sequencing Facility, and sequenced using a GAII-x Illumina sequencer and standard protocols. Bioinformatic analysis of the data

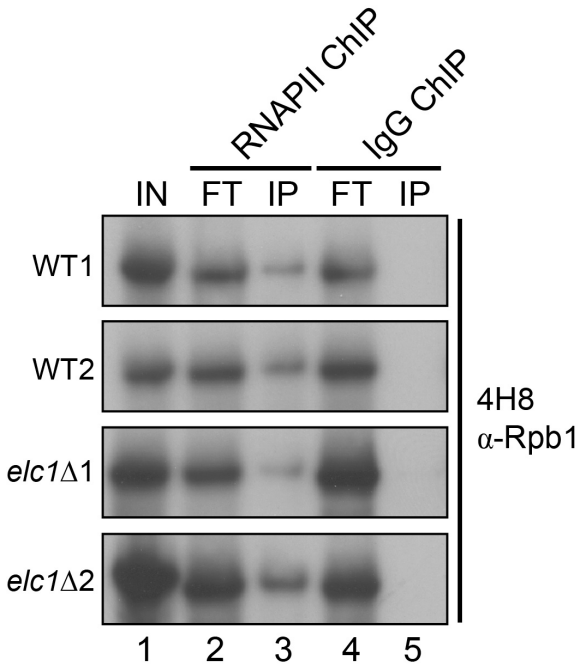

Figure 4.8 Western blot of RNAPII immuno-precipitation.

Western blots from two WT and *elc1*Δ replicate ChIPs probed with anti-Rpb1 antibody (4H8). RNAPII levels for Input (IN), flow-through (FT) and immuno-precipitated sample (IP) following RNAPII or IgG (control) chromatin immuno-precipitation are presented.

was performed in collaboration with Wu Wei and Lars Steinmetz (Genome Biology Unit, EMBL). The sequencing reads obtained were filtered and aligned to the *S. cerevisiae* genome before subtraction of the IgG control signal and grouping of genes into convergent, divergent, or tandem pairs (*see section 2.10*). Finally, intergenic RNAPII density (averaged from two biological replicates) was calculated across all gene pairs.

If collisions do indeed occur between polymerases transcribing natural genes, it is expected that these would most likely be observable at convergent genes in *elc1*Δ cells (Figure 4.9A, schematic). Remarkably, although there was no significant difference in RNAPII density inside the coding regions of genes, a clear increase was indeed observed in the region between convergent genes in the *elc1*Δ mutant (Figure 4.9A). Importantly, when a similar analysis was performed between divergent genes (Figure 4.9B, schematic), which are orientated away from one another and represent areas of low collision probability, RNAPII density is similar between WT and *elc1*Δ (Figure 4.9B). This shows that the *elc1*Δ RNAPII peaks are specific to 3' ends of convergent gene pairs, where collision is most likely to occur.

A difficulty arises when interpreting data concerning RNAPII head-to-head collisions on natural genes. This is due to the frequent, pervasive and randomly distributed nature of non-coding transcription. Additionally, many promoters are bi-directional, and antisense transcripts are also commonly synthesised from the 3' end of genes (Xu et al., 2009, Neil et al., 2009, Katayama et al., 2005). Therefore, divergent gene pairs are the only good example of a negative control, as collisions are also likely to occur at, for example, the promoter-proximal and 3' regions of tandem gene pairs (Figure 4.10A, schematic). In support of this, when RNAPII density was analysed at

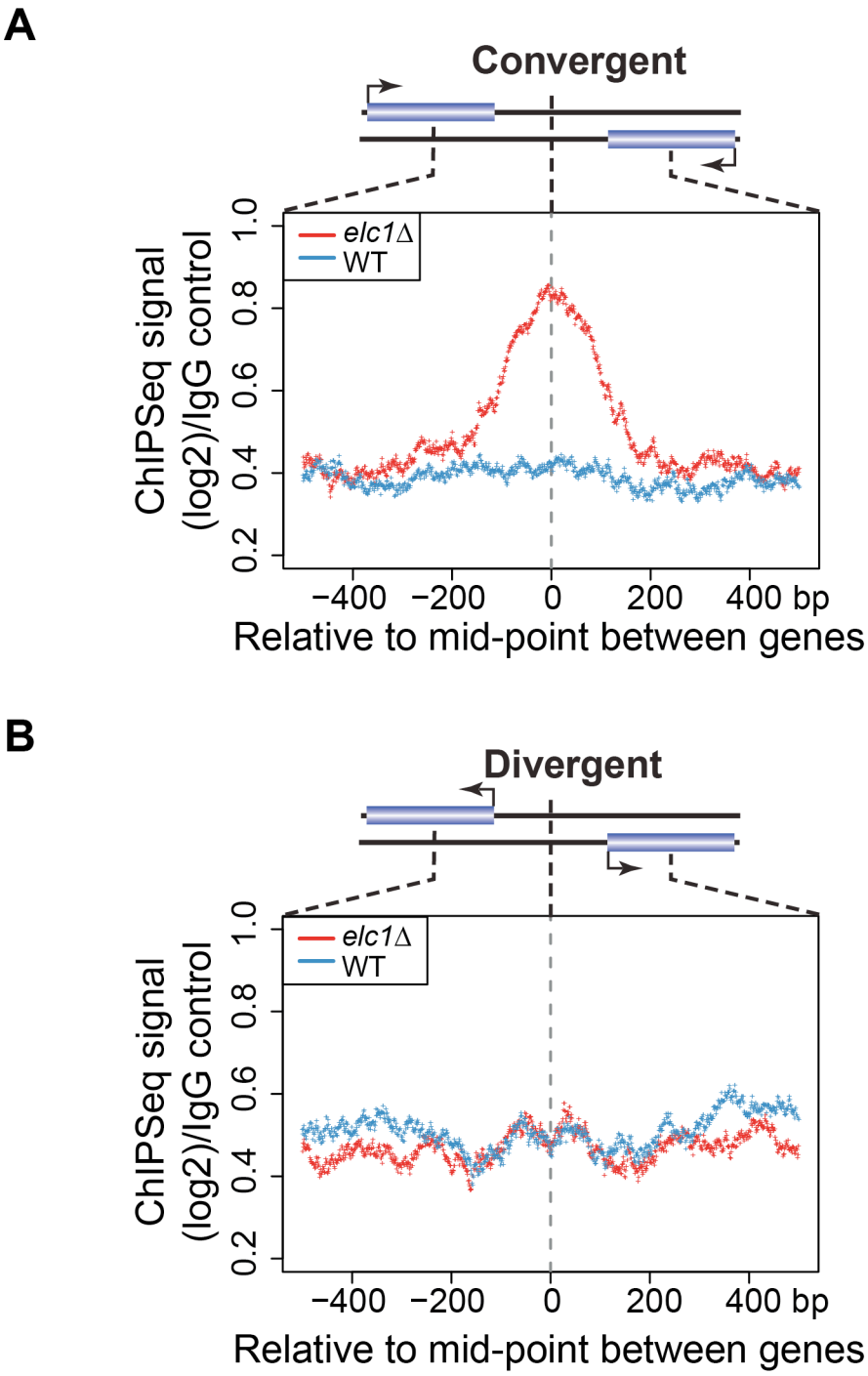

Figure 4.9 RNAPII occupancy on convergent and divergent gene pairs.

A. Upper, schematic representing regions of analysis. Lower, ChIP-Seq data from WT (blue) and *elc1*Δ cells (red) showing RNAPII occupancy at the mid-point between convergent genes. Data plotted is the average of two biological replicates. **B.** As in **A** but data plotted for all divergent genes.

tandem gene pairs, RNAPII accumulated in the 3' intergenic region between the end of one ORF and the promoter of the downstream gene in *elc1*Δ, but not in WT cells (Figure 4.10 A). The increased RNAPII density in *elc1*Δ cells was, however, not as dramatic as for convergent genes (compare with Figure 4.9A), which are likely sites of more frequent head-to-head collisions. Additionally, when tandem gene pairs separated by greater intergenic distances (<600 bp but >400 bp) were analysed specifically, the RNAPII peak observed in *elc1*Δ cells resolved into two distinct peaks (Figure 4.10B, upper). There was not a similar peak distribution of RNAPII in WT cells (Figure 4.10B, upper, compare blue and red lines) and the peaks could not be resolved when genes separated by short intergenic distances were analysed (Figure 4.10B, lower). The two peaks seen at these tandem gene pairs suggest that there are two separate populations of RNAPII arrested in the intergenic region. It is tempting to interpret these as separate collision events; one occurring between sense- and antisense-transcribing polymerases at the 3' end of the first gene, and another between polymerases firing from the bi-directional promoter of the downstream gene and those terminating from the upstream gene (Figure 4.11). The overall probability of these collisions is likely lower than that expected at convergent genes, and the ChIP signal supports this.

Another possible explanation for the accumulation observed at both tandem and convergent gene pairs is that deletion of *ELC1* causes RNAPII termination defects (although the gene has never previously been implicated in this process), so to address this possibility RNA-Seq was performed. Here, it was surmised that general read-through of terminators would be observable as longer RNAs in *elc1*Δ. Total RNA was extracted from WT and *elc1*Δ cells grown to mid-log phase in YPD. Extracted RNA was RiboZero treated to remove rRNA, but poly-A selection was not performed, to

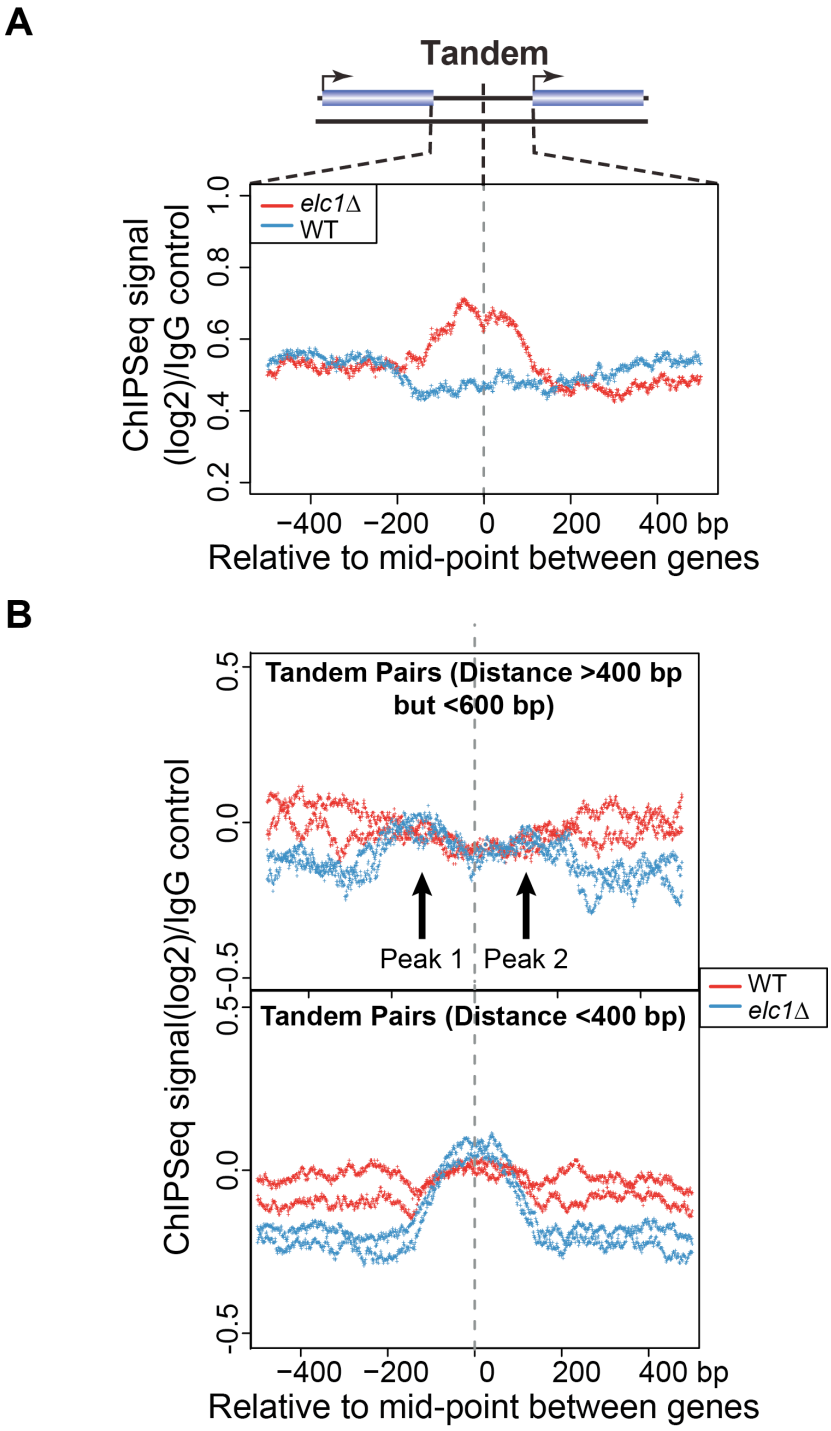

Figure 4.10 RNAPII occupancy on tandem genes.

A. Upper, schematic representing region of analysis. Lower, ChIP-Seq data plotted at the intergenic region between tandem gene pairs in WT (blue) and *elc1*Δ (red). **B.** Filtered ChIP-Seq data showing tandem pairs separated by >400 bp but <600 bp with peaks indicated (upper) and <400 bp only (lower). Note that WT is red and *elc1*Δ blue in this plot.

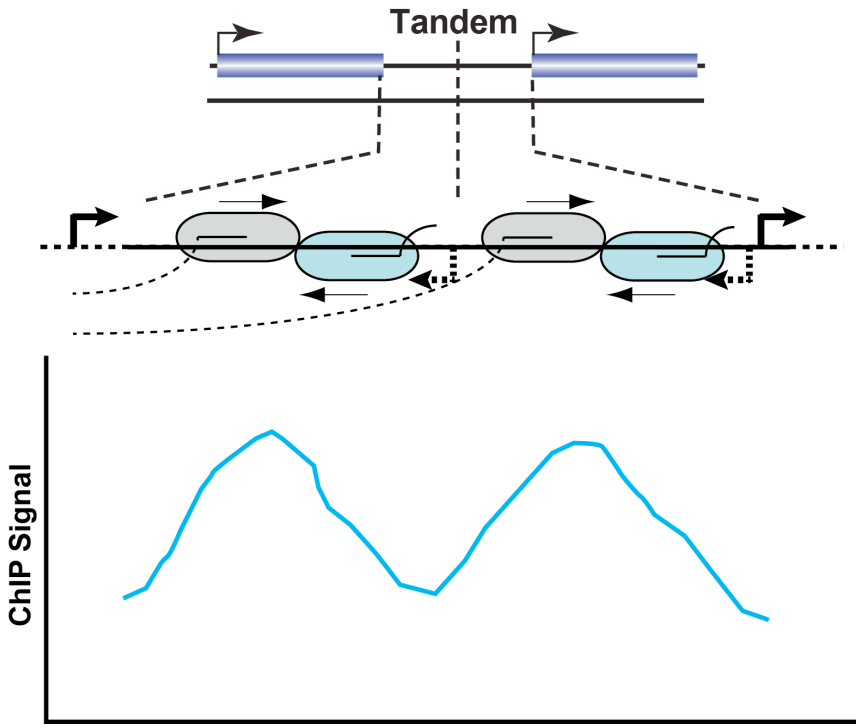

Figure 4.11 Schematic representation of collision at tandem genes.

Upper, schematic representing RNAPII collisions between gene-initiated RNAPII and a 3' antisense-transcribing polymerase (left), or with RNAPII initiated from a bi-directional promoter (right). Sense (gene) promoters represented by solid arrow, dashed arrow represents antisense promoter. Lower, representation of ChIP signal in Figure 4.10B.

prevent a bias towards mRNA. As before, the RNA was sequenced using the GaIIx Illumina platform and reads aligned to the *S. cerevisiae* genome; bioinformatics was again performed in collaboration with Wu Wei and Lars Steinmetz. As before, genes were classified into three groups dependent upon their orientation and the reads at the 3' end of the gene and intergenic region plotted. At both tandem (Figure 4.12A) and convergent (Figure 4.12B) gene pairs, the RNA levels were practically identical in WT and *elc1*Δ cells (compare blue and red lines). This argues against a termination defect in *elc1*Δ mutants, which would be expected to manifest itself as an accumulation of transcripts at the 3' end of genes.

The data presented in this section show that RNAPII accumulates in an *elc1*Δ mutant specifically in regions where head-to-head collisions are expected to occur. WT cells do not display similar patterns of RNAPII build-up, suggesting that in the presence of a fully functioning ubiquitin-proteasome system collided polymerases are efficiently removed. The presence of the Elongin-Cullin pathway indicates why cells tolerate potent, gene-blocking RNAPII collision.

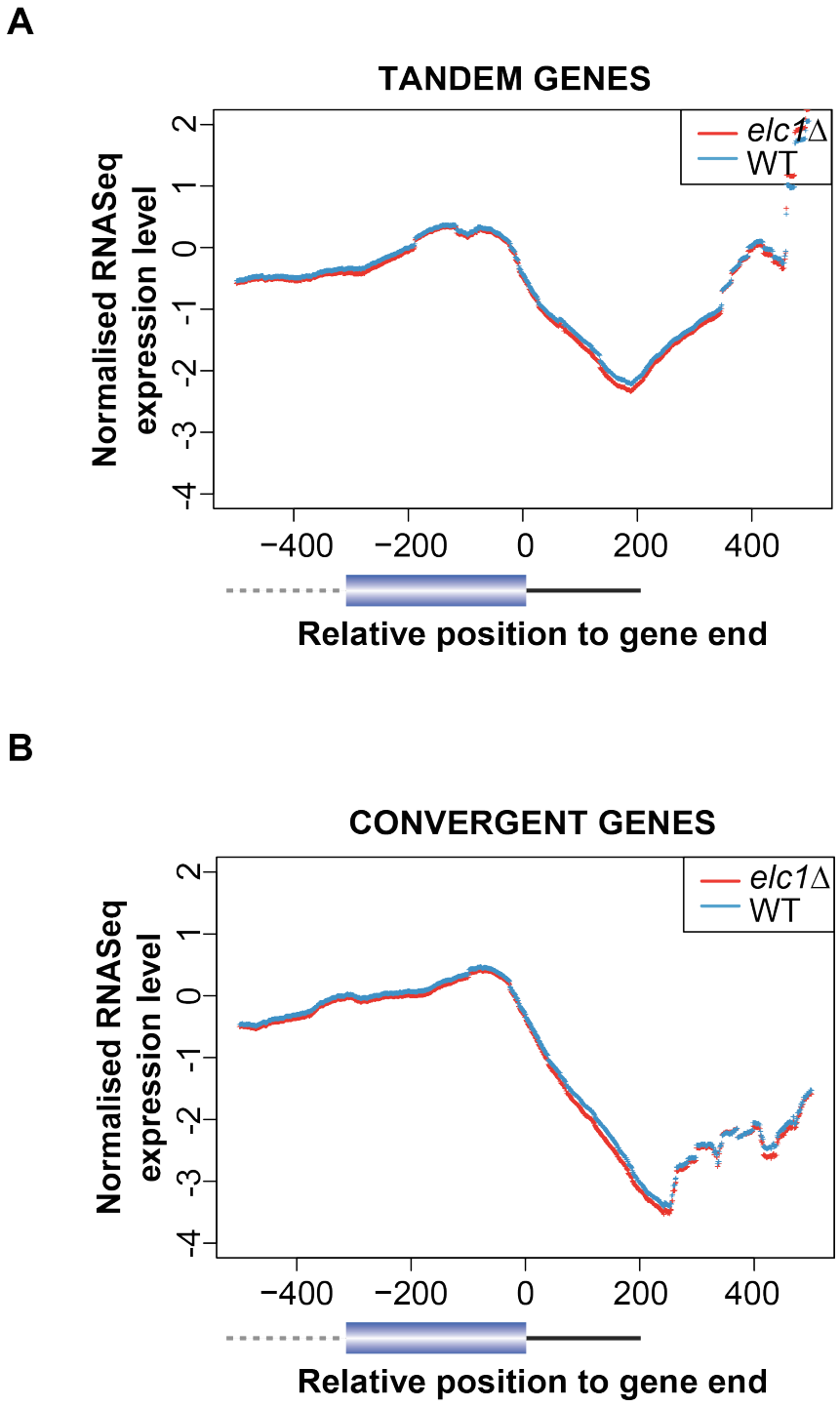

Figure 4.12 RNA-Seq data from WT and *elc1*Δ cells.

A. RNA-Seq data from WT and *elc1*Δ cells represented relative to the end of tandem genes (distance indicated in bp). Data shown is an average of three biological replicates. **B.** As in **A** but for convergent genes.

4.6 Conclusions

The data presented in this chapter address the physiological relevance and nature of RNAPII head-to-head collision *in vivo*. A highly controllable transcription template was created, allowing head-to-head collisions to be induced at a specific, observable region on DNA. Data obtained show that the presence of a convergently transcribing RNAPII molecule represents an insurmountable obstacle to transcription *in vivo*. This is in support of the data obtained *in vitro* (*see Chapter 3*), which showed that RNAPII molecules stop following head-to-head collision and are unable to bypass one another. In stark contrast to the *in vitro* scaffold system (where accessory factors are absent and RNAPII remains firmly bound to the DNA template after collision), RNAPII is removed from the convergent gene construct following head-to-head collision. This was discovered to result, at least partly, from the activity of the RNAPII ubiquitylation machinery. This pathway is responsible for the poly-ubiquitylation and degradation of Rpb1 following irreversible arrest by RNAPII, as polymerase removal was severely compromised in two key mutants of this pathway, *elc1*Δ and *def1*Δ. Interestingly, however, abolishing poly-ubiquitylation had little effect on overall transcription levels from the collision construct *in vivo*, and ubiquitylation of di-ECs had no direct effect on convergent transcription *in vitro*. Possibly, *GAL-GAL* transcription results in levels of collision that overload the ubiquitylation system. In any case, these observations lend further support to a degradation-related role of poly-ubiquitylation following transcriptional collision, rather than a direct mechanistic effect on transcription.

The decrease in RNAPII degradation upon collision in *elc1*Δ cells led to an accumulation of collided polymerases on the convergent gene construct. As a result of

this, RNAPII collisions could be studied in their physiological setting at natural genes. This analysis was important, as the convergent gene construct was highly expressed and arguably not fully representative of the true physiological scenario. However, the genome-wide analysis strongly supports the idea that RNAPII collisions do indeed occur *in vivo*, as RNAPII was shown to accumulate at regions where collision was likely, in an *elc1*Δ mutant but not in WT cells. These data are a further indication of the physiological relevance of RNAPII head-to-head collisions *in vivo*. It seems that collisions are fairly common (but stochastically distributed) and dealt with by poly-ubiquitylation and degradation. Arguably, collision could have a role in regulation of transcription and this will be discussed in the following chapter.

Chapter 5. Discussion

There is a plethora of genome-wide data showing that transcription is surprisingly pervasive in cells (*see section 1.5.2*). Furthermore, antisense transcription often initiates from 3' ends of genes and NFRs, and is commonly concurrent with transcription of protein-coding genes (Xu et al., 2009, Neil et al., 2009, Vallon-Christersson et al., 2007). In conjunction with this, genes are arranged into pairs that can affect one another's regulation (*see section 1.5.1*). Thus, the transcriptional landscape is far more complex than previously envisaged, with the poorly understood element of gene-, and specifically RNAPII-traffic, playing an important role in gene expression. Several studies highlight this importance, showing the act of transcribing ncRNA to have a role in the repression, or activation of genes (Martens et al., 2004, Hongay et al., 2006, Uhler et al., 2007). Additionally, studies addressing the act of tandem (sense) transcription by multiple RNAPII molecules, have provided an insight into this process and especially its ability to aid RNAPII processivity during elongation (Saeki and Svejstrup, 2009, Jin et al., 2010). However, there is a distinct lack of knowledge regarding the fundamental nature of convergent transcription. The main aim of this thesis was to address this shortfall and answer the absolutely fundamental question: can RNAPII molecules transcribe past one another? The data contained, herein, show that both in a basic biochemical assay and *in vivo*, RNAPII molecules cannot transcribe past one another, in fact they represent insurmountable obstacles to further elongation. They stop upon touching, remaining stably bound to the DNA template. Additionally, data obtained in this thesis provide evidence that these collisions do occur *in vivo*, where the collided polymerases are substrates for poly-ubiquitylation

and degradation. In all likelihood, collision is rare at any individual site, but very frequent across the genome.

It might be expected that colliding polymerases, resulting from convergent transcription, could dissociate one another from the DNA template. However, ECs are remarkably stable, with the finding that even head-to-tail RNAPII collision does not lead to dissociation providing further evidence for this (Saeki and Svejstrup, 2009). Although the frequency of RNAPII collision is unclear, when it does happen this results in RNAPII arrest and a block to elongation, which needs to be dealt with. Indeed, polymerases arrested as a result of other circumstances, such as DNA damage, need to be removed because transcriptional arrest blocks gene transcription and can affect other DNA metabolic events (Sigurdsson et al., 2010, Svejstrup, 2003, Svejstrup, 2007a). The findings of this thesis lend support to the theory that all irreversibly stalled RNAPII molecules are likely to be dealt with by poly-ubiquitylation and degradation as a last resort. This process is reliant upon Rpb1 mono-ubiquitylation by Rsp5, followed by Elongin-Cullin-mediated poly-ubiquitylation (occurring in a Def1-dependent manner) and proteasome-mediated degradation (*see section 1.3.3.2*).

The finding that polymerases stop upon head-to-head collision and are unable to invade one another's space was surprising. Furthermore, this is in stark contrast to RNAPII-nucleosome interactions, where DNA is 'stripped' off the surface of the nucleosome by the transcribing polymerase, or head-to-tail (tandem) RNAPII collisions, where temporary, but extensive, structural intermingling was observed (Studitsky et al., 1995, Kireeva et al., 2005, Saeki and Svejstrup, 2009). This shows that the nature of RNAPII head-to-head collision could differ fundamentally from head-to-tail collision. There were two models suggested for the structural invasion and 'recoil' of RNAPII

upon head-to-tail collision. The first, is that invasion is mediated by changes in RNAPII protein conformation, implicating mobile domains in the polymerase structure, such as the jaw lobe and clamp (Saeki and Svejstrup, 2009). Additionally, the ability of polymerase to swivel/rotate on DNA (Herbert et al., 2008) could facilitate this invasion. The protein domains that interact upon head-to-tail collision are of very different shapes and topology. It is, therefore, possible that the protein density of the leading polymerase acts as a 'wedge', separating the mobile domains of the trailing polymerase, allowing invasion. In contrast, during head-to-head collision, the protein domains that interact are very similar and may be unable to separate the mobile domains, thus explaining the lack of invasion. The second model proposed, was that changes in DNA topology account for the perceived protein intermingling. It is possible that the DNA helix between the two transcription bubbles could unwind, or stretch, allowing the trailing polymerase to continue transcribing and for its footprint to decrease, whilst not actually invading the space of the leading polymerase. Furthermore, the change in DNA topology could provide the energy to drive backtracking of the trailing polymerase.

During transcription, the DNA topology changes as RNAPII will generate positive supercoils in front of, and negative supercoils behind, itself (Liu and Wang, 1987, Osborne and Guarente, 1988). In the case of convergent transcription positive supercoiling would thus be expected to accumulate between the polymerases, whilst negative and positive supercoiling would cancel out during tandem transcription. Indeed, accumulation of supercoiling has been shown to drastically reduce the efficiency of convergent transcription in the absence of DNA topoisomerases (Garcia-Rubio and Aguilera, 2011). Supercoiling effects are not an issue in the *in vitro* system used here, as the DNA template has free ends, which can rotate to counteract torsional strain.

Therefore, the inhibition of transcription in the di-EC reaction cannot be accounted for by a build up of torsional strain. Moreover, as mentioned above, DNA topoisomerases efficiently deal with supercoiling arising during transcription *in vivo* (Liu and Wang, 1987, Garcia-Rubio and Aguilera, 2011).

The creation of a convergent gene construct with which head-to-head RNAPII collision could be induced at a specific, examinable locus, provided evidence for polymerase collisions *in vivo*. This system was, however, used simply as a tool to gain more understanding of the process, as arguably, the high transcription levels at these convergent units are probably not physiological. However, one of the key findings of this system was that the Elongin-Cullin complex and Def1 have a role in removing RNAPII from DNA following head-to-head collision. In a more general confirmation of this, on a genome-wide scale RNAPII accumulated between natural convergent gene pairs in an *elc1*Δ mutant.

Although evidence has been obtained for a role of *elc1*Δ and the ubiquitin-proteasome system in removing collided RNAPII from DNA, it is likely that alternative mechanisms do exist. In support of this, *ELC1* and *DEF1* are not required for cell viability. It is an intriguing possibility that an RNAPII bypass factor(s) does exist *in vivo*. An enzyme with helicase/translocase activity would be the most likely candidate. Alternatively, there could be factors that act to dissociate collided ECs following collision. The Rho termination factor (in bacteria) as well as Rat1 and Sen1 (in yeast) have all been implicated in EC destabilisation (*see section 1.4*). In addition to these characterised factors, a yeast factor similar to TTF2 (transcription termination factor 2), a protein in humans which causes EC destabilisation in an ATP-dependent manner (Xie and Price, 1996), might be involved in the process. Such activities might act in a

redundant fashion, with poly-ubiquitylation being a last-resort pathway for coping with collided polymerases. In the experiments presented here, it is likely that any resolution of collision made possible by a bypass/dissociation factor was masked by the high expression levels of the *GAL* collision construct. It was unfortunate that no conclusive evidence for the existence of a bypass factor could be found in the crude extract. The difficulty of producing and purifying a large amount of di-ECs, however, means that it is not a readily amenable substrate for use in biochemical fractionation experiments. For this to be possible, it would be advantageous to set up a new, simpler assay. One possibility is to create a genetic construct to screen for bypass factors *in vivo*. A method for doing this would be to take two genes associated with amino acid biosynthesis, for example *HIS3* and *TRP1*, and clone them in a convergent orientation into a plasmid, with no terminators separating them. Presumably, such a gene organisation would negatively effect expression of both marker genes, which would in theory provide a convenient way of identifying bypass factors, by transforming this construct into the yeast genetic deletion library and screening for survival on selective plates.

5.1 A role for transcriptional collision in gene regulation

The complex interaction between sense and antisense transcription is likely to play an important role in fine-tuning gene expression. In support of this, recent data indicate that genes with antisense transcripts running across them show increased expression level variability (Xu et al., 2011). This means that when activated they show rapid and high expression, whereas when in a repressed state these genes exhibit remarkably low levels of expression. It is very likely that in the sense-antisense gene pairs characterised by high expression levels, the act of transcribing the antisense RNA

helps maintain an open chromatin structure at the promoter and a favourable environment for rapid initiation. Indeed, this has been shown to be the case at the *PHO5* gene (Uhler et al., 2007). Conversely, in genes characterised by extremely tight repression, such as *IME4*, which commits a cell to meiosis (Hongay et al., 2006), this could be mediated by two main effects: promoter occlusion and RNAPII head-to-head collision during elongation.

The findings of this thesis suggest a model in which the tight repression of non-activated genes with an antisense transcript running across them could occur. When in a repressed state, the antisense transcript originating from the 3'-end of the ORF transcribes across the coding region and promoter of the gene (Figure 5.1, (1)). However, if unscheduled initiation does occur at the gene (sense) promoter and RNAPII manages to enter into the processive elongation phase, it would likely collide with antisense-directed RNAPII (Figure 5.1, (2), upper). The subsequent collision would lead to Elongin-Cullin-mediated poly-ubiquitylation and degradation of the polymerases, freeing up DNA so that antisense transcription could resume (Figure 5.1, (2), lower). Destruction of the sense polymerase and its nascent RNA would be highly advantageous if its translation were detrimental for the cell. The method of repression by antisense-transcribing RNAPII is likely to be two-fold; apart from the repression through collision, it likely also displaces transcription factors from the gene promoter, helping eliminate subsequent rounds of unscheduled transcription (Figure 5.1, (3)).

Some studies have found that when the sense gene is activated, the levels of antisense RNA produced decrease to an undetectable level (Hongay et al., 2006, Uhler et al., 2007). This suggests that under activating conditions (Figure 5.2, (1)) the gene promoter is highly dominant and can overcome the repression imposed by RNAPII

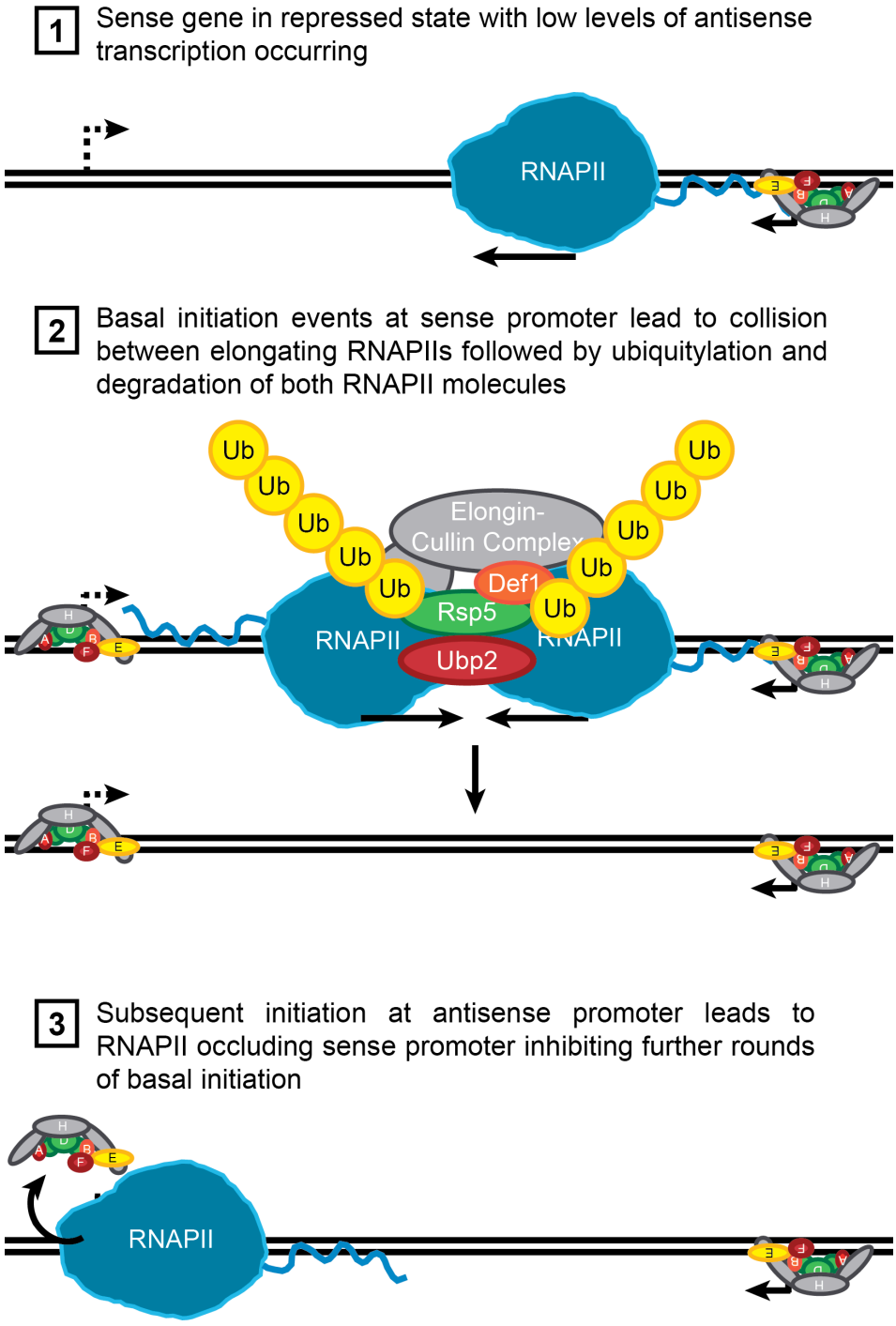

Figure 5.1 The interactions between sense-antisense transcription during sense gene repression.

Low levels of antisense transcription occur during repression of the sense gene (**1**), however, non-activated unscheduled initiation sometimes occurs, leading to head-to-head collision between the sense- and antisense-directed polymerases, which are both poly-ubiquitylated and degraded (**2**). Subsequent re-initiation occurs at the antisense promoter leading to transcription across the sense promoter and displacement of transcription factors (**3**).

collision and promoter occlusion. There is still likely to be collision on such genes, however, at least during early activation where the pioneering RNAPII molecule would meet the antisense-transcribing polymerase (Figure 5.2, (2)). This event, as in the case of unscheduled transcription from a repressed promoter, would likely also lead to poly-ubiquitylation and degradation of the polymerases. However, activated transcription is characterised by high levels of re-initiation. Therefore, it is likely that multiple RNAPII molecules 'stack-up' behind the pioneering polymerase and, following its degradation, keep transcribing downstream, eventually occluding the antisense promoter leading to its repression (Figure 5.2, (3)). Although the model provided is speculative, it is consistent with the data presented in this thesis and would also help explain the genome-wide expression data for sense-antisense pairs (Xu et al., 2011).

In conclusion, the emerging view from numerous genome-wide studies is that RNAPII constantly and stochastically initiates across the genome, often producing antisense ncRNA (Xu et al., 2009, Neil et al., 2009, Vallon-Christersson et al., 2007, He et al., 2008, Katayama et al., 2005). The transcriptional landscape is thus extremely complex, and RNAPII traffic must have an effect on gene expression and regulation. Through several recent studies and the data presented in this thesis, the complexity of gene traffic is beginning to be understood (Saeki and Svejstrup, 2009, Jin et al., 2010). In all likelihood, such collisions are rare at any individual site, but very frequent across the genome. The occurrence of head-to-head collision and the ensuing RNAPII arrest, presents the cell with a situation that must be resolved. The data presented in this thesis provide a basis for an understanding of RNAPII head-to-head collision and its role in gene expression.

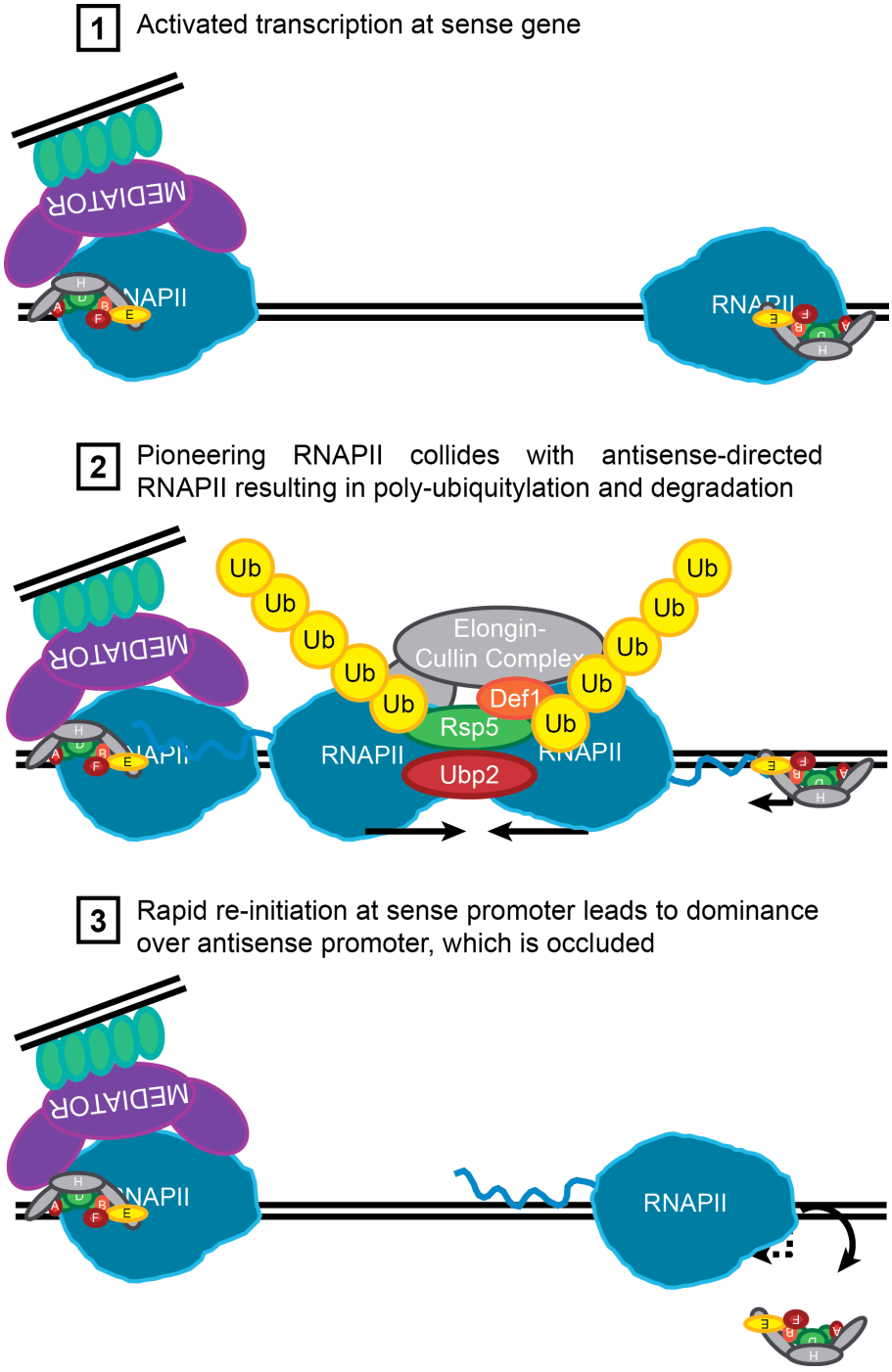

Figure 5.2 The interactions between sense-antisense transcription during sense gene activation.

Activated transcription at the sense gene results in sense transcription and collision with antisense-directed RNAPII (**1**). The antisense-directed polymerase and the pioneering sense RNAPII are poly-ubiquitylated and degraded, whilst frequent re-initiation is occurring upstream at the sense promoter (**2**). A substantially greater number of sense-transcribing RNAPII molecules are present on DNA and eventually occlude the antisense promoter (**3**).

Reference List

ANDRAU, J. C., VAN DE PASCH, L., LIJNZAAD, P., BIJMA, T., KOERKAMP, M. G., VAN DE PEPPEL, J., WERNER, M. & HOLSTEGE, F. C. 2006. Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Mol Cell,* 22**,** 179-92.

ANINDYA, R., AYGUN, O. & SVEJSTRUP, J. Q. 2007. Damage-Induced Ubiquitylation of Human RNA Polymerase II by the Ubiquitin Ligase Nedd4, but Not Cockayne Syndrome Proteins or BRCA1. *Mol Cell,* 28**,** 386-97.

ARNDT, K. M. & CHAMBERLIN, M. J. 1990. RNA chain elongation by Escherichia coli RNA polymerase. Factors affecting the stability of elongating ternary complexes. *J Mol Biol,* 213**,** 79-108.

ASO, T. & CONRAD, M. N. 1997. Molecular cloning of DNAs encoding the regulatory subunits of elongin from Saccharomyces cerevisiae and Drosophila melanogaster. *Biochem Biophys Res Commun,* 241**,** 334-40.

ASO, T., HAQUE, D., BARSTEAD, R. J., CONAWAY, R. C. & CONAWAY, J. W. 1996. The inducible elongin A elongation activation domain: structure, function and interaction with the elongin BC complex. *Embo J,* 15**,** 5557-66.

ASO, T., LANE, W. S., CONAWAY, J. W. & CONAWAY, R. C. 1995. Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. *Science,* 269**,** 1439-43.

AWREY, D. E., SHIMASAKI, N., KOTH, C., WEILBAECHER, R., OLMSTED, V., KAZANIS, S., SHAN, X., ARELLANO, J., ARROWSMITH, C. H., KANE, C. M. & EDWARDS, A. M. 1998. Yeast transcript elongation factor (TFIIS),

structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J Biol Chem,* 273**,** 22595-605.

BAR-NAHUM, G., EPSHTEIN, V., RUCKENSTEIN, A. E., RAFIKOV, R., MUSTAEV, A. & NUDLER, E. 2005. A ratchet mechanism of transcription elongation and its control. *Cell,* 120**,** 183-93.

BARDWELL, V. J., WICKENS, M., BIENROTH, S., KELLER, W., SPROAT, B. S. & LAMOND, A. I. 1991. Site-directed ribose methylation identifies 2'-OH groups in polyadenylation substrates critical for AAUAAA recognition and poly(A) addition. *Cell,* 65**,** 125-33.

BARTKOWIAK, B., LIU, P., PHATNANI, H. P., FUDA, N. J., COOPER, J. J., PRICE, D. H., ADELMAN, K., LIS, J. T. & GREENLEAF, A. L. 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev,* 24**,** 2303-16.

BASEHOAR, A. D., ZANTON, S. J. & PUGH, B. F. 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell,* 116**,** 699-709.

BATTA, K., ZHANG, Z., YEN, K., GOFFMAN, D. B. & PUGH, B. F. 2011. Genome-wide function of H2B ubiquitylation in promoter and genic regions. *Genes Dev,* 25**,** 2254-65.

BEAUDENON, S. L., HUACANI, M. R., WANG, G., MCDONNELL, D. P. & HUIBREGTSE, J. M. 1999. Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in Saccharomyces cerevisiae. *Mol Cell Biol,* 19**,** 6972-9.

BEESE, L. S. & STEITZ, T. A. 1991. Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J,* 10**,** 25-33.

BELL, S. D., KOSA, P. L., SIGLER, P. B. & JACKSON, S. P. 1999. Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A,* 96**,** 13662-7.

BELOTSERKOVSKAYA, R., OH, S., BONDARENKO, V. A., ORPHANIDES, G., STUDITSKY, V. M. & REINBERG, D. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science,* 301**,** 1090-3.

BENGAL, E., FLORES, O., KRAUSKOPF, A., REINBERG, D. & ALONI, Y. 1991. Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Mol Cell Biol,* 11**,** 1195-206.

BERRETTA, J. & MORILLON, A. 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep,* 10**,** 973-82.

BERTONE, P., STOLC, V., ROYCE, T. E., ROZOWSKY, J. S., URBAN, A. E., ZHU, X., RINN, J. L., TONGPRASIT, W., SAMANTA, M., WEISSMAN, S., GERSTEIN, M. & SNYDER, M. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science,* 306**,** 2242-6.

BHOITE, L. T., YU, Y. & STILLMAN, D. J. 2001. The Swi5 activator recruits the Mediator complex to the HO promoter without RNA polymerase II. *Genes Dev,* 15**,** 2457-69.

BIENROTH, S., KELLER, W. & WAHLE, E. 1993. Assembly of a processive messenger RNA polyadenylation complex. *EMBO J,* 12**,** 585-94.

BOOTH, V., KOTH, C. M., EDWARDS, A. M. & ARROWSMITH, C. H. 2000. Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J Biol Chem,* 275**,** 31266-8.

BORTVIN, A. & WINSTON, F. 1996. Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science,* 272**,** 1473-6.

BOURGEOIS, C. F., KIM, Y. K., CHURCHER, M. J., WEST, M. J. & KARN, J. 2002. Spt5 cooperates with human immunodeficiency virus type 1 Tat by preventing premature RNA release at terminator sequences. *Mol Cell Biol,* 22**,** 1079-93.

BRADSHER, J. N., JACKSON, K. W., CONAWAY, R. C. & CONAWAY, J. W. 1993a. RNA polymerase II transcription factor SIII. I. Identification, purification, and properties. *Journal of Biological Chemistry,* 268**,** 25587-93.

BRADSHER, J. N., TAN, S., MCLAURY, H. J., CONAWAY, J. W. & CONAWAY, R. C. 1993b. RNA polymerase II transcription factor SIII. II. Functional properties and role in RNA chain elongation. *J Biol Chem,* 268**,** 25594-603.

BRUECKNER, F., HENNECKE, U., CARELL, T. & CRAMER, P. 2007. CPD damage recognition by transcribing RNA polymerase II. *Science,* 315**,** 859-62.

BRYANT, G. O. & PTASHNE, M. 2003. Independent recruitment in vivo by Gal4 of two complexes required for transcription. *Mol Cell,* 11**,** 1301-9.

BURATOWSKI, S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell,* 36**,** 541-6.

BURKE, T. W. & KADONAGA, J. T. 1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes Dev,* 11**,** 3020-31.

BUSHNELL, D. A., BAMDAD, C. & KORNBERG, R. D. 1996. A minimal set of RNA polymerase II transcription protein interactions. *J Biol Chem,* 271**,** 20170-4.

BUSHNELL, D. A., WESTOVER, K. D., DAVIS, R. E. & KORNBERG, R. D. 2004. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science,* 303**,** 983-8.

CABART, P., UJVARI, A., PAL, M. & LUSE, D. S. 2011. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *Proc Natl Acad Sci U S A,* 108**,** 15786-91.

CAREY, M., LI, B. & WORKMAN, J. L. 2006. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Mol Cell,* 24**,** 481-7.

CARNINCI, P. 2010. RNA dust: where are the genes? *DNA Res,* 17**,** 51-9.

CARNINCI, P., KASUKAWA, T., KATAYAMA, S., GOUGH, J., FRITH, M. C., MAEDA, N., OYAMA, R., RAVASI, T., LENHARD, B., WELLS, C., KODZIUS, R., SHIMOKAWA, K., BAJIC, V. B., BRENNER, S. E., BATALOV, S., FORREST, A. R., ZAVOLAN, M., DAVIS, M. J., WILMING, L. G., AIDINIS, V., ALLEN, J. E., AMBESI-IMPIOMBATO, A., APWEILER, R., ATURALIYA, R. N., BAILEY, T. L., BANSAL, M., BAXTER, L., BEISEL, K. W., BERSANO, T., BONO, H., CHALK, A. M., CHIU, K. P., CHOUDHARY, V., CHRISTOFFELS, A., CLUTTERBUCK, D. R., CROWE, M. L., DALLA, E., DALRYMPLE, B. P., DE BONO, B., DELLA GATTA, G., DI BERNARDO, D., DOWN, T., ENGSTROM, P., FAGIOLINI, M.,

FAULKNER, G., FLETCHER, C. F., FUKUSHIMA, T., FURUNO, M., FUTAKI, S., GARIBOLDI, M., GEORGII-HEMMING, P., GINGERAS, T. R., GOJOBORI, T., GREEN, R. E., GUSTINCICH, S., HARBERS, M., HAYASHI, Y., HENSCH, T. K., HIROKAWA, N., HILL, D., HUMINIECKI, L., IACONO, M., IKEO, K., IWAMA, A., ISHIKAWA, T., JAKT, M., KANAPIN, A., KATOH, M., KAWASAWA, Y., KELSO, J., KITAMURA, H., KITANO, H., KOLLIAS, G., KRISHNAN, S. P., KRUGER, A., KUMMERFELD, S. K., KUROCHKIN, I. V., LAREAU, L. F., LAZAREVIC, D., LIPOVICH, L., LIU, J., LIUNI, S., MCWILLIAM, S., MADAN BABU, M., MADERA, M., MARCHIONNI, L., MATSUDA, H., MATSUZAWA, S., MIKI, H., MIGNONE, F., MIYAKE, S., MORRIS, K., MOTTAGUI-TABAR, S., MULDER, N., NAKANO, N., NAKAUCHI, H., NG, P., NILSSON, R., NISHIGUCHI, S., NISHIKAWA, S., et al. 2005. The transcriptional landscape of the mammalian genome. *Science,* 309**,** 1559-63.

CARNINCI, P., SANDELIN, A., LENHARD, B., KATAYAMA, S., SHIMOKAWA, K., PONJAVIC, J., SEMPLE, C. A., TAYLOR, M. S., ENGSTROM, P. G., FRITH, M. C., FORREST, A. R., ALKEMA, W. B., TAN, S. L., PLESSY, C., KODZIUS, R., RAVASI, T., KASUKAWA, T., FUKUDA, S., KANAMORI-KATAYAMA, M., KITAZUME, Y., KAWAJI, H., KAI, C., NAKAMURA, M., KONNO, H., NAKANO, K., MOTTAGUI-TABAR, S., ARNER, P., CHESI, A., GUSTINCICH, S., PERSICHETTI, F., SUZUKI, H., GRIMMOND, S. M., WELLS, C. A., ORLANDO, V., WAHLESTEDT, C., LIU, E. T., HARBERS, M., KAWAI, J., BAJIC, V. B., HUME, D. A. & HAYASHIZAKI, Y. 2006.

Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet,* 38**,** 626-35.

CARPOUSIS, A. J. & GRALLA, J. D. 1985. Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation. Footprinting, methylation, and rifampicin-sensitivity changes accompanying transcription initiation. *J Mol Biol,* 183**,** 165-77.

CARROLL, K. L., GHIRLANDO, R., AMES, J. M. & CORDEN, J. L. 2007. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA,* 13**,** 361-73.

CARROLL, K. L., PRADHAN, D. A., GRANEK, J. A., CLARKE, N. D. & CORDEN, J. L. 2004. Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol Cell Biol,* 24**,** 6241-52.

CARROZZA, M. J., LI, B., FLORENS, L., SUGANUMA, T., SWANSON, S. K., LEE, K. K., SHIA, W. J., ANDERSON, S., YATES, J., WASHBURN, M. P. & WORKMAN, J. L. 2005. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell,* 123**,** 581-92.

CHALKLEY, G. E. & VERRIJZER, C. P. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J,* 18**,** 4835-45.

CHAMBERLIN, M. J. 1992. New models for the mechanism of transcription elongation and its regulation. *Harvey Lect,* 88**,** 1-21.

CHANG, W. H. & KORNBERG, R. D. 2000. Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. *Cell,* 102**,** 609-13.

CHARLET-BERGUERAND, N., FEUERHAHN, S., KONG, S. E., ZISERMAN, H., CONAWAY, J. W., CONAWAY, R. & EGLY, J. M. 2006. RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *Embo J,* 25**,** 5481-91.

CHEN, H., CONTRERAS, X., YAMAGUCHI, Y., HANDA, H., PETERLIN, B. M. & GUO, S. 2009. Repression of RNA polymerase II elongation in vivo is critically dependent on the C-terminus of Spt5. *PLoS One,* 4**,** e6918.

CHEN, H. T. & HAHN, S. 2003. Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol Cell,* 12**,** 437-47.

CHENG, B. & PRICE, D. H. 2007. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J Biol Chem,* 282**,** 21901-12.

CHERRY, J. M., ADLER, C., BALL, C., CHERVITZ, S. A., DWIGHT, S. S., HESTER, E. T., JIA, Y., JUVIK, G., ROE, T., SCHROEDER, M., WENG, S. & BOTSTEIN, D. 1998. SGD: Saccharomyces Genome Database. *Nucleic Acids Res,* 26**,** 73-9.

CHINCHILLA, K., RODRIGUEZ-MOLINA, J. B., URSIC, D., FINKEL, J. S., ANSARI, A. Z. & CULBERTSON, M. R. 2012. Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae. *Eukaryot Cell,* 11**,** 417-29.

CHO, E. J., TAKAGI, T., MOORE, C. R. & BURATOWSKI, S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA

polymerase II carboxy-terminal domain [see comments]. *Genes Dev,* 11**,** 3319-26.

CHO, H., ORPHANIDES, G., SUN, X., YANG, X. J., OGRYZKO, V., LEES, E., NAKATANI, Y. & REINBERG, D. 1998. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol,* 18**,** 5355-63.

CLAPIER, C. R. & CAIRNS, B. R. 2009. The biology of chromatin remodeling complexes. *Annu Rev Biochem,* 78**,** 273-304.

CLOSE, P., HAWKES, N., CORNEZ, I., CREPPE, C., LAMBERT, C. A., ROGISTER, B., SIEBENLIST, U., MERVILLE, M. P., SLAUGENHAUPT, S. A., BOURS, V., SVEJSTRUP, J. Q. & CHARIOT, A. 2006. Transcription impairment and cell migration defects in elongator-depleted cells: implication for familial dysautonomia. *Mol Cell,* 22**,** 521-31.

COHEN, B. A., MITRA, R. D., HUGHES, J. D. & CHURCH, G. M. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet,* 26**,** 183-6.

COIN, F., BERGMANN, E., TREMEAU-BRAVARD, A. & EGLY, J. M. 1999. Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. *Embo J,* 18**,** 1357-66.

COJOCARU, M., JERONIMO, C., FORGET, D., BOUCHARD, A., BERGERON, D., COTE, P., POIRIER, G. G., GREENBLATT, J. & COULOMBE, B. 2008. Genomic location of the human RNA polymerase II general machinery: evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription. *Biochem J,* 409**,** 139-47.

CONAWAY, R. C. & CONAWAY, J. W. 2002. The von Hippel-Lindau tumor suppressor complex and regulation of hypoxia-inducible transcription. *Adv Cancer Res,* 85**,** 1-12.

CONNELLY, S. & MANLEY, J. L. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev,* 2**,** 440-52.

CORDEN, J. L. 1990. Tails of RNA polymerase II. *Trends Biochem Sci,* 15**,** 383-7.

COSMA, M. P. 2002. Ordered recruitment: gene-specific mechanism of transcription activation. *Mol Cell,* 10**,** 227-36.

COSMA, M. P., PANIZZA, S. & NASMYTH, K. 2001. Cdk1 triggers association of RNA polymerase to cell cycle promoters only after recruitment of the mediator by SBF. *Mol Cell,* 7**,** 1213-20.

COX, J. M., HAYWARD, M. M., SANCHEZ, J. F., GEGNAS, L. D., VAN DER ZEE, S., DENNIS, J. H., SIGLER, P. B. & SCHEPARTZ, A. 1997. Bidirectional binding of the TATA box binding protein to the TATA box. *Proc Natl Acad Sci U S A,* 94**,** 13475-80.

CRAMER, P., BUSHNELL, D. A., FU, J., GNATT, A. L., MAIER-DAVIS, B., THOMPSON, N. E., BURGESS, R. R., EDWARDS, A. M., DAVID, P. R. & KORNBERG, R. D. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science,* 288**,** 640-9.

CRAMER, P., BUSHNELL, D. A. & KORNBERG, R. D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science,* 292**,** 1863-76.

DALAL, R. V., LARSON, M. H., NEUMAN, K. C., GELLES, J., LANDICK, R. & BLOCK, S. M. 2006. Pulling on the nascent RNA during transcription does not alter kinetics of elongation or ubiquitous pausing. *Mol Cell,* 23**,** 231-9.

DANIEL, J. A., PRAY-GRANT, M. G. & GRANT, P. A. 2005. Effector proteins for methylated histones: an expanding family. *Cell Cycle,* 4**,** 919-26.

DARST, S. A., EDWARDS, A. M., KUBALEK, E. W. & KORNBERG, R. D. 1991. Three-dimensional structure of yeast RNA polymerase II at 16 A resolution. *Cell,* 66**,** 121-8.

DARZACQ, X., SHAV-TAL, Y., DE TURRIS, V., BRODY, Y., SHENOY, S. M., PHAIR, R. D. & SINGER, R. H. 2007. In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol,* 14**,** 796-806.

DAVID, L., HUBER, W., GRANOVSKAIA, M., TOEDLING, J., PALM, C. J., BOFKIN, L., JONES, T., DAVIS, R. W. & STEINMETZ, L. M. 2006. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A,* 103**,** 5320-5.

DAVIS, J. A., TAKAGI, Y., KORNBERG, R. D. & ASTURIAS, F. A. 2002. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell,* 10**,** 409-15.

DE LA MATA, M., ALONSO, C. R., KADENER, S., FEDEDA, J. P., BLAUSTEIN, M., PELISCH, F., CRAMER, P., BENTLEY, D. & KORNBLIHTT, A. R. 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell,* 12**,** 525-32.

DICHTL, B. & KELLER, W. 2001. Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J,* 20**,** 3197-209.

DIMITRI, A., BURNS, J. A., BROYDE, S. & SCICCHITANO, D. A. 2008. Transcription elongation past O6-methylguanine by human RNA polymerase II and bacteriophage T7 RNA polymerase. *Nucleic Acids Res,* 36**,** 6459-71.

DONAHUE, B. A., YIN, S., TAYLOR, J. S., REINES, D. & HANAWALT, P. C. 1994. Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. *Proc Natl Acad Sci U S A,* 91**,** 8502-6.

DURNIAK, K. J., BAILEY, S. & STEITZ, T. A. 2008. The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science,* 322**,** 553-7.

DUTROW, N., NIX, D. A., HOLT, D., MILASH, B., DALLEY, B., WESTBROEK, E., PARNELL, T. J. & CAIRNS, B. R. 2008. Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. *Nat Genet,* 40**,** 977-86.

EDWARDS, A. M., KANE, C. M., YOUNG, R. A. & KORNBERG, R. D. 1991. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem,* 266**,** 71-5.

EISENMANN, D. M., ARNDT, K. M., RICUPERO, S. L., ROONEY, J. W. & WINSTON, F. 1992. SPT3 interacts with TFIID to allow normal transcription in Saccharomyces cerevisiae. *Genes Dev,* 6**,** 1319-31.

EISSENBERG, J. C., MA, J., GERBER, M. A., CHRISTENSEN, A., KENNISON, J. A. & SHILATIFARD, A. 2002. dELL is an essential RNA polymerase II elongation factor with a general role in development. *Proc Natl Acad Sci U S A,* 99**,** 9894-9.

EPSHTEIN, V., DUTTA, D., WADE, J. & NUDLER, E. 2010. An allosteric mechanism of Rho-dependent transcription termination. *Nature,* 463**,** 245-9.

EPSHTEIN, V. & NUDLER, E. 2003. Cooperation between RNA polymerase molecules in transcription elongation. *Science,* 300**,** 801-5.

ESNAULT, C., GHAVI-HELM, Y., BRUN, S., SOUTOURINA, J., VAN BERKUM, N., BOSCHIERO, C., HOLSTEGE, F. & WERNER, M. 2008. Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Mol Cell,* 31**,** 337-46.

EZEOKONKWO, C., GHAZY, M. A., ZHELKOVSKY, A., YEH, P. C. & MOORE, C. 2012. Novel interactions at the essential N-terminus of poly(A) polymerase that could regulate poly(A) addition in Saccharomyces cerevisiae. *FEBS Lett,* 586**,** 1173-8.

FEMINO, A. M., FAY, F. S., FOGARTY, K. & SINGER, R. H. 1998. Visualization of single RNA transcripts in situ. *Science,* 280**,** 585-90.

FINKELSTEIN, A., KOSTRUB, C. F., LI, J., CHAVEZ, D. P., WANG, B. Q., FANG, S. M., GREENBLATT, J. & BURTON, Z. F. 1992. A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. *Nature,* 355**,** 464-7.

FISH, R. N. & KANE, C. M. 2002. Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta,* 1577**,** 287-307.

FLEMING, A. B., KAO, C. F., HILLYER, C., PIKAART, M. & OSLEY, M. A. 2008. H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Mol Cell,* 31**,** 57-66.

FU, J., GNATT, A. L., BUSHNELL, D. A., JENSEN, G. J., THOMPSON, N. E., BURGESS, R. R., DAVID, P. R. & KORNBERG, R. D. 1999. Yeast RNA polymerase II at 5 A resolution [see comments]. *Cell,* 98**,** 799-810.

GARCIA-RUBIO, M. L. & AGUILERA, A. 2011. Topological constraints impair RNA polymerase II transcription and causes instability of plasmid-borne convergent genes. *Nucleic Acids Res*.

GARRETT, K. P., ASO, T., BRADSHER, J. N., FOUNDLING, S. I., LANE, W. S., CONAWAY, R. C. & CONAWAY, J. W. 1995. Positive regulation of general transcription factor SIII by a tailed ubiquitin homolog. *Proc Natl Acad Sci U S A,* 92**,** 7172-6.

GEIGER, J. H., HAHN, S., LEE, S. & SIGLER, P. B. 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science,* 272**,** 830-6.

GERBER, M., MA, J., DEAN, K., EISSENBERG, J. C. & SHILATIFARD, A. 2001. Drosophila ELL is associated with actively elongating RNA polymerase II on transcriptionally active sites in vivo. *EMBO J,* 20**,** 6104-14.

GERBER, M. & SHILATIFARD, A. 2003. Transcriptional elongation control and histone methylation. *J Biol Chem*.

GILBERT, C., KRISTJUHAN, A., WINKLER, G. S. & SVEJSTRUP, J. Q. 2004. Elongator interactions with nascent mRNA revealed by RNA immunoprecipitation. *Mol Cell,* 14**,** 457-64.

GNATT, A. 2002. Elongation by RNA polymerase II: structure-function relationship. *Biochim Biophys Acta,* 1577**,** 175-90.

GNATT, A. L., CRAMER, P., FU, J., BUSHNELL, D. A. & KORNBERG, R. D. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science,* 292**,** 1876-82.

GOVIND, C. K., YOON, S., QIU, H., GOVIND, S. & HINNEBUSCH, A. G. 2005. Simultaneous recruitment of coactivators by Gcn4p stimulates multiple steps of transcription in vivo. *Mol Cell Biol,* 25**,** 5626-38.

GOVIND, C. K., ZHANG, F., QIU, H., HOFMEYER, K. & HINNEBUSCH, A. G. 2007. Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. *Mol Cell,* 25**,** 31-42.

GREGER, I. H., ARANDA, A. & PROUDFOOT, N. 2000. Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A,* 97**,** 8415-20.

GRUNBERG, S., WARFIELD, L. & HAHN, S. 2012. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol*.

GU, W. & REINES, D. 1995. Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II. *J Biol Chem,* 270**,** 11238-44.

GUERMAH, M., PALHAN, V. B., TACKETT, A. J., CHAIT, B. T. & ROEDER, R. G. 2006. Synergistic functions of SII and p300 in productive activator-dependent transcription of chromatin templates. *Cell,* 125**,** 275-86.

GUO, Z. & SHERMAN, F. 1995. 3'-end-forming signals of yeast mRNA. *Mol Cell Biol,* 15**,** 5983-90.

HAHN, S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol,* 11**,** 394-403.

HAHN, S. & ROBERTS, S. 2000. The zinc ribbon domains of the general transcription factors TFIIB and Brf: conserved functional surfaces but different roles in transcription initiation. *Genes Dev,* 14**,** 719-30.

HARREMAN, M., TASCHNER, M., SIGURDSSON, S., ANINDYA, R., REID, J., SOMESH, B., KONG, S. E., BANKS, C. A., CONAWAY, R. C., CONAWAY, J. W. & SVEJSTRUP, J. Q. 2009. Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc Natl Acad Sci U S A,* 106**,** 20705-10.

HASSAN, A. H., NEELY, K. E. & WORKMAN, J. L. 2001. Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. *Cell,* 104**,** 817-27.

HAWLEY, D. K., WIEST, D. K., HOLTZ, M. S. & WANG, D. 1993. Transcriptional pausing, arrest, and readthrough at the adenovirus major late attenuation site. *Cell Mol Biol Res,* 39**,** 339-48.

HE, N., LIU, M., HSU, J., XUE, Y., CHOU, S., BURLINGAME, A., KROGAN, N. J., ALBER, T. & ZHOU, Q. 2010. HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Mol Cell,* 38**,** 428-38.

HE, Y., VOGELSTEIN, B., VELCULESCU, V. E., PAPADOPOULOS, N. & KINZLER, K. W. 2008. The antisense transcriptomes of human cells. *Science,* 322**,** 1855-7.

HENRY, K. W., WYCE, A., LO, W. S., DUGGAN, L. J., EMRE, N. C., KAO, C. F., PILLUS, L., SHILATIFARD, A., OSLEY, M. A. & BERGER, S. L. 2003.

Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev,* 17**,** 2648-63.

HERBERT, K. M., GREENLEAF, W. J. & BLOCK, S. M. 2008. Single-molecule studies of RNA polymerase: motoring along. *Annu Rev Biochem,* 77**,** 149-76.

HIRAI, H., SEKIMIZU, K., HORIKOSHI, M., NAKANISHI, Y. & NATORI, S. 1988. Stimulation of transcription from accurate initiation sites by purified S-II. *FEBS Lett,* 238**,** 119-22.

HIROSE, Y. & MANLEY, J. L. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev,* 14**,** 1415-29.

HIRST, M., KOBOR, M. S., KURIAKOSE, N., GREENBLATT, J. & SADOWSKI, I. 1999. GAL4 is regulated by the RNA polymerase II holoenzyme-associated cyclin- dependent protein kinase SRB10/CDK8. *Mol Cell,* 3**,** 673-8.

HO, C. K. & SHUMAN, S. 1999. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell,* 3**,** 405-11.

HOBOR, F., PERGOLI, R., KUBICEK, K., HROSSOVA, D., BACIKOVA, V., ZIMMERMANN, M., PASULKA, J., HOFR, C., VANACOVA, S. & STEFL, R. 2011. Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem,* 286**,** 3645-57.

HONGAY, C. F., GRISAFI, P. L., GALITSKI, T. & FINK, G. R. 2006. Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell,* 127**,** 735-45.

HUISINGA, K. L. & PUGH, B. F. 2004. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Mol Cell,* 13**,** 573-85.

IVAN, M. & KAELIN, W. G., JR. 2001. The von Hippel-Lindau tumor suppressor protein. *Curr Opin Genet Dev,* 11**,** 27-34.

IZBAN, M. G. & LUSE, D. S. 1991. Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev,* 5**,** 683-96.

IZBAN, M. G. & LUSE, D. S. 1992a. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem,* 267**,** 13647-55.

IZBAN, M. G. & LUSE, D. S. 1992b. The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. *Genes Dev,* 6**,** 1342-56.

IZBAN, M. G. & LUSE, D. S. 1993a. The increment of SII-facilitated transcript cleavage varies dramatically between elongation competent and incompetent RNA polymerase II ternary complexes. *J Biol Chem,* 268**,** 12874-85.

IZBAN, M. G. & LUSE, D. S. 1993b. SII-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments. *J Biol Chem,* 268**,** 12864-73.

JACQUIER, A. 2009. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet,* 10**,** 833-44.

JEON, C. & AGARWAL, K. 1996. Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. *Proc Natl Acad Sci U S A,* 93**,** 13677-82.

JEON, C., YOON, H. & AGARWAL, K. 1994. The transcription factor TFIIS zinc ribbon dipeptide Asp-Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. *Proc Natl Acad Sci U S A,* 91**,** 9106-10.

JIN, J., BAI, L., JOHNSON, D. S., FULBRIGHT, R. M., KIREEVA, M. L., KASHLEV, M. & WANG, M. D. 2010. Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. *Nat Struct Mol Biol,* 17**,** 745-52.

JOHNSON, T. L. & CHAMBERLIN, M. J. 1994. Complexes of yeast RNA polymerase II and RNA are substrates for TFIIS- induced RNA cleavage. *Cell,* 77**,** 217-24.

JOSHI, A. A. & STRUHL, K. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell,* 20**,** 971-8.

JOYCE, C. M. & STEITZ, T. A. 1994. Function and structure relationships in DNA polymerases. *Annu Rev Biochem,* 63**,** 777-822.

KAO, C. F., HILLYER, C., TSUKUDA, T., HENRY, K., BERGER, S. & OSLEY, M. A. 2004. Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. *Genes Dev,* 18**,** 184-95.

KAPANIDIS, A. N., MARGEAT, E., HO, S. O., KORTKHONJIA, E., WEISS, S. & EBRIGHT, R. H. 2006. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science,* 314**,** 1144-7.

KAPLAN, C. D., LAPRADE, L. & WINSTON, F. 2003. Transcription elongation factors repress transcription initiation from cryptic sites. *Science,* 301**,** 1096-9.

KATAYAMA, S., TOMARU, Y., KASUKAWA, T., WAKI, K., NAKANISHI, M., NAKAMURA, M., NISHIDA, H., YAP, C. C., SUZUKI, M., KAWAI, J., SUZUKI, H., CARNINCI, P., HAYASHIZAKI, Y., WELLS, C., FRITH, M., RAVASI, T., PANG, K. C., HALLINAN, J., MATTICK, J., HUME, D. A., LIPOVICH, L., BATALOV, S., ENGSTROM, P. G., MIZUNO, Y., FAGHIHI, M. A., SANDELIN, A., CHALK, A. M., MOTTAGUI-TABAR, S., LIANG, Z.,

LENHARD, B. & WAHLESTEDT, C. 2005. Antisense transcription in the mammalian transcriptome. *Science,* 309**,** 1564-6.

KAWAUCHI, J., MISCHO, H., BRAGLIA, P., RONDON, A. & PROUDFOOT, N. J. 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev,* 22**,** 1082-92.

KEE, Y., LYON, N. & HUIBREGTSE, J. M. 2005. The Rsp5 ubiquitin ligase is coupled to and antagonized by the Ubp2 deubiquitinating enzyme. *Embo J,* 24**,** 2414-24.

KELLER, W. & MINVIELLE-SEBASTIA, L. 1997. A comparison of mammalian and yeast pre-mRNA 3'-end processing. *Curr Opin Cell Biol,* 9**,** 329-36.

KEOGH, M. C., KURDISTANI, S. K., MORRIS, S. A., AHN, S. H., PODOLNY, V., COLLINS, S. R., SCHULDINER, M., CHIN, K., PUNNA, T., THOMPSON, N. J., BOONE, C., EMILI, A., WEISSMAN, J. S., HUGHES, T. R., STRAHL, B. D., GRUNSTEIN, M., GREENBLATT, J. F., BURATOWSKI, S. & KROGAN, N. J. 2005. Cotranscriptional set2 methylation of histone h3 lysine 36 recruits a repressive rpd3 complex. *Cell,* 123**,** 593-605.

KERPPOLA, T. K. & KANE, C. M. 1990. Analysis of the signals for transcription termination by purified RNA polymerase II. *Biochemistry,* 29**,** 269-78.

KETTENBERGER, H., ARMACHE, K. J. & CRAMER, P. 2003. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell,* 114**,** 347-57.

KETTENBERGER, H., ARMACHE, K. J. & CRAMER, P. 2004. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell,* 16**,** 955-65.

KHAZAK, V., SADHALE, P. P., WOYCHIK, N. A., BRENT, R. & GOLEMIS, E. A. 1995. Human RNA polymerase II subunit hsRPB7 functions in yeast and influences stress survival and cell morphology. *Mol Biol Cell,* 6**,** 759-75.

KIM, B., NESVIZHSKII, A. I., RANI, P. G., HAHN, S., AEBERSOLD, R. & RANISH, J. A. 2007. The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci U S A,* 104**,** 16068-73.

KIM, J. H., LANE, W. S. & REINBERG, D. 2002. Human Elongator facilitates RNA polymerase II transcription through chromatin. *Proc Natl Acad Sci U S A,* 99**,** 1241-6.

KIM, J. L., NIKOLOV, D. B. & BURLEY, S. K. 1993a. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature,* 365**,** 520-7.

KIM, M., KROGAN, N. J., VASILJEVA, L., RANDO, O. J., NEDEA, E., GREENBLATT, J. F. & BURATOWSKI, S. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature,* 432**,** 517-22.

KIM, T. K., EBRIGHT, R. H. & REINBERG, D. 2000. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science,* 288**,** 1418-22.

KIM, Y., GEIGER, J. H., HAHN, S. & SIGLER, P. B. 1993b. Crystal structure of a yeast TBP/TATA-box complex. *Nature,* 365**,** 512-20.

KIM, Y. J., BJORKLUND, S., LI, Y., SAYRE, M. H. & KORNBERG, R. D. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell,* 77**,** 599-608.

KIREEVA, M. L., HANCOCK, B., CREMONA, G. H., WALTER, W., STUDITSKY, V. M. & KASHLEV, M. 2005. Nature of the nucleosomal barrier to RNA polymerase II. *Mol Cell,* 18**,** 97-108.

KIREEVA, M. L., KOMISSAROVA, N., WAUGH, D. S. & KASHLEV, M. 2000. The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem,* 275**,** 6530-6.

KIREEVA, M. L., WALTER, W., TCHERNAJENKO, V., BONDARENKO, V., KASHLEV, M. & STUDITSKY, V. M. 2002. Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol Cell,* 9**,** 541-52.

KOH, S. S., ANSARI, A. Z., PTASHNE, M. & YOUNG, R. A. 1998. An activator target in the RNA polymerase II holoenzyme. *Mol Cell,* 1**,** 895-904.

KOKUBO, T., SWANSON, M. J., NISHIKAWA, J. I., HINNEBUSCH, A. G. & NAKATANI, Y. 1998. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol Cell Biol,* 18**,** 1003-12.

KORNBERG, R. D. & LORCH, Y. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell,* 98**,** 285-94.

KOSTREWA, D., ZELLER, M. E., ARMACHE, K. J., SEIZL, M., LEIKE, K., THOMM, M. & CRAMER, P. 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature,* 462**,** 323-30.

KOTH, C. M., BOTUYAN, M. V., MORELAND, R. J., JANSMA, D. B., CONAWAY, J. W., CONAWAY, R. C., CHAZIN, W. J., FRIESEN, J. D., ARROWSMITH, C. H. & EDWARDS, A. M. 2000. Elongin from Saccharomyces cerevisiae. *J Biol Chem,* 275**,** 11174-80.

KRAAKMAN, L. S., MAGER, W. H., MAURER, K. T., NIEUWINT, R. T. & PLANTA, R. J. 1989. The divergently transcribed genes encoding yeast

ribosomal proteins L46 and S24 are activated by shared RPG-boxes. *Nucleic Acids Res,* 17**,** 9693-706.

KRISTJUHAN, A. & SVEJSTRUP, J. Q. 2004. Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo. *Embo J,* 23**,** 4243-52.

KRISTJUHAN, A., WALKER, J., SUKA, N., GRUNSTEIN, M., ROBERTS, D., CAIRNS, B. R. & SVEJSTRUP, J. Q. 2002. Transcriptional inhibition of genes with severe histone h3 hypoacetylation in the coding region. *Mol Cell,* 10**,** 925-33.

KROGAN, N. J., KIM, M., TONG, A., GOLSHANI, A., CAGNEY, G., CANADIEN, V., RICHARDS, D. P., BEATTIE, B. K., EMILI, A., BOONE, C., SHILATIFARD, A., BURATOWSKI, S. & GREENBLATT, J. 2003. Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. *Mol Cell Biol,* 23**,** 4207-4218.

KRUMMEL, B. & CHAMBERLIN, M. J. 1989. RNA chain initiation by Escherichia coli RNA polymerase. Structural transitions of the enzyme in early ternary complexes. *Biochemistry,* 28**,** 7829-42.

KRUMMEL, B. & CHAMBERLIN, M. J. 1992. Structural analysis of ternary complexes of Escherichia coli RNA polymerase. Individual complexes halted along different transcription units have distinct and unexpected biochemical properties. *J Mol Biol,* 225**,** 221-37.

KUHN, U., GUNDEL, M., KNOTH, A., KERWITZ, Y., RUDEL, S. & 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. *J Biol Chem,* 284**,** 22803-14.

KVINT, K., UHLER, J. P., TASCHNER, M. J., SIGURDSSON, S., ERDJUMENT-BROMAGE, H., TEMPST, P. & SVEJSTRUP, J. Q. 2008. Reversal of RNA polymerase II ubiquitylation by the ubiquitin protease Ubp3. *Mol Cell,* 30**,** 498-506.

LAGRANGE, T., KAPANIDIS, A. N., TANG, H., REINBERG, D. & EBRIGHT, R. H. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev,* 12**,** 34-44.

LAPRADE, L., ROSE, D. & WINSTON, F. 2007. Characterization of new Spt3 and TATA-binding protein mutants of Saccharomyces cerevisiae: Spt3 TBP allele-specific interactions and bypass of Spt8. *Genetics,* 177**,** 2007-17.

LEE, J. M. & GREENLEAF, A. L. 1997. Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I. *J Biol Chem,* 272**,** 10990-3.

LEE, J. S., GARRETT, A. S., YEN, K., TAKAHASHI, Y. H., HU, D., JACKSON, J., SEIDEL, C., PUGH, B. F. & SHILATIFARD, A. 2012. Codependency of H2B monoubiquitination and nucleosome reassembly on Chd1. *Genes Dev,* 26**,** 914-9.

LEE, J. S., SHUKLA, A., SCHNEIDER, J., SWANSON, S. K., WASHBURN, M. P., FLORENS, L., BHAUMIK, S. R. & SHILATIFARD, A. 2007. Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell,* 131**,** 1084-96.

LEURENT, C., SANDERS, S., RUHLMANN, C., MALLOUH, V., WEIL, P. A., KIRSCHNER, D. B., TORA, L. & SCHULTZ, P. 2002. Mapping histone fold TAFs within yeast TFIID. *EMBO J,* 21**,** 3424-33.

LEUTHER, K. K., BUSHNELL, D. A. & KORNBERG, R. D. 1996. Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation. *Cell,* 85**,** 773-9.

LEVIN, J. R., KRUMMEL, B. & CHAMBERLIN, M. J. 1987. Isolation and properties of transcribing ternary complexes of Escherichia coli RNA polymerase positioned at a single template base. *J Mol Biol,* 196**,** 85-100.

LEVINSON, B., KENWRICK, S., LAKICH, D., HAMMONDS, G., JR. & GITSCHIER, J. 1990. A transcribed gene in an intron of the human factor VIII gene. *Genomics,* 7**,** 1-11.

LI, B., GOGOL, M., CAREY, M., LEE, D., SEIDEL, C. & WORKMAN, J. L. 2007. Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science,* 316**,** 1050-4.

LI, B., HOWE, L., ANDERSON, S., YATES, J. R., 3RD & WORKMAN, J. L. 2003. The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem,* 278**,** 8897-903.

LI, B., JACKSON, J., SIMON, M. D., FLEHARTY, B., GOGOL, M., SEIDEL, C., WORKMAN, J. L. & SHILATIFARD, A. 2009. Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *J Biol Chem,* 284**,** 7970-6.

LI, L., WANG, X., STOLC, V., LI, X., ZHANG, D., SU, N., TONGPRASIT, W., LI, S., CHENG, Z., WANG, J. & DENG, X. W. 2006. Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet,* 38**,** 124-9.

LIN, C., GARRETT, A. S., DE KUMAR, B., SMITH, E. R., GOGOL, M., SEIDEL, C., KRUMLAUF, R. & SHILATIFARD, A. 2011. Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes Dev,* 25**,** 1486-98.

LIN, C., SMITH, E. R., TAKAHASHI, H., LAI, K. C., MARTIN-BROWN, S., FLORENS, L., WASHBURN, M. P., CONAWAY, J. W., CONAWAY, R. C. & SHILATIFARD, A. 2010. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol Cell,* 37**,** 429-37.

LITTLEFIELD, O., KORKHIN, Y. & SIGLER, P. B. 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A,* 96**,** 13668-73.

LIU, D., ISHIMA, R., TONG, K. I., BAGBY, S., KOKUBO, T., MUHANDIRAM, D. R., KAY, L. E., NAKATANI, Y. & IKURA, M. 1998. Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell,* 94**,** 573-83.

LIU, L. F. & WANG, J. C. 1987. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A,* 84**,** 7024-7.

LIU, X., BUSHNELL, D. A., WANG, D., CALERO, G. & KORNBERG, R. D. 2010. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science,* 327**,** 206-9.

LIU, Y., KUNG, C., FISHBURN, J., ANSARI, A. Z., SHOKAT, K. M. & HAHN, S. 2004. Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol Cell Biol,* 24**,** 1721-35.

LIU, Y., WARFIELD, L., ZHANG, C., LUO, J., ALLEN, J., LANG, W. H., RANISH, J., SHOKAT, K. M. & HAHN, S. 2009. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol Cell Biol,* 29**,** 4852-63.

LOGAN, J., FALCK-PEDERSEN, E., DARNELL, J. E., JR. & SHENK, T. 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A,* 84**,** 8306-10.

LORCH, Y., LAPOINTE, J. W. & KORNBERG, R. D. 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell,* 49**,** 203-10.

LU, H., ZAWEL, L., FISHER, L., EGLY, J. M. & REINBERG, D. 1992. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature,* 358**,** 641-5.

LUE, N. F., FLANAGAN, P. M., SUGIMOTO, K. & KORNBERG, R. D. 1989. Initiation by yeast RNA polymerase II at the adenoviral major late promoter in vitro. *Science,* 246**,** 661-4.

LUGER, K. 2006. Dynamic nucleosomes. *Chromosome Res,* 14**,** 5-16.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution [see comments]. *Nature,* 389**,** 251-60.

MA, N. & MCALLISTER, W. T. 2009. In a head-on collision, two RNA polymerases approaching one another on the same DNA may pass by one another. *Journal of Molecular Biology,* 391**,** 808-12.

MAIURI, P., KNEZEVICH, A., DE MARCO, A., MAZZA, D., KULA, A., MCNALLY, J. G. & MARCELLO, A. 2011. Fast transcription rates of RNA polymerase II in human cells. *EMBO Rep,* 12**,** 1280-5.

MALDONADO, E., SHIEKHATTAR, R., SHELDON, M., CHO, H., DRAPKIN, R., RICKERT, P., LEES, E., ANDERSON, C. W., LINN, S. & REINBERG, D. 1996. A human RNA polymerase II complex associated with SRB and DNA-repair proteins [published erratum appears in Nature 1996 Nov 28;384(6607):384]. *Nature,* 381**,** 86-9.

MANDAL, S. S., CHU, C., WADA, T., HANDA, H., SHATKIN, A. J. & REINBERG, D. 2004. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc Natl Acad Sci U S A,* 101**,** 7572-7.

MANDEL, C. R., KANEKO, S., ZHANG, H., GEBAUER, D., VETHANTHAM, V., MANLEY, J. L. & TONG, L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature,* 444**,** 953-6.

MARSHALL, N. F., PENG, J., XIE, Z. & PRICE, D. H. 1996. Control of RNA polymerase II elongation potential by a novel carboxyl- terminal domain kinase. *J Biol Chem,* 271**,** 27176-83.

MARSHALL, N. F. & PRICE, D. H. 1992. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol,* 12**,** 2078-90.

MARTENS, J. A., LAPRADE, L. & WINSTON, F. 2004. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature,* 429**,** 571-4.

MARTENS, J. A., WU, P. Y. & WINSTON, F. 2005. Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. *Genes Dev,* 19**,** 2695-704.

MASON, P. B. & STRUHL, K. 2005. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell,* 17**,** 831-40.

MAX, T., SOGAARD, M. & SVEJSTRUP, J. Q. 2007. Hyperphosphorylation of the C-terminal Repeat Domain of RNA Polymerase II Facilitates Dissociation of Its Complex with Mediator. *J Biol Chem,* 282**,** 14113-20.

MAYER, A., HEIDEMANN, M., LIDSCHREIBER, M., SCHREIECK, A., SUN, M., HINTERMAIR, C., KREMMER, E., EICK, D. & CRAMER, P. 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science,* 336**,** 1723-5.

MAYER, A., LIDSCHREIBER, M., SIEBERT, M., LEIKE, K., SODING, J. & CRAMER, P. 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol,* 17**,** 1272-8.

MCKUNE, K., MOORE, P. A., HULL, M. W. & WOYCHIK, N. A. 1995. Six human RNA polymerase subunits functionally substitute for their yeast counterparts. *Mol Cell Biol,* 15**,** 6895-900.

MCNALLY, J. G., MULLER, W. G., WALKER, D., WOLFORD, R. & HAGER, G. L. 2000. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science,* 287**,** 1262-5.

MEINHART, A. & CRAMER, P. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature,* 430**,** 223-6.

MILLER, T., WILLIAMS, K., JOHNSTONE, R. W. & SHILATIFARD, A. 2000. Identification, cloning, expression, and biochemical characterization of the testis-specific RNA polymerase II elongation factor ELL3. *J Biol Chem,* 275**,** 32052-6.

MOHIBULLAH, N. & HAHN, S. 2008. Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev,* 22**,** 2994-3006.

MORELAND, R. J., HANAS, J. S., CONAWAY, J. W. & CONAWAY, R. C. 1998. Mechanism of action of RNA polymerase II elongation factor Elongin. Maximal stimulation of elongation requires conversion of the early elongation complex to an Elongin-activable form. *J Biol Chem,* 273**,** 26610-7.

MORIN, P. E., AWREY, D. E., EDWARDS, A. M. & ARROWSMITH, C. H. 1996. Elongation factor TFIIS contains three structural domains: solution structure of domain II. *Proc Natl Acad Sci U S A,* 93**,** 10604-8.

MOTEKI, S. & PRICE, D. 2002. Functional coupling of capping and transcription of mRNA. *Mol Cell,* 10**,** 599-609.

MUELLER, D., GARCIA-CUELLAR, M. P., BACH, C., BUHL, S., MAETHNER, E. & SLANY, R. K. 2009. Misguided transcriptional elongation causes mixed lineage leukemia. *PLoS Biol,* 7**,** e1000249.

NAGALAKSHMI, U., WANG, Z., WAERN, K., SHOU, C., RAHA, D., GERSTEIN, M. & SNYDER, M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science,* 320**,** 1344-9.

NAKAO, J., MIYANOHARA, A., TOH-E, A. & MATSUBARA, K. 1986. Saccharomyces cerevisiae PHO5 promoter region: location and function of the upstream activation site. *Mol Cell Biol,* 6**,** 2613-23.

NATORI, S., TAKEUCHI, K., TAKAHASHI, K. & MIZUNO, D. 1973. DNA dependent RNA polymerase from Ehrlich ascites tumor cells. II. Factors stimulating the activity of RNA polymerase II. *J Biochem,* 73**,** 879-88.

NECHAEV, S. & ADELMAN, K. 2011. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta,* 1809**,** 34-45.

NEIL, H., MALABAT, C., D'AUBENTON-CARAFA, Y., XU, Z., STEINMETZ, L. M. & JACQUIER, A. 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature,* 457**,** 1038-42.

NEUMAN, K. C., ABBONDANZIERI, E. A., LANDICK, R., GELLES, J. & BLOCK, S. M. 2003. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell,* 115**,** 437-47.

NIKOLOV, D. B., CHEN, H., HALAY, E. D., USHEVA, A. A., HISATAKE, K., LEE, D. K., ROEDER, R. G. & BURLEY, S. K. 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature,* 377**,** 119-28.

OBRDLIK, A., KUKALEV, A., LOUVET, E., FARRANTS, A. K., CAPUTO, L. & PERCIPALLE, P. 2008. The histone acetyltransferase PCAF associates with actin and hnRNP U for RNA polymerase II transcription. *Mol Cell Biol,* 28**,** 6342-57.

OHKUMA, Y. & ROEDER, R. G. 1994. Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature,* 368**,** 160-3.

OHLER, U., LIAO, G. C., NIEMANN, H. & RUBIN, G. M. 2002. Computational analysis of core promoters in the Drosophila genome. *Genome Biol,* 3**,** RESEARCH0087.

OLMSTED, V. K., AWREY, D. E., KOTH, C., SHAN, X., MORIN, P. E., KAZANIS, S., EDWARDS, A. M. & ARROWSMITH, C. H. 1998. Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region. *J Biol Chem,* 273**,** 22589-94.

ORLOVA, M., NEWLANDS, J., DAS, A., GOLDFARB, A. & BORUKHOV, S. 1995. Intrinsic transcript cleavage activity of RNA polymerase. *Proc Natl Acad Sci U S A,* 92**,** 4596-600.

ORPHANIDES, G., LEROY, G., CHANG, C. H., LUSE, D. S. & REINBERG, D. 1998. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell,* 92**,** 105-16.

OSBORNE, B. I. & GUARENTE, L. 1988. Transcription by RNA polymerase II induces changes of DNA topology in yeast. *Genes Dev,* 2**,** 766-72.

OSLEY, M. A., GOULD, J., KIM, S., KANE, M. Y. & HEREFORD, L. 1986. Identification of sequences in a yeast histone promoter involved in periodic transcription. *Cell,* 45**,** 537-44.

OSSIPOW, V., TASSAN, J. P., NIGG, E. A. & SCHIBLER, U. 1995. A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell,* 83**,** 137-46.

OTERO, G., FELLOWS, J., LI, Y., DE BIZEMONT, T., DIRAC, A. M. G., GUSTAFSSON, C. M., ERDJUMENT-BROMAGE, H., TEMPST, P. & SVEJSTRUP, J. Q. 1999. Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol Cell,* 3**,** 109-18.

PAL, M. & LUSE, D. S. 2002. Strong natural pausing by RNA polymerase II within 10 bases of transcription start may result in repeated slippage and reextension of the nascent RNA. *Mol Cell Biol,* 22**,** 30-40.

PAL, M. & LUSE, D. S. 2003. The initiation-elongation transition: Lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23. *Proc Natl Acad Sci U S A,* 100**,** 5700-5.

PAL, M., PONTICELLI, A. S. & LUSE, D. S. 2005. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell,* 19**,** 101-10.

PAN, G., ASO, T. & GREENBLATT, J. 1997. Interaction of elongation factors TFIIS and elongin A with a human RNA polymerase II holoenzyme capable of promoter-specific initiation and responsive to transcriptional activators. *J Biol Chem,* 272**,** 24563-71.

PARDEE, T. S., BANGUR, C. S. & PONTICELLI, A. S. 1998. The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *J Biol Chem,* 273**,** 17859-64.

PARK, N. J., TSAO, D. C. & MARTINSON, H. G. 2004. The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of

the RNA polymerase II carboxyl-terminal repeat domain. *Mol Cell Biol,* 24**,** 4092-103.

PAVRI, R., ZHU, B., LI, G., TROJER, P., MANDAL, S., SHILATIFARD, A. & REINBERG, D. 2006. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell,* 125**,** 703-17.

PENG, J., MARSHALL, N. F. & PRICE, D. H. 1998a. Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *J Biol Chem,* 273**,** 13855-60.

PENG, J., ZHU, Y., MILTON, J. T. & PRICE, D. H. 1998b. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev,* 12**,** 755-62.

PETERS, M. F. & ROSS, C. A. 2001. Isolation of a 40-kDa Huntingtin-associated protein. *J Biol Chem,* 276**,** 3188-94.

PICKART, C. M. 2001. Mechanisms underlying ubiquitination. *Annu Rev Biochem,* 70**,** 503-33.

POKHOLOK, D. K., HARBISON, C. T., LEVINE, S., COLE, M., HANNETT, N. M., LEE, T. I., BELL, G. W., WALKER, K., ROLFE, P. A., HERBOLSHEIMER, E., ZEITLINGER, J., LEWITTER, F., GIFFORD, D. K. & YOUNG, R. A. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell,* 122**,** 517-27.

POMERANTZ, R. T. & O'DONNELL, M. 2008. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature,* 456**,** 762-6.

POMERANTZ, R. T. & O'DONNELL, M. 2010. Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science,* 327**,** 590-2.

PONJAVIC, J., LENHARD, B., KAI, C., KAWAI, J., CARNINCI, P., HAYASHIZAKI, Y. & SANDELIN, A. 2006. Transcriptional and structural

impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol,* 7**,** R78.

PRATHER, D. M., LARSCHAN, E. & WINSTON, F. 2005. Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in Saccharomyces cerevisiae. *Mol Cell Biol,* 25**,** 2650-9.

PRESCOTT, E. M. & PROUDFOOT, N. J. 2002. Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A,* 99**,** 8796-801.

PROUDFOOT, N. J. 2011. Ending the message: poly(A) signals then and now. *Genes Dev,* 25**,** 1770-82.

PROUDFOOT, N. J., FURGER, A. & DYE, M. J. 2002. Integrating mRNA processing with transcription. *Cell,* 108**,** 501-12.

QIAN, X., GOZANI, S. N., YOON, H., JEON, C. J., AGARWAL, K. & WEISS, M. A. 1993. Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry,* 32**,** 9944-59.

QIU, H., HU, C. & HINNEBUSCH, A. G. 2009. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell,* 33**,** 752-62.

QURESHI, S. A. & JACKSON, S. P. 1998. Sequence-specific DNA binding by the S. shibatae TFIIB homolog, TFB, and its effect on promoter strength. *Mol Cell,* 1**,** 389-400.

RAHL, P. B., LIN, C. Y., SEILA, A. C., FLYNN, R. A., MCCUINE, S., BURGE, C. B., SHARP, P. A. & YOUNG, R. A. 2010. c-Myc regulates transcriptional pause release. *Cell,* 141**,** 432-45.

RANI, P. G., RANISH, J. A. & HAHN, S. 2004. RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Mol Cell Biol,* 24**,** 1709-20.

RANISH, J. A., YUDKOVSKY, N. & HAHN, S. 1999. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev,* 13**,** 49-63.

RAPPAPORT, J., REINBERG, D., ZANDOMENI, R. & WEINMANN, R. 1987. Purification and functional characterization of transcription factor SII from calf thymus. Role in RNA polymerase II elongation. *J Biol Chem,* 262**,** 5227-32.

RASMUSSEN, T. P. & CULBERTSON, M. R. 1998. The putative nucleic acid helicase Sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nucleolar RNAs in Saccharomyces cerevisiae. *Mol Cell Biol,* 18**,** 6885-96.

REEVES, W. M. & HAHN, S. 2003. Activator-independent functions of the yeast mediator sin4 complex in preinitiation complex formation and transcription reinitiation. *Mol Cell Biol,* 23**,** 349-58.

REID, J. & SVEJSTRUP, J. Q. 2004. DNA damage-induced Def1-RNA polymerase II interaction and Def1 requirement for polymerase ubiquitylation in vitro. *J Biol Chem,* 279**,** 29875-29878.

REINBERG, D. & ROEDER, R. G. 1987. Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains. *J Biol Chem,* 262**,** 3331-7.

REINES, D. 1992. Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *J Biol Chem,* 267**,** 3795-800.

REINES, D. & MOTE, J., JR. 1993. Elongation factor SII-dependent transcription by RNA polymerase II through a sequence-specific DNA-binding protein. *Proc Natl Acad Sci U S A,* 90**,** 1917-21.

RENNER, D. B., YAMAGUCHI, Y., WADA, T., HANDA, H. & PRICE, D. H. 2001. A highly purified RNA polymerase II elongation control system. *J Biol Chem,* 11**,** 11.

REVYAKIN, A., LIU, C., EBRIGHT, R. H. & STRICK, T. R. 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science,* 314**,** 1139-43.

RIBAR, B., PRAKASH, L. & PRAKASH, S. 2006. Requirement of ELC1 for RNA polymerase II polyubiquitylation and degradation in response to DNA damage in Saccharomyces cerevisiae. *Mol Cell Biol,* 26**,** 3999-4005.

RIBAR, B., PRAKASH, L. & PRAKASH, S. 2007. ELA1 and CUL3 are required along with ELC1 for RNA polymerase II polyubiquitylation and degradation in DNA-damaged yeast cells. *Mol Cell Biol,* 27**,** 3211-6.

RONDON, A. G., GARCIA-RUBIO, M., GONZALEZ-BARRERA, S. & AGUILERA, A. 2003. Molecular evidence for a positive role of Spt4 in transcription elongation. *EMBO J,* 22**,** 612-20.

RONDON, A. G., MISCHO, H. E., KAWAUCHI, J. & PROUDFOOT, N. J. 2009. Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. *Mol Cell,* 36**,** 88-98.

RUDD, M. D., IZBAN, M. G. & LUSE, D. S. 1994. The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc Natl Acad Sci U S A,* 91**,** 8057-61.

RYAN, K., CALVO, O. & MANLEY, J. L. 2004. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA,* 10**,** 565-73.

SAEKI, H. & SVEJSTRUP, J. Q. 2009. Stability, Flexibility, and Dynamic Interactions of Colliding RNA Polymerase II Elongation Complexes. *Mol Cell,* 35**,** 191-205.

SANDERS, S. L., GARBETT, K. A. & WEIL, P. A. 2002. Molecular characterization of Saccharomyces cerevisiae TFIID. *Mol Cell Biol,* 22**,** 6000-13.

SANDERS, S. L. & WEIL, P. A. 2000. Identification of two novel TAF subunits of the yeast Saccharomyces cerevisiae TFIID complex. *J Biol Chem,* 275**,** 13895-900.

SAWADOGO, M., LESCURE, B., SENTENAC, A. & FROMAGEOT, P. 1981. Native deoxyribonucleic acid transcription by yeast RNA polymerase--P37 complex. *Biochemistry,* 20**,** 3542-7.

SCHAFT, D., ROGUEV, A., KOTOVIC, K. M., SHEVCHENKO, A., SAROV, M., NEUGEBAUER, K. M. & STEWART, A. F. 2003. The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. *Nucleic Acids Res,* 31**,** 2475-82.

SCHNITZLER, G. R. 2008. Control of nucleosome positions by DNA sequence and remodeling machines. *Cell Biochem Biophys,* 51**,** 67-80.

SCHULTZ, P., FRIBOURG, S., POTERSZMAN, A., MALLOUH, V., MORAS, D. & EGLY, J. M. 2000. Molecular structure of human TFIIH. *Cell,* 102**,** 599-607.

SEKIMIZU, K., KOBAYASHI, N., MIZUNO, D. & NATORI, S. 1976. Purification of a factor from Ehrlich ascites tumor cells specifically stimulating RNA polymerase II. *Biochemistry,* 15**,** 5064-70.

SEKIMIZU, K., KUBO, Y., SEGAWA, K. & NATORI, S. 1981. Difference in phosphorylation of two factors stimulating RNA polymerase II of Ehrlich ascites tumor cells. *Biochemistry,* 20**,** 2286-92.

SELTH, L. & SVEJSTRUP, J. Q. 2007. Vps75, A New Yeast Member of the NAP Histone Chaperone Family. *J Biol Chem,* 282**,** 12358-62.

SELTH, L. A., LORCH, Y., OCAMPO-HAFALLA, M. T., MITTER, R., SHALES, M., KROGAN, N. J., KORNBERG, R. D. & SVEJSTRUP, J. Q. 2009. An Rtt109-independent role for Vps75 in transcription-associated nucleosome dynamics. *Mol Cell Biol,* 29**,** 4220–4234.

SELTH, L. A., SIGURDSSON, S. & SVEJSTRUP, J. Q. 2010. Transcript Elongation by RNA Polymerase II. *Annu Rev Biochem,* 79**,** 271-93.

SHILATIFARD, A. 1998. Identification and purification of the Holo-ELL complex. Evidence for the presence of ELL-associated proteins that suppress the transcriptional inhibitory activity of ELL. *J Biol Chem,* 273**,** 11212-7.

SHILATIFARD, A. 2004. Transcriptional elongation control by RNA polymerase II: a new frontier. *Biochim Biophys Acta,* 1677**,** 79-86.

SHILATIFARD, A., DUAN, D. R., HAQUE, D., FLORENCE, C., SCHUBACH, W. H., CONAWAY, J. W. & CONAWAY, R. C. 1997. ELL2, a new member of an ELL family of RNA polymerase II elongation factors. *Proc Natl Acad Sci U S A,* 94**,** 3639-43.

SHILATIFARD, A., LANE, W. S., JACKSON, K. W., CONAWAY, R. C. & CONAWAY, J. W. 1996. An RNA polymerase II elongation factor encoded by the human ELL gene. *Science,* 271**,** 1873-6.

SHIM, E. Y., WALKER, A. K., SHI, Y. & BLACKWELL, T. K. 2002. CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. *Genes Dev,* 16**,** 2135-46.

SHIMASAKI, N. B. & KANE, C. M. 2000. Structural basis for the species-specific activity of TFIIS. *J Biol Chem,* 275**,** 36541-9.

SHPAKOVSKI, G. V., ACKER, J., WINTZERITH, M., LACROIX, J. F., THURIAUX, P. & VIGNERON, M. 1995. Four subunits that are shared by the three classes of RNA polymerase are functionally interchangeable between Homo sapiens and Saccharomyces cerevisiae. *Mol Cell Biol,* 15**,** 4702-10.

SIGURDSSON, S., DIRAC-SVEJSTRUP, A. B. & SVEJSTRUP, J. Q. 2010. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell,* 38**,** 202-10.

SLUDER, A. E., GREENLEAF, A. L. & PRICE, D. H. 1989. Properties of a Drosophila RNA polymerase II elongation factor. *J Biol Chem,* 264**,** 8963-9.

SMALE, S. T. & KADONAGA, J. T. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem,* 72**,** 449-79.

SOBHIAN, B., LAGUETTE, N., YATIM, A., NAKAMURA, M., LEVY, Y., KIERNAN, R. & BENKIRANE, M. 2010. HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Mol Cell,* 38**,** 439-51.

SOMESH, B. P., REID, J., LIU, W. F., SOGAARD, T. M., ERDJUMENT-BROMAGE, H., TEMPST, P. & SVEJSTRUP, J. Q. 2005. Multiple Mechanisms Confining RNA Polymerase II Ubiquitylation to Polymerases Undergoing Transcriptional Arrest. *Cell,* 121**,** 913-23.

SOMESH, B. P., SIGURDSSON, S., SAEKI, H., ERDJUMENT-BROMAGE, H., TEMPST, P. & SVEJSTRUP, J. Q. 2007. Communication between distant sites in RNA polymerase II through ubiquitylation factors and the polymerase CTD. *Cell,* 129**,** 57-68.

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

STEINMETZ, E. J., CONRAD, N. K., BROW, D. A. & CORDEN, J. L. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature,* 413**,** 327-31.

STEITZ, T. A. 1998. A mechanism for all polymerases. *Nature,* 391**,** 231-2.

STOLC, V., GAUHAR, Z., MASON, C., HALASZ, G., VAN BATENBURG, M. F., RIFKIN, S. A., HUA, S., HERREMAN, T., TONGPRASIT, W., BARBANO, P. E., BUSSEMAKER, H. J. & WHITE, K. P. 2004. A gene expression map for the euchromatic genome of Drosophila melanogaster. *Science,* 306**,** 655-60.

STRANEY, D. C. & CROTHERS, D. M. 1987. A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. *J Mol Biol,* 193**,** 267-78.

STUDITSKY, V. M., CLARK, D. J. & FELSENFELD, G. 1995. Overcoming a nucleosomal barrier to transcription. *Cell,* 83**,** 19-27.

SVEJSTRUP, J. Q. 2003. Rescue of arrested RNA polymerase II complexes. *J Cell Sci,* 116**,** 447-51.

SVEJSTRUP, J. Q. 2007a. Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem Sci,* 32**,** 165-71.

SVEJSTRUP, J. Q. 2007b. Elongator complex: how many roles does it play? *Curr Opin Cell Biol,* 19**,** 331-336.

SVEJSTRUP, J. Q., LI, Y., FELLOWS, J., GNATT, A., BJORKLUND, S. & KORNBERG, R. D. 1997. Evidence for a mediator cycle at the initiation of transcription. *Proc Natl Acad Sci U S A,* 94**,** 6075-8.

TAKAGI, Y., CONAWAY, J. W. & CONAWAY, R. C. 1995. A novel activity associated with RNA polymerase II elongation factor SIII. SIII directs promoter-independent transcription initiation by RNA polymerase II in the absence of initiation factors. *J Biol Chem,* 270**,** 24300-5.

TAKAGI, Y., CONAWAY, R. C. & CONAWAY, J. W. 1996. Characterization of elongin C functional domains required for interaction with elongin B and activation of elongin A. *J Biol Chem,* 271**,** 25562-8.

TAN, S., HUNZIKER, Y., SARGENT, D. F. & RICHMOND, T. J. 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature,* 381**,** 127-51.

TENNYSON, C. N., KLAMUT, H. J. & WORTON, R. G. 1995. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat Genet,* 9**,** 184-90.

THOMAS, B. J. & ROTHSTEIN, R. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell,* 56**,** 619-30.

UHLER, J. P., HERTEL, C. & SVEJSTRUP, J. Q. 2007. A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci U S A,* 104**,** 8011–8016.

VALENTINI, S. R., WEISS, V. H. & SILVER, P. A. 1999. Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. *RNA,* 5**,** 272-80.

VALLON-CHRISTERSSON, J., STAAF, J., KVIST, A., MEDSTRAND, P., BORG, A. & ROVIRA, C. 2007. Non-coding antisense transcription detected by conventional and single-stranded cDNA microarray. *BMC Genomics,* 8**,** 295.

VASILJEVA, L. & BURATOWSKI, S. 2006. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell,* 21**,** 239-48.

VASILJEVA, L., KIM, M., MUTSCHLER, H., BURATOWSKI, S. & MEINHART, A. 2008. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol,* 15**,** 795-804.

WADA, T., TAKAGI, T., YAMAGUCHI, Y., FERDOUS, A., IMAI, T., HIROSE, S., SUGIMOTO, S., YANO, K., HARTZOG, G. A., WINSTON, F., BURATOWSKI, S. & HANDA, H. 1998. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes and Development,* 12**,** 343-56.

WAHLE, E. 1991. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell,* 66**,** 759-68.

WALMACQ, C., CHEUNG, A. C., KIREEVA, M. L., LUBKOWSKA, L., YE, C., GOTTE, D., STRATHERN, J. N., CARELL, T., CRAMER, P. & KASHLEV, M. 2012. Mechanism of Translesion Transcription by RNA Polymerase II and Its Role in Cellular Resistance to DNA Damage. *Mol Cell,* 46**,** 18-29.

WANG, A., KURDISTANI, S. K. & GRUNSTEIN, M. 2002. Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science,* 298**,** 1412-4.

WANG, D., BUSHNELL, D. A., HUANG, X., WESTOVER, K. D., LEVITT, M. & KORNBERG, R. D. 2009. Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science,* 324**,** 1203-6.

WANG, D., BUSHNELL, D. A., WESTOVER, K. D., KAPLAN, C. D. & KORNBERG, R. D. 2006. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell,* 127**,** 941-54.

WANG, D. & HAWLEY, D. K. 1993. Identification of a 3'-->5' exonuclease activity associated with human RNA polymerase II. *Proc Natl Acad Sci U S A,* 90**,** 843-7.

WANG, M. D., SCHNITZER, M. J., YIN, H., LANDICK, R., GELLES, J. & BLOCK, S. M. 1998. Force and velocity measured for single molecules of RNA polymerase. *Science,* 282**,** 902-7.

WEAKE, V. M. & WORKMAN, J. L. 2008. Histone ubiquitination: triggering gene activity. *Mol Cell,* 29**,** 653-63.

WEIDEMAN, C. A., NETTER, R. C., BENJAMIN, L. R., MCALLISTER, J. J., SCHMIEDEKAMP, L. A., COLEMAN, R. A. & PUGH, B. F. 1997. Dynamic interplay of TFIIA, TBP and TATA DNA. *J Mol Biol,* 271**,** 61-75.

WEILBAECHER, R. G., AWREY, D. E., EDWARDS, A. M. & KANE, C. M. 2003. Intrinsic transcript cleavage in yeast RNA polymerase II elongation complexes. *J Biol Chem,* 278**,** 24189-99.

WEN, Y. & SHATKIN, A. J. 1999. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev,* 13**,** 1774-9.

WERY, M., SHEMATOROVA, E., VAN DRIESSCHE, B., VANDENHAUTE, J., THURIAUX, P. & VAN MULLEM, V. 2004. Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. *Embo J,* 23**,** 4232-42.

WEST, R. W., JR., YOCUM, R. R. & PTASHNE, M. 1984. Saccharomyces cerevisiae GAL1-GAL10 divergent promoter region: location and function of the upstream activating sequence UASG. *Mol Cell Biol,* 4**,** 2467-78.

WEST, S., GROMAK, N. & PROUDFOOT, N. J. 2004. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature,* 432**,** 522-5.

WESTOVER, K. D., BUSHNELL, D. A. & KORNBERG, R. D. 2004a. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell,* 119**,** 481-9.

WESTOVER, K. D., BUSHNELL, D. A. & KORNBERG, R. D. 2004b. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science,* 303**,** 1014-6.

WHITEHOUSE, I., RANDO, O. J., DELROW, J. & TSUKIYAMA, T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature,* 450**,** 1031-5.

WILHELM, B. T., MARGUERAT, S., WATT, S., SCHUBERT, F., WOOD, V., GOODHEAD, I., PENKETT, C. J., ROGERS, J. & BAHLER, J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature,* 453**,** 1239-43.

WILSON, C. J., CHAO, D. M., IMBALZANO, A. N., SCHNITZLER, G. R., KINGSTON, R. E. & YOUNG, R. A. 1996. RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell,* 84**,** 235-44.

WITTSCHIEBEN, B. O., OTERO, G., DE BIZEMONT, T., FELLOWS, J., ERDJUMENT-BROMAGE, H., OHBA, R., LI, Y., ALLIS, C. D., TEMPST, P. & SVEJSTRUP, J. Q. 1999. A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol Cell,* 4**,** 123-8.

WORKMAN, J. L. & KINGSTON, R. E. 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem,* 67**,** 545-79.

WOUDSTRA, E. C., GILBERT, C., FELLOWS, J., JANSEN, L., BROUWER, J., ERDJUMENT-BROMAGE, H., TEMPST, P. & SVEJSTRUP, J. Q. 2002. A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature,* 415**,** 929-33.

WOYCHIK, N. A., LANE, W. S. & YOUNG, R. A. 1991. Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J Biol Chem,* 266**,** 19053-5.

WOYCHIK, N. A. & YOUNG, R. A. 1989. RNA polymerase II subunit RPB4 is essential for high- and low- temperature yeast cell growth. *Mol Cell Biol,* 9**,** 2854-9.

WU, P. Y., RUHLMANN, C., WINSTON, F. & SCHULTZ, P. 2004. Molecular architecture of the S. cerevisiae SAGA complex. *Mol Cell,* 15**,** 199-208.

XIAO, T., HALL, H., KIZER, K. O., SHIBATA, Y., HALL, M. C., BORCHERS, C. H. & STRAHL, B. D. 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev,* 17**,** 654-63.

XIAO, T., KAO, C. F., KROGAN, N. J., SUN, Z. W., GREENBLATT, J. F., OSLEY, M. A. & STRAHL, B. D. 2005. Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol Cell Biol,* 25**,** 637-51.

XIE, Z. & PRICE, D. H. 1996. Purification of an RNA polymerase II transcript release factor from Drosophila. *J Biol Chem,* 271**,** 11043-6.

XU, Z., WEI, W., GAGNEUR, J., CLAUDER-MUNSTER, S., SMOLIK, M., HUBER, W. & STEINMETZ, L. M. 2011. Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol,* 7**,** 468.

XU, Z., WEI, W., GAGNEUR, J., PEROCCHI, F., CLAUDER-MUNSTER, S., CAMBLONG, J., GUFFANTI, E., STUTZ, F., HUBER, W. & STEINMETZ, L. M. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature,* 457**,** 1033-7.

YAMADA, T., YAMAGUCHI, Y., INUKAI, N., OKAMOTO, S., MURA, T. & HANDA, H. 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell,* 21**,** 227-37.

YAMAGUCHI, Y., TAKAGI, T., WADA, T., YANO, K., FURUYA, A., SUGIMOTO, S., HASEGAWA, J. & HANDA, H. 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell,* 97**,** 41-51.

YAN, Q., MORELAND, R. J., CONAWAY, J. W. & CONAWAY, R. C. 1999. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J Biol Chem,* 274**,** 35668-75.

YOKOYAMA, A., LIN, M., NARESH, A., KITABAYASHI, I. & CLEARY, M. L. 2010. A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell,* 17**,** 198-212.

YONAHA, M. & PROUDFOOT, N. J. 1999. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol Cell,* 3**,** 593-600.

YUDKOVSKY, N., RANISH, J. A. & HAHN, S. 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature,* 408**,** 225-9.

YUE, Z., MALDONADO, E., PILLUTLA, R., CHO, H., REINBERG, D. & SHATKIN, A. J. 1997. Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II [see comments]. *Proc Natl Acad Sci U S A,* 94**,** 12898-903.

ZAWEL, L., KUMAR, K. P. & REINBERG, D. 1995. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev,* 9**,** 1479-90.

ZHANG, C., YAN, H. & BURTON, Z. F. 2003. Combinatorial control of human RNA polymerase II (RNAP II) pausing and transcript cleavage by transcription factor IIF, hepatitis delta antigen, and stimulatory factor II. *J Biol Chem,* 278**,** 50101-11.

ZHELKOVSKY, A., TACAHASHI, Y., NASSER, T., HE, X., STERZER, U., JENSEN, T. H., DOMDEY, H. & MOORE, C. 2006. The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in Saccharomyces cerevisiae. *RNA,* 12**,** 435-45.

ZHOU, K., KUO, W. H., FILLINGHAM, J. & GREENBLATT, J. F. 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci U S A,* 106**,** 6956-61.

ZHOU, Q., LI, T. & PRICE, D. H. 2012. RNA polymerase II elongation control. *Annu Rev Biochem,* 81**,** 119-43.

ZHU, X., WIREN, M., SINHA, I., RASMUSSEN, N. N., LINDER, T., HOLMBERG, S., EKWALL, K. & GUSTAFSSON, C. M. 2006. Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. *Mol Cell,* 22**,** 169-78.