IN VIVO CONSEQUENCES OF ALTERED POL II CATALYSIS

A Dissertation

by

INDRANIL MALIK

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Craig D. Kaplan
Committee Members,	Michael Polymenis
	Hays Rye
	Matthew S. Sachs
Head of Department,	Gregory D. Reinhart

August 2017

Major Subject: Biochemistry

ABSTRACT

Gene transcription by RNA polymerase II (Pol II) is an essential process. Using *Saccharomyces cerevisiae* as a model system, our lab has previously identified and partially characterized Pol II activity mutants that can alter catalysis rate to faster or slower than wild type *in vitro*. In my dissertation research I use a set of these Pol II activity mutants to determine consequences of altered catalysis rate on polymerase functions, co-transcriptional pre-mRNA processing and gene expression *in vivo*. I show that alteration in Pol II catalytic rate, either increase or decrease, leads to a decreased Pol II occupancy and an apparent reduction in elongation rate on a commonly used reporter gene *in vivo*. Measurement of *in vivo* elongation rate on this reporter requires transcriptional shutoff followed by ChIP. I discover that some Pol II catalytic mutants can compromise the kinetics of transcription shutoff by glucose, which is generally assumed to be unaffected by transcription mutants. Further, I show that Pol II catalytic mutants affect model gene expression and the effects on gene expression are exacerbated with increased promoter strength and gene length. My results suggest that gene expression defects in the Pol II mutants may in part result from defective mRNA processing. Additionally, I show that mRNA half-lives for that model gene are increased in Pol II mutant strains and the magnitude of half-life changes correlate both with mutants' growth and the magnitude of reporter gene expression defects. Finally, I test if altered Pol II elongation sensitizes cells to nucleotide depletion and find that Pol II

mutants and several elongation factor mutants respond to GTP starvation similarly to wild type and that putative elongation defects are not likely to drive the cellular response to limiting GTP. Altogether my findings reveal wide-ranging *in vivo* effects of Pol II catalytic mutants, which will be critical for precise use of these Pol II catalytic mutants in gene regulation studies.

DEDICATION

This thesis is dedicated to my parents, my wife Piyali, and to all my teachers.

ACKNOWLEDGEMENTS

I like to thank my research advisor Dr. Kaplan for his continuous support encouragement and mentorship. This work would not have been possible without the myriad opportunities for research that he has provided me for the last five years. His mentorship has helped me to grow both as a scientist and as a human being.

I like to thank my doctoral committee members, Dr. Polymenis, Dr. Rye and Dr. Sachs, for their guidance and support throughout the course of this research.

I would like to thank current members of the Kaplan lab Chenxi, Tingting and Yunye, and former members Dr. Jin and Ping Cui for their help and support during the course of this study. I would also like to extend my gratitude to Chenxi Qiu and Thomas Snavely for their assistance in some of the experiments.

I thank Mary Bryk for providing bead beater for yeast cell breakage and UV cross-linker for blotting. I am grateful to Jennifer Herman and Lanying Zeng for providing microscope time and training. I thank Ry Young/Center for Phage Technology for use of Tecan plate reader. I thank Kaplan, Bryk and Peterson lab members for critical reading of the manuscript and helpful comments.

My time here would have been tough without my friends and colleagues Karthik, Hem, Rajan, Kristina, Quiyan and Vikas. I am thankful to my friends in College Station Danish, Tareq, Lubna, Isita, Rajat, Ananya, Shyamalendu, Debu, Manasi and Riyad, who have been very helpful during this journey. I appreciate continuous support from my school/college friends Rabin, Lalu, Jomidar, Pallab, Salehin, Shubhankar, Atanu,

Sangita and Sarit in both professional and personal aspects. I am always inspired by and greatly indebted to my former mentor Dr. Snezna Rogelj and dear friends Zia, Rituraj Nirupam, and Yindrila.

I thank baba, ma and my family members for their love and support. Finally, I thank my wife Piyali for her love and patience. This journey would not have been possible without her support.

CONTRIBUTORS AND FUNDING SOURCES

The doctoral committee chair, Dr. Craig D. Kaplan, supervised this study with valuable inputs from other committee members Dr. Hays Rye, Dr. Michael Polymenis and Dr. Matthew Sachs. Graduate student in the Kaplan lab, Chenxi Qiu, and former rotation student in the Kaplan lab, Thomas Snavely, have helped in performing some of experiments presented in Chapter II. Study presented in Appendix A, was done in collaboration with the Calero lab at the University of Pittsburgh with the funding from University of Pittsburgh and National Institutes of Health [NIHR01 GM112686] to Dr. Guillermo Calero. My work in the Kaplan lab was funded by the National Institutes of Health, National Institute for General Medical Sciences [R01GM097260 to C.D.K.]; Welch Foundation [A-1763 to C.D.K.].

NOMENCLATURE

Pol II	RNA polymerase II
RNAP	RNA polymerase
msRNAPs	multi subunit RNA polymerases
TL	Trigger loop
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
nt	Nucleotide
kb	kilo base pairs
NTP	Nucleoside triphosphate
ATP	Adenosine triphosphate
GTP	Guanosine triphosphate
TSS	Transcription start site
LOF	Loss of function
GOF	Gain of function
ChIP	Chromatin immunoprecipitation
MPA	Mycophenolic Acid
GFP	Green fluorescent protein

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Cellular RNA polymerases (Pol) are a group of enzymes that catalyze the template-directed synthesis of RNAs. Bacteria and archaea each encode a single RNA polymerase, whereas eukaryotes have multiple RNA polymerases dedicated to specialized functions. The eukaryotic multisubunit RNA polymerase (msRNAPs) family includes Pol I, Pol II and Pol III, which predominantly catalyze the synthesis of ribosomal RNAs (rRNAs), messenger RNAs (mRNAs), and transfer RNAs (tRNAs), respectively. Pol II synthesizes mRNAs and capped non-coding RNAs through the process of transcription, which is comprised of three distinct phases – initiation, elongation and termination (Cheung and Cramer 2012). Two additional classes of RNA polymerase, namely Pol IV and V, have been identified in plants and are involved in synthesizing non-coding RNAs functioning in gene silencing (Landick 2009, Haag and Pikaard 2011). Eukaryotic polymerases contain a ten-subunit core and with each polymerase containing additional subunits related to specialized regulation of each enzyme (Vannini and Cramer 2012). Our lab studies the mechanism and regulation of Pol II enzymatic activity and how it can impact gene expression using budding yeast *Saccharomyces cerevisiae* (addressed as yeast henceforth) as a model system. As msRNAPs are highly similar in their core structural scaffolds, studying the basic mechanisms of Pol II transcription can be relevant to the mechanistic basis of all msRNAPs.

In this section I will briefly walk through each phase of transcription followed by a summary of content of this introductory chapter. The first step of transcription is initiation, which requires up to seventy different proteins in human to form a four million Dalton complex (Hantsche and Cramer 2017). During initiation, Pol II coordinates with general transcription factors (GTFs), which include transcription factor II (TFII) D, TFIIB, TFIIF, TFIIE and TFIIH, to form a pre-initiation complex (PIC) at promoter DNA (Sainsbury, Bernecky et al. 2015). In the presence of nucleotide triphosphates (NTPs) double stranded DNA is unwound by a process called 'promoter melting', which requires TFIIE and TFIIH, to form the 'transcription bubble' with an open promoter complex. For many years promoter melting was attributed to ATP hydrolysis activity of TFIIH subunit XPB (homolog of yeast Ssl2), but a recent report contradicts this notion (Alekseev, Nagy et al. 2017). This study suggests that the ATPase activity of XPB is required to relieve a self-imposed block by XPB itself to initiate transcription. Following promoter opening unwound template DNA is 'scrunched', which denotes that Pol II remains bound to the promoter while downstream DNA is pulled to the active site of Pol II, to facilitate transcription start site (TSS) selection (Fazal, Meng et al. 2015). When a short chain of nascent RNA (9-10 nt) is synthesized to form the RNA-DNA hybrid, initiation factors are proposed to be displaced, enabling 'promoter escape' (Nechaev and Adelman 2011, Luse 2013). Experimentally, it has been shown that an RNA-DNA hybrid of ~ 7 nt partially induces bubble collapse to start promoter escape and an 8 nt hybrid is necessary and sufficient for the formation of a stable transcribing complex (Kireeva, Komissarova et al. 2000, Pal, Ponticelli et al.

2005). However, structural analysis of transcription initiation complexes with different lengths of RNAs have shown that hybrids from 6-9 nts are essentially indistinguishable while a hybrid > 5 nt is suggested to sterically clash with initiation factors to displace them (Bushnell, Westover et al. 2004, Liu, Bushnell et al. 2011). After promoter escape and the formation of a stable elongation complex, Pol II associates with transcription elongation factors, which help in processive and productive elongation. After successful transcript synthesis, RNA chain elongation stops and Pol II dissociates from the template through transcription termination. Termination is coupled to the 3′ end formation of the transcript and there are two prevailing non-mutually exclusive hypotheses for this process, which I discus further.

My dissertation research relates to the study of transcription elongation in yeast. Transcription elongation is a mechano-chemical process that involves: i. catalysis of phosphodiester bond formation through addition of NMPs, derived from NTP substrates, to a nascent RNA chain; ii. the release of PPi after addition of each substrate, and, iii. the subsequent translocation of Pol II by one nucleotide so the process may be repeated (Kaplan 2013, Svetlov and Nudler 2013). Additionally, Pol II is capable of pausing and reverse translocating, which is known as 'back-tracking'(discussed further below). During the elongation phase, eukaryotic mRNAs undergo extensive processing events such as mRNA 5′-capping, splicing, termination and 3′-end formation. It is possible that the Pol II has evolved to maintain an optimum rate of elongation in order to provide a 'window' of opportunity for optimal execution of such co-transcriptional processes (Bentley 2014). In my research, I elucidate how altered Pol II catalysis affect Pol II

elongation and various co-transcriptional processes *in vivo*. In order to study the effects of altered catalysis *in vivo*, I have used perturbation of Pol II catalysis through Pol II active site mutations, a system established in the Kaplan lab to dissect transcription mechanisms (Kaplan, Larsson et al. 2008, Kaplan, Jin et al. 2012, Cabart, Jin et al. 2014, Jin and Kaplan 2014, Cui, Jin et al. 2016). In this introductory chapter, I first describe the basic mechanisms of Pol II catalysis and role of the Pol II trigger loop (TL) domain in catalysis and elongation. Then I discuss different facets of elongating Pol II (such as elongation rate, processivity, pausing and backtracking) *in vivo* and coupling of Pol II elongation with co-transcriptional processes. I also discuss about current methods that are used to study transcription elongation in yeast and possible concerns for interpreting them. Finally, I briefly summarize emerging evidence supporting possible coordination of mRNA synthesis and decay to maintain cellular transcript levels.

Function of Trigger Loop (TL) in Pol II catalysis, translocation and intrinsic cleavage

Yeast RNA Pol II is a ~ 500 kDa large protein complex consisting of twelve subunits with a conserved architecture resembling a crab claw, in which the 'pincers' interact with DNA template (**Figure 1-1**). The catalytic center is formed by the two largest subunits Rpb1 and Rpb2 (β′ and β in bacteria, respectively), which interact with the substrate NTP and participate in catalysis (**Figure 1-1**). The nucleotide addition cycle ("NAC)" underpins transcription elongation and consists of two main steps – catalysis and translocation. In the Pol II elongation complex (EC) downstream DNA is unwound and upstream DNA is rewound to form the so called 'transcription bubble', an

unwound region in between that accommodates the Pol II active site along with a 'RNA-DNA hybrid' formed by base pairing of 8-9 nt nascent RNA hybridized with the template DNA. Forward movement of polymerase is accompanied by melting ahead and rewinding behind. The RNA-DNA hybrid is required for elongation complex stability and is maintained throughout the elongation phase (Kireeva, Komissarova et al. 2000, Brueckner, Armache et al. 2009). During each cycle of nucleotide addition, an NMP derived from the NTP substrate is added to the 3' end of the RNA within the RNA-DNA hybrid, conversely, the RNA at the 5' end of the RNA/DNA hybrid separates from the DNA template to be directed towards the RNA exit channel. Catalysis of phosphodiester bond formation occurs through a two-metal ion mechanism universal to nucleic acid polymerases (Steitz, Smerdon et al. 1994), which in Pol II requires two Mg^{+2} cations coordinated by an active site aspartate triad of Rpb1(D481, 483 and 485 for ion A) and Rpb2 D837 (Rpb1 D481, 483 and Rpb2 D837 for ion B) (Cramer, Bushnell et al. 2001, Westover, Bushnell et al. 2004) (**Figure 1-1**). These two Mg^{2+} ions are critical for deprotonation of 3′-OH of the nascent RNA and nucleophilic attack on the α-phosphate of the incoming NTP, forming the phosphodiester bond. In addition to the Rpb1 aspartate triad, several other Rpb1 residues comprise the active site and help to orchestrate the phosphodiester bond formation and subsequent enzyme translocation. A conserved mobile subdomain of the Pol II largest subunit Rpb1, known as the trigger loop (TL), functions in selection of correct NTP substrates and catalysis, while maintaining transcription fidelity (Wang, Bushnell et al. 2006, Kaplan, Larsson et al. 2008, Kireeva, Nedialkov et al. 2008). Genetic and biochemical analysis from our lab

and others have showed that mutation of TL residues can confer reduced (loss of function/'LOF') or increased (gain of function/'GOF') catalytic activity (**Figure 1-2**) (Malagon, Kireeva et al. 2006, Kaplan, Larsson et al. 2008, Kaplan, Jin et al. 2012). Below I discuss functions of TL in catalysis, translocation and intrinsic RNA cleavage with knowledge gained from structural studies and genetic/biochemical analysis of TL mutants.

TL function in catalysis and translocation

Pol II active site participates in a number of activities, including substrate selection, catalysis, translocation and intrinsic RNA cleavage (Kaplan, Larsson et al. 2008, Kireeva, Nedialkov et al. 2008, Larson, Zhou et al. 2012, Dangkulwanich, Ishibashi et al. 2013, Cabart, Jin et al. 2014). For each of these activities the Rpb1 TL domain accomplishes multiple functions through its flexibility and mobility. The flexibility of TL is apparent in crystal structures of Pol II and bacterial RNAP, which have shown that the TL exists in several conformations – 'open' (in absence of substrate NTP), 'closed' (in presence of the substrate) and other semi-open conformations (in presence of polymerase inhibitor and polymerase interacting proteins) (Wang, Bushnell et al. 2006, Vassylyev, Vassylyeva et al. 2007, Brueckner and Cramer 2008, Kaplan, Larsson et al. 2008, Tagami, Sekine et al. 2010, Barnes, Calero et al. 2015). During the nucleotide addition cycle, the TL closes toward the Pol II active site, presumably to trap matched NTPs in the active site and facilitate catalytic positioning. Individual TL residues are positioned to recognize aspects of the correct, templated NTP (**Figure 1-1**). For example, Q1078 interacts with the 2′ and 3′-OH of the substrate ribose possibly

through interaction with Rpb1 non-TL residue N479 (possible direct interaction in crystal structure), while another non-TL residue R446 lies close proximity to the 2′-OH of the ribose and together these residues help in discrimination of NTPs from 2′-dNTPs (Westover, Bushnell et al. 2004, Wang, Bushnell et al. 2006, Cheung, Sainsbury et al. 2011). L1081, makes hydrophobic contact with the NTP substrate by interaction with the base rings on the substrate, while residue N1082 interacts with 3′-OH of the substrate (Wang, Bushnell et al. 2006, Cheung, Sainsbury et al. 2011). In the 'closed' state, TL residue H1085 interacts with the substrate NTP through salt bridge and hydrogen bonding, and has been shown to participate in substrate selection and catalysis (Wang, Bushnell et al. 2006, Kaplan, Larsson et al. 2008). Analysis of previous structures, a recent structural study from Calero lab with our collaboration, and our own novel 'structural-genomics' dissection of TL domain support a step-wise closing model of TL domain. Wherein, 2′-OH interaction of the TL residue Q1078 promotes the release of another TL residue M1079 from a previously trapped hydrophobic pocket formed by multiple TL (A1076, M1079, T1080, G1097 and L1101) and TL-adjacent residues (Barnes, Calero et al. 2015, Qiu, Erinne et al. 2016). This release of M1079 residue from the hydrophobic pocket presumably promotes TL-closing. However, current unpublished structural analysis of a new, higher resolution closed TL from us with Calero group indicates that TL closure does not require disruption of hydrophobic pocket or flipping of M1079, although our genetic data support that interaction of M1079 with other TL residues is critical. Transition from TL 'closed' state to 'open' state is proposed to promote release of PPi and then allow Pol II translocation (Da, Wang et al. 2012,

Malinen, Turtola et al. 2012). Evidence supporting TL function in translocation comes from single molecule studies with purified Pol II that show a GOF TL mutant *rpb1 E1103G*, which increases catalysis rate compared to WT, decreases the propensity for translocation (Larson, Zhou et al. 2012, Dangkulwanich, Ishibashi et al. 2013). In these single molecule optical trapping experiments, the *rpb1 E1103G* enzyme shows a higher pause free velocity than WT, but it is more sensitive to applied force (assisting force that facilitates translocation) compared to WT. As this mutant is known to promote active site closing (Kireeva, Nedialkov et al. 2008), single molecule findings are consistent with closed conformation of the TL being unfavorable for translocation. Some molecular dynamics simulations on Pol II and other msRNAPs have been able to reproduce limited movement of TL and possible conformational changes in TL during translocation; however, a fully simulated model of TL movement and catalysis correlating all the structural observations has yet to be done (Feig and Burton 2010, Feig and Burton 2010, Kireeva, Opron et al. 2012).

Conservation of TL sequence and function

Interestingly, despite the high structural conservation and similarities in the catalytic mechanism, TL-residues may function differently or have different rate limiting steps in different msRNAPs, suggesting a divergent evolution or regulation of TL sub-domain for different outputs. For example, Pol II TL GOF catalytic mutant *rpb1 E1103G* functions as a LOF mutation when tested in Pol I TL (*rpa190 E1224G*) (Viktorovskaya, Engel et al. 2013). Suitable structural simulations combined with extensive genetic

analysis are required to fully understand the evolution, dynamics and functions of this mobile domain in folding pathways and translocation.

TL function in intrinsic cleavage

Beside catalysis and translocation, msRNAPs are also capable of cleavage/hydrolysis of the phosphodiester bond of nascent RNAs from 3′ to 5′ direction by a process called 'intrinsic cleavage'. Intrinsic cleavage was first observed in bacteria, and bacterial RNAP has capability of hydrolyzing phosphodiester bond in faster rate than yeast Pol II (Surratt, Milan et al. 1991, Zenkin, Yuzenkova et al. 2006, Wang, Bushnell et al. 2009). TL function in intrinsic cleavage was proposed based on the results that the antibiotic Streptolydigin inhibit RNA cleavage in *T. aquaticus*, presumably by blocking the TL (Temiakov, Zenkin et al. 2005). Further biochemical analysis of *T. aquaticus* RNAP showed a central role of the TL in intrinsic cleavage (Yuzenkova and Zenkin 2010). A TL residue in *T. aquaticus* RNAP, β′ H1224 (analogous to Rpb1 TL H1085), was shown to participate in phosphodiester bond hydrolysis by acting as a general base and/or positioning the 3′ end NMP of the nascent transcript for cleavage. In contrast, deletion of the TL in *E. coli* was reported to not affect intrinsic or regulator-assisted transcript cleavage (Zhang, Palangat et al. 2010). However, as *E. coli* RNAP cleavage rate is 1-2 orders of magnitude slower than *T. aquaticus* RNAP, *E. coli* RNAP cleavage activity may differ from *T. aquaticus*. Work from our lab has shown that although yeast Pol II has a very low efficiency for intrinsic cleavage, mutation of the TL residue H1085 to Y1085 leads to a gain of cleavage

activity of the enzyme possibly as the substituted tyrosine may participate in acid-base catalysis (Cabart, Jin et al. 2014).

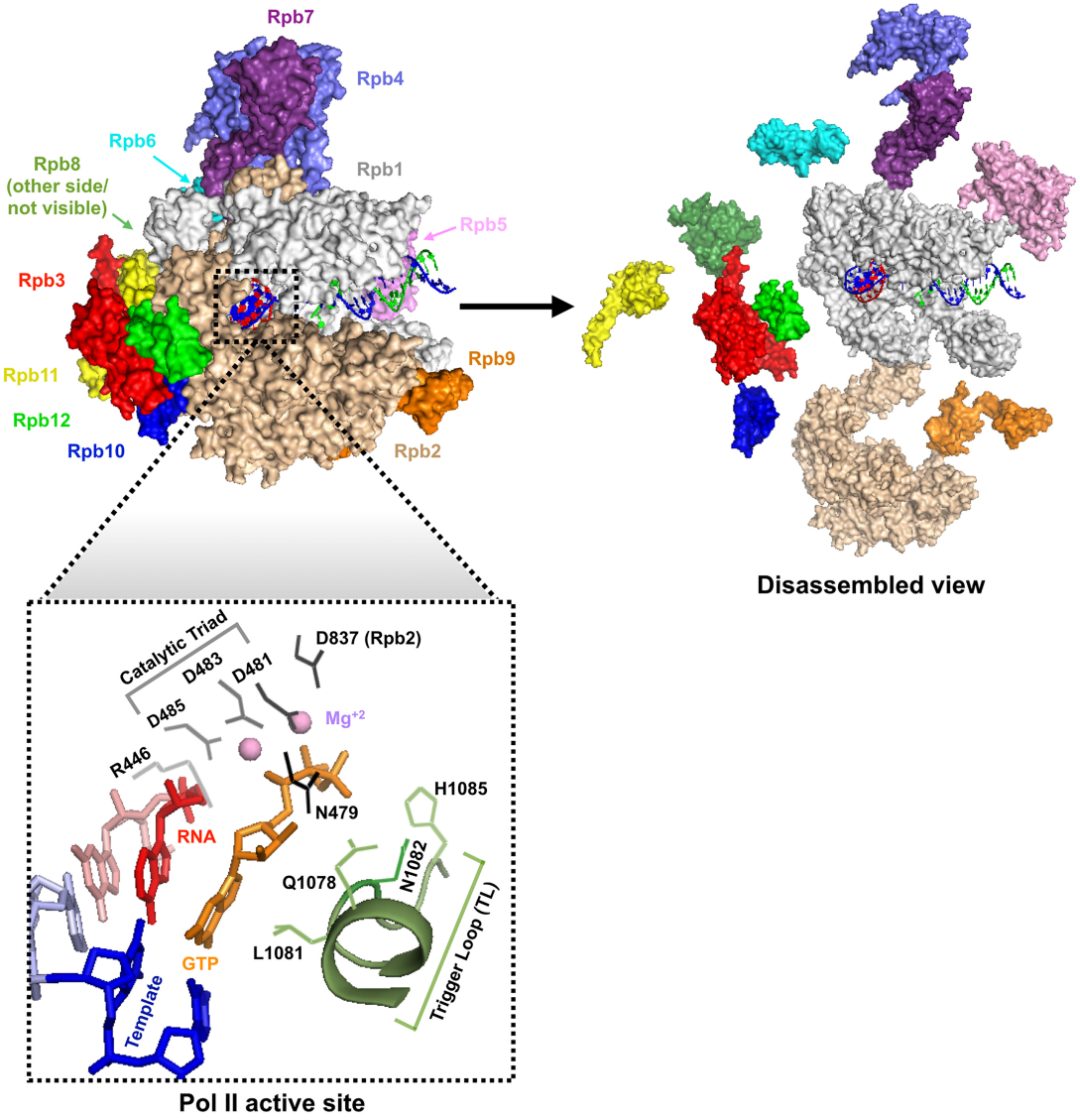


Figure 1-1. Architecture of Pol II and active site. (PDB:2E2H). Yeast Pol II comprises 12 subunits (Left – intact, right – disassembled views). Catalytic core is formed by subunits Rpb1 and Rpb2. (Bottom) Residue composition of Pol II active site.

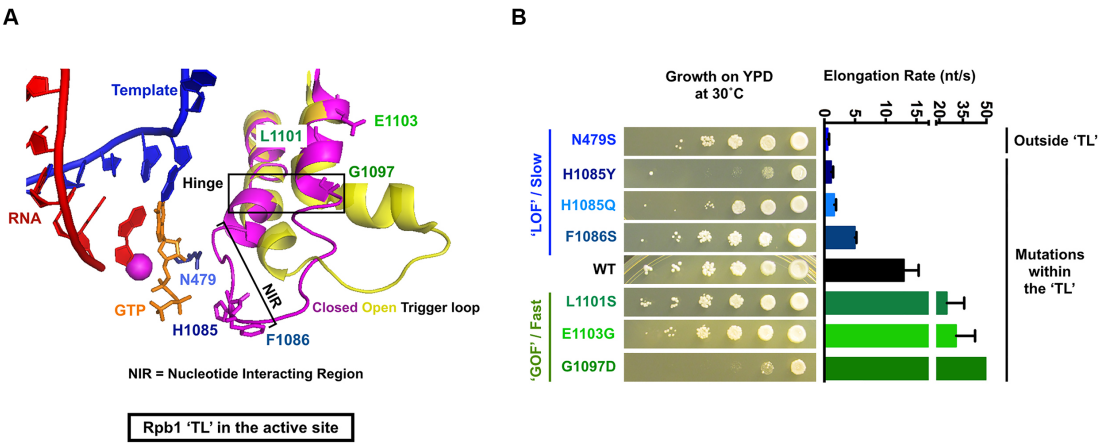


Figure 1-2. Pol II TL mutants alter *in vitro* elongation rate and growth. (A) 'Open' and 'closed' conformation of trigger loop (TL) from PDB:2E2H. Highlighted amino acid residues in the TL and Pol II active site that are substituted to generate catalytic mutants. (B) Pol II mutants show a wide range of *in vitro* elongation rate and growth defects.

Pol II elongation in vivo

In vivo, Pol II elongation happens on a chromatin template and is coincident with chromatin modification and co-transcriptional RNA processing events. Elongating Pol II has several facets, which include – Pol II elongation rate, pausing, backtracking and processivity. Elongation rate is the measure of how quickly Pol II travels over the template, which is frequently interrupted by polymerase pausing and backtracking, while processivity describes the ability of elongating Pol II to complete synthesis of a full transcript. Here I discuss our current knowledge about these functions of Pol II during elongation and how they might impact a productive elongation.

<u>In vivo elongation rate</u>

Pol II in vivo elongation rate represents how fast Pol II moves over the template and describes the approximate rate (kb/min) of RNA chain synthesis inside the cell. Elongation rate comprises both 'on pathway' events and 'off pathway' events. On pathway events include multiple nucleic acid addition cycles (NAC). As noted above, the NAC is composed of Pol II catalysis and translocation, together, which determine the pause free elongation rate. Off pathway events include Pol II pausing and backtracking, which together also impact overall elongation across the template. In vivo elongation rates have been measured for different organisms in a number of ways, mainly using reporter genes that are inducible or responsive to stimulus (**Table 1-1**). In *Drosophila*, Pol II elongation rates have been measured on developmentally regulated or on stimulus-responsive genes by determining the time for 'first wave' of Pol II molecules to reach the 3′ end of the gene (Thummel, Burtis et al. 1990, Shermoen and O'Farrell 1991,

O'Brien and Lis 1993, Yao, Ardehali et al. 2007). Conversely, in yeast, Pol II elongation rates have been measured by monitoring the 'last wave' of Pol II molecules leaving a long reporter template after transcriptional shutoff (Mason and Struhl 2005). These studies reported an approximate average elongation rate of 1-2 kb/min for both yeast and *Drosophila*. In a recent study, by measuring the appearance of transcript 5′ and 3′ ends through binding of fluorescent probes to arrays of RNA binding sites in each part of a reporter transcript in *Drosophila* embryos, a much faster average elongation rate (2.4-3 kb/min) has been reported (Fukaya, Lim et al. 2017). This revised rate supports the transcription of longer genes during the very short cell cycles during *Drosophila* embryo development as being physically possible in between mitoses, where transcription is inhibited. For mammalian cell lines, several techniques such as live cell imaging and fluorescent in situ hybridization have been used to determine a range of average elongation rates in vivo (**Table 1-1**). All these assays measure an overall average elongation rate of Pol II, but journey of Pol II over a template is not even and likely to vary at different positions. Indeed, several genome-wide elongation studies in mammals/murine cells have shown that productive elongation is slower at the 5′ end of gene and the rate increases with the length traversed. Within the initial few kb of a gene Pol II elongation rate can be as low as 0.5 kb/min, which subsequently increases up to 2-5 kb/min on the later part of the gene body (Danko, Hah et al. 2013, Fuchs, Voichek et al. 2014, Jonkers, Kwak et al. 2014, Veloso, Kirkconnell et al. 2014). This disparity of rate over the gene body may arise from chromatin obstacles, differential recruitment of negative or positive factors, or modifications of the transcription machinery. This

speculation is supported by previous observations that nucleosomes are generally well positioned at the 5′ end of genes, while differential phosphorylation of the carboxy-terminal domain of the large subunit, the CTD, increases after a few kb of transcription (Churchman and Weissman 2011, Heidemann, Hintermair et al. 2013, Harlen and Churchman 2017). Nucleosomal obstacles and regulatory factors may impact overall elongation rate through controlling the propensity of Pol II for pausing and backtracking, which I discuss next.

Table 1-1. List of experimentally determined *in vivo* elongation rates

Assay	Organism/ Model system	Average elongation rate (kb/min)	References
Chromatin immunoprecipitation (ChIP) time course after transcription shutoff using *GAL1p*::*YLR454w* reporter	Yeast	~ 2	(Mason and Struhl 2005)
Northern blotting on *E74* (60 kb) gene	*Drosophila*	1.1	(Thummel, Burtis et al. 1990)
In situ hybridization on *Ubx* (74 kb) gene	*Drosophila*	1.4	(Shermoen and O'Farrell 1991)
Nuclear run-on and FISH on Hsp70	*Drosophila*	1.2-1.5	(O'Brien and Lis 1993, Yao, Ardehali et al. 2007)
Dual color fluorescent microscopy	*Drosophila* embryo	2.4-3.0	(Fukaya, Lim et al. 2017)
RT-PCR on human native genes	Human cells	2.4 and 3.8	(Tennyson, Klamut et al. 1995, Singh and Padgett 2009)
Live cell imaging on engineered reporter construct in U20S cells	Human cells	~2.0 and 4.3	(Femino, Fay et al. 1998, Darzacq, Shav-Tal et al. 2007)

Table 1-1 Continued

Assay	Organism/ Model system	Average elongation rate (kb/min)	References
Genome-wide elongation rate	Mammalian cells	Differs on gene body. From 0.5 kb/min at the 5′ end to 2-5 kb/min at the 3′ end	(Danko, Hah et al. 2013, Fuchs, Voichek et al. 2014, Jonkers, Kwak et al. 2014, Veloso, Kirkconnell et al. 2014)

Pausing and backtracking

Pol II elongation is a discontinuous process, interrupted by frequent pausing of and occasional backtracking. We have recently started to understand the biological functions and consequences of Pol II pausing and backtracking during elongation. Evidence of polymerase pausing was first observed in vitro using bacterial polymerase, which was followed by in vivo observation for Pol II using nuclear run-on assay (Maizels 1973, Gariglio, Bellard et al. 1981). Later, using a nuclear run-on assay, the Lis lab remarkably showed that Pol II accumulated at the 5′ end of a *Drosophila* heat shock gene (hsp70) even when the gene is uninduced, a phenomenon termed as 'promoter-proximal pausing' (Gilmour and Lis 1986, Rougvie and Lis 1988). Promoter-proximal pausing in yeast is not common, thus, here I will focus mainly on gene body pausing. Recent advances in genome-wide techniques to map polymerase distributions (see NET-seq further) have enabled measurement of gene body pausing through out the genome (Churchman and Weissman 2011). Such studies have identified consensus sequence and regulatory factors for polymerase pausing in bacterial, yeast and human cells by

determining the positions within the genome that show increased probability of polymerase occupancy through mapping the 3′ ends of nascent RNAs (Churchman and Weissman 2011, Larson, Mooney et al. 2014, Vvedenskaya, Vahedian-Movahed et al. 2014, Mayer, di Iulio et al. 2015, Nojima, Gomes et al. 2015). Factors that can induce polymerase pausing may include DNA sequence, nucleosomes and DNA-binding factors (Mayer, Landry et al. 2017). Conversely, there are *trans-acting* transcription elongation factors such as TFIIS that aid Pol II recovery from extended transcriptional pausing (Reinberg and Roeder 1987, Reines, Chamberlin et al. 1989). Extended pausing induced by a nucleotide mismatch can also lead to 'backtracking' or movement of Pol II backward on the template (Nudler, Mustaev et al. 1997). Backtracked Pol II misaligns the 3′ end of the nascent RNA from the active center of Pol II, which can further lead to a transcriptional 'arrest' (Cheung and Cramer 2011). Arrested Pol II can be rescued with aid of TFIIS, which helps in cleaving the mismatched 3′ end of the nascent RNA to restore the active site alignment of the RNA (Izban and Luse 1992, Reines 1992, Cheung and Cramer 2011). Failure to recover from a backtracked/arrested state can either lead to ubiquitination and degradation of Pol II largest subunit Rpb1 or premature termination of transcription (Somesh, Reid et al. 2005, Park, Kang et al. 2015). Removal of Pol elongation complex from the template leads to a processivity defect or failure to complete transcript synthesis, which I discuss below.

Pol II processivity

A fully processive Pol II will complete uninterrupted synthesis of a transcript, and considering that most yeast genes are relatively shorter in size, it was previously

assumed that Pol II should efficiently traverse the full extent of the template. In a pioneering study by Mason and Struhl, using a galactose-inducible long reporter gene (*GAL1p*::*YLR454w*) in yeast (see further description), apparent processivity was determined by comparing Pol II ChIP ratio between the 5′ and 3′ ends of the reporter (Mason and Struhl 2005). Mutants in a number of transcription-associated factors that showed relative decrease in 3′ end ChIP signal relative to 5′ end and were inferred to be defective for Pol II processivity. Since the Pol II EC is highly stable, it can be assumed that dislodging of Pol II from the template can primarily relate to either premature termination or degradation of Rpb1 followed by EC disassembly. However, a differential speed on the 3′ end of the gene would also lead to an apparent processsivity defect, as ChIP is unable to differentiate between these two possibilities. Interestingly, in yeast, growth conditions can also affect the 5′ to 3′ Pol II occupancy ratio, suggesting a global effect of external perturbation on Pol II elongation. Indeed, we observed that Pol II catalytic mutants show differential occupancy and growth in different media that I present in Chapter II and discuss in Chapter III.

In vivo assays to study Pol II elongation

Pol II occupancy over the coding region of a gene represents the transcriptional status of that gene. With the availability of techniques such as transcription run-on and chromatin immunoprecipitation (ChIP), it was immediately possible to analyze in vivo association of elongating Pol II with a transcribing gene. Subsequently these assays were adapted to genome-wide analysis of Pol II elongation. Beside these molecular techniques several genetic reporter systems have been widely used to assess transcription elongation

defects in yeast. Here I discuss few commonly used methods to study transcription elongation in vivo (**Table 1-2**), with a major focus on the widely used transcriptional shut off assay, which is used to determine apparent in vivo elongation rate in yeast. Further, I discuss commonly used genetic reporter systems to determine transcription elongation defects in yeast.

Molecular techniques to determine elongation defects

The transcription run-on assay has been used as a direct measurement of the density of elongating Pol II (Hirayoshi and Lis 1999). This assay relies on permeabilization of nuclei by sarkosyl, which stops all cellular processes and disrupts most chromatin proteins, while elongation complexes remain stable. Next, transcriptionally engaged Pol II is allowed to run-on in presence of radiolabeled UTP to produce labeled RNAs that can be hybridized to specific probes. Originally developed for higher eukaryotic nuclei, this assay was further adapted for intact yeast cells and for application in a global scale using DNA microarrays (Birse, Lee et al. 1997, Garcia-Martinez, Aranda et al. 2004). Recently, use of ribonucleotide analog [5- bromouridine 5′-triphosphate (BrUTP)] for run-on assay has allowed to pull-down BrU-tag nascent RNAs during the run-on to map transcriptionally engaged RNA polymerases genome-wide (GRO-seq) (Core, Waterfall et al. 2008).

Chromatin immunoprecipitation (ChIP) is commonly used to determine Pol II association with specific genes or genome-wide. Higher Pol II occupancy over a gene is generally assumed to represent active transcriptional status of the gene. As noted above, a remarkable study from Struhl lab established a new system to determine Pol II

elongation defects in yeast (Mason and Struhl 2005). This assay uses chromatin immunoprecipitation (ChIP) to determine Pol II occupancy over a long galactose-inducible hybrid gene, *GAL1p*::*YLR454w*. Steady state Pol II occupancy across this reporter can be compared between WT and mutants. As noted earlier, apparent processivity defects for a mutant is inferred by comparing the 5′ to 3′ Pol II occupancy ratio between WT and mutant. This same reporter can be used to determine apparent in vivo elongation rate by measuring the kinetics of the 'last wave' of Pol II leaving the template. Several Pol II catalytic mutants and elongation factors have been characterized for apparent in vivo elongation rate and apparent processivity defects using this assay (See Chapter II – **Table 2-1**). However, this assay assumes the transcriptional shutoff kinetics by the addition of glucose is same for WT and mutant. We have found that glucose signaling kinetics upstream of transcriptional shutoff can be different for WT and mutants, which may affect the interpretation of this assay (see Chapter II for the findings and Chapter III for further discussion).

There are few assays that determine possible elongation defects or consequences of defective elongation indirectly in yeast. For example, gene length-dependent accumulation of mRNA (GLAM) assay measures reporter gene expression defect based on length of the transcript driven by an identical promoter (Morillo-Huesca, Vanti et al. 2006). This assay works based on the assumption that if the same promoter drives different lengths of reporter transcripts, initiation efficiency between these reporters remains same. Hence, any discrepancy in the mature transcript levels comes from the defects in elongation, which presumably may be exacerbated with increasing length of

the transcript. Indeed, this assay has been successfully used to identify novel regulators of transcription elongation in yeast (Gaur, Hasek et al. 2013, Millan-Zambrano, Rodriguez-Gil et al. 2013). However, as the readout of this assay is defective mRNA expression, which can come from several defects apart from elongation such as mRNA processing or export, and it is difficult to uncouple elongation defects from other associated defects by this assay. Similarly, another assay, comparative dynamic transciptome analysis (cDTA) can measure absolute rate of mRNA synthesis, but alteration in synthesis rate may not be a direct representation of defective elongation (Sun, Schwalb et al. 2012). In cDTA, 4-thiouracil (4tU) is used for non-perturbing metabolic labeling of newly synthesized RNAs in yeast. After desired time course of labeling, total RNA is extracted and labeled RNAs are separated by biotinylation and purification with streptavidin-coated magnetic beads. Total, labeled and unlabeled fractions are analyzed with microarray and a dynamic kinetic modeling is used to determine mRNA synthesis and degradation rate. A defined number of metabolically labeled fission yeast (*S. pombe* labeled with 4-thiouridine) sample is used as an internal standard for accurate comparison and absolute quantification of mRNA synthesis or degradation rate. Several transcription and mRNA processing related factor mutants and two Pol II mutants have shown to alter mRNA synthesis rate in this assay, but, only few of them have shown to alter elongation in the ChIP assay described earlier (Sun, Schwalb et al. 2012, Sun, Schwalb et al. 2013, Schulz, Pirkl et al. 2014).

A recently developed genome-wide technique, nascent elongating transcript sequencing (NET-seq), allows more direct measurement of transcription and

transcription associated processing events in vivo (Churchman and Weissman 2011, Churchman and Weissman 2012). This assay maps 3′ end of the transcriptionally active nascent RNA genome-wide in a single nucleotide resolution, thus, revealing the density RNA polymerase across the genome. A selective advantage of this approach over similar approaches such as ChIP-seq or GRO-seq (discussed above) is that NET-seq exploits the incredible stability of the elongation complex (DNA-RNA-RNA polymerase) to purify polymerase-associated nascent RNAs without cross-linking or perturbing cellular physiology. Originally developed in yeast, this assay has been incredibly useful to determine polymerase occupancy, pausing and backtracking in other systems such as bacteria or mammalian cells (Vvedenskaya, Vahedian-Movahed et al. 2014, Mayer, di Iulio et al. 2015, Nojima, Gomes et al. 2015). Further, this assay can be slightly modified to determine co-transcriptional processing such as 5′-capping, which I discuss in the future direction section (Chapter III).

In vivo phenotypes and genetic screens for Pol II catalytic mutants

Use of genetic reporters has been remarkably useful to identify transcriptional mutants as well as to increase our understanding of transcription process *in vivo*. These genetic reporters are sensitized to altered transcription by different means, which include – insertion of a Ty retrotransposon element into genes or promoters altering host gene transcription (*e.g. lys2-128∂* and *his4-912∂* alleles of *LYS2* and *HIS4* genes, respectively), drug-induced limitation of cellular nucleotide levels (MPA and 6-AU) and mRNA processing defects and transcriptional interference (*gal10Δ56*) (**Figure 1-3**).

Table 1-2. List of assays used to study Pol II elongation in yeast

Assay	Description	Ref
Transcription run-on and genome-wide run-on	Measures density of active Pol II on a specific template or genome-wide	(Hirayoshi and Lis 1999, Garcia-Martinez, Aranda et al. 2004, Core, Waterfall et al. 2008)
Chromatin immunoprecipitation (ChIP)	Determines Pol II association with a target gene or genome wide. In yeast, it is widely used to determine Pol II processivity and elongation on an inducible reporter	(Mason and Struhl 2005)
Gene length dependent accumulation of mRNA (GLAM)	Indirect readout of elongation defect by comparing mRNA expressions of a short vs. long transcript initiating from same promoter.	(Morillo-Huesca, Vanti et al. 2006)
Comparative dynamic transcriptome analysis (cDTA)	Measures genome-wide mRNA synthesis and degradation rate by metabolic labeling	(Sun, Schwalb et al. 2012)
Native elongating transcript sequencing (NET-seq)	Measures genome-wide polymerase density in single nucleotide resolution by mapping 3′ end of the polymerase associated nascent RNA. Can be used to determine Pol II occupancy, pausing, backtracking, fidelity and co-transcriptional pre-mRNA processing	(Churchman and Weissman 2011)
Fluorescent detection of nascent RNA	Measures transcript synthesis in vivo by fluorescent-tracking of 5′ and 3′ stem loops of a reporter	(Larson, Zenklusen et al. 2011, Hocine, Raymond et al. 2013)

Indeed, these genetic screens have been useful to identify or characterize several Pol II mutants (Kaplan 2013). Our lab has used most of these genetic reporter systems extensively not only to probe mechanistic aspects of transcription process but also to expand the understanding of the systems themselves. Here I discuss our current

understanding of these commonly used genetic systems with the major focus on use of a nucleotide-depleting drug MPA to study elongation mutants.

Interaction between adjacent transcription units can lead to complex transcriptional outputs. For example, yeast genomic loci harboring cryptic transcription unit in the 5′ UTR of a gene or within the gene body can interfere with transcription of the native gene. Such transcriptional perturbations in certain amino acid biosynthesis genes render cells to auxotrophic for the relevant amino acid. Insertion of a Ty1 ∂ element (the yeast Ty retrotransposon long terminal repeat is called a ∂ element) 5′ of *HIS4* (*his4-912∂*) or in the 5′ UTR of *LYS2* (*lys2-128∂*) confers auxotrophy for histidine or lysine, respectively (Simchen, Winston et al. 1984, Winston, Chaleff et al. 1984). The Ty1 ∂ element contains Ty element promoter, thus insertion of ∂ element creates a complex transcription unit containing two promoters, one native and one ectopic. Transcription mutants that can suppress transcriptional defects incurred by insertion of Ty element (suppression of Ty, Spt^-) in *his4-912∂* or *lys2-128∂* reporters, allow bypass of the observed auxotrophy. Our lab has utilized the *lys2-128∂* reporter allele to characterized Pol II catalytic mutants with our results indicating that catalytically fast Pol II alleles confer suppression of the *lys2-128∂,* otherwise known as the Spt^- phenotype (Cui, Jin et al. 2016, Qiu, Erinne et al. 2016). Likewise to adjacent transcription units within same gene interfering, impaired transcription of one gene can interplay with transcription of an adjacent gene. As the yeast genome is highly compressed and contains short intergenic sequences, impaired transcription termination of an upstream transcript can cause interference with adjacent downstream transcription.

This phenomenon is the basis for use of the *gal10Δ56* reporter allele, which contains a deletion in the major polyadenylation signal of the *GAL10* gene (Greger and Proudfoot 1998, Kaplan, Holland et al. 2005). As a result, *GAL10* transcription does not terminate properly and transcriptional readthrough interferes with the adjacent downstream *GAL7* gene. This interference reduces *GAL7* transcription resulting into buildup of a toxic intermediate in the galactose metabolic pathway normally metabolized by the *GAL7* gene product, and therefore leading to galactose sensitivity of the cells. Mutations in Pol II and elongation factors can alter this transcriptional interference and suppress *gal10Δ56* galactose sensitivity (Kaplan, Holland et al. 2005, Kaplan, Jin et al. 2012). Generally Pol II LOF mutants, and with a few exceptions of GOF mutants, have been shown to suppress galactose sensitivity of *gal10Δ56* (Qiu, Erinne et al. 2016).

Similar to genetic reporters sensitized to altered transcription in vivo, drugs that deplete the cellular nucleotide pool have also been widely used to characterize transcription mutants in yeast. These drugs include mycophenolic acid (MPA) and 6-azouracil (6-AU), which targets GTP synthesis and both UTP and GTP synthesis pathways, respectively (Sweeney 1977, Archambault, Lacroute et al. 1992, Exinger and Lacroute 1992, Powell and Reines 1996, Reines 2003). It has been widely assumed that growth sensitivity of mutant to MPA or 6-AU is a synergistic effect between elongation defects of the mutant itself and drug-induced elongation defects (such as pausing and backtracking of Pol II) due to limiting nucleotides. Extensive work from the Reines lab has revealed a great detail of the mechanism of how wild type cells confer resistance to these drugs (Shaw and Reines 2000, Shaw, Wilson et al. 2001, Hyle, Shaw et al. 2003,

Jenks and Reines 2005). MPA/6-AU targets inosine monophosphate dehydrogenase (IMPDH), an enzyme in the GTP biosynthesis pathway encoded in the yeast genome by three homologous *IMD* genes (*IMD2, 3* and *4*). The Reines group found that resistance to MPA/6-AU requires up regulation of the drug resistant form of the enzyme encoded by the *IMD2* ortholog, while mutants that fail to induce *IMD2* are generally sensitive to the drug. Further work from the Brow and Reines labs revealed that under GTP-replete conditions, upstream GTP-initiating transcription start sites (transcripts start with G) are used at the *IMD2* promoter, generating cryptic unstable transcripts (CUT) that are attenuated by premature termination (**Figure 1-4**) (Jenks, O'Rourke et al. 2008, Kuehner and Brow 2008). Upon GTP starvation (induced by MPA or 6-AU treatment), a downstream functional 'A' start site (ATP-initiating) is used to produce a functional *IMD2* transcript, allowing expression of the drug resistant form of the enzyme. An analogous regulation of start site utilization also controls expression of *URA2* upon uracil starvation (**Figure 1-3**) (Kwapisz, Wery et al. 2008, Thiebaut, Colin et al. 2008). In the presence of high uracil, an upstream start site is used at *URA2* and is attenuated by premature termination. Upon uracil starvation (such as 6-AU treatment), a distal start site can be used to make productive *URA2* mRNA. Many Pol II mutants derange this drug-induced shift in start site utilization and our lab has extensively studied the regulation TSS utilization at *IMD2* promoter using Pol II catalytic mutants, which I discuss further.

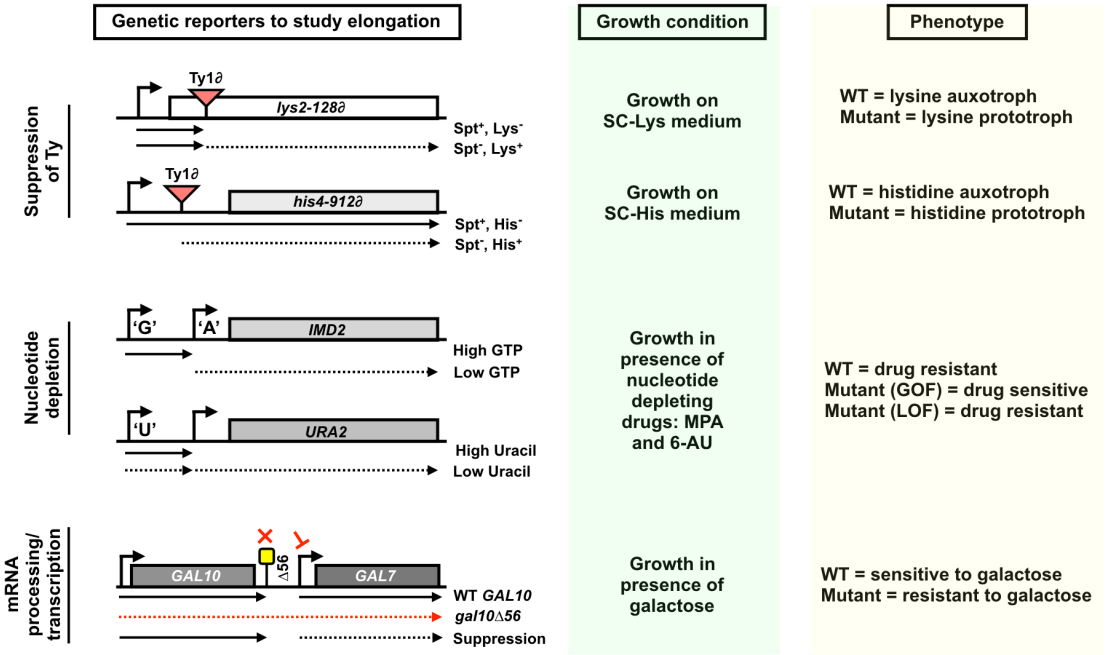
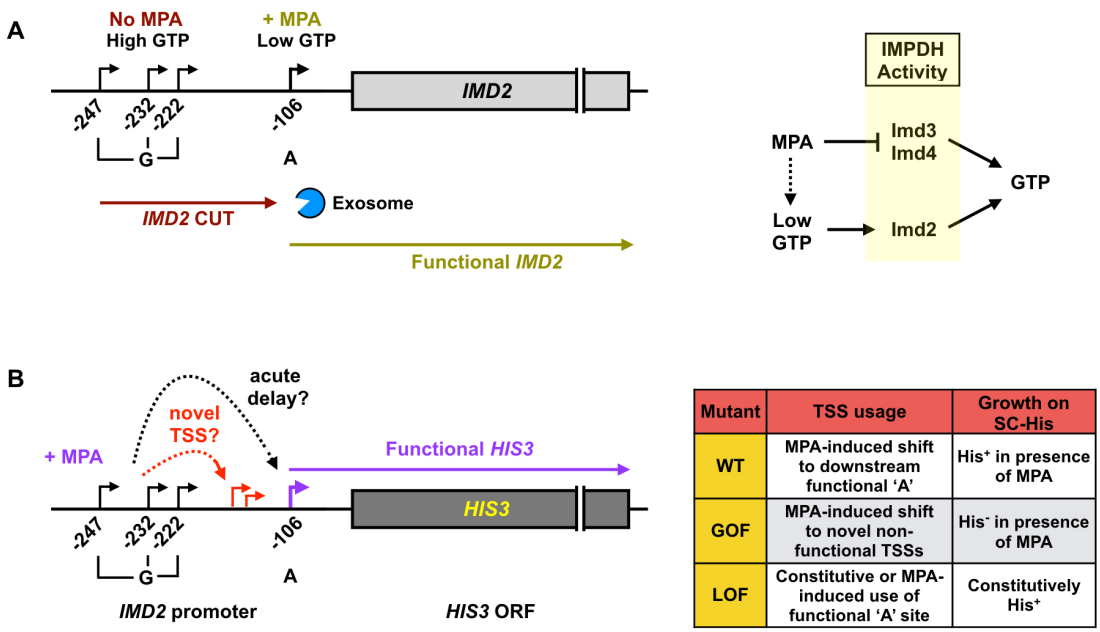


Figure 1-3. Genetic reporters and phenotypes to examine altered Pol II catalysis *in vivo*



Mutant	TSS usage	Growth on SC-His
WT	MPA-induced shift to downstream functional 'A'	His^+ in presence of MPA
GOF	MPA-induced shift to novel non-functional TSSs	His^- in presence of MPA
LOF	Constitutive or MPA-induced use of functional 'A' site	Constitutively His^+

Figure 1-4. Mechanism of TSSs utilization at *IMD2* promoter. (A) (left) *IMD2* gene transcription in presence and absence of MPA. In absence of MPA, *IMD2* is transcribed from upstream 'G' start sites and the resulting non-functional CUTs are degraded by the exosome. In the presence of MPA (GTP depletion), a downstream 'A' start site is utilized to produce functional *IMD2* transcripts. (right) MPA depletes cellular GTP by inhibiting activity of a GTP biosynthesis pathway enzyme IMPDH. In yeast, IMPDH is encoded by paralogous enzymes, Imd2, Imd3 and Imd4. *IMD3* and *4* gene products are sensitive to MPA treatment, however, upon GTP depletion, a functional *IMD2* product makes cell resistant to MPA. (B) (Left) Schematic of the *imd2*Δ::*HIS3* construct and possible mechanism for inability MPA sensitive mutants to shift TSS utilization. (Right) Expected phenotype for Pol II catalytic mutants upon MPA treatment in synthetic medium lacking histidine.

Previous work from our lab has shown that GOF catalytic mutants are generally sensitive to MPA, while LOF mutants are resistant (Kaplan, Jin et al. 2012, Qiu, Erinne et al. 2016). Notably, GOF and LOF catalytic mutants also have propensity to shift start site utilization upstream and downstream, respectively, when tested on reporter genes or genome wide (Kaplan, Jin et al. 2012, Jin 2015). A simple explanation of MPA resistance for LOF mutants is that this class of mutants shifts start site utilization downstream at *IMD2* promoter, thus, constitutively using the functional 'A' TSS to produce drug resistant form of the enzyme. Indeed, downstream functional 'A' TSS at *IMD2* promoter is constitutively used in LOF mutants *in vivo* (Kaplan, Jin et al. 2012). Mechanism of sensitivity for GOF mutants to the drug is complex. We hypothesized that GOF mutants may fail to shift start site at *IMD2* promoter, making cells sensitive to MPA in two possible ways (**Figure 1-4**). First, faster catalytic rate might make GOF mutants insensitive to depleted nucleotide pools and they are thus less responsive to acute GTP starvation after MPA treatment. Second, GOF mutants might shift the start site downstream, but to a non-functional novel start site upstream from the functional 'A' TSS, resulting in production of a cryptic transcript attenuated by premature termination. Our results support the second possibility, which I present in Chapter II. Finally, to exploit the constitutive downstream utilization of functional start site in LOF mutants, we constructed a reporter system where we replaced the *IMD2* ORF with *HIS3* (*imd2*Δ::*HIS3*). This reporter can be utilized to screen for new transcription mutants that shift TSS usage downstream, leading to constitutive expression of *HIS3* (**Figure 1-4**).

Co-transcriptional mRNA processing

Eukaryotic mRNAs are complex and need proper processing and unique packaging to form the messenger ribonucleoproteins (mRNPs), which enables efficient transportation to, and translation in, the cytoplasm. Maturation of pre-mRNAs requires addition of a 7-methylguansine cap to the 5′ end, splicing of introns, and addition of a non-templated poly(A) tail to the processed 3′ end. Although each of these processing events can be sequentially reconstituted in vitro, now it is evident that these maturation processes are coupled to transcription elongation, thus are termed 'co-transcriptional'. Several lines of evidence have established the co-transcriptionality of mRNA processing. Electron microscopy revealed spliced nascent pre-mRNAs in chromatin in close proximity to the Pol II (Beyer and Osheim 1988). Subsequently, several studies have shown that transcription factors and co-factors can influence 5′-capping, splicing, 3′ end formation and polyadenylation (Monsalve, Wu et al. 2000, Chiu, Ho et al. 2002, Rosonina, Bakowski et al. 2003, Schroeder, Zorio et al. 2004, Lenasi, Peterlin et al. 2011, Martins, Rino et al. 2011, Nagaike, Logan et al. 2011, Huang, Li et al. 2012, Ji, Zhou et al. 2013). Moreover, Pol II itself is capable of recruiting pre-mRNA processing factors co-transcriptionally. Recruitment of processing factors can happen through the C-terminal domain (CTD) of largest subunit of Pol II, Rpb1 (Recently reviewed in (Harlen and Churchman 2017)). The Pol II CTD has a heptapeptide repeat (consensus amino acid sequence YSPTSPS), which varies from 26 repeats in yeast to 52 repeats in humans. The Pol II CTD can undergo several posttranslational modifications and is known to influence several stages of transcription. It is proposed that specific

posttranslational modifications of the Pol II CTD act as landing pads for recruitment of chromatin remodelers and pre-mRNA processing factors to facilitate transcription elongation and processing of nascent RNAs, respectively.

As the recruitment of the pre-mRNA processing factors and pre-mRNA processing occurs simultaneously with elongation, it is thought that Pol II has evolved to maintain a certain elongation rate that matches with the rate of processing events, so that both can happen within a coordinated time-frame; hence elongation is said to be 'kinetically' coupled to co-transcriptional processing. Kinetic coupling or kinetic competition suggests that pre-mRNA processing factors compete for the nascent RNA elements while Pol II elongation rate determines when a processing event occurs relative to where Pol II is on the template. However, the precise mechanism or extent of this coupling and how altered Pol II elongation rate affects this coordination is still under investigation. Presumably, failing to coordinate between elongation and pre-mRNA processing can lead to improper processing of the nascent RNAs. Impaired processing of the transcripts results in degradation of the transcripts and may lead to premature termination of elongating Pol II. Thus, cells possess several nuclear quality control mechanisms or 'surveillance' to counter impaired 5′ and 3′ processing of the nascent RNAs (Luo, Johnson et al. 2006, Jiao, Xiang et al. 2010, Chang, Jiao et al. 2012, Kilchert, Wittmann et al. 2016). In this section I discuss how and up to what extent altered Pol II elongation rate may affect co-transcriptional processes, with the focus on 5′ capping, splicing and 3′ formation/termination. Additionally, I also briefly discuss

relevant quality control pathways, which counter impair processing events and may lead a pre-mature termination.

Co-transcriptional capping

Immediately after the 5′ end of the nascent transcript of 20-30 nucleotides length emerges from the RNA exit channel, capping can be detected (Rasmussen and Lis 1993). mRNA capping requires step-wise action of three enzymes, which include a RNA-triphosphatase (Cet1p in yeast), a guanylyltransferase (Ceg1p in yeast) and a guanine N7-methyltransferase (Abd1 in yeast) (**Figure 1-5**) (Ramanathan, Robb et al. 2016). Phosphorylation of Pol II CTD Ser5 and other elongation factors (Spt4/5 in yeast) are required for recruitment and activation of capping enzymes to the nascent RNAs (Cho, Takagi et al. 1997, McCracken, Fong et al. 1997, Wen and Shatkin 1999). The 5′ triphosphate of the nascent RNA is first modified by the triphosphatase, which removes the gamma-phosphate of 5′-ppp to generate 5′-pp. Then the guanylyltransferase transfers a GMP from GTP to the 5′-pp via a lysine-GMP intermediate. Finally, the guanine N7-methyltransferase adds a methyl group to the N7-amine to form the m7G cap. Capping is required to protect the nascent RNAs from degradation. As the cap structure is required for efficient packaging, export to the cytoplasm and translation of the mRNAs, this surveillance mechanism ensures selection of capped mRNAs over uncapped or improperly capped ones. In yeast, 5′ surveillance pathways majorly consist of the 5′ to 3′ exonucleases Rat1p and Xrn1p. Rat1p is an essential 5′ to 3′ nuclear exonuclease involved in transcription termination and nuclear quality control (Kim, Krogan et al. 2004). Rat1p functions in complex with its partner Rai1p, which makes the 5′ end of

partially or defectively capped transcripts competent for Rat1p degradation (Rat1p is specific for RNAs having a 5′-monophosphate) (Jiao, Xiang et al. 2010). Xrn1p is non-essential and a major component of the cytoplasm decay pathway, but proposed to have nuclear function as well. Indirect evidence for of effects of altered Pol II elongation rate on 5′ capping comes from genetic interaction between Pol II mutants with surveillance factor mutants that I discuss further below.

A kinetic coupling model of the capping process, where capping and synthesis have evolved to be of similar rates, would predict that fast Pol II mutants might alter where capping takes place and could in fact be defective if elongation proceeds too far prior to capping occurring. Conversely, slow Pol II elongation would allow more time for capping. In contrast, previous studies have proposed the existence of an mRNA 'capping checkpoint', wherein Pol II pauses/slows down until the capping process is completed (Glover-Cutter, Kim et al. 2008). Hence under such a model, Pol II elongation rate would have little effect on efficiency of capping. In metazoans, Pol II generally pauses at the 5′ end of genes with capping enzymes/cap-binding complex recruited at this pause. This pausing could be a part of the proposed checkpoint mechanism, which ensures that uncapped transcripts are not extended. To date, very few studies have been performed to determine how altered Pol II elongation rate may affect the capping process. A previous study has shown a slow Pol II catalytic mutant in yeast, *rpb1 N488D*, to be defective of capping for a reporter gene (Jimeno-Gonzalez, Haaning et al. 2010). A capping defect of *rpb1 N488D* mutant was inferred from its synthetic sick interaction with a double mutant of a temperature sensitive allele of Rat1 (*rat1-1*) and

*xrn1*Δ. Presumably, in the absence of nuclear quality control exonucleases (*xrn1*Δ/*rat1-1*) inappropriately capped transcripts build up in the Pol II mutant strain (*rpb1 N488D*) to make the cells sick. This notion was further supported by the observation that mutant Pol II showed a decreased level of capped reporter RNA when quantified after capped-only RNA pull-down using an anti-cap antibody. However, since the *rpb1 N488D* mutant has been shown to be catalytically slower than WT (Malagon, Kireeva et al. 2006, Jimeno-Gonzalez, Haaning et al. 2010), the observed results are difficult to square with a simple kinetic coupling model, which would predict slow mutants would have increased time to allow capping, while Pol II fast mutants would have possible reduction in time to cap. We examined Pol II GOF and LOF activity mutants for genetic interactions with *xrn1*Δ, *rat1-1* and *xrn1*Δ/*rat1-1* mutants in Chapter II. It is difficult to predict capping defects from these genetic interactions as Rat1p is also involved in termination (see Chapter II). We propose to determine the extent of co-transcriptional capping by measuring the distance in how far Pol II has elongated when capping occurs (see Chapter III).

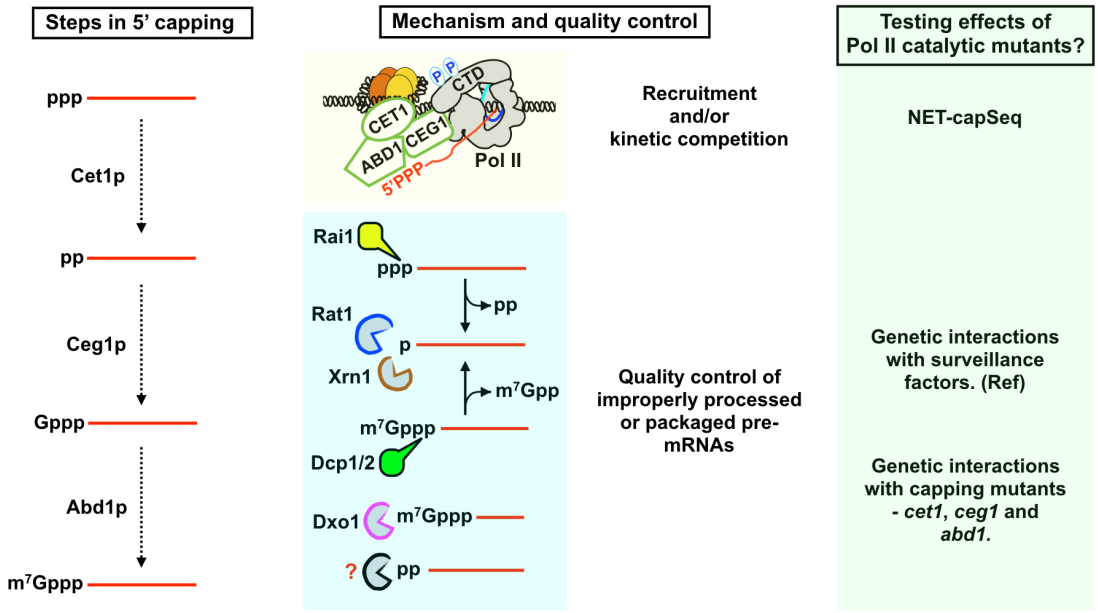


Figure 1-5. pre-mRNA capping and quality control. (Left panel) Stepwise capping mechanism of pre-mRNA and (middle panel) associated quality control pathway to degrade improperly capped message. (Right panel) Experiments to determine *in vivo* capping defect in Pol II mutants (see Chapter III for further discussion)

Co-transcriptional splicing

Among all the studied co-transcriptional processing events, perhaps, the greatest amount of evidence has been reported for the coupling of splicing to Pol II elongation. Due to paucity of introns splicing is less common in yeast than humans, where more than 90% of genes are alternatively spliced and abnormality in splicing can lead to diseases (Wang, Sandberg et al. 2008, Chabot and Shkreta 2016). Spliceosome complexes assemble on each intron in a step-wise manner to accomplish splicing. Recent work has implicated Pol II CTD differential phosphorylation in splicing factor recruitment and alternative splicing, suggesting that Pol II may directly interact with components of the spliceosome (de la Mata and Kornblihtt 2006, David, Boyne et al. 2011, Gu, Eick et al. 2013). Although it is well-established that the majority of intron splicing happens co-transcriptionally, how splicing is coupled to, or sensitive to Pol II elongation rate is unclear. Some evidence supports direct coupling, wherein a splicing 'check-point' exists and splicing requires completion before Pol II can elongate further. Other evidence suggests 'kinetic coupling', which predicts that splicing and transcription rates should match in a way such that splicing catalysis can occur within the timeframe of elongation, but alteration in elongation rate may perturb this**.** An earlier study, using a splicing reporter yeast, showed that Pol II accumulates at the 3′ splice site and defective splicing led to transcription defects in the introns, suggesting crosstalk between splicing and elongation, *i.e.* a 'check-point' might exist (Alexander, Innocente et al. 2010, Chathoth, Barrass et al. 2014). Two recent studies, using nascent elongating transcript sequencing (NET-seq) in mammalian cells, have shown that Pol II has propensity to

pause at 3′ and 5′ splice sites (Mayer, di Iulio et al. 2015, Nojima, Gomes et al. 2015). However, Pol II pausing at introns might support either model. For example, Pol II may have evolved to pause or slow down at introns to better couple two processes 'kinetically' or, conversely, the splicing process can itself act as a 'check-point' that induces Pol II pausing.

The kinetic coupling of co-transcriptional splicing is supported by previous studies on splicing in Pol II catalytic mutants, with the prediction that elongation rate can influence where the spliceosome components and splicing regulators bind on nascent RNAs relative to the position of Pol II. According to this prediction, faster elongation will lead to an enhanced competition between upstream and downstream splice sites for alternative splicing. Conversely, slow Pol II mutant will enhance the time duration between appearance upstream and downstream splice sites. Thus, slow mutants will decrease competition and increase the chances of recognition of upstream alternative splice sites. Indeed, a slow Pol II mutant in human cells inhibits exon skipping by favoring suboptimal upstream splice site (de la Mata, Alonso et al. 2003). Further, consistent with the kinetic predictions, a slow Pol II mutant globally increased intron removal in *Drosophila* cells, suggesting increased efficiency of splicing for inefficient introns due to longer duration of Pol II in regions with relevant signals (Khodor, Rodriguez et al. 2011). Studies from our lab, using LOF and GOF Pol II mutants in yeast, have shown that slower catalytic Pol II mutants generally favor splicing by decreasing intron retention, while faster elongation leads to increased intron retention (Braberg, Jin et al. 2013). A recent study in yeast showed that Pol II GOF mutant *rpb1*

E1103G shifted Pol II position further downstream on the gene compared to WT at the time of nascent pre-mRNA splicing, suggesting that *rpb1 E1103G* Pol II travels further while splicing is happening, which fits the model of kinetic coupling of events assuming that *rpb1 E1103G* is indeed a fast elongator *in vivo* as *in vitro* (Carrillo Oesterreich, Herzel et al. 2016). However, using analogous mutants in human cells, the Bentley lab has shown that slow and fast elongation often both increase or decrease inclusion of particular exons or retained introns, suggesting that an optimal elongation rate might required for efficient splicing, or that there are confounding secondary effects (Fong, Kim et al. 2014). In this regard, it is notable that use of Pol II mutants to test the kinetic coupling model cannot rule out the possibility of delayed splicing kinetics as secondary effects in catalytically fast Pol II mutants. For example, some other factor (such as recruitment of splicing components) can be impaired in fast mutants leading to slow splicing kinetics. Further studies are needed to examine the exact length of splicing intermediates associated with elongating Pol II in an experimental step where we can rule out the possibility of delayed splicing kinetics and determine the true effect of Pol II elongation rate on splicing efficiency (see discussion and future direction in Chapter III).

3′ end formation/termination

After complete synthesis of a transcript, transcription terminates in a way that is functionally connected to 3′ end formation. Productive 3′ end formation is required for proper termination, and 3′ end formation/termination both are critical for transcriptional directionality and controlling cryptic transcription. Failing to terminate can lead to Pol II running over the neighboring gene, leading to detrimental consequences (Greger and

Proudfoot 1998). After decades of research, the mechanism of Pol II transcription termination is still a matter of debate. Coding RNA transcription termination can possibly occur through two non-mutually exclusive pathways; these are so-called 'torpedo' and 'allosteric' modes of termination. According to the 'torpedo' model, once the Pol II counters a poly(A) signal, the nascent RNA is endonucleolytically cleaved for further 3′ end formation (polyadenylation), generating a 5′-monophoshate end on the nascent RNA still attached to Pol II. Subsequently, the 5′ to 3′ exonuclease Rat1p with its activating partner Rai1p attacks the unprotected (non-capped 5′ triphosphate) RNA, degrading the RNA and removing it from Pol II, allowing Pol II to disengage from the template. The 'allosteric' model describes that after the Pol II encounters a poly(A) signal, termination occurs through a physical change in the Pol II that disengage it from the template. As stated earlier, beside mRNAs Pol II also transcribes non-coding RNAs. Non-coding transcripts can overlap and interfere with coding genes, thus efficient termination is required. Non-coding RNA transcription is terminated by the Nrd1/Nab3/Sen1(NNS) pathway in yeast (Reviewed in (Porrua and Libri 2015)). The NNS proteins are recruited by Ser5 phosphorylation of Pol II CTD at the early stage of transcription. Nrd1 and Nab3 are RNA binding proteins, which bind to target nascent RNAs with appropriate sequence motifs and elongation complexes are dissociated from the template by the Sen1 helicase. NNS termination is generally followed by degradation of the transcript by the nuclear exosome complex.

The co-transcriptional nature of 3′ end formation is supported by several observations. First, most coding RNA transcription termination is poly(A) signal

dependent, indicating elongating Pol II remains still engaged to the template while polyadenylation occurs (Zaret and Sherman 1982, Logan, Falck-Pedersen et al. 1987). Second, polyadenylation and cleavage factors associate with the 3′ end of the genes, suggesting they are recruited co-transcriptionally (Kim, Erickson et al. 2010, Mayer, Lidschreiber et al. 2010). Third, if co-transcriptional polyadenylation/cleavage is kinetically coupled to elongating Pol II, then it is expected that slow elongating Pol II may have increased time to allow processing at upstream poly-A sites, with termination closer to such sites. Indeed, proximal poly(A) sites are favored by a slow elongating Pol II in *Drosophila* compared to WT and, in yeast, slow elongating Pol II enhances processing of crippled poly(A) site (Kaplan, Holland et al. 2005, Pinto, Henriques et al. 2011).

Kinetic coupling of termination to elongation has been shown for both non-coding and coding RNA transcription. Using a LOF and a GOF catalytic mutant, it has been shown in yeast that Sen1p helicase mediates termination of non-coding RNAs through competition with elongating Pol II (Hazelbaker, Marquardt et al. 2013). A GOF catalytic mutant, *rpb1-E1103G*, assumed to have faster in vivo elongation rate, indeed travels further on a template prior to termination, while a LOF catalytic mutant, *rpb1-N488D*, terminates after a shorter distance travelled. Further, termination and growth defects of *sen1* mutant cells are partially suppressed by a LOF catalytic mutant *rpb1-N488D* and enhanced by a Pol II GOF catalytic mutant *rpb1-E1103G*, consistent with kinetic competition between termination and elongation. Similarly, for coding gene termination, kinetic competition is evident for ‘torpedo’ model of termination in human

cells, where inactivation of the 5′ exonuclease encoding *XRN2* (yeast *RAT1* homolog) leads to longer and shorter run-off for a faster (POLR2A E1126G) and a slower Pol II mutant (POLR2A R739H), respectively (Fong, Brannan et al. 2015). These observations are consistent with genetic-interaction results in yeast, that LOF/slow catalytic mutants partially recue lethality of *rat1-1* ts- mutant (defective termination) at the non-permissive temperature, while GOF/fast catalytic mutants are synthetic sick with *rat1-1* mutant at the permissive temperature. Taken together these results support a co-transcriptional kinetic coupling for termination, wherein slower elongation provide a longer temporal 'window of opportunity' for termination in termination defective cells (*rat1-1*), while faster elongation makes termination ineffective by running off faster and/or a temporally reduced 'window of opportunity'. Interestingly, a recent study on the 'allosteric' mode of termination showed that, in vitro, the elongation complex gets disassembled when encounters a poly(A) signal (Zhang, Rigo et al. 2015). This disassembly process does not require transcript cleavage, but requires specific Pol II conformation that can be blocked by α-amanitin. This observation led to an intriguing possibility that Rpb1 TL might be involved in the so-called allosteric change, as the TL is the target of α-amanitin (Brueckner and Cramer 2008, Kaplan, Larsson et al. 2008). Possible involvement of TL in the proposed allosteric change for termination complicates the use of TL mutants to infer kinetic completion during 'torpedo' mode of termination, as the termination defects observed in GOF mutant (*rpb1 E1103G*) may derive from 'allosteric' changes as well. Despite the debate over these two modes of termination it is possible that these mutually non-exclusive pathways may coexist in vivo.

Coordination of mRNA synthesis and mRNA decay

Cellular mRNA levels are governed by variation in synthesis rate and degradation rate. According to this view, any perturbation in transcription or mRNA degradation should result in a significant change in cellular mRNA level. However, in contrast, several recent works suggest that cells may have an unidentified mechanism to buffer the mRNA levels following impaired mRNA synthesis or decay. This 'feedback' of gene expression control between mRNA synthesis and decay rates is proposed to maintain a certain concentration of mRNA in the cells even if synthesis/degradation is impaired. An early connection between mRNA synthesis and decay was established by the Choder group with experimental evidences that Pol II subunit Rpb4p mediates decay of a specific set of mRNAs and impaired mRNA synthesis had very little impact on these mRNA levels (Goler-Baron, Selitrennik et al. 2008). Further studies from the Choder and Cramer labs have suggested that mRNA buffering could be a global phenomenon and that the coupling of synthesis and degradation may have evolved coordinately in yeast (Sun, Schwalb et al. 2012, Haimovich, Choder et al. 2013, Haimovich, Medina et al. 2013, Sun, Schwalb et al. 2013). The exact mechanism of this coupling is still under investigation; although, several putative factors have been identified and proposed to maintain the coordination between mRNA synthesis and degradation. The Choder group proposed that the Pol II subunits Rpb4/7, which form a heterodimer, shuttle between nucleus and cytoplasm and stimulate mRNA decay process to buffer impaired synthesis (Goler-Baron, Selitrennik et al. 2008, Harel-Sharvit, Eldad et al. 2010). In contrast, studies from Cramer group, using metabolic RNA labeling and

comparative Dynamic Transcriptome Analysis (cDTA), have suggested that Rpb4p mainly functions in transcription (Schulz, Pirkl et al. 2014). Comparing global synthesis and degradation rates of 46 mutant factors in mRNA metabolism/degradation pathways, the Cramer group showed that deletion of 5′ to 3′ exonuclease Xrn1p abolishes the mRNA buffering and proposed that Xrn1p maintain the feedback between mRNA synthesis and degradation (Sun, Schwalb et al. 2013). However, mutants tested in this study or earlier studies, may affect cellular growth rates and recent studies suggest that growth rate can also control mRNA decay and overall mRNA abundance (Neymotin, Ettorre et al. 2015, Garcia-Martinez, Delgado-Ramos et al. 2016). Thus, it is difficult deconvolute between two possibilities whether the proposed mRNA feedback is controlled by certain factors or changes occur through cellular growth rate. As our Pol II catalytic mutants have growth defects, I test the effects of growth on mRNA decay rate of a reporter, which I present in Chapter II.

CHAPTER II

WIDE-RANGING AND UNEXPECTED CONSEQUENCES OF POL II CATALYTIC ACTIVITY IN VIVO[1]

Disclaimer for Chapter II

Chapter II is reprint of a publication, which I am the lead author on. I performed all of the work presented in this chapter, expect for growth assays presented in **Figure 2-12** and **2-14**, which are done by Dr. Kaplan and Thomas Snavely. Chenxi Qiu helped me in microscopy experiments presented in **Figure 2-10**. Summary section of this chapter is the abstract of the publication, rest are as in publication.

---

[1]Reprinted with permission from "Wide-ranging and unexpected consequences of altered Pol II catalytic activity in vivo" by Malik et al. 2017. Nucleic Acids Research, Vol. 45, 4431–4451, Copyright ©The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.

Summary

Here we employ a set of RNA Polymerase II (Pol II) activity mutants to determine the consequences of increased or decreased Pol II catalysis on gene expression in *Saccharomyces cerevisiae*. We find that alteration of Pol II catalytic rate, either fast or slow, leads to decreased Pol II occupancy and apparent reduction in elongation rate *in vivo*. However, we also find that determination of elongation rate *in vivo* by chromatin IP can be confounded by the kinetics and conditions of transcriptional shutoff in the assay. We identify promoter and template-specific effects on severity of gene expression defects for both fast and slow Pol II mutants. We show that mRNA half-lives for a reporter gene are increased in both fast and slow Pol II mutant strains and the magnitude of half-life changes correlate both with mutants' growth and reporter expression defects. Finally, we tested a model that altered Pol II activity sensitizes cells to nucleotide depletion. In contrast to model predictions, mutated Pol II retains sensitivity to altered nucleotide levels. Our experiments establish a framework for understanding the diversity of transcription defects derived from altered Pol II activity mutants, essential for their use as probes of transcription mechanisms.

Introduction

Gene transcription by RNA polymerase II (Pol II) is an essential process and involves three distinct phases: initiation, elongation and termination. Transcription elongation proceeds through an iterative cycle of substrate selection, catalysis of phosphodiester bond formation, and enzyme translocation [Reviewed in (Kaplan 2013,

Svetlov and Nudler 2013)]. Pausing, backtracking, and arrest of Pol II can occur during elongation. Pol II elongation factors are proposed to promote Pol II elongation by modulating these processes or otherwise enabling Pol II to overcome obstacles. Coordinated with elongating Pol II, several co-transcriptional events occur to control the fate of nascent RNAs and ensure proper gene expression [Reviewed in (Bentley 2014)]. Thus, it is likely that perturbation of Pol II elongation will have multi-faceted effects *in vivo*. To understand the complexity of Pol II activity-mediated control of gene expression it is necessary to understand how alteration of Pol II catalytic activity relates to specific gene expression defects.

Studies from our lab and others have identified several Pol II catalytic mutants that can alter elongation rate *in vitro* (Malagon, Kireeva et al. 2006, Kaplan, Larsson et al. 2008, Kaplan, Jin et al. 2012). Based on these mutants' ability to increase or decrease transcription elongation rate relative to wild type (WT), we term them "gain of function" (GOF) or "loss of function" (LOF) mutants, respectively (Kaplan, Jin et al. 2012). These classes of mutant confer distinct gene expression profiles, genetic interaction profiles, splicing, and transcription initiation defects *in vivo*, consistent with their having distinct biochemical defects *in vitro* (Kaplan, Jin et al. 2012, Braberg, Jin et al. 2013, Jin and Kaplan 2014). Most of these mutants reside in a highly conserved, mobile sub-domain of the largest Pol II subunit Rpb1, known as the trigger loop (TL) (Malagon, Kireeva et al. 2006, Kaplan, Larsson et al. 2008, Kaplan, Jin et al. 2012, Qiu, Erinne et al. 2016). The TL is a component of the Pol II catalytic center, and can directly interact with incoming NTPs, undergoing conformational changes to promote rapid catalysis (Wang, Bushnell

et al. 2006, Kaplan, Larsson et al. 2008, Kireeva, Nedialkov et al. 2008). TL mutants have been shown to affect a number of Pol II biochemical properties including catalysis, substrate selection and transcription fidelity (Kaplan, Larsson et al. 2008, Kireeva, Nedialkov et al. 2008). In addition, TL mutants have been also shown to affect Pol II translocation, pausing and intrinsic cleavage properties (Larson, Zhou et al. 2012, Dangkulwanich, Ishibashi et al. 2013, Cabart, Jin et al. 2014).

The rate of transcription elongation has likely evolved to facilitate and enhance the efficiency of pre-mRNA processing and maturation [Reviewed in (Bentley 2014)]. Maturation of pre-mRNA requires addition of a 7-methyl guanosine cap at the 5′-end of the transcript, splicing of introns, and addition of a poly(A) tail to the 3′-end of the transcript. Further, the pre-mRNA is uniquely packaged with protein components into a mature mRNA granule, which facilitates export and efficient translation. Impaired processing leads to degradation of pre-mRNAs by nuclear surveillance pathways. Mechanistic coupling of transcription and pre-mRNA processing is achieved through recruitment of factors by C-terminal domain (CTD) and by kinetic competition between transcription and processing [Reviewed in (Bentley 2014)]. Using Pol II catalytic mutants, it has been shown that kinetic competition functions in the efficiency of pre-mRNA splicing (de la Mata, Alonso et al. 2003, Braberg, Jin et al. 2013, Fong, Kim et al. 2014, Carrillo Oesterreich, Herzel et al. 2016). At least one Pol II catalytic mutant has been reported to be defective in 5′-capping, leading to the degradation of transcript by 5′ to 3′ nuclear exonuclease (Jimeno-Gonzalez, Haaning et al. 2010). Additionally, a number of findings suggest kinetic competition between transcription termination and

elongation in both yeast and human cells (Hazelbaker, Marquardt et al. 2013, Fong, Brannan et al. 2015). Recently, a competition-independent pathway has been proposed for termination that occurs through a conformational change to Pol II (Zhang, Rigo et al. 2015). Furthermore, it has been proposed that overall RNA synthesis rate is connected to mRNA decay rate through feedback between synthesis and degradation (Haimovich, Choder et al. 2013, Haimovich, Medina et al. 2013, Sun, Schwalb et al. 2013, Braun, Vaga et al. 2014, Braun and Young 2014). Growth rate is also proposed to control mRNA decay and overall mRNA abundance, but linkages between RNA synthesis rate, growth, and mRNA decay have not been determined at the molecular level (Garcia-Martinez, Delgado-Ramos et al. 2016, Neymotin, Ettorre et al. 2016).

In order to study Pol II elongation *in vivo*, a number of methods have been implemented that either directly measure apparent Pol II elongation rate or determine indirect consequences of elongation rate (Hirayoshi and Lis 1999, Garcia-Martinez, Aranda et al. 2004, Mason and Struhl 2005, Morillo-Huesca, Vanti et al. 2006) (for a of list methods and estimated *in vivo* Pol II elongation rates see here (Ardehali and Lis 2009)). Two methods are generally used to study elongation in yeast, one to study elongation properties and the other to genetically implicate factors in elongation control. The first utilizes chromatin immunoprecipitation (ChIP) to determine Pol II occupancy across a long galactose-inducible gene, *GAL1p::YLR454w*, either in steady state or after transcription shutoff by addition of glucose (Mason and Struhl 2005). Apparent processivity is inferred from comparison of steady state Pol II occupancy for wild type and transcription mutants, while kinetics of the 'last wave' of Pol II leaving the template

can be used to determine the apparent elongation rate. Indeed, Pol II catalytic mutants and several factors mutants have shown altered apparent *in vivo* elongation rate and apparent processivity defects in this assay (**Table 2-1**). Interpretation of apparent elongation rate differences based on transcriptional shutoff makes assumptions that signaling and kinetics of the shutoff are identical between WT and mutant strains. Similarly, in a second widely used approach, genetic detection of elongation defects through use of nucleotide-depleting drugs makes assumptions that drug effects are identical between WT and mutant strains, and this issue is discussed below.

Nucleotide-depleting drugs, such as mycophenolic acid (MPA), which limits cellular GTP levels by inhibiting IMPDH activity, are assumed to elicit transcription elongation defects by enhancing pausing, arrest, or backtracking due to limitation in substrate levels (Sweeney 1977, Archambault, Lacroute et al. 1992, Powell and Reines 1996, Reines 2003, Kaplan 2013). Growth sensitivity to MPA for Pol II or presumptive elongation factor mutants has been widely interpreted as a synergistic effect between MPA treatment and impaired elongation due to the mutant. The notion that limiting nucleotide is the major determinant of drug phenotypes was further strengthened by the observation that guanine supplementation suppresses sensitivity to the drug, along with the observation of elongation defects due to drug treatment (Desmoucelles, Pinson et al. 2002, Mason and Struhl 2005). However, it has been shown subsequently that many MPA-sensitive transcription mutants are defective for upregulation of the MPA-resistant IMPDH activity encoded by the *IMD2* gene (Shaw and Reines 2000).

Table 2-1. List of mutants/conditions that used glucose shut-off experiment to determine apparent *in vivo* elongation rate on *GAL1*p::*YLR454w* reporter.

Mutant/Condition tested	Apparent *In vivo* elongation	Reference
rpb2-10 *hpr1*, *thp2*, *mft1*, *cdc73*, *rtf1*, *spt4*, *ctk1*, *ctk2* and *ppr2*.	*rpb2-10* apparent slower than WT no detectable effect	(Mason and Struhl 2005)
asf1	no detectable effect	(Schwabish and Struhl 2006)
set2, pob3 and set2/pob3	no detectable effect	(Biswas, Dutta-Biswas et al. 2006)
swi2	no detectable effect	(Schwabish and Struhl 2007)
gcn5	*gcn5* apparent slower than WT	(Govind, Zhang et al. 2007)
esa1 *gcn5* *esa1/gcn5*	*esa1* apparent slower than WT *gcn5* no detectable effect *esa1*/ *gcn5* apparent slower than WT	(Ginsburg, Govind et al. 2009)
chd1 *spt5-242*	*chd1* apparent slower than WT *spt5-242* apparent slower than WT	(Quan and Hartzog 2010)
rpb1 N488D	*rpb1* N488D apparent slower than WT	(Jimeno-Gonzalez, Haaning et al. 2010)
dhh1 *ccr4* *not4*	*dhh1* apparent slower than WT *ccr4* apparent slower than WT *not4* apparent slower than WT	(Kruk, Dutta et al. 2011)
Temperature	apparent elongation rate increases with temperature	(Miguel, Monton et al. 2013)
pfd1 *dst1*	*pfd1* apparent slower than WT *dst1* no detectable effect	(Millan-Zambrano, Rodriguez-Gil et al. 2013)
rpb1 E1103G	*rpb1* E1103G apparent faster than WT	(Hazelbaker, Marquardt et al. 2013)
rat1-1	*rat1-1* apparent faster than WT	(Jimeno-Gonzalez, Schmid et al. 2014)

Table 2-1 Continued

Mutant/Condition tested	Apparent *In vivo* elongation	Reference
rpb2 K864G/K865G/Δ866-871	no detectable effect	(Barnes, Calero et al. 2015)

Table 2-2. Summary of Pol II mutants' phenotypes

Experiment/Phenotype	LOF/Slow mutant phenotype	GOF/Fast mutant phenotype
Steady state Pol II occupancy (over *GAL1*p::*YLR454w*)	Decreased overall Pol II occupancy	Decreased overall Pol II occupancy
Apparent Pol II processivity defect at the 3′ end of *GAL1*p::*YLR454w*	Apparent 3′ end processivity defect in *rpb1* H1085Y	Apparent 3′ end processivity defect in *rpb1* G1097D
Steady state Pol II occupancy in glucose vs galactose media	Subtle increase in overall Pol II occupancy in galactose for *rpb1* H1085Y	Decrease in overall Pol II occupancy in galactose for *rpb1* G1097D
In vivo gene expression (Reporters- *GAL1p*::*YLR454w*, *TEF1p*::*YLR454w*, *TEF1* and *GAL1*)	Gene expression decreases; most robust effect for *rpb1* H1085Y	Gene expression decreases; most robust effect for *rpb1* G1097D
*GAL1*p induction kinetics	Induction is delayed	Induction is delayed
Genetic interaction with pre-mRNA processing factors	Suppresion of Ts- phenotype of *rat1-1* and *xrn1Δ*/*rat1-1* mutants	Synthetic sick interactions with *xrn1Δ* and *rat1-1*
mRNA decay rate of *GAL1p*::*YLR454w*	Decay rate decreases	Decay rate decreases
In vivo elongation rate over *GAL1*p::*YLR454w* template	Apparent elongation rate slower than WT on *GAL1*p::*YLR454w* (with caveats noted in main text)	Apparent elongation rate slower than WT on *GAL1*p::*YLR454w* (with caveats noted in main text, especially for G1097D)
snR33 termination window	Shorter termination window than WT	Longer termination window than WT

Table 2-2. Continued

Experiment/Phenotype	LOF/Slow mutant phenotype	GOF/Fast mutant phenotype
Ability of GTP-sensing in absence of endogenous *IMD2*	Retain ability to sense GTP; MPA sensitivity is suppressed by addition of guanine	Retain ability to sense GTP; MPA sensitivity poorly suppressed by addition of guanine
Response to GTP depletion in absence of endogenous *IMD2*	Constitutively use downstream functional 'A' start site at *IMD2* promoter; sense GTP depletion and can further shift TSS downstream to functional 'A' site upon MPA treatment	Sense GTP depletion but cannot shift TSS downstream to functional 'A' site; instead use novel TSSs that produce non-functional *IMD2* CUT

Under GTP-replete conditions, an upstream TSS at *IMD2* is used (where transcripts initiate using GTP), generating an unstable transcript that terminates within the *IMD2* promoter (Jenks, O'Rourke et al. 2008, Kuehner and Brow 2008). Upon GTP starvation elicited by MPA treatment, a downstream TSS is utilized (where transcripts initiate using ATP), allowing for expression of functional *IMD2*. Regulation of *IMD2* expression, presumably mediated by this TSS switch, is defective in a wide range of transcription elongation mutants, as well as for mutants that alter Pol II catalytic activity (Desmoucelles, Pinson et al. 2002, Riles, Shaw et al. 2004, Kaplan, Jin et al. 2012). The mechanism for defects in *IMD2* TSS switching has not been determined for any mutant. An attractive model for Pol II mutant effects at *IMD2* is that Pol II catalytic defects mimic changes to NTP substrate levels, and suggest that the Pol II active site may directly communicate GTP levels to the *IMD2* promoter through initiation efficiency at different transcription start sites.

Here, we present a detailed molecular analysis of how alteration to Pol II catalysis rate through the use of specific *rpb1* alleles, conferring slower or faster catalysis than wild type *in vitro*, affects several facets of transcription and gene expression *in vivo* (a summary of observed mutant effects is in **Table 2-2**). Using the widely used *GAL1*p::*YLR454w* reporter, we show that both slow and fast Pol II catalytic mutants decrease Pol II occupancy and reporter gene expression *in vivo*. We find that the reporter gene expression defects in slow or fast Pol II mutants negatively correlate with increased reporter mRNA half-lives. Pol II catalytic mutants show genetic interaction with pre-mRNA processing factors, and we present evidence for an mRNA processing defect in fast catalytic mutants. Finally, we have critically evaluated the two major widely used systems for studying transcription elongation in yeast – chromatin IP of the 'last wave' of Pol II upon transcription inhibition, and response to GTP starvation. For both assays we uncover underlying biological complexities that differentially affect WT and presumptive elongation mutant cells, confounding interpretation of these assays by simple models. Our results provide a useful framework for future utilization of Pol II catalytic mutants to probe biological processes and gene expression mechanisms.

Materials and Methods

Yeast strains, plasmid, media and growth

Yeast media are prepared following standard (Amberg, Burke et al. 2005) and previously described protocols (Kaplan, Jin et al. 2012). Yeast extract (1% w/v; BD), peptone (2% w/v; BD) and 2% bacto-agar (BD), supplemented with adenine (0.15mM) and tryptophan (0.4mM) (Sigma-Aldrich) comprised YP solid medium. YPD plates

contained dextrose (2% w/v, VWR), YPRaf plates contained raffinose (2% w/v, Amresco), YP-Raf/Gal plates contained raffinose (2% w/v) plus galactose (1% w/v, Amresco) and YPGal plates contained galactose (2% w/v) as carbon sources. YPRaf, YPGal and YPRaf/Gal plates also contained Antimycin A (1 mg/ml; Sigma-Aldrich). Minimal media plates were prepared with synthetic complete (SC) or 'Hopkins mix' with appropriate amino acid(s) dropped out as described in (Amberg, Burke et al. 2005), with slight modifications as described in (Kaplan, Jin et al. 2012). For studies with mycophenolic acid (MPA, Sigma-Aldrich), a stock solution (10 mg/ml, in 100% ethanol) of MPA was added to solid or liquid media to achieve desired concentration. NaOH, HCl and Guanine were added to solid media to achieve desired concentration as indicated. Liquid YPD, YPRaf, YPGal and YPRafGal media are prepared with yeast extract (1% w/v), peptone (2% w/v) and 2% (w/v) carbon source (dextrose, raffinose, galactose or raffinose plus galactose), with no supplementary adenine, tryptophan or Antimycin A. Antimycin A is a standard additive to yeast plates when growth on fermentable carbon sources such as galactose or raffinose is being assayed. Prevention of respiration enforces growth by fermentation. These assays can be performed in the absence of Antimycin A (which is a highly toxic substance), but they become less sensitive as the growth differenced between WT *gal10*Δ*56* strains on YPRafGal and mutants that suppress galactose toxicity on YPRafGal are lessened.

Yeast phenotyping assays were performed by serial dilution and spotting onto plates as described earlier (Kaplan, Jin et al. 2012). Doubling times for the mutants in liquid medium (YPGal) were determined using Tecan plate-reader as described earlier

(Lenstra, Benschop et al. 2011), with minor modifications. Overnight grown saturated cultures were diluted to an OD_{600} of ~ 0.1 in fresh YPGal medium and grown in triplicate at 30˚C in a 96-well plate in a Tecan Infinite F200 plate reader under continuous shaking**.** Data obtained from each plate were considered as a single biological replicate and was analyzed in Graphpad Prism using an exponential-growth fitting function.

Chromatin immunoprecipitation

All the strains used for ChIP experiments contained a C-terminal epitope tag on the Rpb3 subunit of RNA Pol II (RPB3::3XFLAG::KANMX; see strain list). ChIP experiments were performed as described previously (Kaplan, Holland et al. 2005). Briefly, 100 ml of mid-log phase cells (~$1x10^7$ cells/ml) were cross-linked with 1% formaldehyde (final) for 20 min, and then quenched with 15 ml of 2.5 M glycine for 5 min. Cross-linked cells were washed twice with cold 1X TBS buffer at 3500 rpm for 3-4 min at 4˚C using a JS-5.3 rotor (Avanti J-26 XP centrifuge, Beckman-Coulter) and were disrupted by bead beating with glass beads in lysis buffer (0.1 M Tris pH 8.0, glycerol 20%, 1 mM PMSF). Cross-linked cell lysates were subjected to a low speed spin (1500 rpm, 1 min at 4˚C) to remove cell-debris, followed by centrifugation of chromatin pellets, subsequent washing of pellets (twice) with 1 ml FA buffer (50 mM HEPES-KOH pH 7.5, 300 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 0.01% sodium deoxycholate, 0.1% SDS and 1 mM PMSF) at 14000 rpm for 20 min at 4˚C using F45-30-11 rotor (Centrifuge 5417R, Eppendorf). Chromatin pellets were resuspended in 1 ml of FA buffer and sonicated at 4˚C using a Diagenode Bioruptor (45 cycles – 3 x 15

cycles; 30 sec ON/ 45 sec OFF) to generate ~300-500 bp chromatin fragments (verified on 1% agarose gel). Approximately 100 µl sonicated chromatin was used for each immunoprecipitation (IP) with anti-FLAG antibody (FLAG M2 magnetic beads, Sigma-Aldrich). Surfactant and detergent composition in buffers were changed according to manufacturer's recommendation for compatibility with M2 FLAG antibody, and all buffers contained 1 mM PMSF. For Pol II occupancy determination, the amount of chromatin used for WT or mutant IPs was normalized by starting cell number and chromatin concentration (estimated by spectrophotometer and agarose gel-lane profile). Magnetically-captured FLAG beads were washed twice with FA buffer, once with FA buffer with 500 mM NaCl, once with wash buffer (10 mM Tris-HCl pH 8.0, 0.25 M LiCl, 0.5% NP-40, 1 mM EDTA) and once with TE. Immunoprecipitated chromatin was eluted by two-step elution at 65˚C, with 100 ul and 150 ul elution buffers (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% SDS) for 15 min and 30 min, respectively. Both eluates were pooled and incubated at 65˚C (> 6 hrs) for cross-linking reversal. 10 ul of sonicated chromatin (10% of material added to IP) plus 240 ul of elution buffer were treated identically and served as IP input control. Input or immunoprecipitated DNA was purified by standard phenol-chloroform extraction and ethanol precipitated in presence of pellet paint (MilliporeSigma) or glycoblue (ThermoFisher). Immunoprecipitated DNA and 10% of corresponding input DNA (1:10 diluted in nuclease free water) were used for qPCR with SsoAdvanced or SsoAdvanced Universal SYBR Green supermix (Bio-Rad) using CFX 96 (Bio-Rad). Fold enrichment for target amplicon over non-transcribed region was determined by the $\Delta\Delta C_T$ method (Livak and Schmittgen 2001).

For *in vivo* elongation assays, WT or mutant cells were grown in YPGal to mid-log phase and a pre- glucose shutoff sample was taken for the 0 minute time point. Then glucose (4% final) was added to inhibit transcription and aliquots were removed after 2, 4, 6 and 8 minutes (optional longer time points were taken for some strains). Alternatively, after isolation of 0 min sample as above, remainder of culture was centrifuged and washed with SC medium lacking carbon source as described earlier (Hazelbaker, Marquardt et al. 2013), then inoculated in YPD (4% glucose) media or SC media lacking carbon source to isolate glucose shutoff samples or galactose-depletion samples, respectively, at indicated time points. Formaldehyde cross-linking, chromatin preparation and subsequent steps were performed as described above.

RNA isolation, Northern blotting

For gene expression analysis, RNAs were isolated from mid-log phase cells grown in appropriate medium as described in the main text. Post shutoff samples were collected by a quick centrifugation for 1 min and immediate freezing of the cell pellets by placing at -80C. Centrifugation time was included while calculating shutoff time. Total RNA was purified using hot phenol-chloroform method as described previously (Schmitt, Brown et al. 1990).

Northern blotting was performed essentially as described in the manual for GeneScreen hybridization membranes (Perkin-Elmer) with minor modifications as described earlier (Kaplan, Jin et al. 2012). In brief, 20 µg of total RNA, treated with Glyoxal sample-load dye (Ambion), was separated on 1% agarose gel and transferred to membrane by capillary blotting. Pre-hybridization solution contained 50% formamide,

5X Denhardt's solution, 10% Dextran sulfate, 1 M NaCl, 50 mM Tris-HCl pH 7.5, 0.1% sodium pyrophosphate, 0.1% SDS and 500 μg/ml denatured salmon sperm DNA. PCR generated DNA double-stranded probes for northern blots were radiolabeled with ^{32}P-dATP using the Decaprime II kit (Ambion) according to manufacturer's instructions. After overnight hybridization of the blot at 42˚C, washes were done twice each in 2X SSC for 15 minutes at 42˚C, in 5X SSC with 0.5% SDS for 30 minutes at 65˚C, and in 0.2X SSC for 30 minutes at room temperature. Blots were visualized by phosphorimaging (Bio-Rad or GE Healthcare) and quantified using Quantity One (Bio-Rad).

Northern blotting for mapping the termination window of the pre-processed *snR33* transcript was performed essentially as described in (Marquardt, Hazelbaker et al. 2011, Hazelbaker, Marquardt et al. 2013) with minor modifications. Briefly, 5-8 μg of RNA were separated on 6% polyacrylamide-7M urea gel. RNAs were transferred from polyacrylamide gel to a membrane (GeneScreen Plus, PerkinElmer) with a Bio-Rad Trans-Blot apparatus at a setting of 45W for 1.5 hrs. RNAs were cross-linked to the membrane by UV light. Pre-hybridization of the membrane, probe synthesis and membranes-hybridization were performed as described in (Marquardt, Hazelbaker et al. 2011). Membranes were washed twice with low stringency wash buffer (0.1x SSC, 0.1% SDS) and visualized by phosphorimaging (Bio-Rad or GE Healthcare). Lane traces were determined for each sample using ImageQuant (GE Healthcare).

Primer extension

For primer extension (PE) analysis, RNA was isolated from mid-log phase cells grown in appropriate media, optionally treated with desired concentration of MPA for indicated time periods. Primer extension analysis was done essentially as described earlier (Ranish and Hahn 1991) with modification described in (Kaplan, Jin et al. 2012). Briefly, 30 µg of total RNA was annealed with ^{32}P end-labeled oligo. M-MLV Reverse Transcriptase (Fermentas) was used for reverse transcription, in the presence of RNase Inhibitor (Fermentas). Primer extension products were ethanol precipitated overnight and separated on 8% polyacrylamide gels (19:1 acrylamide:bisacrylamide, Bio-Rad) containing 1X TBE and 7M urea. PE gels were visualized by phosphorimaging (Bio-Rad or GE Healthcare) and quantified using Image Lab software (Bio-Rad). For *IMD2* TSS annotation we have considered 'A' of the start codon (ATG) as +1, so that bases upstream start at -1, as in our previous publication (Kaplan, Jin et al. 2012).

Microscopy and image analysis

Mig1p-GFP tagged strain was made by integrating GFP C-terminal tag at the genomic locus (see strain description). Microscopy was performed as described previously (Miermont, Waharte et al. 2013), with modifications. Briefly, cells grown overnight in SC medium (2% galactose) at 30˚C, diluted in fresh SC media and grown till mid-log at 30˚C before microscopy. Perfusion chamber gasket (ThermoFisher, 4 chamber: 19 mm X 6 mm) was used for changing medium. Chamber was treated with Con A (2 mg/ml, MP Biomedicals) for 10-15 min, and then cells were injected into the chamber and allowed 10-15 min to adhere. Medium exchange from SC (2% galactose)

to SC (2% galactose) + glucose (4% final) was done by pipetting quickly, while the chamber is fixed on the microscope stage. Pre-glucose sample was considered as time 0 and glucose repression time points were taken immediately after exchange of the medium.

For Mig1p-GFP nuclear localization kinetics, microscopy was performed with an inverted epifluorescence microscope (Ti-E, Nikon, Tokyo, Japan) using a 100x objective (Plan Fluo, NA 1.40, oil immersion) and standard filter sets. Images were acquired using a cooled EMCCD camera (iXon3 897, Andor, Belfast, United Kingdom). For delayed Mig1p response (longer time points), microscopy was performed with a Nikon Ti-E microscope equipped with a CFI Plan Apo lambda DM 100x objective and a Prior Scientific Lumen 200 Illumination system. All images were acquired using NIS Element software and data analysis was done using Quantity One (Bio-Rad). Ratio of nuclear/cytoplasmic GFP intensity was calculated as [(nuclear GFP intensity – background) / (cytoplasmic GFP intensity – background)]. Obtained values for glucose repression time points were normalized to pre-glucose (time 0, t0) value. Non-responding cells were quantified similarly, except for the fact that there was no visible nuclear foci, thus a random, central area was selected and measured for fluorescence intensity. Position and area of the measurement was kept identical for all the time points for the same cell. We reasoned that a non-responding cell would not have a distinct change over time in nuclear/cytoplasmic GFP intensity, hence quantifying a random area as nucleus should not have an overall affect on the interpretation of the data.

Results

Pol II catalytic mutants decrease steady state Pol II occupancy and apparent processivity

To explore *in vivo* consequences of altered Pol II activity, we examined a subset of Pol II active site mutants in the largest subunit Rpb1. These mutants were previously characterized biochemically, molecularly, and genetically as falling into two classes (Kaplan, Larsson et al. 2008, Kaplan, Jin et al. 2012, Jin and Kaplan 2014, Qiu, Erinne et al. 2016). These two classes of Pol II mutants appear to derive from either increase ("GOF") or decrease ("LOF") in Pol II catalytic activity, as determined using *in vitro* transcription. First, we examined Pol II occupancy by chromatin IP over an inducible long reporter gene, *GAL1p*::*YLR454w* (**Figure 2-1**A and (Mason and Struhl 2005)). We observed decreased overall polymerase occupancy for both GOF and LOF mutants compared to WT (**Figure 2-1**B). Among the mutants tested, H1085Y (strong LOF) and G1097D (strong GOF) showed the most severe defects in Pol II occupancy, consistent with their severe growth defects *in vivo* (Kaplan, Jin et al. 2012). When we normalized the Pol II fold-enrichment over the 5′ end of the gene body ("1kb" amplicon) followed by normalization of mutant values to WT, catalytic mutants show selective decrease in occupancy over the 3′ end of the reporter (most obvious for G1097D) (**Figure 2-1**C). Decreased steady state Pol II occupancy at the 3' end is consistent with a processivity defect, wherein Pol II that begins synthesis is unable to complete the transcript, leading to polar effects on occupancy from 5′ to 3′. Generally, greater Pol II occupancy over 5′ ends of transcription units relative to 3′ ends are interpreted as processivity defects when observed (Mason and Struhl 2005); however, a faster elongation rate over a gene's 3' end

would lead to a similar phenotype. Because differential elongation rates over different parts of genes has been observed in mammalian cells (Danko, Hah et al. 2013, Jonkers, Kwak et al. 2014, Veloso, Kirkconnell et al. 2014), we are careful to assert that what we observe is differential apparent processivity for Pol II mutants as ChIP is unable to distinguish between the two possibilities.

Additional factors may determine Pol II occupancy levels over genes. For example, Pol II occupancy over a template may be gene class-dependent or regulon-specific, given that external perturbations such as changes in temperature or carbon source have been shown to affect transcription (Pelechano, Jimeno-Gonzalez et al. 2009, Miguel, Monton et al. 2013). Therefore, we examined Pol II occupancy over the same reporter driven by a constitutive promoter (*TEF1p*::*YLR454w*) with strains grown in glucose or galactose for the most severely defective Pol II LOF and GOF mutants (**Figure 2-1**D and E). We observed an increased Pol II occupancy for H1085Y in galactose relative to glucose, with the converse for G1097D, compared to WT (**Figure 2-1**E), although a 3′ occupancy defect was still apparent for both mutants. WT Pol II showed more subtle changes in occupancy in glucose relative to galactose. Furthermore, the G1097D 3′ end was enhanced in glucose relative to galactose, suggesting growth conditional effects on transcription elongation.

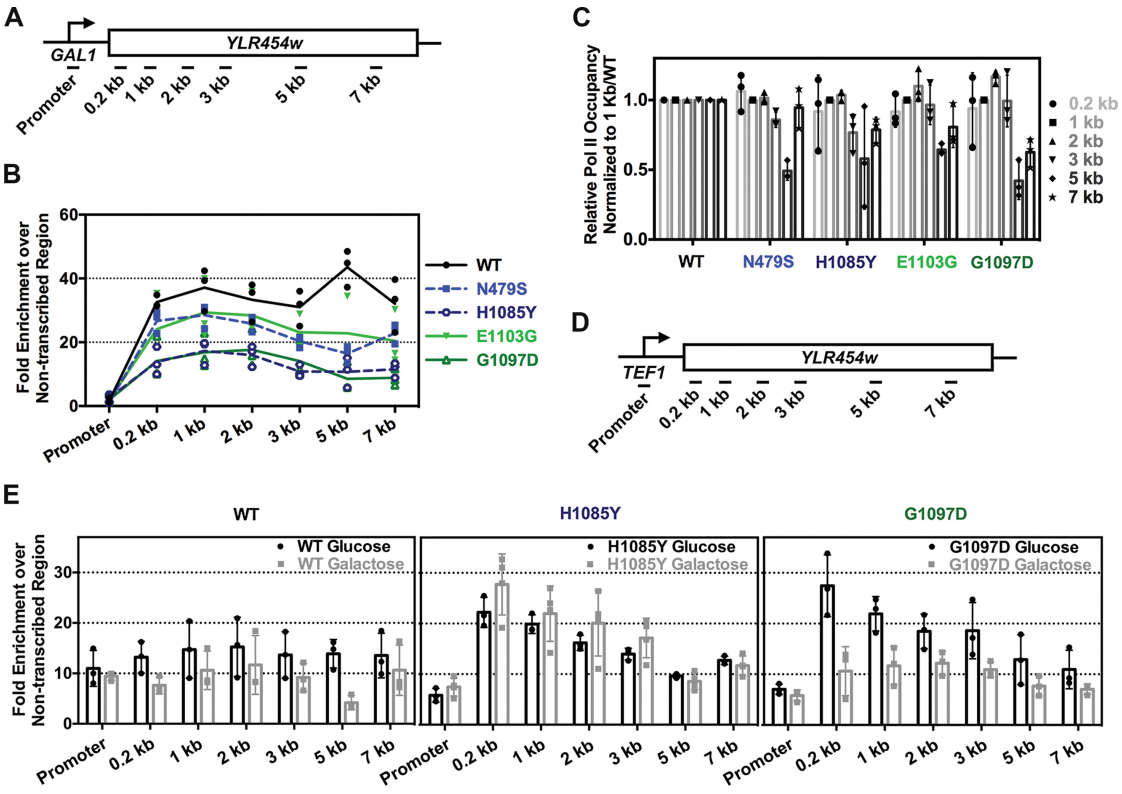


Figure 2-1. Pol II catalytic mutants decrease Pol II occupancy and show apparent processivity defects. (**A**) Schematic of the galactose inducible reporter *GAL1p*::*YLR454w* annotated with positions of PCR amplicons used for ChIP experiments. Promoter amplicon derived from primers specific to the *kanmx::GAL1p* integrated promoter. (**B**) Pol II occupancy over *GAL1p*::*YLR454w* reporter is decreased for both GOF and LOF mutants compared to WT. Line graph shows steady-state Pol II occupancy for WT (black line), GOF (green lines) and LOF (blue dashed lines) mutants under galactose induction. Green and blue color-coding is used to annotate GOF and LOF mutants, respectively. This color-coding is used throughout. (**C**) Pol II catalytic mutants show apparent 3′ processivity defects. Apparent Pol II processivity defects determined for each mutant by normalizing mutant ChIP signal to signal at 1 kb followed by normalization to WT. (**D**) Schematic of constitutively expressed reporter gene *TEF1p*::*YLR454w* annotated with positions of PCR amplicons used for ChIP experiments. Promoter amplicon derived from primers specific to the *kanmx::TEF1p* integrated promoter. (**E**) Carbon source may differentially affect Pol II mutants' occupancy. Comparison of steady-state Pol II occupancy at *TEF1p*::*YLR454w* for WT and Pol II catalytic mutants grown in galactose- or glucose-containing medium. Individual data points from at least three biological repeats are shown with bars showing average +/− the standard deviation of the mean.

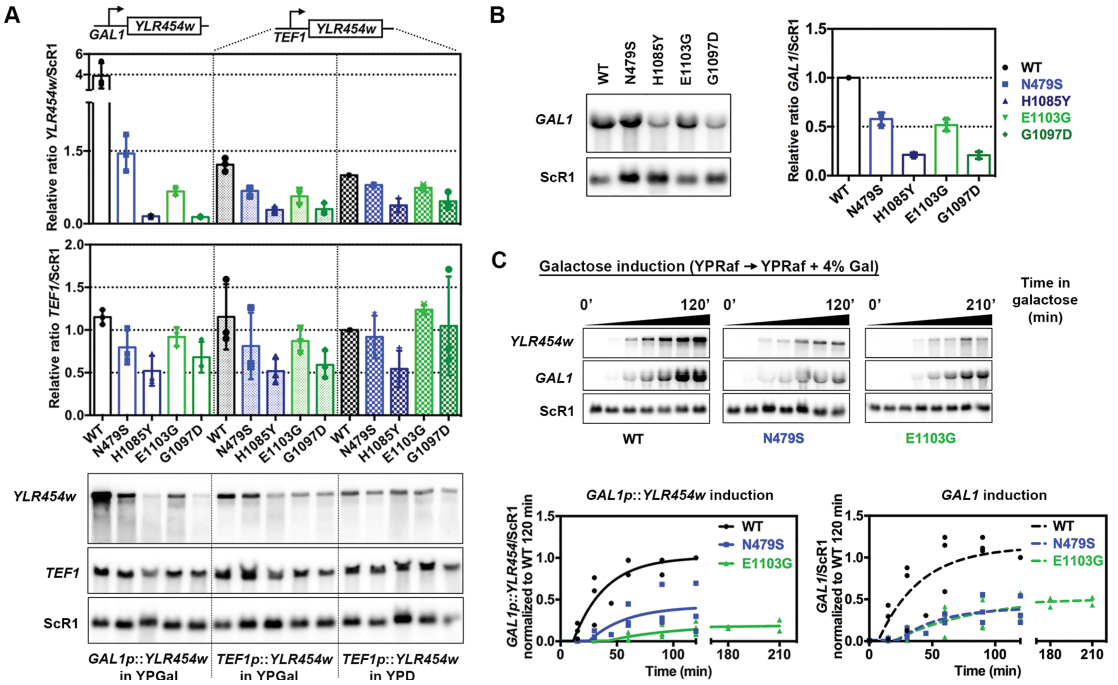


Figure 2-2. Pol II catalytic mutants generally decrease *in vivo* gene expression. (A) Pol II mutant effects on gene expression are exacerbated with increasing promoter strength. Steady-state RNA levels of reporter genes used for Pol II occupancy experiments (*GAL1p*::*YLR454w* and *TEF1p*::*YLR454w*) and endogenous *TEF1* levels from cells grown in galactose- or glucose-containing medium as indicated. Values were normalized to WT *TEF1p*::*YLR454w* expression level in YPD. The most severe expression defects are evident at *GAL1p*::*YLR454w* for both GOF and LOF mutants compared to WT. Individual data points from at least three biological repeats are shown with error bars indicating average +/− standard deviation of the mean. (B) Endogenous *GAL1* mRNA expression level is decreased in Pol II catalytic mutant strains compared to WT, yet *GAL1* expression defects in mutants are less severe than *GAL1p*::*YLR454w* expression defects as showed in A. Values were normalized to WT *GAL1*mRNA level. Error bars as in A. (C) Pol II catalytic mutants delay induction of the *GAL1*promoter. Time courses showing induction of *GAL1p*::*YLR454w* and endogenous *GAL1*mRNA in WT and Pol II catalytic mutants. Overnight grown cells were inoculated into fresh YPRaf medium and grown until mid-log phase at 30°C, subsequently galactose was added (4% final concentration) to induce *GAL* gene expression. RNAs isolated prior (time 0) and after galactose addition were used for northern blotting to determine accumulation of mature *GAL1p*::*YLR454w* and endogenous *GAL1*. Data normalized to WT 120 min value and plotted using non-linear regression using GraphPad Prism. Individual data points from at least three biological repeats are shown.

Pol II catalytic mutants show decreases in reporter gene expression that are exacerbated by promoter strength and template length

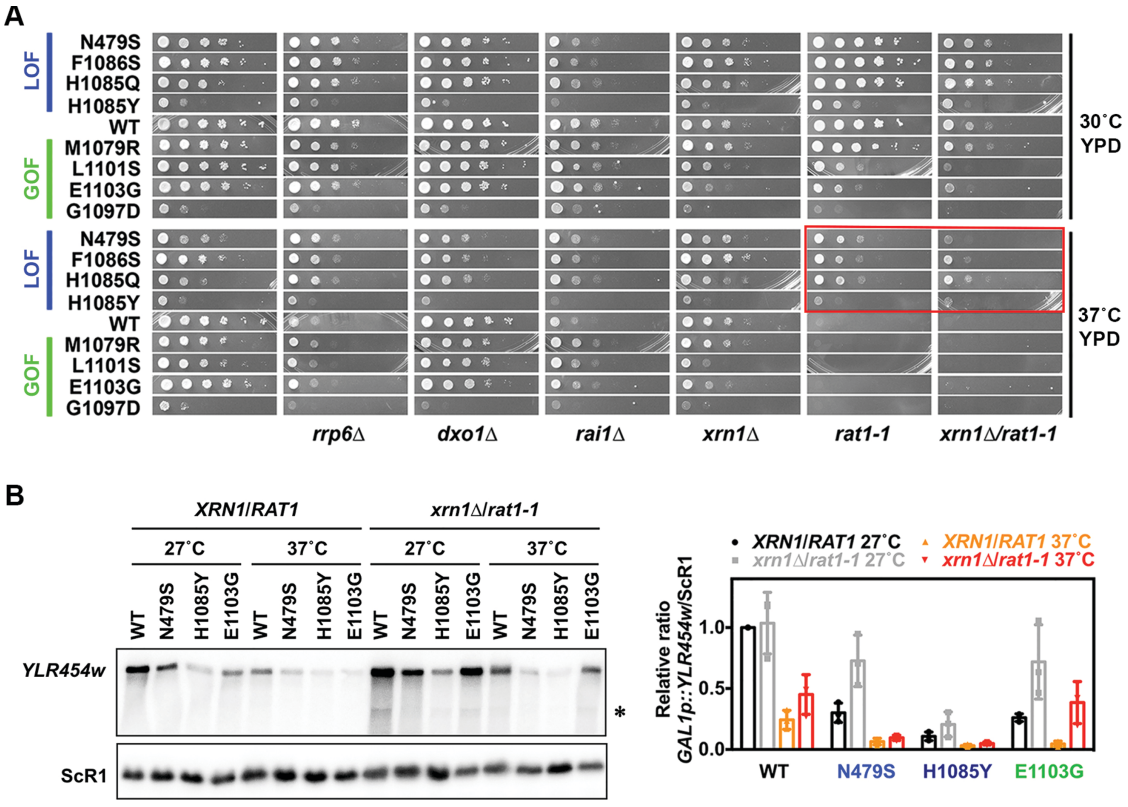
Defective initiation or elongation might lead to decreased Pol II occupancy (Barnes, Calero et al. 2015). Pol II activity mutants do perturb initiation as most or all show defects in TSS selection (Braberg, Jin et al. 2013, Jin and Kaplan 2014). Mutants affecting Pol II elongation have also been shown to affect reporter gene expression depending on length of the transcription unit driven by identical promoters, using the so-called "GLAM" assay (Morillo-Huesca, Vanti et al. 2006, Millan-Zambrano, Rodriguez-Gil et al. 2013). Therefore, we determined if Pol II catalytic mutants show gene expression defects based on gene length and/or promoter strength. We measured Pol II catalytic mutants' effects on expression of *YLR454w* reporters in two different promoter contexts, driven by the strong, inducible GAL1 promoter and the constitutive TEF1 promoter by Northern blotting (**Figure 2-2**A). Additionally, we also measured expressions of the endogenous *GAL1* and *TEF1* genes for comparison with *YLR454w* driven by the same promoters. Pol II mutants confer the strongest effects on *GAL1*p::*YLR454w* expression compared to WT, with H1085Y and G1097D having the most severe effects (**Figure 2-2**A). Furthermore, we observed greater gene expression defects when examining the highly expressed *GAL1*p::*YLR454w* relative to the more lowly expressed *TEF1*p::*YLR454w*. We also found that expression defects for the *YLR454w* reporters were greater than defects observed for the native genes (shorter in length) under control of their respective promoters (**Figure 2-2**A and **2-2**B). These results are consistent with previous observations showing Pol II occupancy- and gene-

length-dependent effects on reporter expression for several elongation mutants (Morillo-Huesca, Vanti et al. 2006, Millan-Zambrano, Rodriguez-Gil et al. 2013, Barnes, Calero et al. 2015). Given reduced Pol II occupancy and gene expression, it is likely that initiation defects might also be in play. When we examined the kinetics of activation for *GAL1p*::*YLR454w* or *GAL1*, we found that activation of expression was greatly delayed in Pol II mutant strains, consistent with defects in activation or initiation (**Figure 2-2**C).

Pol II catalytic mutants show allele-specific genetic interactions with pre-mRNA processing factors

We observed greater expression defects for the Pol II mutants compared to occupancy defects over the same reporter *GAL1p*::*YLR454w* (compare **Figure 2-1**B with **Figure 2-2**A). These results suggested that there could be additional defects beside Pol II initiation or elongation defects contributing to decreased abundance of the reporter mRNA in the mutants. As the abundance of cellular mRNAs is determined by synthesis and degradation rates, we reasoned that overall expression defects observed for Pol II mutants might stem from changes to mRNA stability and/or RNA processing. To start, we determined if Pol II catalytic mutants show *GAL1*p::*YLR454w* processing defects. We took a two-pronged approach to address processing defects. First, we examined genetic interactions between Pol II mutants and pre-mRNA processing/degradation factors; second, we asked if blocking pre-mRNA degradation pathways could rescue gene expression defects of Pol II mutants.

Figure 2-3. Allele-specific genetic interactions between Pol II catalytic mutants and pre-mRNA and mRNA processing factors.
(A) Pol II catalytic mutants show allele-specific genetic interactions with 5′ exonuclease mutants. 10-fold serial dilutions of saturated cultures of Pol II catalytic mutants alone or in combination with pre-mRNA and mRNA processing factor mutants plated on YPD for comparison of growth at 30°C and 37°C. Suppression of *rat1-1* lethality at restrictive temperature is highlighted in red box. (B) Inactivation of both Xrn1 and Rat1 rescue *GAL1p*::*YLR454w* expression in the Pol II GOF mutant E1103G. (Left) Representative gel showing *GAL1p*::*YLR454w* reporter expression in Pol II catalytic mutants in WT and *xrn1*Δ/*rat1-1* mutant background. Overnight grown cells were inoculated in fresh YPRafGal media to amplify until mid-log at permissive temperature (27°C) then shifted to restrictive temperature (37°C) to inactivate Rat1-1p. RNAs isolated from the half of the culture before shifting temperature used as 27°C sample. The other half was washed and resuspended in pre-warmed YPRafGal media to grow for another 2 h at 37°C to isolate 37°C sample for northern blotting. Asterisk indicates an internal cryptic *GAL1p*::*YLR454w*transcript in *xrn1*Δ/*rat1-1* mutant background, as reported earlier for an *spt16* mutant (Mason and Struhl 2003). (Right) Quantification of *xrn1Δ/rat1-1* effects. Values were normalized to WT *GAL1p*::*YLR454w* mRNA level (*XRN1/RAT1*) at 27°C. Data shown are average of three biological repeats with error bars representing standard deviation of the mean.


A
LOF
GOF
N479S
F1086S
H1085Q
H1085Y
WT
M1079R
L1101S
E1103G
G1097D
30˚C
YPD
37˚C
YPD
rrp6Δ
dxo1Δ
rai1Δ
xrn1Δ
rat1-1
xrn1Δ/rat1-1
B
XRN1/RAT1
xrn1Δ/rat1-1
27˚C
37˚C
WT
N479S
H1085Y
E1103G
YLR454w
ScR1
*
XRN1/RAT1 27˚C
XRN1/RAT1 37˚C
xrn1Δ/rat1-1 27˚C
xrn1Δ/rat1-1 37˚C
Relative ratio
GAL1p::YLR454w/ScR1
1.0
0.5
0.0
WT
N479S
H1085Y
E1103G

We reasoned that an mRNA capping defect or premature termination could lead to degradation of pre-mRNA by nuclear 5′ or 3′-exonucleases, respectively, resulting in decreased abundance of *GAL1*p::*YLR454w*. Thus, we tested if Pol II mutants showed genetic interactions with known nuclear exonucleases that function in surveillance. Deletion of Rrp6, a component of the 3′-exonuclease exosome complex, showed a slight negative genetic interaction with Pol II GOF mutants (Figure 2-3A). Inactivation of Rat1 and Xrn1, two 5′-exonucleases, showed strong genetic interactions with GOF mutants (Figure 2-3A). Rat1 is essential and functions in termination of Pol II at protein-coding genes (Kim, Krogan et al. 2004, Luo, Johnson et al. 2006). Pol II LOF mutants partially suppressed the Ts^- phenotype of *rat1-1*, a conditionally viable allele of *RAT1*. In contrast, Pol II GOF alleles showed synthetic sick interactions with *rat1-1*, consistent with prior findings and model proposed for kinetic competition between termination and elongation (Hazelbaker, Marquardt et al. 2013, Jimeno-Gonzalez, Schmid et al. 2014, Fong, Brannan et al. 2015). Rat1 works with an activating partner Rai1, which processes partially capped mRNA allowing Rat1 access (Jiao, Xiang et al. 2010). Growth defects of *rai1*Δ were slightly suppressed by both LOF and GOF mutants at 37˚C, while *dxo1*Δ, a recently described nuclear 5′-exonuclase (Chang, Jiao et al. 2012), showed no strong genetic interactions with Pol II alleles (Figure 2-3A).

We then asked if the mRNA processing mutants that showed genetic interactions with Pol II mutants were able to modulate Pol II alleles' gene expression defects. As predicted from the lack of strong genetic interactions, deletion of Rrp6 did not rescue full-length *GAL1*p::*YLR454w* expression level significantly; furthermore the nuclear

exosome would be expected to mainly degrade prematurely-terminated transcripts and not-full length ones (**Figure 2-4**A). Deletion of *XRN1* slightly rescued *GAL1*p::*YLR454w* expression level in Pol II GOF mutant E1103G (**Figure 2-4**A). Inactivation of Rat1 (at the restrictive temperature, 37˚C) severely decreased *GAL1*p::*YLR454w* expression levels in both WT and catalytic mutants possibly due the *rat1-1* growth defect, although relative levels of *GAL1*p::*YLR454w* in WT and Pol II mutants didn't differ greatly between permissive (27˚C) and restrictive temperature (37˚C) (**Figure 2-4**B**)**.

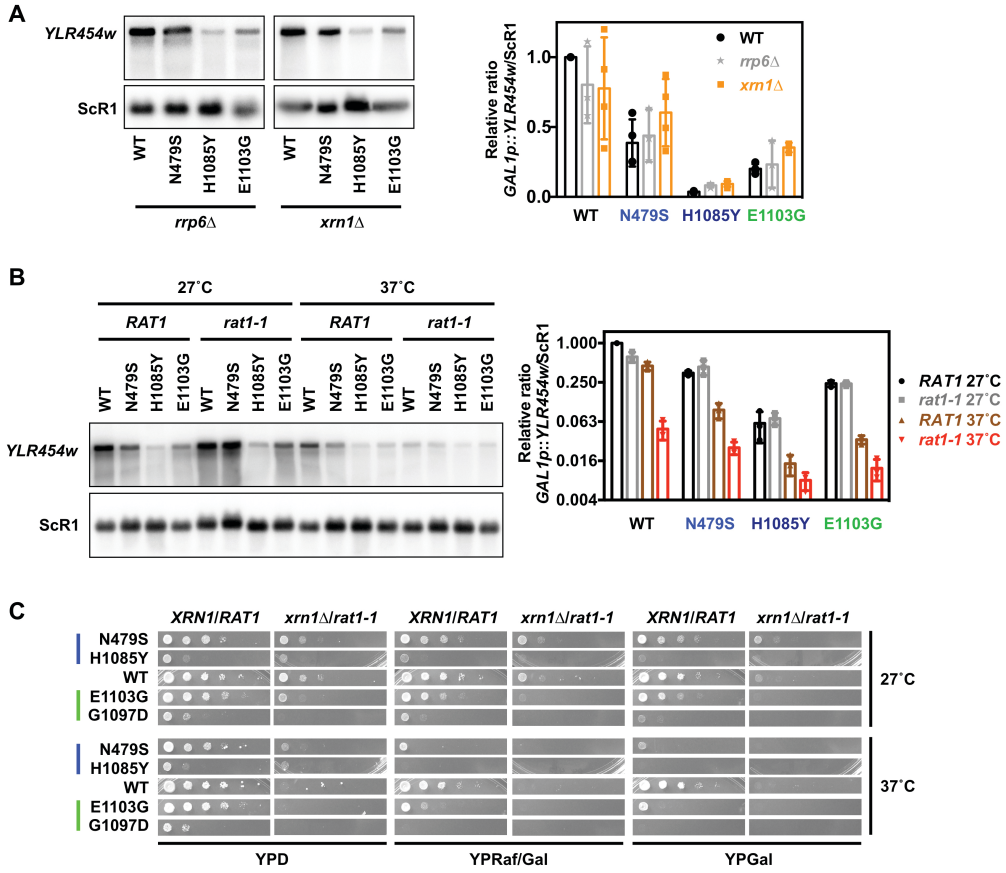


Figure 2-4. Modulation of reporter gene expression defect in pre-mRNA processing factor mutants. (A) *GAL1p*::*YLR454w* reporter expression level in Pol II catalytic mutants in WT, *xrn1*Δ, or *rrp6*Δ backgrounds determined by Northern blotting. WT Pol II sample (*XRN1/RRP6*) was run in parallel with *xrn1*Δ or *rrp6*Δ mutant samples on each blot for normalization purposes (**Figure 2A**) (B) *GAL1p*::*YLR454w* reporter expression level in Pol II catalytic mutants in WT and in *rat1-1* background at permissive (27˚C) or restrictive (37˚C) temperatures. Overnight grown cells were inoculated in fresh YPGal media to amplify to mid-log growth at 27˚C, then shifted to 37˚C to inactivate Rat1p. RNAs were isolated from half of each culture prior to temperature shift (27˚C samples). Remaining cultures were washed and resuspended in pre-warmed YPGal media to grow for another 2 hrs at 37˚C prior to RNA isolation. Relative *GAL1p*::*YLR454w* expression levels were normalized to WT (*RAT1*) 27˚C value. Data shown are average of three biological repeats with error bars representing the standard deviation (SD) of the mean. Note: values were normalized to WT (*RAT1*) at 27˚C, and are presented on logarithmic scale. (C) 10-fold serial dilutions of saturated cultures of Pol II catalytic mutants alone or in combination with *xrn1*Δ/*rat1-1* mutant were plated for growth at 27˚C or 37˚C on different media as indicated. Blue and green bars indicate LOF and GOF mutants, respectively.

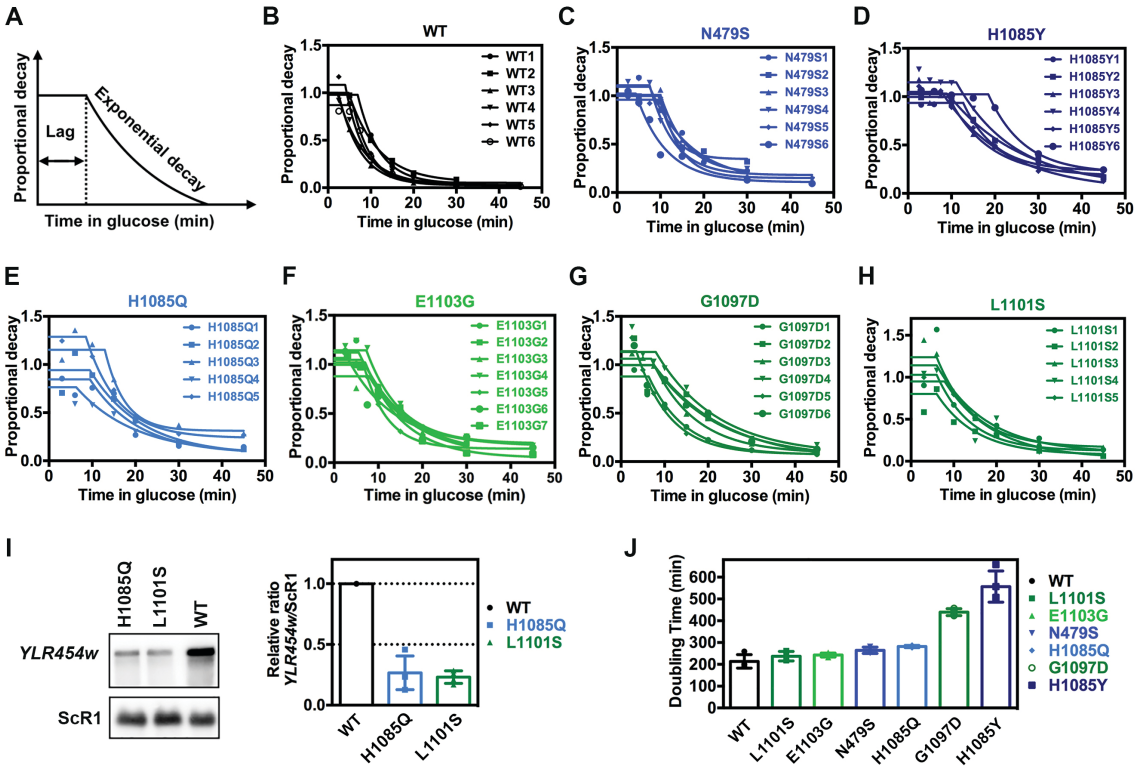


Figure 2-5. Pol II catalytic mutants alter *GAL1*p::*YLR454w* mRNA decay and growth in YPGal media. (A) Schematic of the mRNA decay curve fitting showing 'lag' and 'exponential decay' periods after transcriptional shut-off. (B-H) Individual decay curve for *GAL1*p::*YLR454w* decay in Pol II catalytic mutants plotted using non-linear regression using GraphPad prism. (I) *GAL1*p::*YLR454w* expression level in H1085Q and L1101S determined by Northern blotting. Values normalized to WT *GAL1*p::*YLR454w* expression level. Data shown are average of three biological repeats with error bars representing the standard deviation (SD) of the mean. (J) Doubling time of Pol II mutants in YPGal media determined using Tecan plate reader. Error bars represent average +/- SD of four replicate cultures.

We found that *xrn1*Δ/*rat1-1* double mutant strains were extremely sick on YPGal even at permissive temperature (**Figure 2-4**C). Hence, for determination of *GAL1*p::*YLR454w* expression level we used YPRaf/Gal liquid medium, which allows growth of Pol II mutants in the *xrn1*Δ/*rat1-1* background. Intriguingly, inactivation of both Xrn1 and Rat1 together completely rescued the *GAL1*p::*YLR454w* expression level in E1103G (*xrn1*Δ/*rat1-1*) compared to WT (*xrn1*Δ/*rat1-1*) at 37˚C (**Figure 2-3**B), suggesting possible 5′ end processing-defective or 5′-cap defective transcripts in the GOF mutant E1103G are stabilized by inactivation of Xrn1 and Rat1. Taken together these genetic interaction results suggest Pol II GOF mutants may have 5′ end processing or capping defects, rendering transcripts sensitive to Rat1/Xrn1.

Both slow and fast Pol II catalytic mutants confer increased mRNA stability to a reporter gene

We reasoned that in addition to myriad Pol II transcription defects and pre-mRNA processing defects discussed above, compromised mRNA stability of the *GAL1*p::*YLR454w* transcript may also contribute to its lower abundance in Pol II catalytic mutants. In order to determine if Pol II mutants alter mRNA stability, we performed transcriptional shutoff followed by measurement of *GAL1*p::*YLR454w* mRNA over a time course by Northern blotting. mRNA half-lives were determined by fitting decay curves with a lag followed by exponential decay (**Figure 2-5**A). We found that all GOF and LOF mutants conferred an increase in both the *YLR454w* half-life and lag periods prior to exponential decay relative to WT (**Figure 2-6**A**, Figure 2-5**B-H). Further, the increases in mRNA half-lives positively correlated with the overall decrease

in expression level of the reporter in the mutants (**Figure 2-6**B**, Figure 2-5**I). These results suggest a number of conclusions. First, they indicate defects in mRNA synthesis rates for all tested Pol II mutants, as steady state expression levels were reduced relative to wild type even though we found that the *GAL1-YLR454w* mRNA showed increased stability. Second, longer lag periods prior to exponential decay may indicate Pol II mutants confer a delayed response for transcriptional shutoff, slower elongation rate on this template, or delayed mRNA export (addressed below). Such lags have been observed in genome-wide experiments for mRNA stability where transcription inhibitor blocks initiation but elongation is allowed to proceed (Chen, Shiroguchi et al. 2015). The lag can result from the time it takes polymerases that have initiated prior to the block to finish synthesis of their messages. Given that mRNA decay is strongly coupled to translation, and translational demand is coupled to growth rate, it may be difficult to deconvolute Pol II mutant effects on mRNA decay through changes to growth rates versus direct effects on mRNA decay. Indeed, we observed that mRNA half-lives positively correlated with strain doubling times (**Figure 2-6**C**, Figure 2-5**J).

Slow and fast Pol II catalytic mutants exhibit slower apparent *in vivo* elongation in a commonly used ChIP assay

Our observation of longer lag times than WT for all Pol II mutants tested, both slow and fast, was unexpected under the presumption that Pol II mutants that are fast elongaters *in vitro* are also fast *in vivo*. In our mRNA decay experiments, longer lag times prior to exponential decay for LOF mutants were consistent with delayed clearance of the template and predicted slow elongation kinetics. Conversely, GOF

mutants would be predicted to run off the template more quickly. In contrast to this prediction, GOF mutants also showed longer lag times prior to exponential decay (**Figure 2-6**A). One GOF mutant tested here, E1103G, was previously suggested to be faster than WT *in vivo* based on chromatin IP analysis of the same *GAL1*p::*YLR454w* reporter used here (Hazelbaker, Marquardt et al. 2013). Therefore, we examined elongation of E1103G and additional Pol II alleles *in vivo* using this commonly used ChIP assay. In this assay, transcription of *GAL1*p::*YLR454w* is shutoff by addition of glucose to the medium, and the kinetics of the last wave of Pol II transcription clearing the gene is measured using ChIP (**Figure 2-7**A and B) (Mason and Struhl 2005). As predicted, LOF mutants N479S and H1085Y showed extensive delay in Pol II run-off kinetics following the transcriptional shut-off (**Figure 2-7**C and D). The apparent *in vivo* elongation rate of H1085Y was slower than N479S, which is consistent with its stronger growth defects and *in vivo* phenotypes relative to N479S. Surprisingly, both test GOF mutants, E1103G and G1097D, also showed reduced kinetics of Pol II runoff subsequent to glucose addition (**Figure 2-7**C). This result was unexpected, especially as E1103G was reported to have a faster apparent elongation rate than WT in a version of this assay (Hazelbaker, Marquardt et al. 2013). Furthermore, analogous substitution in human POLR2A E1126G confers a slightly faster than WT elongation rate *in vivo* (Fong, Kim et al. 2014).

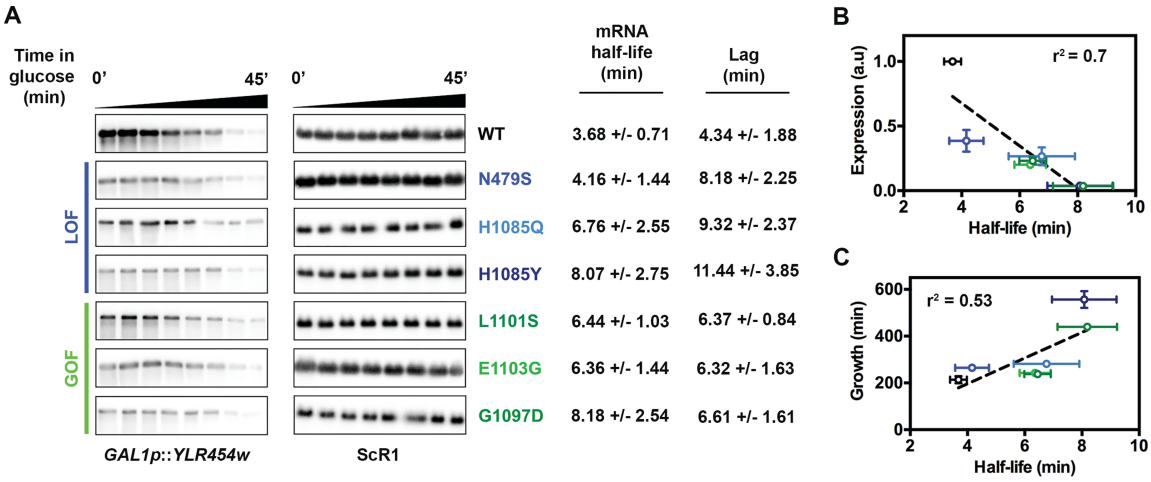
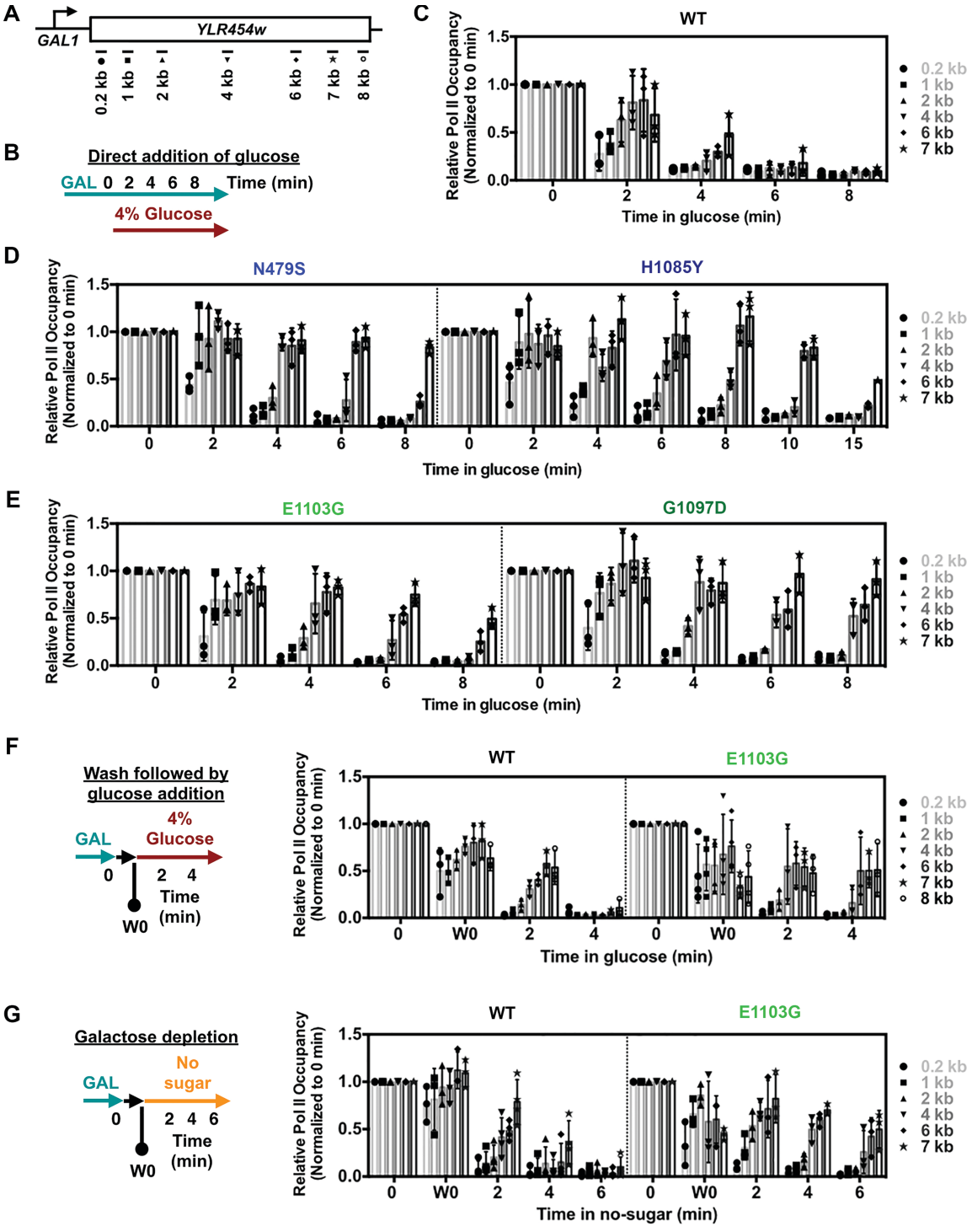


Figure 2-6. Pol II catalytic mutants increase half-life of an mRNA reporter. (A) Both GOF and LOF mutants show reduced degradation rate of the *YLR454w* transcript compared to WT. Representative blots showing *GAL1p*::*YLR454w* mRNA decay, determined by glucose shutoff. Overnight grown cells were inoculated in fresh YPGal medium and grown until mid-log at 30°C. RNAs isolated from pre-glucose addition (time 0) and post-glucose addition time points were used for northern blotting. mRNA half-lives and lags were determined using plateau followed by one-phase decay curve fitting in GraphPad Prism (see **Figure 2-5**). Values indicate average of a minimum of five biological repeats +/− standard deviation of the mean. (B and C) Correlation of *GAL1p*::*YLR454w* half-life in different mutants with the expression level (**Figure 2-1**A and **Figure 2-5**I) and with the mutants' growth defects (**Figure 2-5**J) determined by linear regression using GraphPad Prism.

Figure 2-7. Pol II catalytic mutants show slower *in vivo* elongation rate in a glucose shutoff ChIP assay. (A) Schematic of the *GAL1p*::*YLR454w* with positions of PCR amplicons used for ChIP experiments. (B) Schematic of the regular glucose shutoff experiment used in (C) and (D). *In vivo* apparent elongation rate for WT (C), LOF (D; N479S and H1085Y) and GOF (E; E1103G and G1097D) catalytic mutants determined by ChIP upon glucose shutoff of transcription by direct addition of glucose (to 4% final concentration) to the mid-log grown culture in YPGal at 30°C. Values were normalized to pre-glucose addition (0 min) and error bars represent standard deviation of the mean for at least three independent experiments. For H1085Y, longer time point (10 and 15 min) values obtained from two repeats, with error bars indicating the range of the two experiments. (F) Glucose shutoff assay to compare apparent *in vivo* elongation rate between WT and fast catalytic mutant (E1103G) following the Hazelbaker *et al*. protocol (Hazelbaker, Marquardt et al. 2013). Pre-glucose addition sample (0 min) was isolated as described in Figure 5A; subsequently cells were washed in synthetic complete medium lacking carbon source and inoculated in YPD (4% glucose) to shutoff the transcription. One wash 0 (W0 min) sample was isolated after the washing and before shutoff to determine the effect of washing. Values were normalized to pre-glucose addition (0 min) and error bars represent standard deviation of the mean of at least three independent repeats. (G) Galactose depletion to determine apparent elongation rate in WT and fast catalytic mutant (E1103G). Pre-glucose (0 min) and wash 0 (W0 min) samples were taken as described in Figure 5F, followed by inoculation of cells into synthetic medium lacking any sugar to incur transcriptional shutoff due to galactose depletion. Values were normalized to pre-glucose addition (0 min) and error bars represent standard deviation of the mean for at least three independent repeats.


A
GAL1
YLR454w
0.2 kb
1 kb
2 kb
4 kb
6 kb
7 kb
8 kb
B
Direct addition of glucose
GAL 0 2 4 6 8 Time (min)
4% Glucose
C
WT
Relative Pol II Occupancy (Normalized to 0 min)
Time in glucose (min)
0.2 kb
1 kb
2 kb
4 kb
6 kb
7 kb
D
N479S
H1085Y
Relative Pol II Occupancy (Normalized to 0 min)
Time in glucose (min)
E
E1103G
G1097D
Relative Pol II Occupancy (Normalized to 0 min)
Time in glucose (min)
F
Wash followed by glucose addition
GAL
4% Glucose
0 2 4
Time (min)
W0
WT
E1103G
Relative Pol II Occupancy (Normalized to 0 min)
Time in glucose (min)
8 kb
G
Galactose depletion
GAL
No sugar
0 2 4 6
Time (min)
W0
WT
E1103G
Relative Pol II Occupancy (Normalized to 0 min)
Time in no-sugar (min)

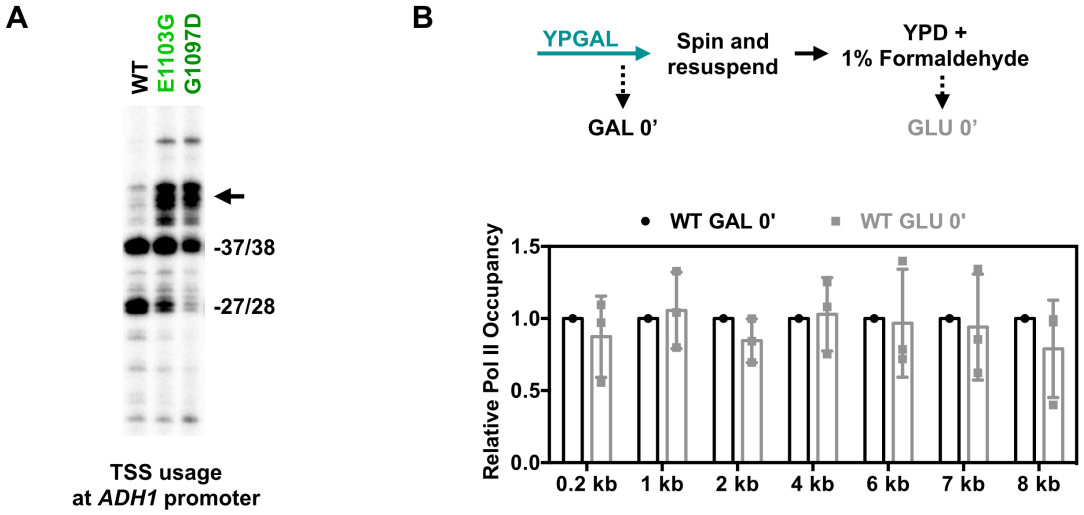


Figure 2-8. Validation of GOF mutant phenotypes and formaldehyde crosslinking kinetics. (A) Primer extension analysis of start site usage at *ADH1* in WT and GOF catalytic mutants shows expected upstream shift (arrow) for GOF mutants. (B) Evaluating formaldehyde cross-linking efficiency during shut-off of *GAL1*p::*YLR454w* transcription by addition of 4% glucose. WT culture was grown to mid-log phase at 30˚C and half of the culture was isolated and fixed with 1% formaldehyde as time 0 in YPGal (GAL 0'). Then, the rest half was quickly centrifuged and resuspended in YPD (4% dextrose) containing 1% formaldehyde to obtain time 0 in YPD + formaldehyde (Glu 0'). ChIP was performed to determine Pol II occupancy over *GAL1*p::*YLR454w* (see Fig. 5A) in GAL 0' versus Glu 0'. This comparative analysis allows us to determine if there is any apparent repression during cross-linking in the presence of glucose to inhibit *GAL1*p::*YLR454w* transcription. Bar graphs show average of three biological repeats +/- SD.

Given the discrepancy between our results and those of Hazelbaker et al (Hazelbaker, Marquardt et al. 2013), we felt it was essential to closely examine Pol II mutant behavior in the elongation assay and identify if there is any hidden or confounding variable. We first extensively validated our GOF mutant strains in three ways – i. confirmation of introduced mutations by sequencing, ii. confirmation of transcriptional growth phenotypes as described in (Kaplan, Jin et al. 2012), and iii. confirmation of mutants' ability to shift transcriptional start sites (TSSs) at the *ADH1* promoter as described in (Kaplan, Jin et al. 2012, Braberg, Jin et al. 2013, Jin and Kaplan 2014) (**Figure 2-8**A). In all cases, results were consistent with strains showing expected phenotypes, leading us to speculate that there might be confounding variable(s) in the ChIP elongation assay itself. While we employed the most commonly used method to shutoff *GAL* transcription by direct addition of glucose to the media (Mason and Struhl 2005), in Hazelbaker et al (Hazelbaker, Marquardt et al. 2013), cells were washed in the absence of any carbon source prior to addition of glucose. We repeated the method of (Hazelbaker, Marquardt et al. 2013) with the addition of a critical control, taking a time point just post wash (W0) to determine how the wash step affected Pol II occupancy. In our hands, GOF E1103G was still slower exiting the template than WT (**Figure 2-7**F), consistent with both our direct glucose shutoff without washing, and our measurement of increased lag in mRNA decay experiments (**Figure 2-7**E **and 2-6**A). Additionally, we observed two more phenomena during the wash. First, the W0 time point showed evidence of transcriptional shutoff for both WT and E1103G, presumably due to galactose depletion (see below for further discussion). We confirmed that Pol II

run off was not due to delayed formaldehyde cross-linking, as we did not detect any difference in Pol II occupancy when crosslinking was simultaneous with the addition of glucose (**Figure 2-8**B). Second, GOF E1103G showed a faster loss of Pol II over the 3′ end of *GAL1p::YLR454w* early in the time course (W0) that was not observed in the direct glucose shutoff assay (compare **Figure 2-7**E with **2-7**F). Together, these results indicate that the mode of transcription inhibition could alter Pol II transcriptional properties in unexpected ways, additionally confounding ability to estimate apparent Pol II elongation rate.

The absence of galactose itself is a distinct mechanism of transcription shutoff for *GAL* genes. Depletion of galactose will lead to reestablishment of Gal80-mediated repression of Gal4, the critical activator of *GAL* transcription. To examine this mode of transcription shutoff further, without the mixing of two distinct modes of repression, we performed a galactose starvation/depletion time course where no sugar was added. We found similar 3′ specific runoff in E1103G early in the time course (W0) when a wash was employed (**Figure 2-7**G). As Pol II runs into the gene 3′ end at later time points (presumably), the effect diminishes. Additionally, in this iteration of the experiment there appears to be faster loss of E1103G than WT from the 5′ end of the reporter early, though E1103G is maintained much longer over the 3′ end later in the time course (**Figure 2-7**G). Taken together, we find that Pol II GOF mutants, where tested, appear to be retained longer on the template, when examined in the *GAL1*p*::YLR454w* system.

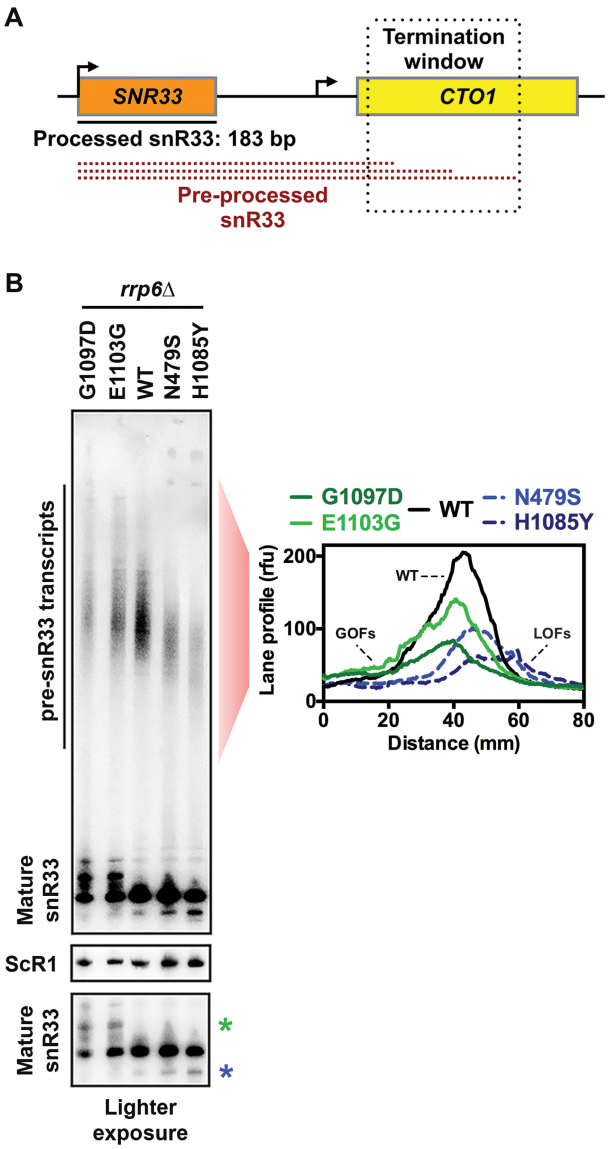


Figure 2-9. Mapping of termination window for preprocessed snR33 in WT and Pol II catalytic mutants. (A) Schematic of the snR33 termination window. (B) GOF and LOF mutants increase and decrease average length of pre-processed snR33, respectively. Northern blotting for snR33 from WT, slow (N479S and H1085Y) and fast (E1103G and G1097D) catalytic mutant strains containing deletion of the nuclear exosome subunit gene *RRP6*. Pol III transcript ScR1 was used as loading control. Average densitometric values of indicated lanes from three independent repeats are presented in the graph. A lighter exposure of the mature snR33 full-length transcript shows alternative longer (green asterisk) and shorter (blue asterisk) products for GOF and LOF mutants, respectively, that are proposed to be generated by shifts in TSS usage.

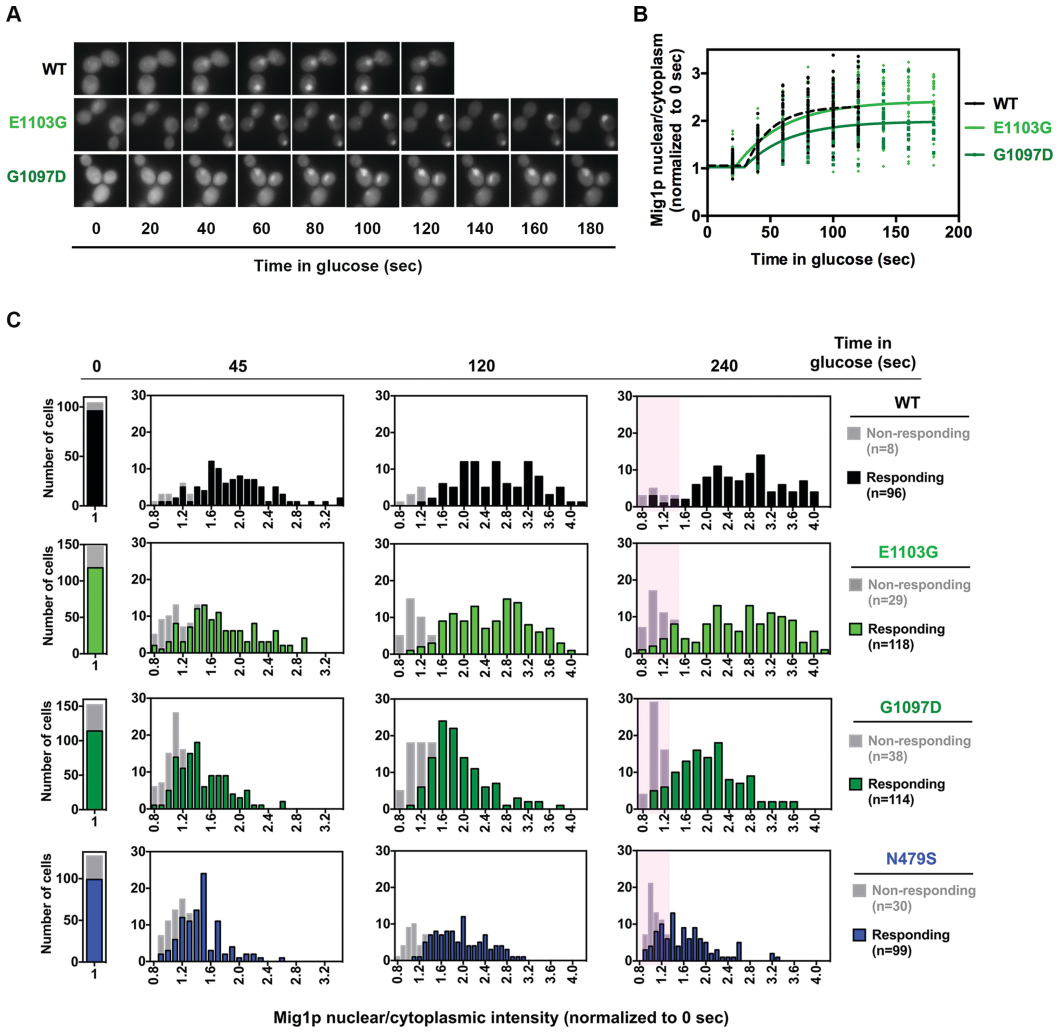


Figure 2-10. Mig1 nuclear translocation is aberrant in Pol II catalytic mutants. (A) Representative images of nuclear localization of Mig1p-GFP in WT and catalytic mutants (E1103G and G1097D) upon glucose addition (4% final concentration). Overnight grown cells were inoculated into fresh synthetic complete medium containing galactose (SC-2% Gal) and grown until mid-log phase at 30°C. Pre-glucose repression samples were used as time 0 (t0), followed by replacing the medium with SC-2% Gal + 4% glucose to induce repression. (B) Mig1p nuclear localization kinetics is delayed in GOF mutant G1097D. Normalized (to pre-glucose treatment; t0) data from glucose-responding cells ($n > 35$) plotted using non-linear regression in GraphPad Prism. (C) Histograms of Mig1p nuclear localization show the distribution of nuclear Mig1p fluorescence (normalized to t0) intensity over indicated time points. Cells that do not show any traces of fluorescence accumulation are designated as non-responding cells (see 'Materials and Methods' for quantification details). Highlighted areas indicate population of cells that are either non-responding or responding cells with decreased Mig1p nuclear localization.

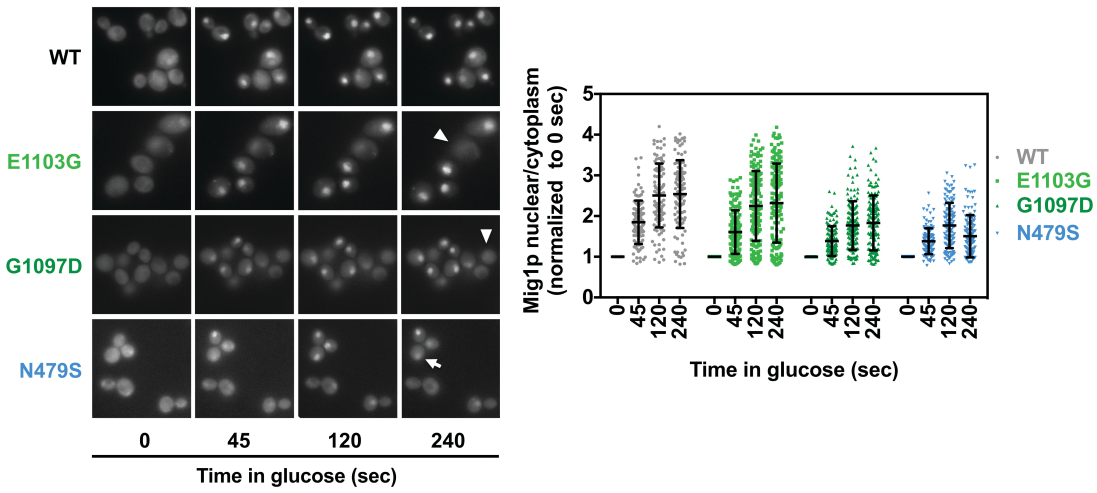


Figure 2-11. Impaired Mig1p nuclear import in G1097D. Representative images of nuclear localization of Mig1p-GFP in WT and catalytic mutants upon glucose (4% final) addition (left) and average Mig1p nuclear localization +/- SD (right). Data from multiple cells (WT, n=104 cells from 6 individual experiments; E1103G, n= 147 cells from 8 individual experiments; G1097D, n= 152 from 5 individual experiments; N479S, n=129 from 9 individual experiments) used to make the histogram in **Figure 7**. Arrowhead showing example of non-responding cell that does not show Mig1p-GFP foci accumulation upon glucose addition and arrow showing cell that decays Mig1p-GFP signal over time.

We observe this under three distinct transcription shutoff conditions and in all cases our results contradict previously published results for Pol II GOF E1103G. However, our studies suggest that E1103G effects on the 3′ end of *GAL1*p*::YLR454w* upon galactose starvation (during wash) and unintended variability in wash time between strains could account for the observed differences in our experiments and previous observations.

Finally, we examined alternative *in vivo* evidence for putative fast and slow elongation of Pol II GOF and LOF mutants. We analyzed how Pol II mutants altered termination at *SNR33*. It has been proposed previously that kinetic competition between elongating Pol II and Sen1 helicase determines the termination of the snR33 transcript (Hazelbaker, Marquardt et al. 2013). Examination of snR33 intermediate transcripts prior to complete processing allows detection of the snR33 termination window (**Figure 2-9**A). Both GOF mutants E1103G and G1097D increase the average length of pre-processed snR33, which is in agreement with previous observations for E1103G (**Figure 2-9**B and (Hazelbaker, Marquardt et al. 2013)). Conversely, LOF mutants N479S and H1085Y decreased the average length of pre-processed snR33 (**Figure 2-9**B). We also observed alternative mature snR33 transcripts longer or shorter in length for GOF and LOF mutants, respectively, consistent with increased upstream or downstream utilization of transcription start site (TSS) in these mutants as predicted for alterations in elongation rate.

Glucose signal transduction defects in Pol II alleles may partly explain apparent slow elongation by ChIP

One possibility for apparent slower elongation of GOF mutants on the *GAL1*p*::YLR454w* template could be delayed transcriptional shutoff. We reasoned that if glucose repression (or galactose starvation) were delayed in GOF mutants, it would contribute to apparent slow kinetics of Pol II exiting the gene. To this end, we investigated the kinetics of nuclear localization of Mig1p, which upon addition of glucose to medium is imported into the nucleus due to regulation by dephosphorylation, binds to target promoters of glucose-repressed genes and recruits co-repressors (Nehlin, Carlberg et al. 1991, Gancedo 1998). As a proxy for early events in this cascade, we determined the kinetics of Mig1p nuclear localization by monitoring import of Mig1p-GFP upon glucose addition in WT or GOF catalytic mutants We observed signal for presumptive Mig1p-GFP localization to the nucleus saturate within ~2 min for WT cells, as previously reported (De Vit, Waddle et al. 1997). However, for at least one GOF mutant, the signal did not appear to reach maximum nuclear fluorescence within 2 min of glucose addition, thus we measured extended time points and analyzed the data from individual cells using non-linear regression (**Figure 2-10**A and B). GOF G1097D showed both a delay in average time of Mig1p-GFP nuclear import and reduced average maximum nuclear fluorescence compared to WT, suggesting *in vivo* shutoff kinetics may be delayed due to deranged Mig1p signaling in this mutant at least (**Figure 2-10**B). Next, we measured extended time points of Mig1p translocation upon glucose repression over populations of individual cells for the GOF mutants analyzed in **Figure 2-10**A/**B**,

while also including one LOF mutant, N479S. A noticeable subset of delayed responding cells was observed for GOF mutant G1097D, while higher fractions of non-responding cells were observed for both GOF mutants and the LOF mutant N479S compared to WT (**Figure 2-10**C and **Figure 2-11**). Both GOF mutant G1097D and LOF mutant N479S showed lower average maximum nuclear Mig1p-GFP accumulation compared to WT, with the LOF mutant N479S showing an increase, then decrease, in Mig1p-GFP signal over time (**Figure 2-11**). Altogether, these results indicate that Mig1p nuclear translocation in Pol II catalytic mutants can be aberrant, potentially affecting interpretation of transcription kinetics upon glucose shutoff.

Pol II catalytic mutants do not confer sensitization of cells to GTP starvation through hypothesized global elongation defects

To understand the relationship between altered Pol II activity and growth defects derived from putative global elongation defects, we turned to pharmacological reduction of a Pol II substrate, GTP. Nucleotide depleting drugs such as MPA or 6-AU have been interpreted as specifically exacerbating global elongation defects of Pol II or transcription factor mutants, and therefore have been widely used as pharmacological probes for transcription elongation *in vivo*. Such interpretations of MPA sensitivity assume that WT and mutant strains have similar GTP starvation upon MPA treatment. We and others have noted that WT and MPA-sensitive mutants can experience different levels of GTP starvation due to different expression levels of the *IMD2* gene product, which is required for resistance to MPA (Shaw, Wilson et al. 2001, Hyle, Shaw et al. 2003, Jenks and Reines 2005). These differences in *IMD2* expression explain why

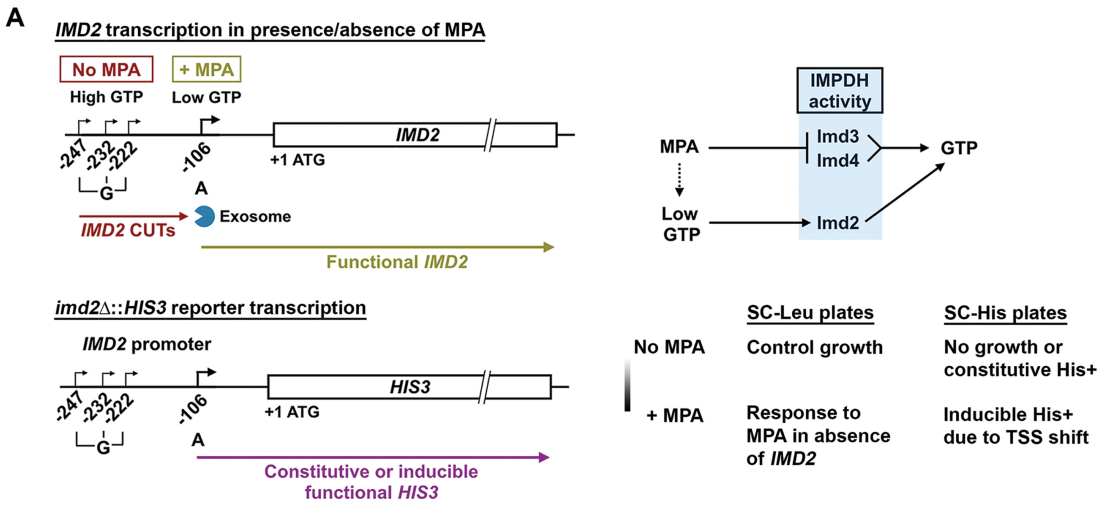
increased catalytic activity Pol II GOF mutants are MPA sensitive – they do not express *IMD2*. In contrast, Pol II LOF mutants constitutively express *IMD2*, and do not show MPA sensitivity. We chose to evaluate how Pol II responds to GTP starvation and determine if Pol II transcription is a critical determinant for cellular sensitivity to GTP starvation. To do so, we rendered GTP starvation independent of *IMD2* expression by deleting *IMD2*. Deleting *IMD2* is predicted to normalize GTP starvation between WT and Pol II mutants, allowing us to examine the assertion that GTP depletion exacerbates putative Pol II elongation defects *in vivo* (**Figure 2-12**). Furthermore, to enable an *in vivo* readout for *IMD2* expression, we replaced the *IMD2* ORF with *HIS3* (**Figure 2-12**A).

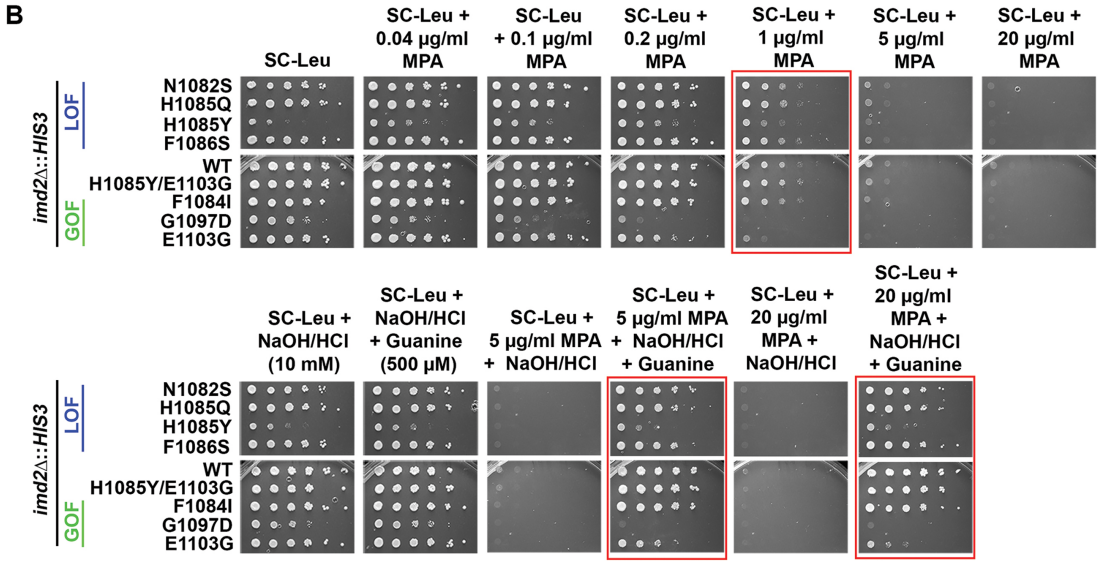
In the absence of *IMD2*, otherwise wild type yeast strains become sensitive to MPA treatment as expected (**Figure 2-12**B) (Shaw, Wilson et al. 2001, Hyle, Shaw et al. 2003). If Pol II transcription were a major determinant of GTP starvation, it would be predicted that elongation-defective LOF Pol II alleles would become hypersensitive to MPA, while GOF increased catalytic activity mutants would show relative resistance. In contrast to these predictions, deletion of *IMD2* strongly blunted the Pol II allele-specific effects of MPA treatment on yeast cells (**Figure 2-12**B). We did observe mild MPA sensitivity for Pol II GOF alleles E1103G and G1097D. These results suggest that the major determinant for Pol II allele sensitivity to MPA is differential *IMD2* expression, and not necessarily Pol II elongation defects. As stated earlier, MPA effects are presumed to function through GTP starvation. This presumption can be tested by addition of guanine to the growth medium can, which can suppress MPA-sensitivity by

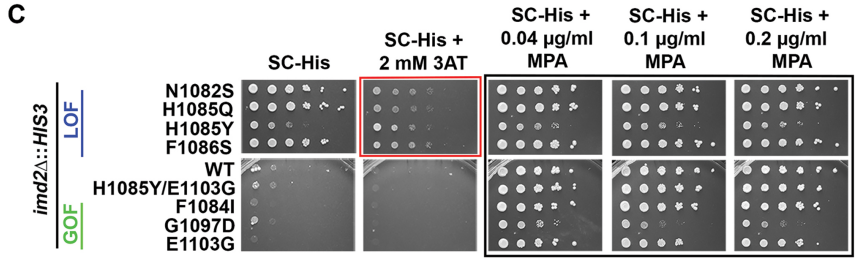
supporting an alternate route for GTP synthesis. In testing this presumption, we found that use of 10 mM NaOH to solubilize guanine in the medium on its own reduced effects of MPA on most of the Pol II mutants (**Figure 2-13**). Notably, some of these Pol II mutants have shown allele-specific phenotypes upon change to pH of media (Cabart, Jin et al. 2014). We determined that addition of equivalent HCl to neutralize NaOH effects allowed examination of guanine suppression of MPA sensitivity (**Figure 2-12**B).

We observed that addition of guanine suppresses MPA sensitivity for most Pol II mutants except the GOF mutants E1103G and G1097D, which show mild or no suppression of MPA sensitivity by guanine, respectively. These results suggest that either the *IMD2*-independent MPA effects on GOF mutants are independent of GTP starvation, or these mutants have alterations in their GTP salvage synthesis pathways. We next determined if the ability to induce the *IMD2* promoter was maintained in Pol II mutants when GTP starvation conditions were normalized, using growth on medium lacking histidine to detect expression of *imd2*Δ*::HIS3* (**Figure 2-12**C). WT cells and GOF mutants were almost completely His^- in the absence of induction of the *IMD2* promoter (controlling *HIS3*), while LOF mutants were His^+, indicative of constitutive expression of *imd2*Δ*::HIS3*. All strains showed induction of His^+ phenotype in the presence of MPA (inducer of *IMD2* promoter), suggesting that productive transcription from the *IMD2* promoter was possible for all tested Pol II mutants (discussed further below).

Figure 2-12. Pol II catalytic mutants defective for *IMD2* transcription do not abolish GTP sensing. (A) (Top left) Schematic of *IMD2* gene transcription in presence and absence of MPA. In GTP-replete cells, *IMD2* is transcribed from upstream 'G' start sites producing non-functional CUTs, which are degraded by the RNA exosome. In the presence of MPA (GTP depletion), a downstream 'A' start site is utilized to produce functional *IMD2* transcripts. (top right) MPA depletes GTP by inhibition of IMPDH activity, present in yeast in two paralogous enzymes, Imd3 and Imd4. Upon GTP depletion, expression of an MPA-resistant IMPDH enzyme, Imd2 is induced. (bottom) Schematic of the *imd2*Δ::*HIS3* construct and expected phenotype upon MPA treatment in synthetic medium lacking leucine or histidine. All strains lack endogenous *IMD2* ORF, which is replaced with *HIS3* (*imd2*Δ::*HIS3*), hence rendering them highly sensitive to MPA treatment. (B) In absence of endogenous *IMD2*, Pol II catalytic mutants respond to GTP depletion similarly to WT. 10-fold serial dilutions of saturated cultures of Pol II catalytic mutants plated on synthetic medium lacking leucine for comparison of growth in various concentration of MPA treatment at 30°C. (Upper panel) For the Pol II catalytic mutants' response to MPA treatment, the most obvious comparison with WT can be made at 1 μg/ml MPA treatment (highlighted with red box). (Lower panel) Addition of guanine suppresses MPA sensitivity of Pol II mutants (highlighted with red boxes). (C) 10-fold serial dilutions of saturated cultures of Pol II catalytic mutants plated on synthetic medium lacking histidine for comparison of growth at various concentrations of MPA treatment at 30°C. 3-aminotriazole (3-AT) is a competitive inhibitor of His3p and can be used to assess level of *HIS3* expression. Slow catalytic mutants constitutively use the downstream, productive 'A' site at the *IMD2* promoter, which produces a functional transcript (see text for details). Hence, His^+ colonies are observed even in the absence of MPA (red highlighted box). All Pol II catalytic mutants show inducible His^+ phenotype upon low level of MPA treatment (highlighted black box), indicating retention of GTP sensing ability.


A
IMD2 transcription in presence/absence of MPA
No MPA
High GTP
+ MPA
Low GTP
IMD2
-247
-232
-222
-106
+1 ATG
G
A
IMD2 CUTs
Exosome
Functional IMD2
IMPDH activity
MPA
Imd3
Imd4
GTP
Low GTP
Imd2
imd2Δ::HIS3 reporter transcription
IMD2 promoter
HIS3
Constitutive or inducible functional HIS3
SC-Leu plates
SC-His plates
No MPA
Control growth
No growth or constitutive His+
+ MPA
Response to MPA in absence of IMD2
Inducible His+ due to TSS shift


B
SC-Leu
SC-Leu + 0.04 µg/ml MPA
SC-Leu + 0.1 µg/ml MPA
SC-Leu + 0.2 µg/ml MPA
SC-Leu + 1 µg/ml MPA
SC-Leu + 5 µg/ml MPA
SC-Leu + 20 µg/ml MPA
imd2Δ::HIS3
LOF
GOF
N1082S
H1085Q
H1085Y
F1086S
WT
H1085Y/E1103G
F1084I
G1097D
E1103G
SC-Leu + NaOH/HCl (10 mM)
SC-Leu + NaOH/HCl + Guanine (500 µM)
SC-Leu + 5 µg/ml MPA + NaOH/HCl
SC-Leu + 5 µg/ml MPA + NaOH/HCl + Guanine
SC-Leu + 20 µg/ml MPA + NaOH/HCl
SC-Leu + 20 µg/ml MPA + NaOH/HCl + Guanine


C
SC-His
SC-His + 2 mM 3AT
SC-His + 0.04 µg/ml MPA
SC-His + 0.1 µg/ml MPA
SC-His + 0.2 µg/ml MPA
imd2Δ::HIS3
LOF
GOF
N1082S
H1085Q
H1085Y
F1086S
WT
H1085Y/E1103G
F1084I
G1097D
E1103G

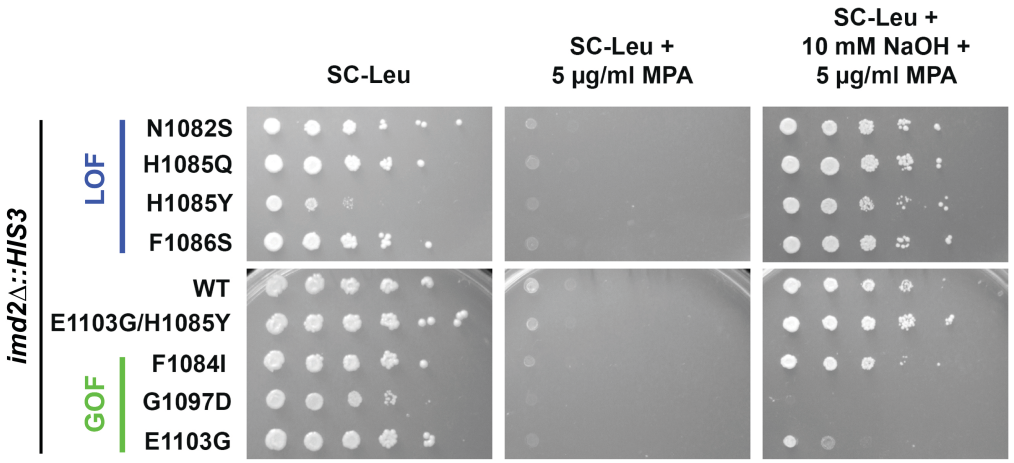
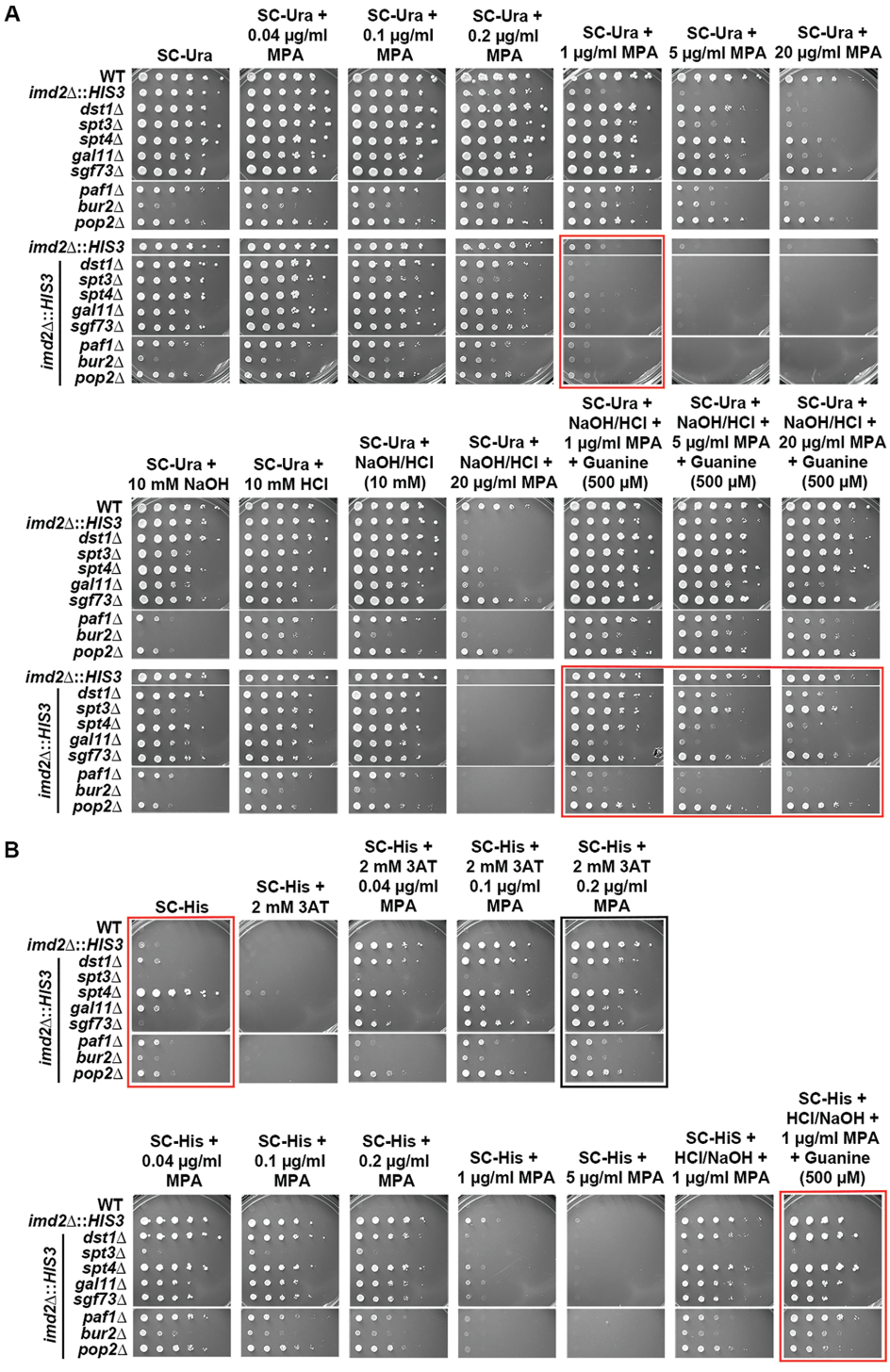


Figure 2-13. Addition of NaOH in the media abrogates MPA effects. 10-fold serial dilutions of saturated cultures of Pol II catalytic mutants were plated for growth at 30˚C on synthetic media containing NaOH. All strains contain endogenous *IMD2* deletion (*imd2*Δ::*HIS3*), rendering them highly sensitive to MPA. Only G1097D (strongest GOF) shows sensitivity to MPA in presence of NaOH in the medium.

Figure 2-14. Transcription related factor mutants respond similarly to WT upon MPA treatment in absence of endogenous *IMD2*. (A) 10-fold serial dilutions of saturated cultures of indicated factor mutants plated on synthetic medium lacking uracil for comparison of growth at various concentrations of MPA treatment at 30°C. Upper and lower panels contain factor mutant strains with and without endogenous *IMD2* ORF (*imd2*Δ::*HIS3*), respectively. For convenience of comparison, spots of WT strain with *imd2*Δ::*HIS3* (second on the upper panel) is repeatedly presented next to factor mutants with *imd2*Δ::*HIS3* (top of the lower panel). In the absence of endogenous *IMD2*, most tested factor mutants show sensitivity to MPA at 1 μg/ml MPA treatment, which is comparable to WT sensitivity (upper panel highlighted red box). Addition of guanine suppresses MPA sensitivity for most of the factor mutants (highlighted red box, lower panel). (B) 10-fold serial dilutions of saturated cultures of indicated factor mutants plated on synthetic medium lacking histidine for comparison of growth at various concentrations of MPA treatment at 30°C. 3-AT treatment as in Figure 8 to assess level of *HIS3* expression. *spt4*Δ confers constitutive His^+ phenotype (upper panel highlighted red box), while *spt3*Δ cannot show a His^+ phenotype at the highest concentration of MPA used (upper panel highlighted black box). Addition of guanine suppresses MPA sensitivity of all factor mutants except *spt3*Δ (lower panel highlighted red box).


A
SC-Ura
SC-Ura + 0.04 µg/ml MPA
SC-Ura + 0.1 µg/ml MPA
SC-Ura + 0.2 µg/ml MPA
SC-Ura + 1 µg/ml MPA
SC-Ura + 5 µg/ml MPA
SC-Ura + 20 µg/ml MPA
WT
imd2Δ::HIS3
dst1Δ
spt3Δ
spt4Δ
gal11Δ
sgf73Δ
paf1Δ
bur2Δ
pop2Δ
imd2Δ::HIS3
dst1Δ
spt3Δ
spt4Δ
gal11Δ
sgf73Δ
paf1Δ
bur2Δ
pop2Δ
imd2Δ::HIS3
SC-Ura + 10 mM NaOH
SC-Ura + 10 mM HCl
SC-Ura + NaOH/HCl (10 mM)
SC-Ura + NaOH/HCl + 20 µg/ml MPA
SC-Ura + NaOH/HCl + 1 µg/ml MPA + Guanine (500 µM)
SC-Ura + NaOH/HCl + 5 µg/ml MPA + Guanine (500 µM)
SC-Ura + NaOH/HCl + 20 µg/ml MPA + Guanine (500 µM)
WT
imd2Δ::HIS3
dst1Δ
spt3Δ
spt4Δ
gal11Δ
sgf73Δ
paf1Δ
bur2Δ
pop2Δ
imd2Δ::HIS3
dst1Δ
spt3Δ
spt4Δ
gal11Δ
sgf73Δ
paf1Δ
bur2Δ
pop2Δ
B
SC-His
SC-His + 2 mM 3AT
SC-His + 2 mM 3AT 0.04 µg/ml MPA
SC-His + 2 mM 3AT 0.1 µg/ml MPA
SC-His + 2 mM 3AT 0.2 µg/ml MPA
WT
imd2Δ::HIS3
dst1Δ
spt3Δ
spt4Δ
gal11Δ
sgf73Δ
paf1Δ
bur2Δ
pop2Δ
imd2Δ::HIS3
SC-His + 0.04 µg/ml MPA
SC-His + 0.1 µg/ml MPA
SC-His + 0.2 µg/ml MPA
SC-His + 1 µg/ml MPA
SC-His + 5 µg/ml MPA
SC-HiS + HCl/NaOH + 1 µg/ml MPA
SC-His + HCl/NaOH + 1 µg/ml MPA + Guanine (500 µM)
WT
imd2Δ::HIS3
dst1Δ
spt3Δ
spt4Δ
gal11Δ
sgf73Δ
paf1Δ
bur2Δ
pop2Δ
imd2Δ::HIS3

A number of transcription factor mutants have been described as either MPA or 6-AU sensitive with these sensitivities interpreted as indicative of Pol II elongation defects. Such mutants include deletions in known elongation factor genes *dst1*Δ (encodes TFIIS, (Exinger and Lacroute 1992, Riles, Shaw et al. 2004)), *spt4*Δ (subunit of Spt4/Spt5 DSIF, (Hartzog, Wada et al. 1998, Gaillard, Tous et al. 2009, Gaur, Hasek et al. 2013)), *paf1*Δ (Paf1C complex member, (Desmoucelles, Pinson et al. 2002, Squazzo, Costa et al. 2002)), *bur2*Δ (P-TEFb subunit homolog, (Gaillard, Tous et al. 2009)), *pop2*Δ (encodes Ccr4/NOT complex member, (Denis, Chiang et al. 2001, Riles, Shaw et al. 2004)) or deletions in genes encoding subunits of transcriptional coactivator complexes *spt3*Δ*, sgf73*Δ (SAGA, (Desmoucelles, Pinson et al. 2002, Riles, Shaw et al. 2004)) or *gal11*Δ (Mediator, (Riles, Shaw et al. 2004)). It has previously been demonstrated for some of these mutants, including *dst1*Δ, *spt3*Δ and *paf1*Δ, that *IMD2* induction is defective (Desmoucelles, Pinson et al. 2002, Riles, Shaw et al. 2004), indicating differential GTP starvation upon drug treatment for these strains relative to WT, just as for the Pol II mutants described above. We characterized these strains' sensitivity to MPA treatment in presence or in the absence of endogenous *IMD2* (**Figure 2-14**). We observed that some of the mutants including *sgf73*Δ and *pop2*Δ were only slightly sensitive to MPA in presence of *IMD2* (**Figure 2-14**A, top panels). However, in the absence of *IMD2,* only *spt3*Δ showed MPA hypersensitivity relative to WT, with *dst1*Δ showing only slight sensitivity relative to WT. We next examined suppression of observed MPA sensitivities by guanine supplementation (**Figure 2-14**A, bottom panels). We observed that *spt3*Δ was entirely suppressed by guanine, *dst1*Δ mostly suppressed,

and *spt4*Δ, *paf1*Δ, and *gal11*Δ much less so. We hypothesize that *spt3*Δ hypersensitivity relates to defects in *TPO1* expression, which is known to modulate MPA sensitivity (Desmoucelles, Pinson et al. 2002). Taken together, these results are inconsistent with putative global Pol II transcription defects being a critical determinant for cell growth under GTP limitation. We further examined if *IMD2* promoter function in response to MPA were intact in the cohort of transcription factor mutants tested (**Figure 2-14**B). We observed a His^+ phenotype consistent with constitutive expression of *imd2*Δ*::HIS3* in *spt4*Δ. This phenotype suggests possible altered initiation in *spt4* mutants as observed in LOF Pol II mutants (**Figure 2-12**C). Furthermore, we observed that the His^+ phenotype was inducible in the presence of MPA for all factor mutants except *spt3*Δ. These results indicate that presumptive GTP sensing is maintained in most mutants, but that *spt3*Δ cells have a distinct defect not observed for other factor mutants or Pol II mutants. This defect indicates an almost complete inability to induce the *IMD2* promoter::*HIS3* reporter, although *spt3*Δ is known to express *IMD2* upon MPA treatment at a very low level in presence of endogenous *IMD2* (Riles, Shaw et al. 2004).

Pol II catalytic mutants do not abolish the response to GTP depletion but derange TSS usage at *IMD2* promoter

By removing the possibility of differential *IMD2* expression complicating the MPA response (using *imd2*Δ), we showed above that many factor mutants or Pol II catalytic mutants have similar responses to MPA treatment. These results indicated Pol II transcription is not especially sensitive to presumptive GTP starvation *in vivo* relative to other pathways that rely on GTP (**Figure 2-12** and **2-14**). To better understand the

mechanism of *IMD2* expression defects in the absence of differential GTP starvation, we further analyzed transcriptional responses at the *IMD2* promoter upon MPA treatment. Using the same *imd2*Δ::*HIS3* reporter construct used above, we analyzed the kinetics of TSS utilization upon addition of a concentration of MPA that induces a TSS shift at the *IMD2* promoter (**Figure 2-15**A). Pol II has been proposed to directly sense GTP levels through its active site, and through this sensing Pol II catalytic activity controls *IMD2* TSS usage (Kuehner and Brow 2008). Under this model, it would be predicted that LOF Pol II mutants might show precocious downstream TSS usage; indeed, this behavior has previously been observed (Kaplan, Jin et al. 2012). Conversely, under this model, catalytically GOF Pol II mutants would show delayed TSS shifting kinetics due to their relative insensitivity to reduced GTP levels.

We observed that both LOF (H1085Y) and GOF (E1103G) mutants gradually lose upstream 'G' sites upon MPA treatment (**Figure 2-15**B), indicating retention of ability to respond to GTP-depletion. As expected, WT cells rapidly lose upstream 'G' TSSs and subsequently gain the downstream functional 'A' TSS (**Figure 2-15**C). H1085Y constitutively uses the downstream functional 'A' TSS and shows increased usage of this site upon MPA treatment, while showing decreased kinetics of loss of upstream TSSs. This result is unexpected from the "defective GTP sensing" model where kinetics of upstream TSS loss are predicted to be faster than WT, not slower for Pol II LOF mutants (**Fig 2-15**B and C). E1103G shows reduced kinetics of loss of upstream site, consistent with defective GTP sensing, but shifts TSS usage to novel TSSs predicted to be nonfunctional due to their position upstream of the *IMD2* Nab3/Nrd1-

dependent terminator (**Figure 2-15**B and C). These novel TSSs are upstream of the normal -106 A *IMD2* TSS, and their appearance in GOF E1103G are consistent with E1103G shifting TSSs upstream at other genes (Kaplan, Jin et al. 2012). If terminated by the Nab/Nrd pathway, putative non-functional novel TSSs should produce cryptic unstable transcripts (CUTs) that are degraded by the nuclear exosome (Davis and Ares 2006, Steinmetz, Warren et al. 2006). Such *IMD2* CUTs have previously been observed in *sen1*, *nab3* and *rpb11* mutants (Kuehner and Brow 2008). Indeed, deletion of the exosome subunit Rrp6 also stabilized CUTs derived from the novel TSSs we observe, with significant utilization of these TSSs in E1103G compared to WT (**Figure 2-15**D and **Figure 2-16**A-B). Stabilization of these CUTs in *rrp6*Δ strains did not provide MPA resistance for MPA sensitive Pol II GOF mutants, as these transcripts do not likely produce functional *IMD2* transcript required for resistance even if stabilized (**Figure 2-16**C). We note that although E1103G generally appears unable to utilize downstream functional 'A' TSS upon MPA treatment, enough functional transcript was made to produce a His^+ phenotype in the presence of *imd2*Δ*::HIS3* at low concentrations of MPA treatment (**Figure 2-14**B). Our results here provide an initiation-based mechanism for the MPA sensitivity of Pol II GOF mutants.

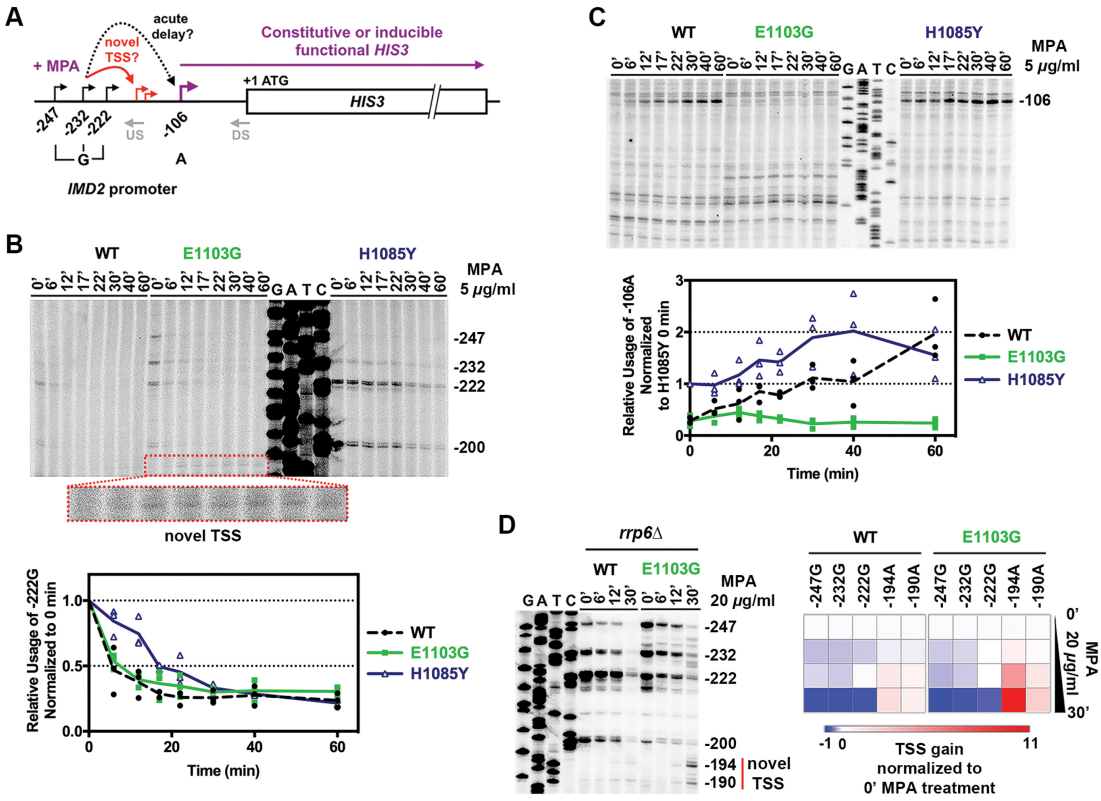
Discussion

Examination of Pol II catalytic mutants by our lab and others has shown Pol II activity-sensitive aspects of transcription initiation, elongation, co-transcriptional processes, and termination (Kaplan, Jin et al. 2012, Braberg, Jin et al. 2013, Hazelbaker, Marquardt et al. 2013, Fong, Kim et al. 2014, Jin and Kaplan 2014, Fong, Brannan et al.

2015). Our current study utilizes a set of yeast Pol II catalytic mutants to probe broad aspects of gene expression *in vivo*. We show that Pol II catalytic defects lead to a decrease in overall Pol II occupancy with reduction in 3′ end occupancy on a galactose inducible reporter gene, *GAL1*p::*YLR454w* (**Figure 2-1**). Further, using a constitutively expressed *TEF1*p::*YLR454w* reporter we find that growth on different carbon sources may alter Pol II occupancy profiles (**Figure 2-1**).

Previously, several observations suggested alterations of Pol II activity and gene expression due to external perturbations such as carbon source or temperature (Pelechano, Jimeno-Gonzalez et al. 2009, Miguel, Monton et al. 2013). Likewise, environmental stress has also been shown to affect Pol II activity and gene expression (Canadell, Garcia-Martinez et al. 2015). Notably, it has also been shown that RAS/PKA signaling pathway, which controls aspects of glucose signaling, can target proteins associated with general transcription machinery, can putatively regulate elongating Pol II by targeting Spt5/4 (Howard, Chang et al. 2001, Howard, Hester et al. 2003), and shows interactions with Nab3/Nrd1 termination factors (Darby, Serebreni et al. 2012). In addition to differential effects of carbon source on steady state Pol II occupancy, we also observe impaired galactose induction and glucose repression kinetics in both LOF and GOF mutants (**Figure 2-2 and 2-10**). Interestingly, both classes of catalytic mutants show impaired growth at 37˚C on galactose likely arising from the sum of a number of distinct individual defects in initiation, elongation, and termination (**Figure 2-4**B). Taken together, these results support growth condition-dependent modulation Pol II elongation *in vivo.*

Figure 2-15. Pol II catalytic mutants do not abolish response to GTP depletion. (A) Schematic of the *imd2*Δ::*HIS3* construct used to determine WT and Pol II catalytic mutants' response to nucleotide depletion in absence of endogenous *IMD2*. Architecture of the *IMD2* promoter indicating the upstream 'G' start sites that are used under normal GTP levels to produce non-functional *IMD2* cryptic unstable transcripts (CUTs). Downstream −106 'A' is used upon presumptive GTP-depletion in the presence of MPA, producing a functional transcript that confers MPA resistance if the *IMD2* ORF is present. Slow catalytic mutants constitutively show use of both upstream and downstream start sites. Gray 'US' (Upstream) and 'DS' (Downstream) indicate positions of primers used for primer extension (PE) experiments. Two models for MPA-sensitivity of Pol II mutants are shown: TSS shift to inappropriate novel TSS, or acute delay in shift to appropriate TSS. (B) Both GOF and LOF mutants lose upstream 'G' TSSs upon MPA treatment. Time courses showing the usage of upstream 'G' sites upon MPA treatment (US primer used for PE). Cells grown to mid-log phase in synthetic complete medium at 30°C were treated with MPA (5 μg/ml final concentration) and RNA was isolated at indicated time points. Pre-treatment sample was used for time 0. Graph shows the relative usage of −222G site for WT and mutants upon MPA treatment normalized to 0 min (no-treatment). (C) E1103G is unable to shift TSS usage to the downstream functional 'A' site upon GTP depletion. Same RNAs from Figure 10B used for PE experiment with DS primer. Graph shows the relative gain of the −106A TSS for WT and mutants upon MPA treatment normalized to H1085Y 0 min (no-treatment), which is used constitutively. (D) Fast catalytic mutants gradually lose upstream 'G' sites and shift TSS usage to novel 'A' sites (putatively −194A and −190A) in response to nucleotide depletion. Usage of these sites is likely to produce non-functional CUTs, shown to be stabilized here by the deletion of exosome subunit gene *RRP6*. Heat map represents the average loss/gain of indicated TSS upon MPA treatment obtained from three independent repeats (see **Figure 2-16**).


A
acute delay?
novel TSS?
+ MPA
Constitutive or inducible functional HIS3
+1 ATG
HIS3
-247
-232
-222
-106
US
DS
G
A
IMD2 promoter
B
WT
E1103G
H1085Y
MPA 5 μg/ml
G A T C
-247
-232
-222
-200
novel TSS
Relative Usage of -222G Normalized to 0 min
Time (min)
C
-106
Relative Usage of -106A Normalized to H1085Y 0 min
D
rrp6Δ
MPA 20 μg/ml
-194
-190
novel TSS
-247G
-232G
-222G
-194A
-190A
0'
30'
TSS gain normalized to 0' MPA treatment

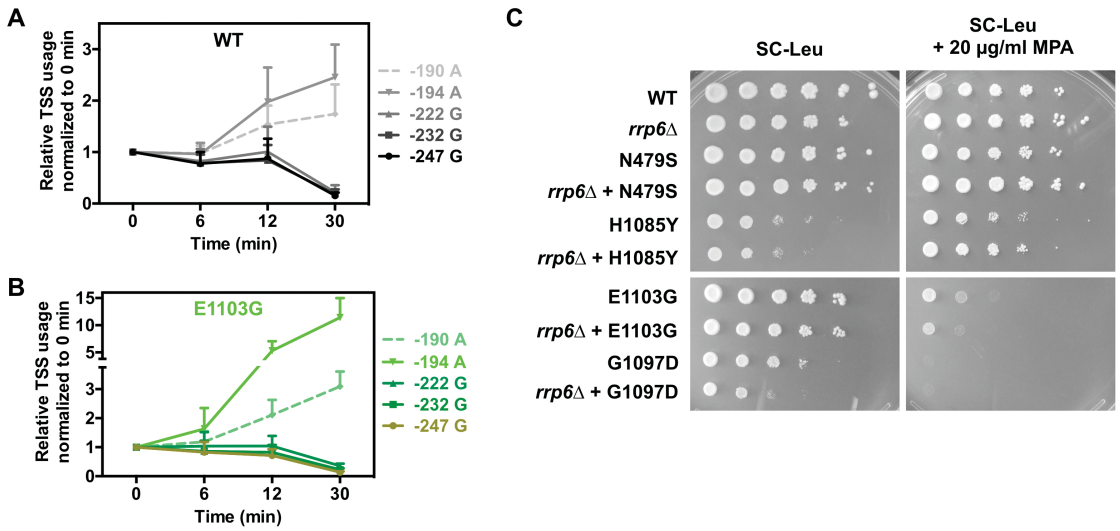


Figure 2-16. Stabilization of *IMD2* CUTs allows *IMD2* TSSs at intermediate downstream positions to be observed. Kinetics of upstream 'G' start site loss and subsequently gain of presumptive novel 'A' sites upon MPA (20 µg/ml) treatment in WT (A) and E1103G (B) cells containing *rrp6*Δ. Error bars represent average of three biological repeats +/- SD. Note difference in scale between (A) and (B). (C) Deletion of *RRP6* does not confer MPA resistance to MPA-sensitive mutants.10-fold serial dilution of saturated cultures of Pol II mutants strains alone and in combination with *rrp6*Δ mutant plated on synthetic complete medium lacking leucine (SC-Leu) and SC-Leu medium containing 20 µg/ml MPA (final) to determine MPA sensitivity.

Pol II mutant *rpb1* E1103G, which was reported previously to be fast for elongation *in vivo* relative to WT (Hazelbaker, Marquardt et al. 2013), appears slower than WT in our study. *In vitro* biochemical studies, including direct observations of individual polymerases at the single molecule level, have repeatedly shown faster elongation rate for E1103G compared to WT (Malagon, Kireeva et al. 2006, Kaplan, Jin et al. 2012, Larson, Zhou et al. 2012, Dangkulwanich, Ishibashi et al. 2013). Several lines of evidence suggest E1103G *in vitro* GOF activity manifests itself *in vivo* through transcriptional effects distinct from LOF mutants. First, primer extension analysis shows that E1103G shifts TSS usage upstream at a number of promoters, consistent with a model that increase in catalytic activity increases initiation probability during Pol II promoter scanning, resulting in observed upstream TSS shifts (Kaplan, Jin et al. 2012, Braberg, Jin et al. 2013). Second, E1103G and other Pol II GOF mutants show allele-specific genetic interactions with a number of factors, suggesting their defects are distinct from LOF mutants (Braberg, Jin et al. 2013). For example, the synthetic sick genetic interaction of E1103G with the RNA processing factor alleles *rat1-1* and *xrn1*Δ/*rat1-1* is consistent with exacerbation of termination defects through faster elongation (Jimeno-Gonzalez, Haaning et al. 2010, Hazelbaker, Marquardt et al. 2013, Jimeno-Gonzalez, Schmid et al. 2014, Fong, Brannan et al. 2015), and in contrast to the suppression of *rat1-1* and *rat1-1/xrn1*Δ by LOF Pol II alleles. Indeed, delayed termination for GOF E1103G was evident through observation of increased length of pre-processed snR33 (**Figure 2-9**). We further show that an even stronger GOF mutant, G1097D, shows a corresponding increase in pre-processed snR33 length relative to

E1103G or WT. While increased length of pre-processed snR33 is consistent with increased elongation rate in GOF mutants, defective termination due to altered probability of a Pol II conformational change should not be ruled out as possible mechanism (Zhang, Rigo et al. 2015). Finally, a recent study demonstrated that an E1103G strain displays a shift in co-transcriptional splicing towards downstream positions, consistent Pol II traveling further downstream prior to splicing, though delayed splicing kinetics cannot be ruled out (Carrillo Oesterreich, Herzel et al. 2016). Altogether, these data are consistent with increased elongation kinetics in GOF catalytic mutants *in vivo* under a parsimonious view, though they do not explain the delayed Pol II runoff from *YLR454w* observed here.

Our results suggest that the commonly used glucose shutoff assay for determining *in vivo* elongation rate may be confounded by previously unappreciated limitations and biological complexity. We also identify a possible artifact of prior experimental design that may have contributed to discrepant results reported for GOF mutant E1103G. First, we have shown that galactose starvation induces as fast a transcriptional shutoff at *GAL1*p::*YLR454w* as glucose addition (**Figure 2-7**F). Second, this galactose-depleted transcriptional shutoff affects Pol II E1103G differently than WT. During the wash, E1103G Pol II runs off both the 5′ and 3′ ends of the reporter faster than WT (**Figure 2-7**F-G). 5′ runoff is consistent with E1103G elongating faster than WT early in the time course, or E1103G being more sensitive to galactose starvation. The increased runoff kinetics from the 3′ end of the reporter in E1103G is perplexing. Furthermore, after these early effects observed for E1103G relative to WT

under galactose starvation, E1103G Pol II shows a delay in running off the template at longer time points regardless of mechanism of shutoff, just as purported LOF Pol II alleles do under glucose inhibition. From these results we can only conclude that Pol II could be differentially affected while on the 3′ end of the gene relative to the middle under different growth conditions. Further, we identified slow Mig1p nuclear translocation upon glucose exposure as a possible confounding variable for use of *GAL1*p::*YLR454w* to determine *in vivo* elongation rate. There may be additional untested steps in inhibition specific to the *GAL* system that are defective, or defects specific to the long *YLR454w* template for Pol II GOF strains. Regardless, careful consideration of possible variables in assay behavior is urged given the results we present.

In order to exploit the existing transcription reporter systems to probe transcription mechanisms carefully, we need a detailed understanding of each phenotypic system, such as our analyses to dissect the Spt$^-$ phenotype of Pol II GOF alleles (Cui, Jin et al. 2016). Here we investigated Pol II catalytic mutants and several transcription factors mutants' response to the nucleotide-depleting drug MPA under conditions where differential *IMD2* expression was obviated. Most tested Pol II mutants and many transcription factor mutants behave similarly to WT upon MPA treatment of *imd2*Δ strains (**Figure 2-12** and **2-14**). Our results also suggest *IMD2*-independent mechanisms of MPA sensitivity, which have been observed in previous large-scale deletion screens for MPA or 6-AU sensitivity, and have identified Pol II mutants and transcription factor mutants that do not affect *IMD2* transcription (Desmoucelles, Pinson et al. 2002, Riles, Shaw et al. 2004). Our previous analyses identified a correlation

between MPA sensitivity of Pol II GOF alleles and upstream shifts in TSS usage at *ADH1* (Kaplan, Jin et al. 2012, Braberg, Jin et al. 2013). As *IMD2* regulation proceeds by initiation shifting from upstream non-productive TSSs to a downstream productive one, it is conceivable that initiation defects underlie Pol II mutant MPA sensitivity. However, in the presence of endogenous *IMD2* it is difficult to determine mechanisms of MPA sensitivity of GOF mutants, as they would be differentially starved for GTP due to differential expression of functional *IMD2* upon MPA treatment. One model for MPA sensitivity based on differential sensitivity to reduced GTP is that increased catalysis rate of GOF alleles might buffer reduction in substrate (GTP) levels. This could lead to delay in sensing low GTP and switching *IMD2* upstream GTP-initiated TSSs to downstream ATP-initiated TSSs, causing acute GTP starvation possibly beyond a critical threshold for growth/viability. Our results disfavor this model. Here we present evidence that MPA sensitivity of the Pol II GOF allele E1103G correlates with usage of novel TSSs that are intermediately positioned between the known productive -106 A TSS at *IMD2* and upstream non-functional starts. We conclude that *IMD2* defects in Pol II GOF E1103G are likely to derive from initiation defects (**Figure 2-15**).

Our Pol II catalytic mutants and pre-mRNA processing factor show allele-specific genetic interactions (**Figure 2-3**). Synthetic sick interactions of GOF with *xrn1*Δ and stabilization of the *GAL1*p::*YLR454w* transcript in the *xrn1*Δ/*rat1-1* double mutant background is suggestive of a 5′-end processing defect in GOF mutants, possibly a capping defect. A defect in capping would be predicted to expose mRNA to the action of 5′ exonucleases Xrn1 and Rat1. Capping of nascent transcripts occurs co-

transcriptionally with Pol II elongation proposed to be coupled to successful capping, though a direct "capping checkpoint" has yet to be shown (Rasmussen and Lis 1993, Glover-Cutter, Kim et al. 2008). Potential sensitivity of capping to Pol II GOF mutants is suggestive of either a defective checkpoint in Pol II mutant strains or the absence of one. We also observe correlations of reporter transcript half-life with expression levels in Pol II mutants, and with Pol II mutants' growth defects (**Figure 2-6**). Observed correlations between mRNA decay rate with both expression and growth rate is consistent with recent findings (Garcia-Martinez, Delgado-Ramos et al. 2016). 'Feedback' of gene expression control between the abundance of mRNA and mRNA decay rates is proposed to occur in response to defects in transcription synthesis (Haimovich, Medina et al. 2013, Sun, Schwalb et al. 2013). Alternatively, as growth efficiency is connected to overall translational demand, mRNA decay 'feedback' may occur upon alteration of overall translation rate, as most mRNA decay happens co-translationally (Hu, Sweet et al. 2009, Hu, Petzold et al. 2010, Pelechano, Wei et al. 2015). Given that Pol II mutants show reduced apparent initiation based on overall Pol II occupancy and reduced growth rates, it is difficult to deconvolute primary and secondary global effects. Although we looked at both phenomena using a single reporter transcript, mutants described here could be used further to probe these proposed mechanisms in large scale.

Here we extensively characterize a set of Pol II catalysis mutants for *in vivo* consequences. We show that altered Pol II catalysis affects Pol II occupancy, putative elongation, and reporter gene expression and decay rate *in vivo*. Notably, we interrogate two widely used elongation reporter systems, raising caveats about their use and

interpretation. For use of nucleotide depleting drugs MPA or 6-AU, we constructed and tested a useful novel reporter system (*imd2*Δ::*HIS3*), which can be further utilized to characterize or screen for new mutants that shift TSS usage downstream, leading to constitutively expression of *imd2*Δ*::HIS3*. Development of approaches allowing more direct determination of *in vivo* elongation rate will bypass issues identified here. Recent advances in high-resolution microscopy have enabled real-time observation of all transcription phases on endogenous genes using fluorescently labeled proteins that bind to nascent transcript (Larson, Zenklusen et al. 2011, Hocine, Raymond et al. 2013). This approach could be to address how any number of variables might modulate elongation such as template sequence, RNA secondary structure *etc*, and likely represents the next steps toward understanding transcription elongation and co-transcriptional processes *in vivo*.

CHAPTER III

CONCLUSIONS AND FUTURE DIRECTIONS

Recently, RNA polymerase II (Pol II) catalytic mutants have been used extensively to probe transcription mechanisms and co-transcriptional processes. In order to exploits these mutants in an unbiased way, we need a detailed understanding about how these mutants behave *in vivo* and how the experimental systems used to study these mutants work. In this dissertation, I presented molecular analyses of how alterations to Pol II catalysis rate affect several facets of transcription and gene expression in budding yeast (*Saccharomyces cerevisiae*). I explored the influence of Pol II catalytic activity *in vivo* through genetic and molecular approaches utilizing specific alleles of the Pol II largest subunit, Rpb1, that confer slower or faster catalytic rate than WT in vitro. Finally, based on my observations, we proposed to exploit a new tool (single cell/gene system) for studying transcription elongation in yeast.

Conclusions

My studies have shown that alterations to Pol II catalytic activity lead to a wide range of consequences *in vivo*. Using the widely-used galactose inducible *GAL1p::YLR454w* reporter system (Mason and Struhl 2005), I have shown that alteration in Pol II catalytic activity, both slower and faster than WT, leads to a decrease in overall *YLR454w* Pol II occupancy with relatively greater reduction at the 3' end of *YLR454w*, suggesting an apparent processivity defect (Chapter II). Further, I determined that Pol II occupancy could be differentially affected in Pol II mutants when grown on

different carbon sources, suggesting external perturbations such as carbon source might have global effects on transcription. I found that expression of *GAL1p::YLR454w* reporter is decreased to different extent in different Pol II catalytic mutants, although the reporter mRNA is relatively more stable in all tested Pol II catalytic mutants. I found that fast catalytic mutants generally show synthetic sick genetic interactions with pre-mRNA processing/surveillance factors. Further, in fast mutants, depletion of nuclear 5' exonucleases/surveillance factors led to possible stabilization of impaired or improperly processed pre-mRNAs, which may be otherwise degraded in the presence of these surveillance factors. Together, these observations led to the conclusion that fast mutants may have impaired pre-mRNA processing defects, hypothetically a capping defect. Finally, I comprehensively examined the two widely used systems for studying transcription elongation in yeast – a ChIP assay to determine 'last wave' of Pol II after transcriptional shutoff, and elongation mutant sensitivity to a nucleotide depleting drug. For both assays, I found previously unaccounted variables that differentially affect WT and presumptive elongation mutants.

Although two classes of mutants (fast and slow) tested in this study have distinct genetic and biochemical properties, both classes of mutant showing similar phenotypes for Pol II occupancy and processivity indicates that different defects can lead to similar *in vivo* outcome. Such defects may include – i. Pol II pausing, stalling and degradation of the elongation complex, ii. premature termination and degradation of the elongation complex, and iii. differential elongation rate on different part of the template. First, fast Pol II catalytic mutants are known to increase misincorporation rate, which denotes

selecting and incorporating mismatched substrate, both *in vitro* and *in vivo*. Although cells have the ability to counter misincorporations by backtracking the polymerase from the active site and cleaving the nascent RNA to restore a properly aligned elongation complex (discussed in Chapter I), increased misincorporation can lead to extended pausing or stalling of the elongation complex on the template. Indeed, in the absence of TFIIS, a factor involved in rescuing Pol II from the backtracked state, the frequency of genome wide Pol II pausing increases in yeast as shown by increased probability of Pol II occupancy in NET-seq experiment (James, Gamba et al. 2017). As noted earlier (Chapter I), stalled elongation complexes (EC) can be disassembled by ubiquitination and degradation of the largest subunit of Pol II, Rpb1. Hence, it can be hypothesize as possible that the decrease in Pol II occupancy at 3' end of the reporter in fast mutants is due to increased misinorcoration induced Pol II stalling and degradation of EC. Slow mutants can also possibly lead to an increase in Pol II pausing or stalling, but by a different mean. For example, a recent study has shown that, in human cells, slow Pol II elongation repositions CTD Ser2 phosphorylation from 3′ end towards 5′ end of the genes (Fong, Saldi et al. 2017). This differential phosphorylation may presumably lead to uncoupling of elongation associated factors from slow elongating Pol II, leading to Pol II pausing or stalling. Secondly, a recent *in vitro* study suggests that misincorporation-induced Pol II pausing can lead to termination of Pol II by Rat1p/Rai1p, and the efficiency of such termination increases with the increasing length of RNA (Park, Kang et al. 2015). Thus, it is possible that increased misincorporation in Pol II catalytic mutants leads to pausing and termination of Pol II at the 3' end of the

reporter, resulting in decreased 3' occupancy. Further, this mode of termination can possibly also occur due to synthesis of an improperly capped pre-mRNA, wherein Rat1p/Rai1p degrades improperly capped nascent RNA and dissociates Pol II from the template. Both of these possibilities (misincorporation and improper-capping driven termination) indicate a novel quality control mechanism by Rat1p and can be impaired in the Pol II mutants. Finally, ChIP occupancy changes can be interpreted in multiple ways due to the fact that ChIP cannot differentiate between actively transcribing Pol II from paused/stalled Pol II. For example, a faster elongation rate at the 3' end of a gene will lead to a lower steady state Pol II occupancy at the 3' end compared to the 5', resulting in an apparent (but not real) processivity defect. Similarly, overall occupancy also relates to the average time Pol II spends on the template (*in vivo* elongation rate). Thus, if the fast catalytic mutants transcribe much faster than WT across the template, the steady state Pol II occupancy on the template is expected to be lower for the fast mutants. However, it is not possible to distinguish between decreased occupancy and faster elongation rate, as the system to measure *in vivo* elongation rate in yeast has confounding variables, which I discuss further below.

As noted earlier (Chapter I), a widely used assay to determine *in vivo* elongation rate of Pol II in yeast uses glucose repression of a galactose inducible reporter, *GAL1p::YLR454w* (**Figure 3-1**). In my study, I observed an apparent slower elongation rate for all tested fast mutants compared to WT in this glucose shutoff assay. Glucose repression of *GAL* genes functions through nuclear localization and binding of transcription factor Mig1p to the promoters of glucose-repressed genes (**Figure 3-1**).

Snf1 kinase is a key player in this pathway. In the absence of glucose, active Snf1p kinase phosphorylates Mig1p, leading to the cytoplasmic localization of Mig1p. Upon addition of glucose to the medium, Glc7p/Reg1p dephosphorylates and inactivates the Snf1 kinase, leading to dephosphorylation of cytoplasmic Mig1p, which is then imported to the nucleus. In the nucleus, Mig1p binds to the target promoters of glucose-repressed genes and recruits co-repressors to repress *GAL* gene transcription. I found that, in at least some Pol II catalytic mutants, glucose-induced nuclear localization of Mig1p is impaired, possibly leading to a delayed shutoff of the *GAL1p::YLR454w* reporter. This observation revealed a limitation of the shutoff assay, but only partially explained apparent slower *in vivo* elongation rate of the fast mutants in this system. The impaired glucose signaling is expected to delay the shutoff, leading to a delayed Pol II runoff from the 5′ end of the reporter. However, I was unable to measure the kinetics of Pol II runoff at the 5′ end of *GAL1p::YLR454w* reporter, because the time required for Pol II leaving the 5′ (for example, 1 kb amplicon in ChIP assay) is very short and there were not enough data points to model the kinetics**.** Although, impaired Mig1p signaling can delay the shutoff at 5′ end, why fast Pol II mutant occupancy remained high at the 3' end of the reporter in later time points (when shutoff presumably completed) could not be explained by delayed Mig1p signaling. I observed further distinct phenomenon at the 3′ end of the reporter for a fast mutant when I tested shutoff by galactose depletion, which functions to shut off transcription in a way distinct from glucose addition (**Figure 3-1**). The activator Gal4p transcriptionally regulates galactose structural genes, such as *GAL1* and *GAL10*. Gal4p remains bound to the upstream activation sites of these *GAL* genes

and recruits transcriptional machinery to activate them. In the absence of galactose, the transcriptional repressor Gal80p forms a dimer with Gal4p, preventing it from recruiting factors to activate *GAL* genes. However, in the presence of galactose, Gal3p sequesters Gal80p in the cytoplasm, leading to release of Gal4p inhibition and activation of *GAL* genes. To induce galactose starvaion, I washed cells and resuspended them in medium lacking sugar. I observed that fast catalytic mutant E1103G appeared to exist from both the 5′ and 3′ ends early in the time course during the wash, appearing faster than WT as predicted from in vitro biochemistry. However, in the later time points E1103G Pol II accumulated at the 3′ end of *GAL1p::YLR454w* reporter, exiting at much slower rate than WT. Following the shutoff, polymerase would be expected to exit the template in a polar fashion from 5' end to 3' end irrespective of carbon source, however my observation that E1103G exits from both end of the reporter during the galactose wash indicates that Pol II might be differentially affected at the 3' end of this reporter than 5′ in presence of galactose. The Nrd1/Nab3/Sen1(NNS) termination pathway (discussed in Chapter I) can be hypothesized to be involved in such differential regulation. NNS pathway associates with elongating Pol II and controls spurious transcriptions in yeast genome by terminating pervasive transcriptions. Termination of Pol II by NNS can presumably affect Pol II processivity. Interestingly, some previous studies have shown that different carbon sources differentially affect the binding of Nrd1/Nab3 to the target short RNAs, indicating differential affect on NNS function and therefore possibly on strength of NNS surveillance on Pol II transcription. Interestingly, in my experiments, I observed Pol II catalytic mutants show a differential Pol II occupancy in different

growth conditions and both LOF and GOF mutants show impaired growth on galactose medium at 37˚C. These results along with the 3' end specific effects I observed in the glucose shutoff assay indicate carbon sources may have a number of distinct individual defects on Pol II mutants.

Genetic interactions between Pol II mutants and RNA exonucleases indicate fast catalytic mutants may have 5′ RNA processing and/or termination defects. Pol II fast catalytic mutants show synthetic sick genetic interaction with *rat1-1*, *xrn1*Δ and *rat1-1/ xrn1*Δ double mutants. A synthetic sick interaction of fast mutants with *rat1-1* is consistent with a termination defect. Additionally, synthetic sick interaction can be due to pre-mRNA 5′ processing defect, as Rat1p is involved in both termination and 5′ surveillance processes. However, as Xrn1p is only involved in 5′ surveillance, synthetic sick interactions of fast catalytic mutants with *xrn1*Δ suggest that fast mutants may have 5′ processing defect, possibly a capping defect. However, the relationship of Pol II with RNA exonucleases could be more complex. For example, a cap-defective transcript is expected to have a decreased mRNA stability, as it is less stable than a capped transcript. In contrast, I observed an increased mRNA stability of a reporter in all tested fast catalytic mutants. However, defective capping can be linked to other processes such as translation. Thus, we can think of possibilities that the mRNA degradation rate of this reporter is decreased due to cellular mRNA 'buffering' phenomenon through slower translation or growth rate (See Chapter I for description). Since genetic interaction with pre-mRNA processing factors is an indirect evidence for capping defect, a more direct

assessment is needed to determine a possible capping defect in fast catalytic mutants, which I present in the future direction section.

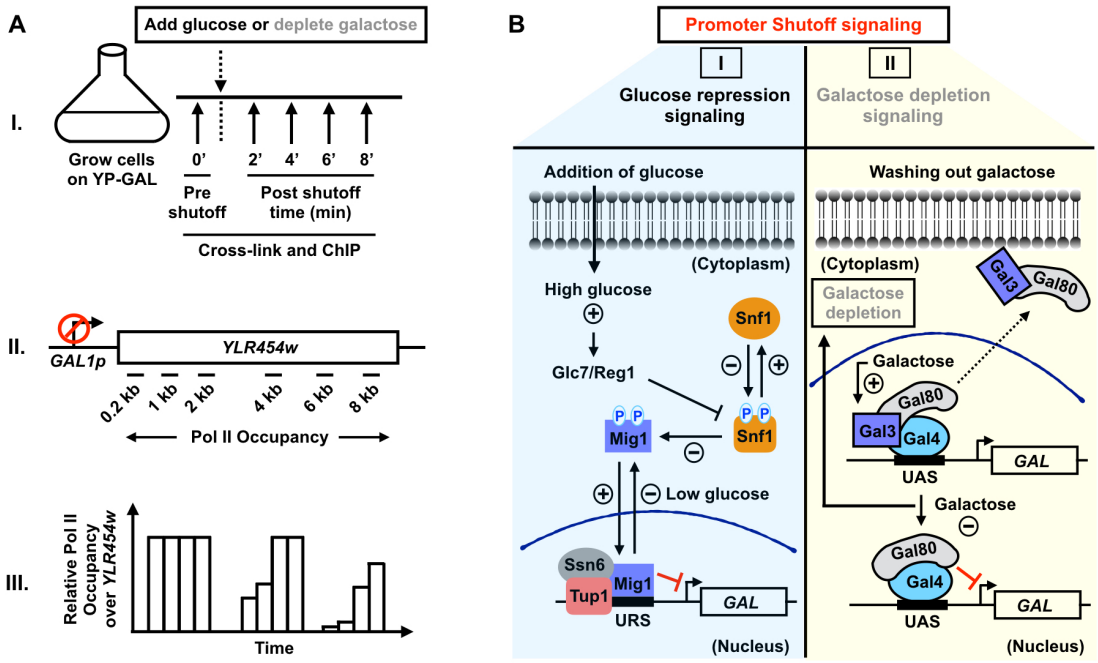


Figure 3-1. Schematic of the shutoff assays and signaling pathways that drive the shutoff of galactose inducible reporters.

Future Directions

In the present study, I examined consequences of altered Pol II catalytic rates using molecular/genetic reporter systems and my results shape the basis for future studies that we can perform on a genome-wide scale or the single gene/cell level. Currently we have multiple future studies envisioned, including – I. Determination of the relationship of capping to Pol II catalytic rate and measuring capping defects in Pol II catalytic mutants on genome-wide scale, and II. Developing single gene/cell technique to study transcription elongation *in vivo*.

Determination of the relationship of capping to Pol II catalytic rate and measuring capping defects in Pol II catalytic mutants

We can take a two-pronged approach to address possible capping defects in Pol II catalytic mutants. First, we plan to test genetic interactions of Pol II catalytic mutants with capping mutants. The addition of a 5' m^7G cap to the nascent RNAs requires recruitment and action three essential enzymes (see Chapter I) - Cet1p (triphosphatase), Ceg1p (guanylultransferase) and Abd1p (guanine N7-methyltransferase). We can make conditional ts^- mutant alleles of these enzymes and test how they genetically interact with Pol II catalytic mutants. Second, we will perform a modified version of the Native Elongating Transcript sequencing (NET-seq) assay to determine the extent of capping defects in Pol II mutants (**Figure 3-2**). We would like to purify Pol II associated total nascent RNAs, then separate capped mRNAs present in Pol II IPs (15-200 nt in length) and specifically map 5′ and 3′ ends of the transcript. We will combine sequencing of 5′ RNA ends with regular NET-seq that maps the 3′ ends (**Figure 3-2**). From the 5′

sequencing data of capped RNAs we will determine Pol II position in the gene when capped 5′ ends appear.

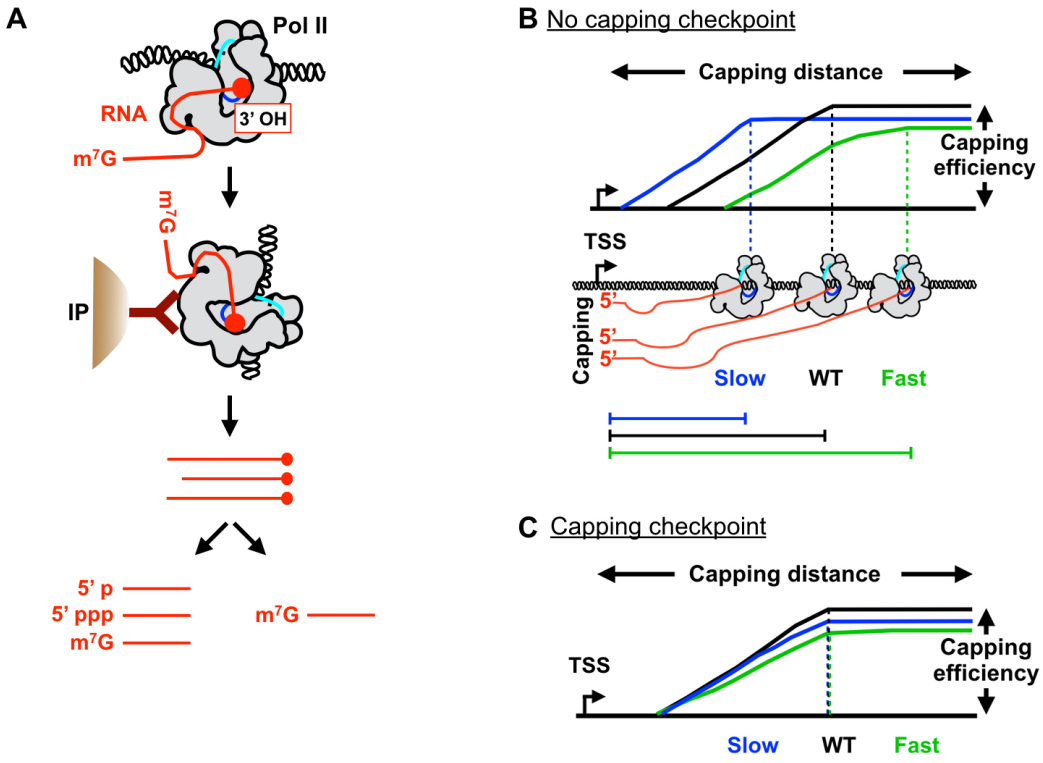


Figure 3-2. Schematic of the modified NET-seq experiment. Schematic (A) and expected position of Pol II on the gene for WT, slow and fast catalytic mutants if the Pol II elongation rate determines where capping occur (B, no capping checkpoint) or if there is capping checkpoint (C).

This measured distance will reveal whether alteration in Pol II activity changes the window within genes where capping occurs. We would like to test if there is any evidence for a capping checkpoint (where mRNA capping will occur in the same window for WT/slow/fast mutants), or Pol II elongation rate determines where capping occurs (where the length of the window for mRNA capping will be shorter or longer for slow and fast mutants, respectively, compared to WT). We can also measure the efficiency of capping by determining the ratio between capped and uncapped mRNAs for the mutants.

Employing single gene/cell technique to study transcription elongation *in vivo*.

Since I found confounding variables that can affect the interpretation of the widely used glucose shutoff assay to study *in vivo* elongation in yeast, we will adopt a more direct method to study elongation in live yeast cells (**Figure 3-3**). Using a system where nascent RNAs can be marked by binding of fluorescent proteins, we will employ a reporter that can bind different fluorescent proteins to 5′ and 3′ binding sites. For this purpose, we will utilize a reporter system with viral RNA stem loops PP7 and MS2 engineered into the 5′ and 3′ UTRs of the reporter, respectively. In this system synthesis of nascent RNAs at the site of transcription can be detected by binding of fluorescently labeled viral coat proteins to the stem loops. We are collaborating with the Larson lab at NIH, who has been extensively working on the PP7/MS2 system in yeast (Larson, Zenklusen et al. 2011, Lenstra, Coulon et al. 2015). Using Dr. Larson's construct we have demonstrated the ability to detect position of dual labeling of a single nuclear spot upon activation of *GAL* genes with galactose, indicating that we will be able to obtain

data using this system at Texas A&M (**Figure 3-3**). We are currently constructing a version of this reporter that contains *YLR454w* gene under *GAL10* promoter. Using the *GAL10p*-PP7-*YLR454w*-MS2 reporter we aim to determine: the transcription initiation/activation frequency (appearance of 5' fluorescent signal over time), Pol II elongation rate (length of the transcript/time between appearance of 5' and 3' fluorescent signals), and Pol II processivity (frequency of 5' fluorescent signal vs 3' signal) in WT and catalytic mutants.

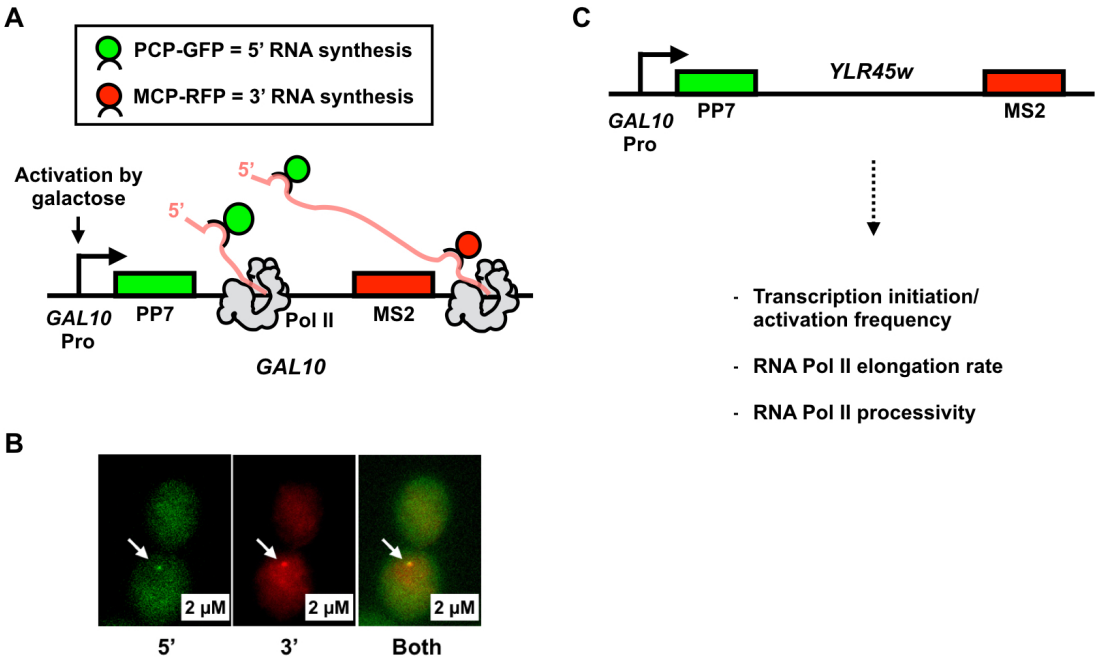


Figure 3-3. Schematic of the dual fluorescent reporter for the single gene/cell transcription assay. Schematic of the reporter (A) and proof of principle experiment that detects 5' and 3' transcript synthesis (B). Schematic of the proposed reporter to study *in vivo* elongation in Pol II catalytic mutants.

REFERENCES

Alekseev, S., Z. Nagy, J. Sandoz, A. Weiss, J. M. Egly, N. Le May and F. Coin (2017). "Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression." Mol Cell **65**(3): 504-514 e504.

Alexander, R. D., S. A. Innocente, J. D. Barrass and J. D. Beggs (2010). "Splicing-dependent RNA polymerase pausing in yeast." Mol Cell **40**(4): 582-593.

Amberg, D. C., D. J. Burke and J. N. Strathern (2005). "Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 Edition (Cold Spring)."

Archambault, J., F. Lacroute, A. Ruet and J. D. Friesen (1992). "Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II." Mol Cell Biol **12**(9): 4142-4152.

Ardehali, M. B. and J. T. Lis (2009). "Tracking rates of transcription and splicing in vivo." Nat Struct Mol Biol **16**(11): 1123-1124.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc Natl Acad Sci U S A **98**(18): 10037-10041.

Bar-Nahum, G., V. Epshtein, A. E. Ruckenstein, R. Rafikov, A. Mustaev and E. Nudler (2005). "A ratchet mechanism of transcription elongation and its control." Cell **120**(2): 183-193.

Barnes, C. O., M. Calero, I. Malik, B. W. Graham, H. Spahr, G. Lin, A. E. Cohen, I. S. Brown, Q. Zhang, F. Pullara, M. A. Trakselis, C. D. Kaplan and G. Calero (2015). "Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble." Mol Cell **59**(2): 258-269.

Bentley, D. L. (2014). "Coupling mRNA processing with transcription in time and space." Nat Rev Genet **15**(3): 163-175.

Beyer, A. L. and Y. N. Osheim (1988). "Splice site selection, rate of splicing, and alternative splicing on nascent transcripts." Genes Dev **2**(6): 754-765.

Birse, C. E., B. A. Lee, K. Hansen and N. J. Proudfoot (1997). "Transcriptional termination signals for RNA polymerase II in fission yeast." EMBO J **16**(12): 3633-3643.

Biswas, D., R. Dutta-Biswas, D. Mitra, Y. Shibata, B. D. Strahl, T. Formosa and D. J. Stillman (2006). "Opposing roles for Set2 and yFACT in regulating TBP binding at promoters." EMBO J **25**(19): 4479-4489.

Blanc, E., P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea and G. Bricogne (2004). "Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2210-2221.

Braberg, H., H. Jin, E. A. Moehle, Y. A. Chan, S. Wang, M. Shales, J. J. Benschop, J. H. Morris, C. Qiu, F. Hu, L. K. Tang, J. S. Fraser, F. C. Holstege, P. Hieter, C. Guthrie, C. D. Kaplan and N. J. Krogan (2013). "From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II." Cell **154**(4): 775-788.

Braun, K. A., S. Vaga, K. M. Dombek, F. Fang, S. Palmisano, R. Aebersold and E. T. Young (2014). "Phosphoproteomic analysis identifies proteins involved in transcription-coupled mRNA decay as targets of Snf1 signaling." Sci Signal **7**(333): ra64.

Braun, K. A. and E. T. Young (2014). "Coupling mRNA synthesis and decay." Mol Cell Biol **34**(22): 4078-4087.

Brueckner, F., K. J. Armache, A. Cheung, G. E. Damsma, H. Kettenberger, E. Lehmann, J. Sydow and P. Cramer (2009). "Structure-function studies of the RNA polymerase II elongation complex." Acta Crystallogr D Biol Crystallogr **65**(Pt 2): 112-120.

Brueckner, F. and P. Cramer (2008). "Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation." Nat Struct Mol Biol **15**(8): 811-818.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T.

Simonson and G. L. Warren (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-921.

Bushnell, D. A., K. D. Westover, R. E. Davis and R. D. Kornberg (2004). "Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms." Science **303**(5660): 983-988.

Cabart, P., H. Jin, L. Li and C. D. Kaplan (2014). "Activation and reactivation of the RNA polymerase II trigger loop for intrinsic RNA cleavage and catalysis." Transcription **5**(3): e28869.

Canadell, D., J. Garcia-Martinez, P. Alepuz, J. E. Perez-Ortin and J. Arino (2015). "Impact of high pH stress on yeast gene expression: A comprehensive analysis of mRNA turnover during stress responses." Biochim Biophys Acta **1849**(6): 653-664.

Carrillo Oesterreich, F., L. Herzel, K. Straube, K. Hujer, J. Howard and K. M. Neugebauer (2016). "Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II." Cell **165**(2): 372-381.

Chabot, B. and L. Shkreta (2016). "Defective control of pre-messenger RNA splicing in human disease." J Cell Biol **212**(1): 13-27.

Chang, J. H., X. Jiao, K. Chiba, C. Oh, C. E. Martin, M. Kiledjian and L. Tong (2012). "Dxo1 is a new type of eukaryotic enzyme with both decapping and 5'-3' exoribonuclease activity." Nat Struct Mol Biol **19**(10): 1011-1017.

Chathoth, K. T., J. D. Barrass, S. Webb and J. D. Beggs (2014). "A splicing-dependent transcriptional checkpoint associated with prespliceosome formation." Mol Cell **53**(5): 779-790.

Chen, H., K. Shiroguchi, H. Ge and X. S. Xie (2015). "Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli." Mol Syst Biol **11**(5): 808.

Cheung, A. C. and P. Cramer (2011). "Structural basis of RNA polymerase II backtracking, arrest and reactivation." Nature **471**(7337): 249-253.

Cheung, A. C. and P. Cramer (2012). "A movie of RNA polymerase II transcription." Cell **149**(7): 1431-1437.

Cheung, A. C., S. Sainsbury and P. Cramer (2011). "Structural basis of initial RNA polymerase II transcription." EMBO J **30**(23): 4755-4763.

Chiu, Y. L., C. K. Ho, N. Saha, B. Schwer, S. Shuman and T. M. Rana (2002). "Tat stimulates cotranscriptional capping of HIV mRNA." Mol Cell **10**(3): 585-597.

Cho, E. J., T. Takagi, C. R. Moore and S. Buratowski (1997). "mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain." Genes Dev **11**(24): 3319-3326.

Chung, W. H., J. L. Craighead, W. H. Chang, C. Ezeokonkwo, A. Bareket-Samish, R. D. Kornberg and F. J. Asturias (2003). "RNA polymerase II/TFIIF structure and conserved organization of the initiation complex." Mol Cell **12**(4): 1003-1013.

Churchman, L. S. and J. S. Weissman (2011). "Nascent transcript sequencing visualizes transcription at nucleotide resolution." Nature **469**(7330): 368-373.

Churchman, L. S. and J. S. Weissman (2012). "Native elongating transcript sequencing (NET-seq)." Curr Protoc Mol Biol **Chapter 4**: Unit 4 14 11-17.

Core, L. J., J. J. Waterfall and J. T. Lis (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." Science **322**(5909): 1845-1848.

Cramer, P., D. A. Bushnell and R. D. Kornberg (2001). "Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution." Science **292**(5523): 1863-1876.

Cui, P., H. Jin, M. R. Vutukuru and C. D. Kaplan (2016). "Relationships Between RNA Polymerase II Activity and Spt Elongation Factors to Spt- Phenotype and Growth in Saccharomyces cerevisiae." G3 (Bethesda) **6**(8): 2489-2504.

Da, L. T., D. Wang and X. Huang (2012). "Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II." J Am Chem Soc **134**(4): 2399-2406.

Dangkulwanich, M., T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev and C. J. Bustamante (2013). "Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism." Elife **2**: e00971.

Danko, C. G., N. Hah, X. Luo, A. L. Martins, L. Core, J. T. Lis, A. Siepel and W. L. Kraus (2013). "Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells." Mol Cell **50**(2): 212-222.

Darby, M. M., L. Serebreni, X. Pan, J. D. Boeke and J. L. Corden (2012). "The Saccharomyces cerevisiae Nrd1-Nab3 transcription termination pathway acts in opposition to Ras signaling and mediates response to nutrient depletion." Mol Cell Biol **32**(10): 1762-1775.

Darzacq, X., Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair and R. H. Singer (2007). "In vivo dynamics of RNA polymerase II transcription." Nat Struct Mol Biol **14**(9): 796-806.

David, C. J., A. R. Boyne, S. R. Millhouse and J. L. Manley (2011). "The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex." Genes Dev **25**(9): 972-983.

Davis, C. A. and M. Ares, Jr. (2006). "Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **103**(9): 3262-3267.

de la Mata, M., C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein, F. Pelisch, P. Cramer, D. Bentley and A. R. Kornblihtt (2003). "A slow RNA polymerase II affects alternative splicing in vivo." Mol Cell **12**(2): 525-532.

de la Mata, M. and A. R. Kornblihtt (2006). "RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20." Nat Struct Mol Biol **13**(11): 973-980.

De Vit, M. J., J. A. Waddle and M. Johnston (1997). "Regulated nuclear translocation of the Mig1 glucose repressor." Mol Biol Cell **8**(8): 1603-1618.

DeLaBarre, B. and A. T. Brunger (2006). "Considerations for the refinement of low-resolution crystal structures." Acta Crystallogr D Biol Crystallogr **62**(Pt 8): 923-932.

Denis, C. L., Y. C. Chiang, Y. Cui and J. Chen (2001). "Genetic evidence supports a role for the yeast CCR4-NOT complex in transcriptional elongation." Genetics **158**(2): 627-634.

Desmoucelles, C., B. Pinson, C. Saint-Marc and B. Daignan-Fornier (2002). "Screening the yeast "disruptome" for mutants affecting resistance to the immunosuppressive drug, mycophenolic acid." J Biol Chem **277**(30): 27036-27044.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Exinger, F. and F. Lacroute (1992). "6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae." Curr Genet **22**(1): 9-11.

Fazal, F. M., C. A. Meng, K. Murakami, R. D. Kornberg and S. M. Block (2015). "Real-time observation of the initiation of RNA polymerase II transcription." Nature **525**(7568): 274-277.

Feig, M. and Z. F. Burton (2010). "RNA polymerase II flexibility during translocation from normal mode analysis." Proteins **78**(2): 434-446.

Feig, M. and Z. F. Burton (2010). "RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation." Biophys J **99**(8): 2577-2586.

Femino, A. M., F. S. Fay, K. Fogarty and R. H. Singer (1998). "Visualization of single RNA transcripts in situ." Science **280**(5363): 585-590.

Fiedler, U. and H. T. Timmers (2001). "Analysis of the open region of RNA polymerase II transcription complexes in the early phase of elongation." Nucleic Acids Res **29**(13): 2706-2714.

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

Fong, N., H. Kim, Y. Zhou, X. Ji, J. Qiu, T. Saldi, K. Diener, K. Jones, X. D. Fu and D. L. Bentley (2014). "Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate." Genes Dev **28**(23): 2663-2676.

Fong, N., T. Saldi, R. M. Sheridan, M. A. Cortazar and D. L. Bentley (2017). "RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction." Mol Cell.

Fouqueau, T., M. E. Zeller, A. C. Cheung, P. Cramer and M. Thomm (2013). "The RNA polymerase trigger loop functions in all three phases of the transcription cycle." Nucleic Acids Res **41**(14): 7048-7059.

Fuchs, G., Y. Voichek, S. Benjamin, S. Gilad, I. Amit and M. Oren (2014). "4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells." Genome Biol **15**(5): R69.

Fukaya, T., B. Lim and M. Levine (2017). "Rapid Rates of Pol II Elongation in the Drosophila Embryo." Curr Biol **27**(9): 1387-1391.

Gaillard, H., C. Tous, J. Botet, C. Gonzalez-Aguilera, M. J. Quintero, L. Viladevall, M. L. Garcia-Rubio, A. Rodriguez-Gil, A. Marin, J. Arino, J. L. Revuelta, S. Chavez and A. Aguilera (2009). "Genome-wide analysis of factors affecting transcription elongation and DNA repair: a new role for PAF and Ccr4-not in transcription-coupled repair." PLoS Genet **5**(2): e1000364.

Gancedo, J. M. (1998). "Yeast carbon catabolite repression." Microbiol Mol Biol Rev **62**(2): 334-361.

Garcia-Martinez, J., A. Aranda and J. E. Perez-Ortin (2004). "Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms." Mol Cell **15**(2): 303-313.

Garcia-Martinez, J., L. Delgado-Ramos, G. Ayala, V. Pelechano, D. A. Medina, F. Carrasco, R. Gonzalez, E. Andres-Leon, L. Steinmetz, J. Warringer, S. Chavez and J. E. Perez-Ortin (2016). "The cellular growth rate controls overall mRNA turnover, and modulates either transcription or degradation rates of particular gene regulons." Nucleic Acids Res **44**(8): 3643-3658.

Gariglio, P., M. Bellard and P. Chambon (1981). "Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes." Nucleic Acids Res **9**(11): 2589-2598.

Gaur, N. A., J. Hasek, D. G. Brickner, H. Qiu, F. Zhang, C. M. Wong, I. Malcova, P. Vasicova, J. H. Brickner and A. G. Hinnebusch (2013). "Vps factors are required for efficient transcription elongation in budding yeast." Genetics **193**(3): 829-851.

Gilmour, D. S. and J. T. Lis (1986). "RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells." Mol Cell Biol **6**(11): 3984-3989.

Ginsburg, D. S., C. K. Govind and A. G. Hinnebusch (2009). "NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5." Mol Cell Biol **29**(24): 6473-6487.

Glover-Cutter, K., S. Kim, J. Espinosa and D. L. Bentley (2008). "RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes." Nat Struct Mol Biol **15**(1): 71-78.

Gnatt, A. L., P. Cramer, J. Fu, D. A. Bushnell and R. D. Kornberg (2001). "Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution." Science **292**(5523): 1876-1882.

Goler-Baron, V., M. Selitrennik, O. Barkai, G. Haimovich, R. Lotan and M. Choder (2008). "Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes." Genes Dev **22**(15): 2022-2027.

Govind, C. K., F. Zhang, H. Qiu, K. Hofmeyer and A. G. Hinnebusch (2007). "Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions." Mol Cell **25**(1): 31-42.

Greger, I. H. and N. J. Proudfoot (1998). "Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae." EMBO J **17**(16): 4771-4779.

Gu, B., D. Eick and O. Bensaude (2013). "CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo." Nucleic Acids Res **41**(3): 1591-1603.

Haag, J. R. and C. S. Pikaard (2011). "Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing." Nat Rev Mol Cell Biol **12**(8): 483-492.

Haimovich, G., M. Choder, R. H. Singer and T. Trcek (2013). "The fate of the messenger is pre-determined: a new model for regulation of gene expression." Biochim Biophys Acta **1829**(6-7): 643-653.

Haimovich, G., D. A. Medina, S. Z. Causse, M. Garber, G. Millan-Zambrano, O. Barkai, S. Chavez, J. E. Perez-Ortin, X. Darzacq and M. Choder (2013). "Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis." Cell **153**(5): 1000-1011.

Hantsche, M. and P. Cramer (2017). "Conserved RNA polymerase II initiation complex structure." Curr Opin Struct Biol **47**: 17-22.

Harada, Y., O. Ohara, A. Takatsuki, H. Itoh, N. Shimamoto and K. Kinosita, Jr. (2001). "Direct observation of DNA rotation during transcription by Escherichia coli RNA polymerase." Nature **409**(6816): 113-115.

Harel-Sharvit, L., N. Eldad, G. Haimovich, O. Barkai, L. Duek and M. Choder (2010). "RNA polymerase II subunits link transcription and mRNA decay to translation." Cell **143**(4): 552-563.

Harlen, K. M. and L. S. Churchman (2017). "The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain." Nat Rev Mol Cell Biol **18**(4): 263-273.

Hartzog, G. A., T. Wada, H. Handa and F. Winston (1998). "Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae." Genes Dev **12**(3): 357-369.

Hazelbaker, D. Z., S. Marquardt, W. Wlotzka and S. Buratowski (2013). "Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination." Mol Cell **49**(1): 55-66.

He, Y., J. Fang, D. J. Taatjes and E. Nogales (2013). "Structural visualization of key steps in human transcription initiation." Nature **495**(7442): 481-486.

Heidemann, M., C. Hintermair, K. Voss and D. Eick (2013). "Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription." Biochim Biophys Acta **1829**(1): 55-62.

Hirayoshi, K. and J. T. Lis (1999). "Nuclear run-on assays: assessing transcription by measuring density of engaged RNA polymerases." Methods Enzymol **304**: 351-362.

Hocine, S., P. Raymond, D. Zenklusen, J. A. Chao and R. H. Singer (2013). "Single-molecule analysis of gene expression using two-color RNA labeling in live yeast." Nat Methods **10**(2): 119-121.

Howard, S. C., Y. W. Chang, Y. V. Budovskaya and P. K. Herman (2001). "The Ras/PKA signaling pathway of Saccharomyces cerevisiae exhibits a functional interaction with the Sin4p complex of the RNA polymerase II holoenzyme." Genetics **159**(1): 77-89.

Howard, S. C., A. Hester and P. K. Herman (2003). "The Ras/PKA signaling pathway may control RNA polymerase II elongation via the Spt4p/Spt5p complex in Saccharomyces cerevisiae." Genetics **165**(3): 1059-1070.

Hu, W., C. Petzold, J. Coller and K. E. Baker (2010). "Nonsense-mediated mRNA decapping occurs on polyribosomes in Saccharomyces cerevisiae." Nat Struct Mol Biol **17**(2): 244-247.

Hu, W., T. J. Sweet, S. Chamnongpol, K. E. Baker and J. Coller (2009). "Co-translational mRNA decay in Saccharomyces cerevisiae." Nature **461**(7261): 225-229.

Huang, Y., W. Li, X. Yao, Q. J. Lin, J. W. Yin, Y. Liang, M. Heiner, B. Tian, J. Hui and G. Wang (2012). "Mediator complex regulates alternative mRNA processing via the MED23 subunit." Mol Cell **45**(4): 459-469.

Hyle, J. W., R. J. Shaw and D. Reines (2003). "Functional distinctions between IMP dehydrogenase genes in providing mycophenolate resistance and guanine prototrophy to yeast." J Biol Chem **278**(31): 28470-28478.

Izban, M. G. and D. S. Luse (1992). "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**(7): 1342-1356.

James, K., P. Gamba, S. J. Cockell and N. Zenkin (2017). "Misincorporation by RNA polymerase is a major source of transcription pausing in vivo." Nucleic Acids Res **45**(3): 1105-1113.

Jenks, M. H., T. W. O'Rourke and D. Reines (2008). "Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast." Mol Cell Biol **28**(12): 3883-3893.

Jenks, M. H. and D. Reines (2005). "Dissection of the molecular basis of mycophenolate resistance in Saccharomyces cerevisiae." Yeast **22**(15): 1181-1190.

Ji, X., Y. Zhou, S. Pandit, J. Huang, H. Li, C. Y. Lin, R. Xiao, C. B. Burge and X. D. Fu (2013). "SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase." Cell **153**(4): 855-868.

Jiao, X., S. Xiang, C. Oh, C. E. Martin, L. Tong and M. Kiledjian (2010). "Identification of a quality-control mechanism for mRNA 5'-end capping." Nature **467**(7315): 608-611.

Jimeno-Gonzalez, S., L. L. Haaning, F. Malagon and T. H. Jensen (2010). "The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II." Mol Cell **37**(4): 580-587.

Jimeno-Gonzalez, S., M. Schmid, F. Malagon, L. L. Haaning and T. H. Jensen (2014). "Rat1p maintains RNA polymerase II CTD phosphorylation balance." RNA **20**(4): 551-558.

Jin, H. (2015). Influence of RNA Polymerase II Catalytic Activity on Transcription Start Site Selection.

Jin, H. and C. D. Kaplan (2014). "Relationships of RNA polymerase II genetic interactors to transcription start site usage defects and growth in Saccharomyces cerevisiae." G3 (Bethesda) **5**(1): 21-33.

Jonkers, I., H. Kwak and J. T. Lis (2014). "Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons." Elife **3**: e02407.

Kapanidis, A. N., E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss and R. H. Ebright (2006). "Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism." Science **314**(5802): 1144-1147.

Kaplan, C. D. (2013). "Basic mechanisms of RNA polymerase II activity and alteration of gene expression in Saccharomyces cerevisiae." Biochim Biophys Acta **1829**(1): 39-54.

Kaplan, C. D., M. J. Holland and F. Winston (2005). "Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus." J Biol Chem **280**(2): 913-922.

Kaplan, C. D., H. Jin, I. L. Zhang and A. Belyanin (2012). "Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo." PLoS Genet **8**(4): e1002627.

Kaplan, C. D., K. M. Larsson and R. D. Kornberg (2008). "The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin." Mol Cell **30**(5): 547-556.

Karplus, P. A. and K. Diederichs (2012). "Linking crystallographic model and data quality." Science **336**(6084): 1030-1033.

Kashkina, E., M. Anikin, F. Brueckner, E. Lehmann, S. N. Kochetkov, W. T. McAllister, P. Cramer and D. Temiakov (2007). "Multisubunit RNA polymerases melt

only a single DNA base pair downstream of the active site." J Biol Chem **282**(30): 21578-21582.

Kettenberger, H., K. J. Armache and P. Cramer (2003). "Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage." Cell **114**(3): 347-357.

Kettenberger, H., K. J. Armache and P. Cramer (2004). "Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS." Mol Cell **16**(6): 955-965.

Khodor, Y. L., J. Rodriguez, K. C. Abruzzi, C. H. Tang, M. T. Marr, 2nd and M. Rosbash (2011). "Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila." Genes Dev **25**(23): 2502-2512.

Kilchert, C., S. Wittmann and L. Vasiljeva (2016). "The regulation and functions of the nuclear RNA exosome complex." Nat Rev Mol Cell Biol **17**(4): 227-239.

Kim, H., B. Erickson, W. Luo, D. Seward, J. H. Graber, D. D. Pollock, P. C. Megee and D. L. Bentley (2010). "Gene-specific RNA polymerase II phosphorylation and the CTD code." Nat Struct Mol Biol **17**(10): 1279-1286.

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

Kireeva, M. L., N. Komissarova, D. S. Waugh and M. Kashlev (2000). "The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex." J Biol Chem **275**(9): 6530-6536.

Kireeva, M. L., Y. A. Nedialkov, G. H. Cremona, Y. A. Purtov, L. Lubkowska, F. Malagon, Z. F. Burton, J. N. Strathern and M. Kashlev (2008). "Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation." Mol Cell **30**(5): 557-566.

Kireeva, M. L., K. Opron, S. A. Seibold, C. Domecq, R. I. Cukier, B. Coulombe, M. Kashlev and Z. F. Burton (2012). "Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase." BMC Biophys **5**: 11.

Klein, B. J., D. Bose, K. J. Baker, Z. M. Yusoff, X. Zhang and K. S. Murakami (2011). "RNA polymerase and transcription elongation factor Spt4/5 complex structure." Proc Natl Acad Sci U S A **108**(2): 546-550.

Korzheva, N., A. Mustaev, M. Kozlov, A. Malhotra, V. Nikiforov, A. Goldfarb and S. A. Darst (2000). "A structural model of transcription elongation." Science **289**(5479): 619-625.

Kostrewa, D., M. E. Zeller, K. J. Armache, M. Seizl, K. Leike, M. Thomm and P. Cramer (2009). "RNA polymerase II-TFIIB structure and mechanism of transcription initiation." Nature **462**(7271): 323-330.

Kruk, J. A., A. Dutta, J. Fu, D. S. Gilmour and J. C. Reese (2011). "The multifunctional Ccr4-Not complex directly promotes transcription elongation." Genes Dev **25**(6): 581-593.

Kuehner, J. N. and D. A. Brow (2008). "Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation." Mol Cell **31**(2): 201-211.

Kwapisz, M., M. Wery, D. Despres, Y. Ghavi-Helm, J. Soutourina, P. Thuriaux and F. Lacroute (2008). "Mutations of RNA polymerase II activate key genes of the nucleoside triphosphate biosynthetic pathways." EMBO J **27**(18): 2411-2421.

Landick, R. (2009). "Functional divergence in the growing family of RNA polymerases." Structure **17**(3): 323-325.

Larson, D. R., D. Zenklusen, B. Wu, J. A. Chao and R. H. Singer (2011). "Real-time observation of transcription initiation and elongation on an endogenous yeast gene." Science **332**(6028): 475-478.

Larson, M. H., R. A. Mooney, J. M. Peters, T. Windgassen, D. Nayak, C. A. Gross, S. M. Block, W. J. Greenleaf, R. Landick and J. S. Weissman (2014). "A pause sequence

enriched at translation start sites drives transcription dynamics in vivo." Science **344**(6187): 1042-1047.

Larson, M. H., J. Zhou, C. D. Kaplan, M. Palangat, R. D. Kornberg, R. Landick and S. M. Block (2012). "Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II." Proc Natl Acad Sci U S A **109**(17): 6555-6560.

Lenasi, T., B. M. Peterlin and M. Barboric (2011). "Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive transcription elongation factor b (P-TEFb)." J Biol Chem **286**(26): 22758-22768.

Lenstra, T. L., J. J. Benschop, T. Kim, J. M. Schulze, N. A. Brabers, T. Margaritis, L. A. van de Pasch, S. A. van Heesch, M. O. Brok, M. J. Groot Koerkamp, C. W. Ko, D. van Leenen, K. Sameith, S. R. van Hooff, P. Lijnzaad, P. Kemmeren, T. Hentrich, M. S. Kobor, S. Buratowski and F. C. Holstege (2011). "The specificity and topology of chromatin interaction pathways in yeast." Mol Cell **42**(4): 536-549.

Lenstra, T. L., A. Coulon, C. C. Chow and D. R. Larson (2015). "Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription." Mol Cell **60**(4): 597-610.

Liu, C. and C. T. Martin (2001). "Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase." J Mol Biol **308**(3): 465-475.

Liu, X., D. A. Bushnell, D. A. Silva, X. Huang and R. D. Kornberg (2011). "Initiation complex structure and promoter proofreading." Science **333**(6042): 633-637.

Liu, X., D. A. Bushnell, D. Wang, G. Calero and R. D. Kornberg (2010). "Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism." Science **327**(5962): 206-209.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.

Logan, J., E. Falck-Pedersen, J. E. Darnell, Jr. and T. Shenk (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**(23): 8306-8310.

Luo, W., A. W. Johnson and D. L. Bentley (2006). "The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model." Genes Dev **20**(8): 954-965.

Luse, D. S. (2013). "Promoter clearance by RNA polymerase II." Biochim Biophys Acta **1829**(1): 63-68.

Maizels, N. M. (1973). "The nucleotide sequence of the lactose messenger ribonucleic acid transcribed from the UV5 promoter mutant of Escherichia coli." Proc Natl Acad Sci U S A **70**(12): 3585-3589.

Malagon, F., M. L. Kireeva, B. K. Shafer, L. Lubkowska, M. Kashlev and J. N. Strathern (2006). "Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil." Genetics **172**(4): 2201-2209.

Malinen, A. M., M. Turtola, M. Parthiban, L. Vainonen, M. S. Johnson and G. A. Belogurov (2012). "Active site opening and closure control translocation of multisubunit RNA polymerase." Nucleic Acids Res **40**(15): 7442-7451.

Marquardt, S., D. Z. Hazelbaker and S. Buratowski (2011). "Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species." Transcription **2**(3): 145-154.

Martinez-Rucobo, F. W., S. Sainsbury, A. C. Cheung and P. Cramer (2011). "Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity." EMBO J **30**(7): 1302-1310.

Martins, S. B., J. Rino, T. Carvalho, C. Carvalho, M. Yoshida, J. M. Klose, S. F. de Almeida and M. Carmo-Fonseca (2011). "Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes." Nat Struct Mol Biol **18**(10): 1115-1123.

Mason, P. B. and K. Struhl (2003). "The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo." Mol Cell Biol **23**(22): 8323-8333.

Mason, P. B. and K. Struhl (2005). "Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo." Mol Cell **17**(6): 831-840.

Mayer, A., J. di Iulio, S. Maleri, U. Eser, J. Vierstra, A. Reynolds, R. Sandstrom, J. A. Stamatoyannopoulos and L. S. Churchman (2015). "Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution." Cell **161**(3): 541-554.

Mayer, A., H. M. Landry and L. S. Churchman (2017). "Pause & go: from the discovery of RNA polymerase pausing to its functional implications." Curr Opin Cell Biol **46**: 72-80.

Mayer, A., M. Lidschreiber, M. Siebert, K. Leike, J. Soding and P. Cramer (2010). "Uniform transitions of the general RNA polymerase II transcription complex." Nat Struct Mol Biol **17**(10): 1272-1278.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, S. Shuman and D. L. Bentley (1997). "5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II." Genes Dev **11**(24): 3306-3318.

Meyer, P. A., P. Ye, M. H. Suh, M. Zhang and J. Fu (2009). "Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data." J Biol Chem **284**(19): 12933-12939.

Miermont, A., F. Waharte, S. Hu, M. N. McClean, S. Bottani, S. Leon and P. Hersen (2013). "Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding." Proc Natl Acad Sci U S A **110**(14): 5725-5730.

Miguel, A., F. Monton, T. Li, F. Gomez-Herreros, S. Chavez, P. Alepuz and J. E. Perez-Ortin (2013). "External conditions inversely change the RNA polymerase II elongation rate and density in yeast." Biochim Biophys Acta **1829**(11): 1248-1255.

Millan-Zambrano, G., A. Rodriguez-Gil, X. Penate, L. de Miguel-Jimenez, M. Morillo-Huesca, N. Krogan and S. Chavez (2013). "The prefoldin complex regulates chromatin dynamics during transcription elongation." PLoS Genet **9**(9): e1003776.

Monsalve, M., Z. Wu, G. Adelmant, P. Puigserver, M. Fan and B. M. Spiegelman (2000). "Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1." Mol Cell **6**(2): 307-316.

Morillo-Huesca, M., M. Vanti and S. Chavez (2006). "A simple in vivo assay for measuring the efficiency of gene length-dependent processes in yeast mRNA biogenesis." FEBS J **273**(4): 756-769.

Muhlbacher, W., S. Sainsbury, M. Hemann, M. Hantsche, S. Neyer, F. Herzog and P. Cramer (2014). "Conserved architecture of the core RNA polymerase II initiation complex." Nat Commun **5**: 4310.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-255.

Nagaike, T., C. Logan, I. Hotta, O. Rozenblatt-Rosen, M. Meyerson and J. L. Manley (2011). "Transcriptional activators enhance polyadenylation of mRNA precursors." Mol Cell **41**(4): 409-418.

Naryshkin, N., A. Revyakin, Y. Kim, V. Mekler and R. H. Ebright (2000). "Structural organization of the RNA polymerase-promoter open complex." Cell **101**(6): 601-611.

Nechaev, S. and K. Adelman (2011). "Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation." Biochim Biophys Acta **1809**(1): 34-45.

Nehlin, J. O., M. Carlberg and H. Ronne (1991). "Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response." EMBO J **10**(11): 3373-3377.

Neymotin, B., V. Ettorre and D. Gresham (2015). "Global determinants of mRNA degradation rates in Saccharomyces cerevisiae." bioRxiv.

Neymotin, B., V. Ettorre and D. Gresham (2016). "Multiple Transcript Properties Related to Translation Affect mRNA Degradation Rates in Saccharomyces cerevisiae." G3 (Bethesda) **6**(11): 3475-3483.

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

Nudler, E., A. Mustaev, E. Lukhtanov and A. Goldfarb (1997). "The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase." Cell **89**(1): 33-41.

O'Brien, T. and J. T. Lis (1993). "Rapid changes in Drosophila transcription after an instantaneous heat shock." Mol Cell Biol **13**(6): 3456-3463.

Pal, M., A. S. Ponticelli and D. S. Luse (2005). "The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II." Mol Cell **19**(1): 101-110.

Park, J., M. Kang and M. Kim (2015). "Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1." Nucleic Acids Res **43**(5): 2625-2637.

Pelechano, V., S. Jimeno-Gonzalez, A. Rodriguez-Gil, J. Garcia-Martinez, J. E. Perez-Ortin and S. Chavez (2009). "Regulon-specific control of transcription elongation across the yeast genome." PLoS Genet **5**(8): e1000614.

Pelechano, V., W. Wei and L. M. Steinmetz (2015). "Widespread Co-translational RNA Decay Reveals Ribosome Dynamics." Cell **161**(6): 1400-1412.

Pinto, P. A., T. Henriques, M. O. Freitas, T. Martins, R. G. Domingues, P. S. Wyrzykowska, P. A. Coelho, A. M. Carmo, C. E. Sunkel, N. J. Proudfoot and A. Moreira (2011). "RNA polymerase II kinetics in polo polyadenylation signal selection." EMBO J **30**(12): 2431-2444.

Porrua, O. and D. Libri (2015). "Transcription termination and the control of the transcriptome: why, where and how to stop." Nat Rev Mol Cell Biol **16**(3): 190-202.

Powell, W. and D. Reines (1996). "Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro." J Biol Chem **271**(12): 6866-6873.

Pullara, F., J. Guerrero-Santoro, M. Calero, Q. Zhang, Y. Peng, H. Spahr, G. L. Kornberg, A. Cusimano, H. P. Stevenson, H. Santamaria-Suarez, S. L. Reynolds, I. S. Brown, S. P. Monga, B. Van Houten, V. Rapic-Otrin, G. Calero and A. S. Levine (2013). "A general path for large-scale solubilization of cellular proteins: from membrane receptors to multiprotein complexes." Protein Expr Purif **87**(2): 111-119.

Qiu, C., O. C. Erinne, J. M. Dave, P. Cui, H. Jin, N. Muthukrishnan, L. K. Tang, S. G. Babu, K. C. Lam, P. J. Vandeventer, R. Strohner, J. Van den Brulle, S. H. Sze and C. D. Kaplan (2016). "High-Resolution Phenotypic Landscape of the RNA Polymerase II Trigger Loop." PLoS Genet **12**(11): e1006321.

Quan, T. K. and G. A. Hartzog (2010). "Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of Saccharomyces cerevisiae Spt4-Spt5 in transcription." Genetics **184**(2): 321-334.

Ramanathan, A., G. B. Robb and S. H. Chan (2016). "mRNA capping: biological functions and applications." Nucleic Acids Res **44**(16): 7511-7526.

Ranish, J. A. and S. Hahn (1991). "The yeast general transcription factor TFIIA is composed of two polypeptide subunits." J Biol Chem **266**(29): 19320-19327.

Rasmussen, E. B. and J. T. Lis (1993). "In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes." Proc Natl Acad Sci U S A **90**(17): 7923-7927.

Reinberg, D. and R. G. Roeder (1987). "Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains." J Biol Chem **262**(7): 3331-3337.

Reines, D. (1992). "Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II." J Biol Chem **267**(6): 3795-3800.

Reines, D. (2003). "Use of RNA yeast polymerase II mutants in studying transcription elongation." Methods Enzymol **371**: 284-292.

Reines, D., M. J. Chamberlin and C. M. Kane (1989). "Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro." J Biol Chem **264**(18): 10799-10809.

Revyakin, A., C. Liu, R. H. Ebright and T. R. Strick (2006). "Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching." Science **314**(5802): 1139-1143.

Riles, L., R. J. Shaw, M. Johnston and D. Reines (2004). "Large-scale screening of yeast mutants for sensitivity to the IMP dehydrogenase inhibitor 6-azauracil." Yeast **21**(3): 241-248.

Rosonina, E., M. A. Bakowski, S. McCracken and B. J. Blencowe (2003). "Transcriptional activators control splicing and 3'-end cleavage levels." J Biol Chem **278**(44): 43034-43040.

Rougvie, A. E. and J. T. Lis (1988). "The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged." Cell **54**(6): 795-804.

Sainsbury, S., C. Bernecky and P. Cramer (2015). "Structural basis of transcription initiation by RNA polymerase II." Nat Rev Mol Cell Biol **16**(3): 129-143.

Sainsbury, S., J. Niesser and P. Cramer (2013). "Structure and function of the initially transcribing RNA polymerase II-TFIIB complex." Nature **493**(7432): 437-440.

Schmitt, M. E., T. A. Brown and B. L. Trumpower (1990). "A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae." Nucleic Acids Res **18**(10): 3091-3092.

Schroeder, S. C., D. A. Zorio, B. Schwer, S. Shuman and D. Bentley (2004). "A function of yeast mRNA cap methyltransferase, Abd1, in transcription by RNA polymerase II." Mol Cell **13**(3): 377-387.

Schulz, D., N. Pirkl, E. Lehmann and P. Cramer (2014). "Rpb4 subunit functions mainly in mRNA synthesis by RNA polymerase II." J Biol Chem **289**(25): 17446-17452.

Schwabish, M. A. and K. Struhl (2006). "Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II." Mol Cell **22**(3): 415-422.

Schwabish, M. A. and K. Struhl (2007). "The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo." Mol Cell Biol **27**(20): 6987-6995.

Schwinghammer, K., A. C. Cheung, Y. I. Morozov, K. Agaronyan, D. Temiakov and P. Cramer (2013). "Structure of human mitochondrial RNA polymerase elongation complex." Nat Struct Mol Biol **20**(11): 1298-1303.

Sevostyanova, A. and I. Artsimovitch (2010). "Functional analysis of Thermus thermophilus transcription factor NusG." Nucleic Acids Res **38**(21): 7432-7445.

Shaw, R. J. and D. Reines (2000). "Saccharomyces cerevisiae transcription elongation mutants are defective in PUR5 induction in response to nucleotide depletion." Mol Cell Biol **20**(20): 7427-7437.

Shaw, R. J., J. L. Wilson, K. T. Smith and D. Reines (2001). "Regulation of an IMP dehydrogenase gene and its overexpression in drug-sensitive transcription elongation mutants of yeast." J Biol Chem **276**(35): 32905-32916.

Shermoen, A. W. and P. H. O'Farrell (1991). "Progression of the cell cycle through mitosis leads to abortion of nascent transcripts." Cell **67**(2): 303-310.

Simchen, G., F. Winston, C. A. Styles and G. R. Fink (1984). "Ty-mediated gene expression of the LYS2 and HIS4 genes of Saccharomyces cerevisiae is controlled by the same SPT genes." Proc Natl Acad Sci U S A **81**(8): 2431-2434.

Singh, J. and R. A. Padgett (2009). "Rates of in situ transcription and splicing in large human genes." Nat Struct Mol Biol **16**(11): 1128-1133.

Somesh, B. P., J. Reid, W. F. Liu, T. M. Sogaard, H. Erdjument-Bromage, P. Tempst and J. Q. Svejstrup (2005). "Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest." Cell **121**(6): 913-923.

Squazzo, S. L., P. J. Costa, D. L. Lindstrom, K. E. Kumer, R. Simic, J. L. Jennings, A. J. Link, K. M. Arndt and G. A. Hartzog (2002). "The Paf1 complex physically and functionally associates with transcription elongation factors in vivo." EMBO J **21**(7): 1764-1774.

Steinmetz, E. J., C. L. Warren, J. N. Kuehner, B. Panbehi, A. Z. Ansari and D. A. Brow (2006). "Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase." Mol Cell **24**(5): 735-746.

Steitz, T. A., S. J. Smerdon, J. Jager and C. M. Joyce (1994). "A unified polymerase mechanism for nonhomologous DNA and RNA polymerases." Science **266**(5193): 2022-2025.

Stivers, J. T. (1998). "2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects." Nucleic Acids Res **26**(16): 3837-3844.

Sun, M., B. Schwalb, N. Pirkl, K. C. Maier, A. Schenk, H. Failmezger, A. Tresch and P. Cramer (2013). "Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels." Mol Cell **52**(1): 52-62.

Sun, M., B. Schwalb, D. Schulz, N. Pirkl, S. Etzold, L. Lariviere, K. C. Maier, M. Seizl, A. Tresch and P. Cramer (2012). "Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation." Genome Res **22**(7): 1350-1359.

Surratt, C. K., S. C. Milan and M. J. Chamberlin (1991). "Spontaneous cleavage of RNA in ternary complexes of Escherichia coli RNA polymerase and its significance for the mechanism of transcription." Proc Natl Acad Sci U S A **88**(18): 7983-7987.

Svetlov, V. and E. Nudler (2013). "Basic mechanism of transcription by RNA polymerase II." Biochim Biophys Acta **1829**(1): 20-28.

Sweeney, M. J. (1977). "Mycophenolic acid and its mechanism of action in cancer and psoriasis." Jpn J Antibiot **30 Suppl**: 85-92.

Tagami, S., S. Sekine, T. Kumarevel, N. Hino, Y. Murayama, S. Kamegamori, M. Yamamoto, K. Sakamoto and S. Yokoyama (2010). "Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein." Nature **468**(7326): 978-982.

Temiakov, D., N. Zenkin, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, E. Kashkina, M. Savkina, S. Zorov, V. Nikiforov, N. Igarashi, N. Matsugaki, S. Wakatsuki, K. Severinov and D. G. Vassylyev (2005). "Structural basis of transcription inhibition by antibiotic streptolydigin." Mol Cell **19**(5): 655-666.

Tennyson, C. N., H. J. Klamut and R. G. Worton (1995). "The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced." Nat Genet **9**(2): 184-190.

Thiebaut, M., J. Colin, H. Neil, A. Jacquier, B. Seraphin, F. Lacroute and D. Libri (2008). "Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae." Mol Cell **31**(5): 671-682.

Thummel, C. S., K. C. Burtis and D. S. Hogness (1990). "Spatial and temporal patterns of E74 transcription during Drosophila development." Cell **61**(1): 101-111.

Treutlein, B., A. Muschielok, J. Andrecka, A. Jawhari, C. Buchen, D. Kostrewa, F. Hog, P. Cramer and J. Michaelis (2012). "Dynamic architecture of a minimal RNA polymerase II open promoter complex." Mol Cell **46**(2): 136-146.

Vannini, A. and P. Cramer (2012). "Conservation between the RNA polymerase I, II, and III transcription initiation machineries." Mol Cell **45**(4): 439-446.

Vassylyev, D. G., M. N. Vassylyeva, A. Perederina, T. H. Tahirov and I. Artsimovitch (2007). "Structural basis for transcription elongation by bacterial RNA polymerase." Nature **448**(7150): 157-162.

Veloso, A., K. S. Kirkconnell, B. Magnuson, B. Biewen, M. T. Paulsen, T. E. Wilson and M. Ljungman (2014). "Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications." Genome Res **24**(6): 896-905.

Viktorovskaya, O. V., K. L. Engel, S. L. French, P. Cui, P. J. Vandeventer, E. M. Pavlovic, A. L. Beyer, C. D. Kaplan and D. A. Schneider (2013). "Divergent contributions of conserved active site residues to transcription by eukaryotic RNA polymerases I and II." Cell Rep **4**(5): 974-984.

Vvedenskaya, I. O., H. Vahedian-Movahed, J. G. Bird, J. G. Knoblauch, S. R. Goldman, Y. Zhang, R. H. Ebright and B. E. Nickels (2014). "Interactions between RNA polymerase and the "core recognition element" counteract pausing." Science **344**(6189): 1285-1289.

Wang, D., D. A. Bushnell, X. Huang, K. D. Westover, M. Levitt and R. D. Kornberg (2009). "Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution." Science **324**(5931): 1203-1206.

Wang, D., D. A. Bushnell, K. D. Westover, C. D. Kaplan and R. D. Kornberg (2006). "Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis." Cell **127**(5): 941-954.

Wang, E. T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth and C. B. Burge (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.

Wen, Y. and A. J. Shatkin (1999). "Transcription elongation factor hSPT5 stimulates mRNA capping." Genes Dev **13**(14): 1774-1779.

Westover, K. D., D. A. Bushnell and R. D. Kornberg (2004). "Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center." Cell **119**(4): 481-489.

Winston, F., D. T. Chaleff, B. Valent and G. R. Fink (1984). "Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae." Genetics **107**(2): 179-197.

Yao, J., M. B. Ardehali, C. J. Fecko, W. W. Webb and J. T. Lis (2007). "Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation." Mol Cell **28**(6): 978-990.

Yuzenkova, Y. and N. Zenkin (2010). "Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis." Proc Natl Acad Sci U S A **107**(24): 10878-10883.

Zaret, K. S. and F. Sherman (1982). "DNA sequence required for efficient transcription termination in yeast." Cell **28**(3): 563-573.

Zenkin, N., Y. Yuzenkova and K. Severinov (2006). "Transcript-assisted transcriptional proofreading." Science **313**(5786): 518-520.

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

Zhang, J., M. Palangat and R. Landick (2010). "Role of the RNA polymerase trigger loop in catalysis and pausing." Nat Struct Mol Biol **17**(1): 99-104.

Zhang, Y., Y. Feng, S. Chatterjee, S. Tuske, M. X. Ho, E. Arnold and R. H. Ebright (2012). "Structural basis of transcription initiation." Science **338**(6110): 1076-1080.

Zuo, Y. and T. A. Steitz (2015). "Crystal structures of the E. coli transcription initiation complexes with a complete bubble." Mol Cell **58**(3): 534-540.

APPENDIX A

Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble[2]

Disclaimer for Appendix A

Appendix A is reprint of a publication, which I am the second author on. This study was performed in collaboration with the Calero lab at University of Pittsburgh. Christopher O. Barnes and Monica Calero performed major experiments in the Calero lab. I performed experiments presented in **Figure A-5** and **A-6**. Summary section of this chapter is the abstract of the publication, rest are as in publication.

---

[2]Reprinted with permission from "Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble" by Barnes et al. 2015 Mol Cell, Vol. 59, 258-69. Copyright © 2013 Elsevier Inc.

Summary

Notwithstanding numerous published structures of RNA Polymerase II (Pol II), structural details of Pol II engaging a complete nucleic acid scaffold have been lacking. Here, we report the structures of TFIIF-stabilized transcribing Pol II complexes, revealing the upstream duplex and full transcription bubble. The upstream duplex lies over a wedgeshaped loop from Rpb2 that engages its minor groove, providing part of the structural framework for DNA tracking during elongation. At the upstream transcription bubble fork, rudder and fork loop 1 residues spatially coordinate strand annealing and the nascent RNA transcript. At the downstream fork, a network of Pol II interactions with the nontemplate strand forms a rigid domain with the trigger loop (TL), allowing visualization of its open state. Overall, our observations suggest that ‘‘open/ closed’’ conformational transitions of the TL may be linked to interactions with the non-template strand, possibly in a synchronized ratcheting manner conducive to polymerase translocation.

Introduction

The pre-initiation stage of transcription requires concerted interactions between RNA Polymerase II (Pol II) and the general transcription factors TFIIB, TFIID, TFIIF, TFIIE, and TFIIH. During initial promoter melting, TFIIH generates an unwound region of 7–9 base pairs. Subsequently, this transcription bubble is unwound to approximately 18–25 bases, and a short DNA-RNA hybrid is synthesized. Transcripts of ten or more nucleotides result in promoter escape and stabilization of a mature bubble (Liu, Bushnell et al. 2011, Nechaev and Adelman 2011, Luse 2013). The number of nucleotides

unwound in a mature bubble is still a matter of debate, since sizes ranging from 8 to 22 nucleotides have been reported for bacterial, archaeal, and eukaryotic polymerases (Naryshkin, Revyakin et al. 2000, Fiedler and Timmers 2001, Pal, Ponticelli et al. 2005). In addition, the size of the bubble might not be fixed but depend on Pol II's transcriptional stage; evidence of a scrunched state—where template and non-template strand bases are compacted in space relative to relaxed conformations—has been proposed for the early stages of transcription initiation in bacteria (Kapanidis, Margeat et al. 2006, Revyakin, Liu et al. 2006). Similarly, other transcriptional events such as backtracking or interactions with elongation or termination factors might alter the number of bases (and location) in the mature bubble (Fiedler and Timmers 2001). Notwithstanding numerous Pol II structures published to date, structural details of a complete transcribing complex, including upstream and downstream DNA duplexes and a full transcription bubble, have yet to be revealed. Here we report the crystal structures of Pol II in complex with a complete nucleic acid scaffold that illustrates the architecture of a Pol II transcribing complex.

Results

Design, Assembly, and Crystallization of Pol II Transcribing Complexes

Assembly of a *Saccharomyces cerevisiae* Pol II transcribing complex was achieved by mixing Pol II with pre-assembled nucleic acid scaffolds (see Experimental procedures). The main scaffold used for our experiments (scaffold 1) consisted of two synthetic DNA oligonucleotides (53 nucleotides long), featuring upstream and downstream duplexes, a non-complementary stretch of 15 nucleotides to generate a

synthetic transcription bubble, and a 9-mer RNA complementary to the template strand to form a DNA-RNA hybrid (**Figure A-1**A). The number of non-complementary bases used in the design of the bubble was based on crystal structures of partial Pol II transcribing complexes, including PDB: 1Y1W, 2NVZ (Kettenberger, Armache et al. 2004, Wang, Bushnell et al. 2006). These structures show, at the downstream end, base complementarity at positions i+3 and i+5, respectively (where i+1 indicates the nucleotide addition site and i−1 the first base of the nascent RNA transcript), and at the upstream end, a partial template strand reaches position i−9 below arch residues (comprising rudder [$Rpb1^{312-319}$] and fork loop 1 [FL1, $Rpb2^{470-480}$]) (**Figure A-1**B). However, steric clashes with arch residues at this position suggested that at least two additional nucleotides are required to allow template and non-template strand annealing. Collectively, these observations suggested an artificial bubble size with a minimum of 14 nucleotides for in vitro structural studies of a transcribing Pol II.

Initially, crystals of Pol II bound to our scaffolds showed weak electron density for the upstream duplex but none for the non-template strand (**Figure A-1**C). In search for factors that could contribute to a stabilized transcription bubble, we assembled Pol II or ten-subunit Pol II (lacking Rpb4 and Rpb7 subunits, Δ4/7) transcribing complexes with TFIIF or its 45 kDa β subunit, Tfg2 (**Figure A-1**D **and** E). Two sets of crystals were obtained using PEG 4000 and low salt (**Table A-1**). The first transcribing complex comprises Δ4/7-TFIIF (hereafter referred to as Δ4/7-TC for simplicity) and the second comprises Pol II-Tfg2 (hereafter referred to as Pol II-TC).

Figure A-1. Design and reconstitution of Pol II transcribing complex. (A) Schematic representation of scaffolds used in our structural studies, where the open circles represent the non-complementary region to generate artificial transcription bubbles (template strand: cyan, non-template strand: green, RNA: red). Structural analysis presented in this paper utilized scaffold 1 in complex with Pol II and either TFIIF or its β-subunit, Tfg2. The use of a non-complementary non-template strand to generate a transcription bubble is analogous to use of chain-terminators or non-hydrolysable substrates to examine Pol II-substrate interactions or interpretation of Pol II inhibitor-bound complexes as translocation intermediates. Our current structure interpreted in similar light still allows for a number of valuable insights. (B) A conservative model of the size of the bubble could be estimated using PDB:ID 1Y1W (Kettenberger, Armache et al. 2004), where complementarity was observed at i+3. Upstream clashes with the "arch" would require 2 additional bases before annealing could occur. (C) Final refined 2*Fobs-Fcalc* electron density map contoured at 0.8σ of a crystal comprising Pol II + scaffold 3 in the absence of Tfg2 or TFIIF. Partial electron density is observed for the upstream duplex and for the downstream fork and duplex, however no density is observed for the non-template strand within the transcription bubble (template: cyan, non-template: green RNA: red). (D) SDS-PAGE of Pol II-TC complex.
(E) SDS-PAGE of Δ4/7-TC after coomassie staining (left panel), also depicting scaffold 1 after ethidium bromide staining (red arrow, NAS, right panel). (F) Crystals of Δ4/7-TC were transferred to mother liquor to remove excess protein from crystals. After four transfers, 8X loading buffer was added to the drop containing crystals and subjected to SDS-PAGE. The four largest proteins are easily discernible from the gel confirming the presence of TFIIF in the crystals. Faint bands corresponding to smaller Pol II subunits are also present in the gel. (G) Example of B-factor sharpening on Pol II residues at the Rpb2 wedge, contoured at 1.0σ. Application of negative B-factors results in increased clarity for high resolution features, especially for large side chains (ie Arg, Lys, Phe, Trp, Tyr) at a relatively low resolution (DeLaBarre and Brunger 2006).

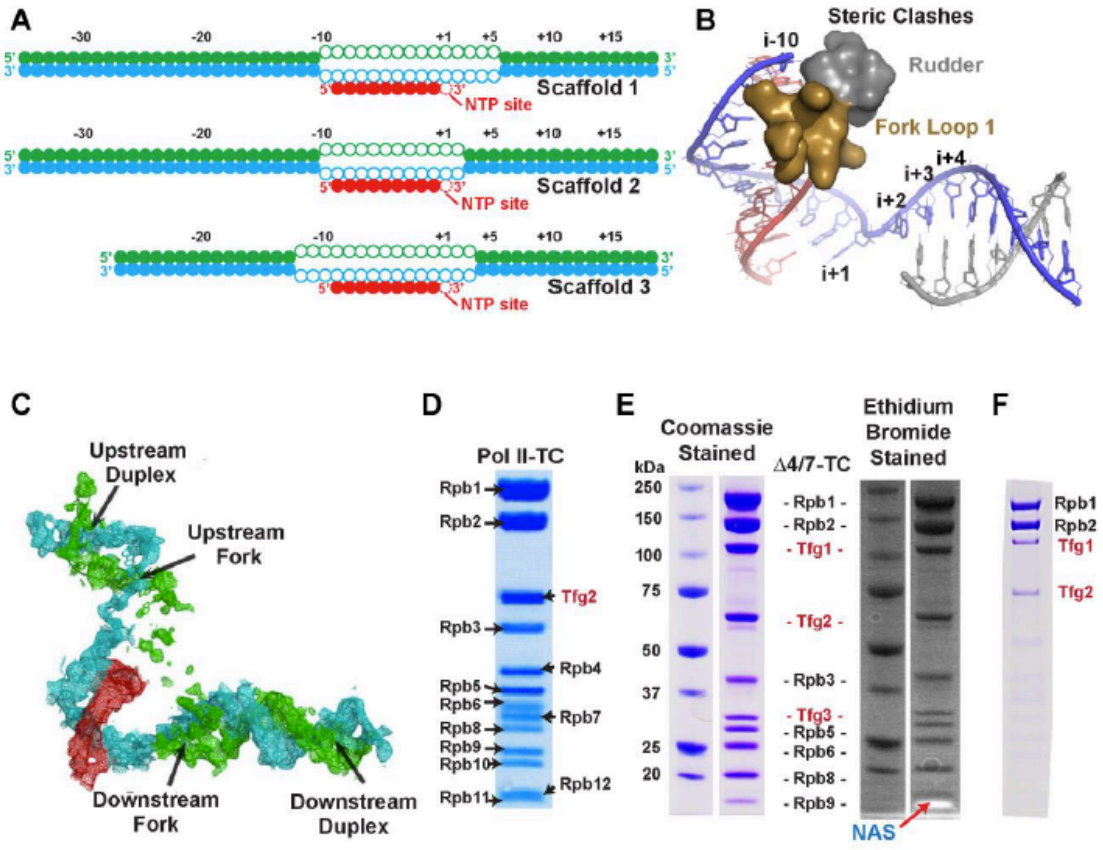

A
-30
-20
-10
+1
+5
+10
+15
Scaffold 1
NTP site
Scaffold 2
Scaffold 3
B
Steric Clashes
i-10
Rudder
Fork Loop 1
i+4
i+3
i+2
i+1
C
Upstream Duplex
Upstream Fork
Downstream Fork
Downstream Duplex
D
Pol II-TC
Rpb1
Rpb2
Tfg2
Rpb3
Rpb4
Rpb5
Rpb6
Rpb7
Rpb8
Rpb9
Rpb10
Rpb12
Rpb11
E
Coomassie Stained
Δ4/7-TC
Ethidium Bromide Stained
kDa
250
150
100
75
50
37
25
20
- Rpb1 -
- Rpb2 -
- Tfg1 -
- Tfg2 -
- Rpb3 -
- Tfg3 -
- Rpb5 -
- Rpb6 -
- Rpb8 -
- Rpb9 -
NAS
F
Rpb1
Rpb2
Tfg1
Tfg2
G
Met868
Ser869
Gly867
Tyr866
Lys864
B-factor Sharpening = 0
B-factor Sharpening = -50
B-factor Sharpening = -150

Table A-1. Crystallographic Data and Refinement Statistics

	Pol II-Scaffold 1 (i+5)[a] (APS-GM/CA)	Δ4/7-TC (i+5) (APS-GM/CA)	Pol II-TC (i+5) (SSRL-11.2)	Pol II-Tfg2-Scaffold 2 (i+2) (SSRL-11.2)	Pol II-Scaffold 3 (i+3) (APS-GM/CA)
Data Collection[b]					
PDB ID	5C44	5C4J	5C4X	5C4A	5C3E
Nucleic acid	Scaffold 1	Scaffold 1	Scaffold 1	Scaffold 2	Scaffold 3
Space group	C2221	C21	C2221	C2221	C2221
Unit cell (Å)	220.2, 391.8, 282.3	280.7, 223.3, 156.4	220.7, 393.3, 281.6	219.8, 396.7, 273.6	219, 390.9, 278
α, β, γ (°)	90, 90, 90	90, 98.1, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.03	0.979	0.979	0.978	1.03
Resolution (Å)[c]	120–3.9	174–4	50–4	200–4.2	178–3.7
Unique reflections	106,511	80,485	103,153	80,062	126,815
Completeness (%)	96 (93.8)	97 (96.1)	97.3 (96.4)	99 (98.3)	94.88 (78)
Redundancy	4.3 (3.2)	5.1 (4.3)	3.8 (3.4)	4.5 (3.7)	3.6 (2.8)
$<I/\sigma I>$	8.2 (1.2)	10.3 (3.7)	12.5 (1.95)	14.2 (2.1)	8.5 (1.1)
Mosaicity (°)	0.6	0.6	0.35	0.55	0.35
R_{merge} (%)	13.8 (49)	14.7 (61)	9.1 (42)	8.5 (47)	12.5 (56)
Refinement					
No. atoms	33,086	30,259	33,673	32,281	33,086
R_{cryst}/R_{free} (%)	22.3/25.4	21/26	21.5/23.2	23.53/27.47	23.2/27.62
Refinement program	Buster/CNS	Refmac/ CNS	Buster/ CNS	Refmac/ CNS	Refmac/ CNS

[a]Indicates the position of complementary base pairing in the downstream bubble.
[b]Numbers in parentheses correspond to the highest-resolution shell.
[c]Resolution limits were extended to include weak intensity data (Karplus and Diederichs, 2012). Using the traditional criterion of I/sI > 2.0, resolution limits are 4.15 A˚ and 3.9 A˚ for Pol II-Scaffold 1 and Pol II-Scaffold 3 complexes, respectively.

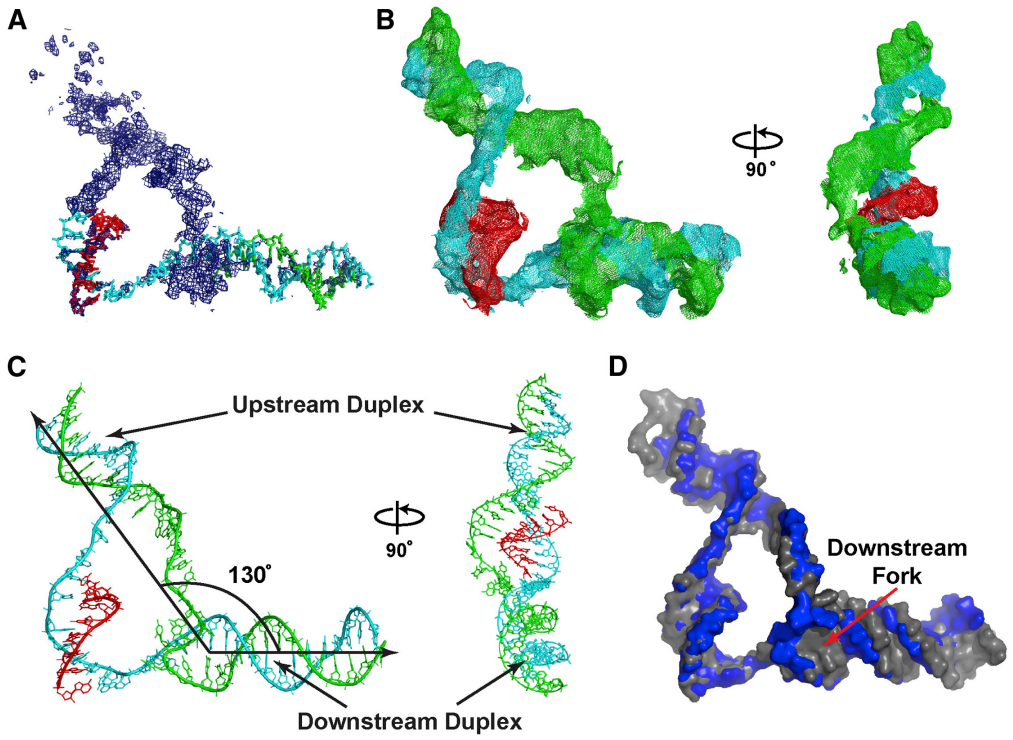


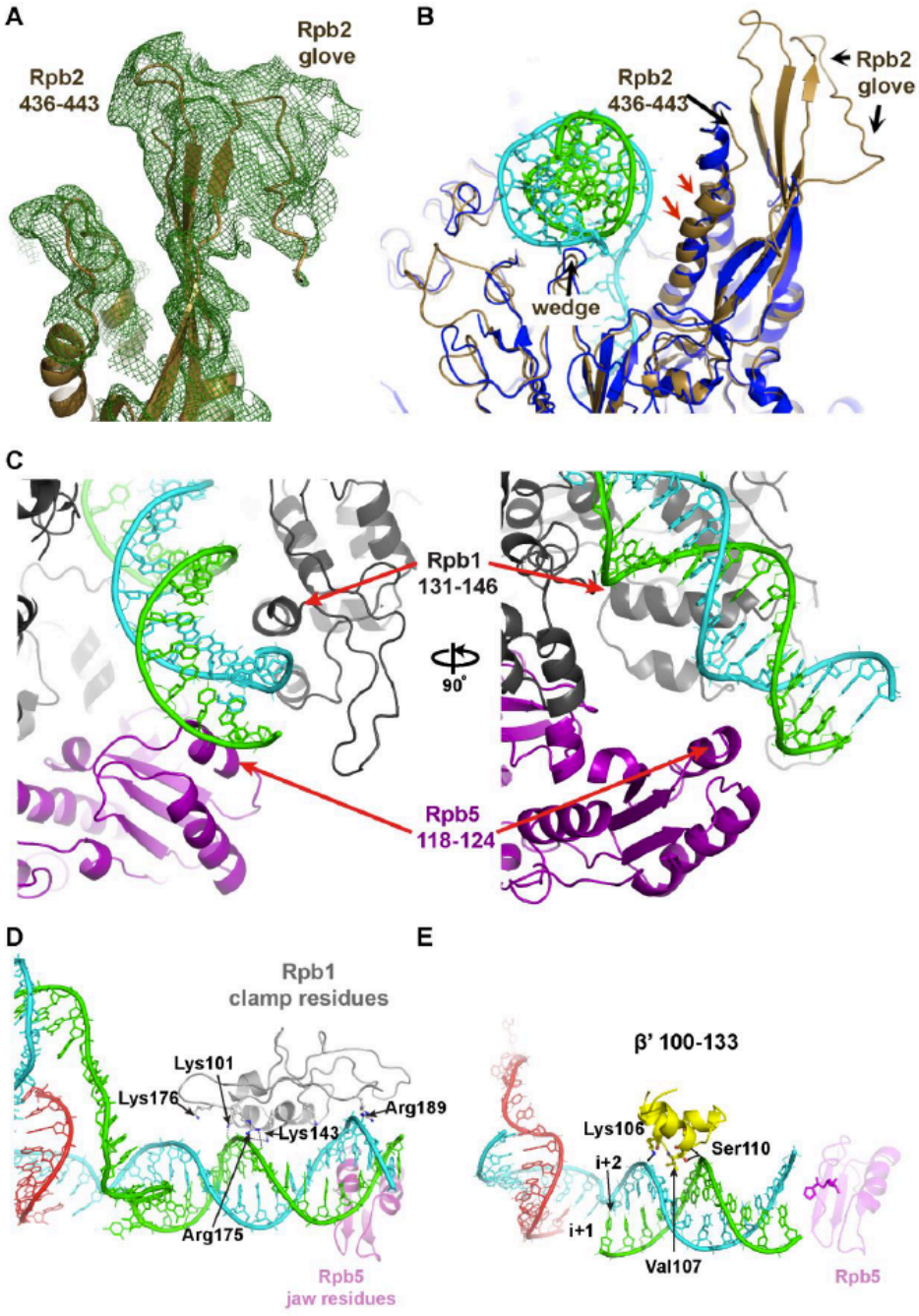
Figure A-2. Architecture of the Complete Nucleic Acid Scaffold.
(A) Difference *Fobs− Fcalc* electron density map contoured at 2 σ. The following color scheme will be used throughout: cyan, template strand; green, non-template strand; red, RNA transcript. (B) Front and side views of the *2Fobs− Fcalc* map for the final refined map contoured at 1.0 σ. (C) Cartoon representation of the 38-nucleotide refined nucleic acid scaffold; downstream and upstream duplexes form an angle of approximately 130°. (D) Surface representation of the overlay between Δ4/7-TC (blue) and Pol II-TC (gray). The two structures overlay remarkably well; minor structural differences occur at the downstream fork (see also **Figure A-1**)

Structures were solved by molecular replacement in Phaser (McCoy, Grosse-Kunstleve et al. 2007) using Δ4/7 or Pol II as search models (see Experimental procedures). An initial unbiased *Fobs− Fcalc* map revealed the presence of extra density corresponding to upstream dsDNA, the non-template strand (**Figure A-2**A), and three previously disordered regions of Rpb2 (**Figure A-3**A **and** B). Unfortunately, any additional density for Tfg2 or TFIIF was non-interpretable due to either high mobility or partial occupancy in the crystals. Nevertheless, the presence of TFIIF within the crystals (**Figure A-1**F) was essential to reveal the full structure of the transcription bubble in our complexes. Interestingly, low-resolution cryo-EM data of the pre-initiation complex suggests a role of TFIIF, or the Tfg2 winged-helix domain specifically, in stabilizing the upstream duplex, which could not be visualized prior to TFIIF addition within these structures (He, Fang et al. 2013, Muhlbacher, Sainsbury et al. 2014). While strikingly similar, without clear extra density, we can only speculate that TFIIF or Tfg2 is acting in an analogous fashion within our structures.

The molecular replacement models were refined using the programs Buster (Blanc, Roversi et al. 2004), Refmac (Murshudov, Vagin et al. 1997), and manual building with B factor sharpening in *Coot*, which clarified side chain positioning, thus allowing model refinement (**Figure A-1**G and **Table A-1**) (Emsley, Lohkamp et al. 2010). The nucleic acid scaffold was built into the electron density using the characteristic features of the DNA-RNA hybrid as register (**Figure A-2**A). The final refined *2Fobs− Fcalc* map for Δ4/7-TC is illustrated in Figure A-2B. The full observable DNA scaffold (38 nucleotides long) spans the length of Pol II and comprises:

the downstream duplex, the DNA-RNA hybrid and two previously uncharacterized regions, an upstream duplex, and the full transcription bubble including the non-template strand (**Figure A-2**B and **4-1**C). Overlay between the DNA scaffolds from Δ4/7-TC and Pol II-TC show minor differences, mainly located at the downstream fork (**Figure A-2**D). Pol II regions involved in DNA binding include: (1) the previously described Rpb5 "jaw" and Rpb1 "clamp" residues that interact with the downstream duplex (Gnatt, Cramer et al. 2001), (2) Rpb2 "wedge" residues ($Rpb2^{862-874}$) that interact with the upstream duplex, and (3) "arch" residues that interact with the upstream fork (**Figure A-3**A and B). For clarity, the presentation of our structural findings is based on the ten-subunit Pol II (Δ4/7) transcribing complex comprising TFIIF and scaffold 1, which has been labeled as Δ4/7-TC unless otherwise noted.

Figure A-3. Pol II contacts with upstream and downstream duplexes. (A) *2Fobs-Fcalc* electron density map contoured at 1.0 s _to illustrate the presence of traceable electron density for previously disordered Rpb2 regions. (B) Cartoon representation of the overlay of our structure with PDB:ID 1Y1W (illustrated in sand and blue, respectively); template and non-template strands are illustrated in cyan and green respectively. Presence of TFIIF or Tfg2 allowed tracing of two previously unfolded regions of Rpb2, the first one folds into a 3-strand beta sheet that resembles a catcher's glove (Rpb2 glove) and includes residues 71-90 and 131-164. The second one includes loop 436-443 from Rpb2 protrusion domain. Overlay between our structure and PDB:ID 1Y1W illustrate conformational changes in helix 443-465 (red arrows) possibly elicited by the close proximity of the upstream duplex. (C) Cartoon representation of Pol II stabilizing the downstream duplex illustrates that Rpb1 (silver) and Rpb5 (magenta) helices "thread" the major and minor grooves of the DNA respectively.
(D) Cartoon and ball & stick representation of Pol II interactions with the downstream duplex, where positively charged residues of Rpb1 clamp domain (silver) contact the i+7 to i+9 phosphates of the non-template strand. (E) Contacts resembling head-clamp interactions in *T. thermophilus* polymerase (PDB:ID 2PPB) are located on a helix-loop-helix domain in the b´ subunit. These include: 1) Val107 which is inserted in the minor groove of the double helix, 2) the backbone nitrogen of Ser110 interacting with i+8 phosphate of the non-template strand and 3) Lys106 interacting with phosphate template strand i+4.


A
Rpb2
glove
Rpb2
436-443
B
Rpb2
436-443
Rpb2
glove
wedge
C
Rpb1
131-146
90°
Rpb5
118-124
D
Rpb1
clamp residues
Lys101
Lys176
Arg189
Lys143
Arg175
Rpb5
jaw residues
E
β' 100-133
Lys106
Ser110
i+2
i+1
Val107
Rpb5

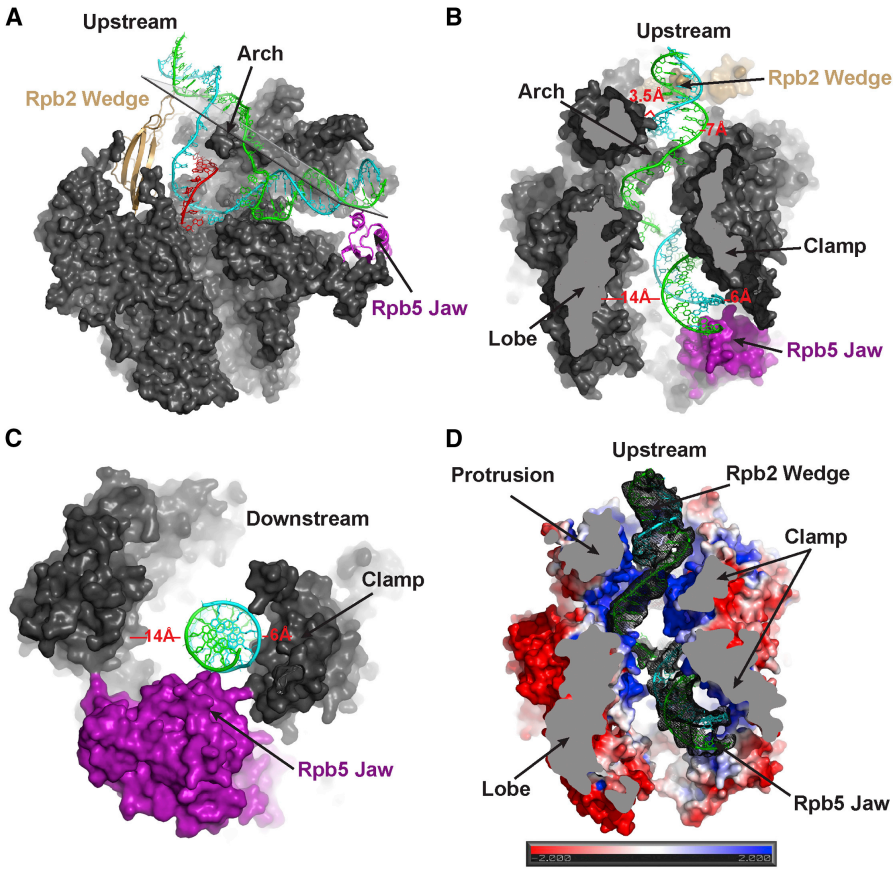


Figure A-4. Overall Structure of Δ4/7-TC
(A) Surface representation (side view, Rpb2 removed) illustrating the position of the scaffold inside Pol II. Wedge, jaw, and arch interactions with the scaffold lie on almost a perfect plane, possibly to minimize strain during elongation. (B and C) The scaffold binds asymmetrically inside Pol II's cleft, more prominently at the downstream end where observed clamp-DNA distances of ≈6 Å versus lobe DNA distances of ≈14 Å are due to interactions with clamp and jaw residues (see also Figure S2). (D) Surface electrostatic representation calculated using the APBS (Baker, Sept et al. 2001) suite in PyMOL to illustrate how the non-template strand follows a path of positively charged residues inside Pol II's cleft (lobe and protrusion). The final refined *2Fobs− Fcalc* map (gray) contoured at 1.0 σ is also illustrated to show the continuous density for the DNA scaffold.

Pol II Interacts with the Minor Groove of Upstream and Downstream DNA Duplexes Using Two Domains Located 90 Å Apart

Our refined models show that Pol II interacts with the upstream duplex, which appears as if propped by a hairpin loop or "wedge" ($Rpb2^{862-874}$)—a vertical extension of wall residues ($Rpb2^{855-861}$)—that engages the minor groove of the double helix (**Figure A-4**A). At the tip of the wedge, Met^{868} lies between template and non-template strands, while the amide backbone of Gly^{867} appears to form hydrogen bonds (H-bonds) with two contiguous phosphates on the non-template strand (**Figure A-5**A). No crystal contacts that could potentially stabilize the conformation of wedge residues or the upstream duplex were present in the two different crystal forms (**Figure A-6**A and B). The framework of the wedge—a hairpin loop within a long concave five-strand β sheet—is conserved in all multi-subunit DNA-directed RNA polymerases (**Figure A-5**B). Moreover, overlay of the human mitochondrial polymerase elongation complex (mtRNAP) (Schwinghammer, Cheung et al. 2013) and our structure shows that the upstream duplex and fork adopt similar conformations (**Figure A-5**C). This is particularly interesting given the scarcity of conserved structural elements between the two structures (**Figure A-6**C). Furthermore, overlay of the archaeal RNAP clamp domain in complex with the heterodimer interface of the Spt4/5 complex (PDB: 3QQC) (Martinez-Rucobo, Sainsbury et al. 2011) with the Rpb1 clamp domain of Δ4/7-TC shows that the upstream duplex is situated between the wedge domain and position of the Spt5 NusG domain (**Figure A-6**E). While speculative, the observed location of the duplex in our structure is consistent with previous biochemical data and Pol II-Spt4/5

elongation models (Sevostyanova and Artsimovitch 2010, Klein, Bose et al. 2011, Martinez-Rucobo, Sainsbury et al. 2011).

To assess a role of Rpb2 wedge residues during in vivo transcription, we constructed a number of alleles and characterized them for growth phenotypes consistent with transcription defects (**Figure A-6**D, left panel). We observed sensitivity to mycophenolic acid (MPA), which can be indicative of *IMD2* transcriptional phenotypes (Kaplan 2013). Indeed we found that Rpb2 wedge alleles were generally defective for induction of *IMD2* gene expression in the presence of MPA (**Figure A-6**D, right panel). In order to more directly assess the role of the Rpb2 wedge, we examined the K864G/K865G/Δ866-871 *rpb2* wedge allele for in vitro or in vivo elongation phenotypes (**Figure A-5**D and E). We found that this particular *rpb2* wedge allele did not confer robust elongation defects in vitro or in vivo, though it did confer very strong defects in steady-state Pol II reporter gene occupancy and expression (**Figure A-5**F and G). The defects observed are consistent with strong defects in initiation, and we observed altered transcription start site selection consistent with altered initiation in vivo (**Figure A-5**H).

Pol II contacts with the downstream duplex encompass Rpb5 (jaw) and Rpb1 (clamp-head) domains (**Figure A-4**B and C). Specifically, Rpb5 jaw residue Pro^{118} (from helix 118–124) is positioned inside the minor groove of the DNA double helix, and Thr^{117} (from loop 112–117) and Ser^{119} locate within H-bond distance to non-template strand positions i+15 and i+16 (**Figure A-3**C, also observed in PDB: 1R9T, 2NVQ, 2NVZ). Rpb1 clamp-head residues Lys^{100}, Lys^{101}, Lys^{143}, and Arg^{175} locate

within H-bond distance of the phosphate chain of non-template strand positions i+8 to i+10 (**Figure A-3**D). As a result of these interactions, the downstream duplex is asymmetrically positioned inside the cleft (**Figure A-4**B and C). Interestingly, contacts resembling clamp-head interactions in *T. thermophilus* polymerase are located on a helix-loop-helix domain in its β′ subunit (**Figure A-3**E) forming H-bonds with non-template strand i+8 to i+9. However, interactions with jaw residues are observed only in archaeal and eukaryotic polymerases, since bacterial polymerases lack Rpb5 homologs.

Strand Annealing and Three-Way Coordination of the Template Strand, Non-template Strand, and RNA at the Upstream Fork

The transcription bubble in our complexes lies within upstream and downstream duplexes and was enforced by non-complementarity between the template and non-template strands at positions i+5 through i−10 (**Figure A-1**A, scaffold 1). At the upstream "closing" end of the bubble, rudder and FL1 residues come in close proximity to form an "arch" located 25–30 Å above the bridge helix (**Figure A-4**A). The arch is situated in between the template and non-template strands and physically marks the upstream boundary of the bubble (**Figure A-7**). Arch residues adopt unique conformations that allow simultaneous coordination of the nucleic acid scaffold (**Figure A-7**A).

Figure A-5. Rpb2 Wedge Residues: Structure, Conservation, and Function.
(A) Interaction of wedge residues with the minor groove of the upstream duplex. Sequence conservation of tip residues across species is nearly universal for Gly^{867}, whose amide bond interacts with the phosphate chain of the non-template strand; Met^{868} is conserved in yeast (*S. cerevisiae* and *S. pombe*) and is substituted by a bulky hydrophobic residue in other species. A refined *2Fobs− Fcalc* map contoured at 1.0 σ is also illustrated. (B) Conservation of wedge structure: yellow, *T. thermophilus* (Vassylyev et al., 2007); hot pink, archaea (*S. sulfactarius*) (Hirata et al., 2008); sand, *S. cerevisiae*; blue, *S. pombe* (Spåhr et al., 2009); and purple, 14-subunit Pol I (*S. cerevisiae*) (Engel et al., 2013). (C) Overlay of mtRNAP-upstream duplex (PDB: 4BOC) with our structure about template strand i+1 and i+2 (see also Figure S3). (D) In vitro elongation rate of WT Pol II and an *rpb2* wedge deletion mutant (K864G/K865G/Δ866-871). Elongation rate determined on nucleic acid scaffolds at a number of NTP concentrations followed by non-linear regression of the rates for determination of maximum elongation rates (bar graph, error bars indicate range of 95% confidence interval). (E) In vivo apparent elongation rates for WT Pol II and the *rpb2* wedge deletion mutant at a galactose-inducible reporter gene determined by ChIP upon glucose shutoff of transcription (schematic of reporter in F). Values are normalized to 0 min of glucose and error bars represent SD of the mean for three independent experiments. (F) Steady-state occupancy for WT Pol II and the *rpb2* wedge deletion mutant at a galactose-inducible reporter gene under galactose induction determined by ChIP (schematic of reporter with positions of PCR amplicons shown below) (n = 3 independent experiments). (G) Steady-state RNA levels of reporter used in (F) for WT Pol II and the *rpb2* wedge deletion mutant. Values were normalized to SCR1 levels (a Pol III transcript) and averaged (n = 3) with error bars representing SD.(H) Primer extension analysis of *ADH1* transcripts for various *rpb2* wedge alleles showing average change in fraction of *ADH1* starts in various positions relative to wild-type, with error bars representing SD of the mean (n = 3).

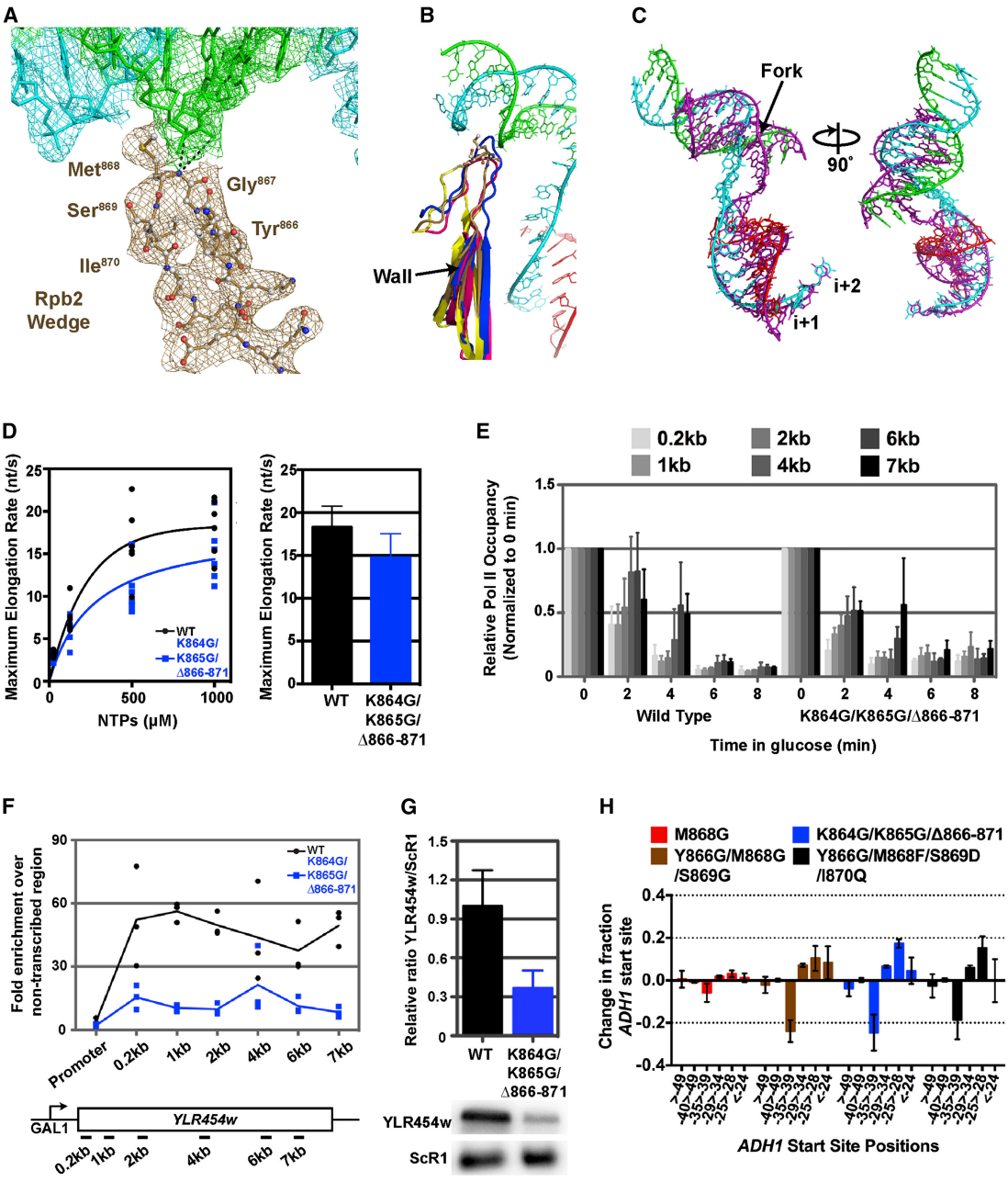
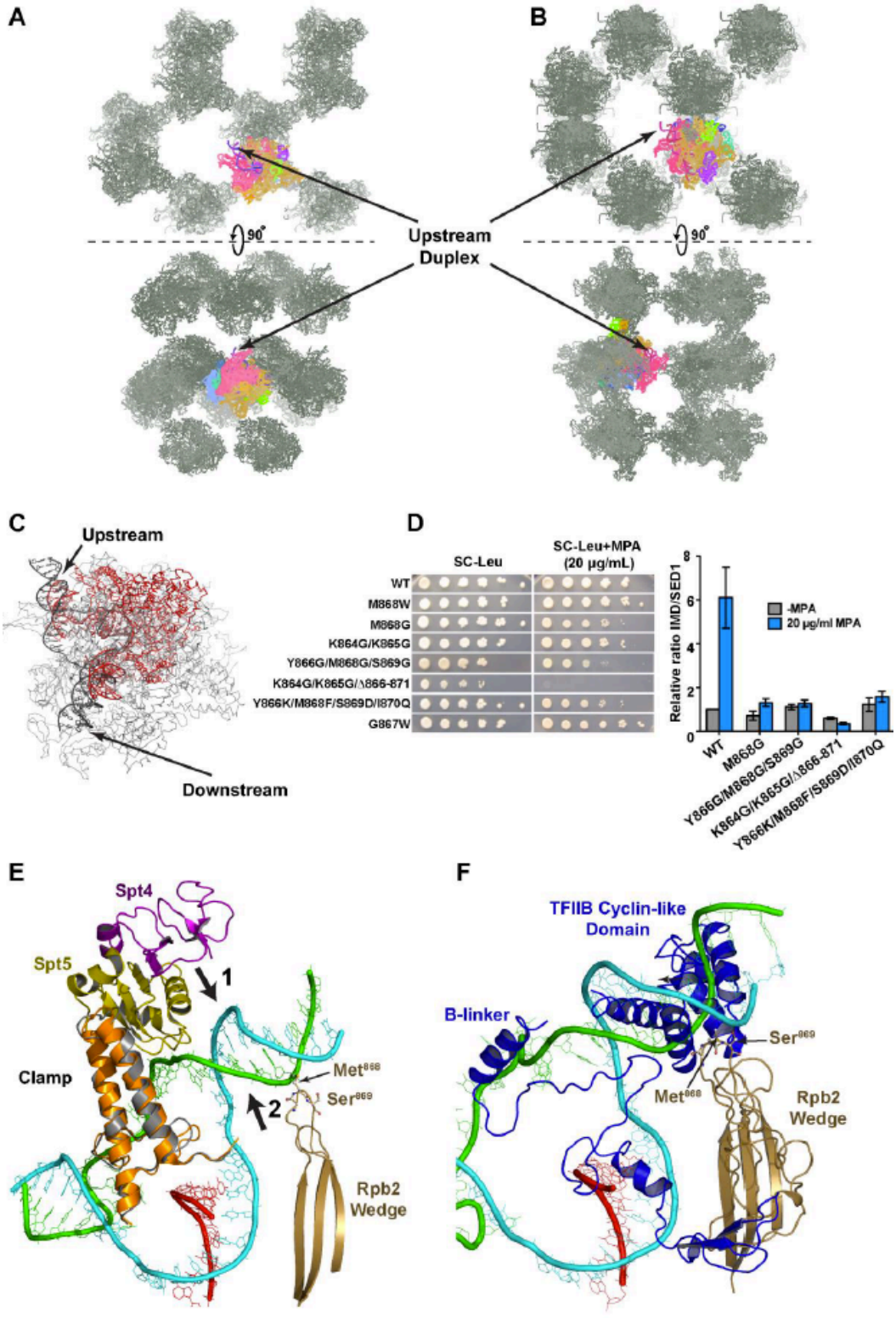

A
Met868
Gly867
Ser869
Tyr866
Ile870
Rpb2
Wedge
B
Wall
C
Fork
90˚
i+2
i+1
D
Maximum Elongation Rate (nt/s)
NTPs (μM)
WT
K864G/
K865G/
Δ866-871
WT
K864G/
K865G/
Δ866-871
E
0.2kb
1kb
2kb
4kb
6kb
7kb
Relative Pol II Occupancy
(Normalized to 0 min)
Wild Type
K864G/K865G/Δ866-871
Time in glucose (min)
F
Fold enrichment over
non-transcribed region
Promoter
0.2kb
1kb
2kb
4kb
6kb
7kb
GAL1
YLR454w
G
Relative ratio YLR454w/ScR1
YLR454w
ScR1
H
M868G
Y866G/M868G
/S869G
K864G/K865G/Δ866-871
Y866G/M868F/S869D
/I870Q
Change in fraction
ADH1 start site
ADH1 Start Site Positions

Figure A-6. Pol II transcribing complex crystal contacts and the role of Rpb2 wedge residues in upstream duplex stabilization.
(A) Crystal symmetry of Δ4/7-TC crystallized in C21 space group with novel unit cell dimensions 280x223x156 and angles 90°, 98°, 90°, never observed in previous PolΔ4/7 structures. Symmetry molecules were visualized in PyMol with upstream DNA positions highlighted (black arrow). The upstream duplex is solvent exposed. No contacts between molecules at the upstream end of the DNA indicate that the DNA architecture was not an artifact of crystal packing. (B) Crystal symmetry of Pol II-TC in space group C2221 with unit cell dimensions 220x393x282 and angles 90°, 90°, 90°. Symmetry molecules were visualized in PyMol with upstream DNA positions highlighted (black arrow). The upstream duplex is solvent exposed. No contacts between molecules at the upstream end of the DNA indicate that the DNA architecture was not an artifact of crystal packing. (C) Ribbon and cartoon representation of the overlay between the elongating human mitochondrial RNAP (red) PDB:ID 4BOC (Schwinghammer et al., 2013) and our structure (grey) about positions i+1 and i+2 on the template strand. (D) Left panel – Serial dilutions of yeast strains containing Rpb2 wedge alleles (mutations as designated on figure) on various media. Right panel – Quantitative analysis of Northern blotting for *IMD* gene expression (normalized to *SED1* as loading control) in presence or absence of 20μg/ml mycophenolic acid (MPA). Error bars are standard deviation of the mean (n=3). MPA sensitive alleles of Pol II can be of a number of classes, with disparate effects on *IMD2* expression, with a major class being defective for *IMD2* induction and also exhibiting upstream shifts in start site selection (Braberg et al., 2013; Kaplan et al., 2012). Also, we show that Rpb2 wedge alleles confer a downstream shift in start site utilization at *ADH1* (see main Fig. 3H), consistent with defective initiation and a shared phenotype with mutants in Pol II subunits and TFIIB (Kaplan, 2013). (E) Overlay of Δ4/7-TC with the archaea *P. furiosus* clamp-Spt4/5 complex, PDB:ID 3QQC (Martinez-Rucobo et al., 2011) about the clamp coiled coil domains of Pol II Rpb1 (grey) and RNAP A′ (orange), suggests that the upstream duplex is positioned between the NusG domain of Spt5 (olive) (arrow 1) and Rpb2 wedge domain (arrow 2). While speculative, our model suggests that Spt4/5 locks the upstream duplex in place and may play a role in maintaining bubble integrity during elongation. This is in agreement with previously published models (Klein et al., 2011; Martinez-Rucobo et al., 2011). (F) Structure overlay with initially transcribing Pol II-TFIIB structure, PDB ID: 4BBS (Sainsbury et al., 2012) illustrate possible interactions between wedge residues (copper) and Cyclin-like domain residues (blue). Comparisons between a model of the closed promoter complex (Kettenberger et al., 2003; Wang et al., 2009) and the structure of the transcribing complex suggest that transition from closed to open promoter entails a clockwise rotation (approximately 90º degrees) of the upstream end of the DNA. Wedge residues Met868 and Ser869 interact with TFIIB residues 148-153 and 178-183 from the cyclin-like domain, and such interactions could participate in Pol II duplex loading. It is possible that this rotational motion induces steric clashes that ultimately result in ejection of TFIIB and TBP allowing the wedge to engage the minor groove. The close proximity of these two regions could allow loading of the duplex on the wedge during the initiation to elongation transition.


A
B
Upstream
Duplex
90°
90°
C
Upstream
Downstream
D
SC-Leu
SC-Leu+MPA
(20 µg/mL)
WT
M868W
M868G
K864G/K865G
Y866G/M868G/S869G
K864G/K865G/Δ866-871
Y866K/M868F/S869D/I870Q
G867W
Relative ratio IMD/SED1
-MPA
20 µg/ml MPA
WT
M868G
Y866G/M868G/S869G
K864G/K865G/Δ866-871
Y866K/M868F/S869D/I870Q
E
Spt4
Spt5
Clamp
1
2
Met868
Ser869
Rpb2
Wedge
F
TFIIB Cyclin-like
Domain
B-linker
Ser869
Met868
Rpb2
Wedge

As the template strand separates from the RNA transcript (i−8) and emerges from Pol II's active site, non-specific packing interactions with FL1 residues and potential salt bridges between the phosphate chain and arch residues guide the template strand in a straight conformation toward its junction with the non-template strand (**Figure A-7**A, B, and **Figure A-8**A). Once above the arch, template strand nucleotides anneal with the non-template strand at i−12 (**Figure A-7**B and C). Stabilization of the nucleic acid scaffold by arch residues include interactions between template strand i−8, i−9, and i−11 with FL1 residues and between non-template strand i−11 and i−12 with rudder residue Lys^{317} (**Figure A-7** and **Figure A-8**B) (Treutlein, Muschielok et al. 2012). In addition to contacts that arch residues make with the template and non-template strands, rudder residue Arg^{320} reaches within H-bond distance of the 2′ hydroxyl of the 8^{th} RNA base of the nascent transcript (**Figure A-7**A and **Figure A-8**B, observed also in PDB: IYIW). This interaction is conserved in bacterial polymerases, where rudder Arg^{598} forms a H-bond with the 2′ hydroxyl group of the 7^{th} base on the nascent transcript (**Figure A-8**C) (Vassylyev, Vassylyeva et al. 2007) and has been corroborated by crosslinking experiments (Korzheva, Mustaev et al. 2000).

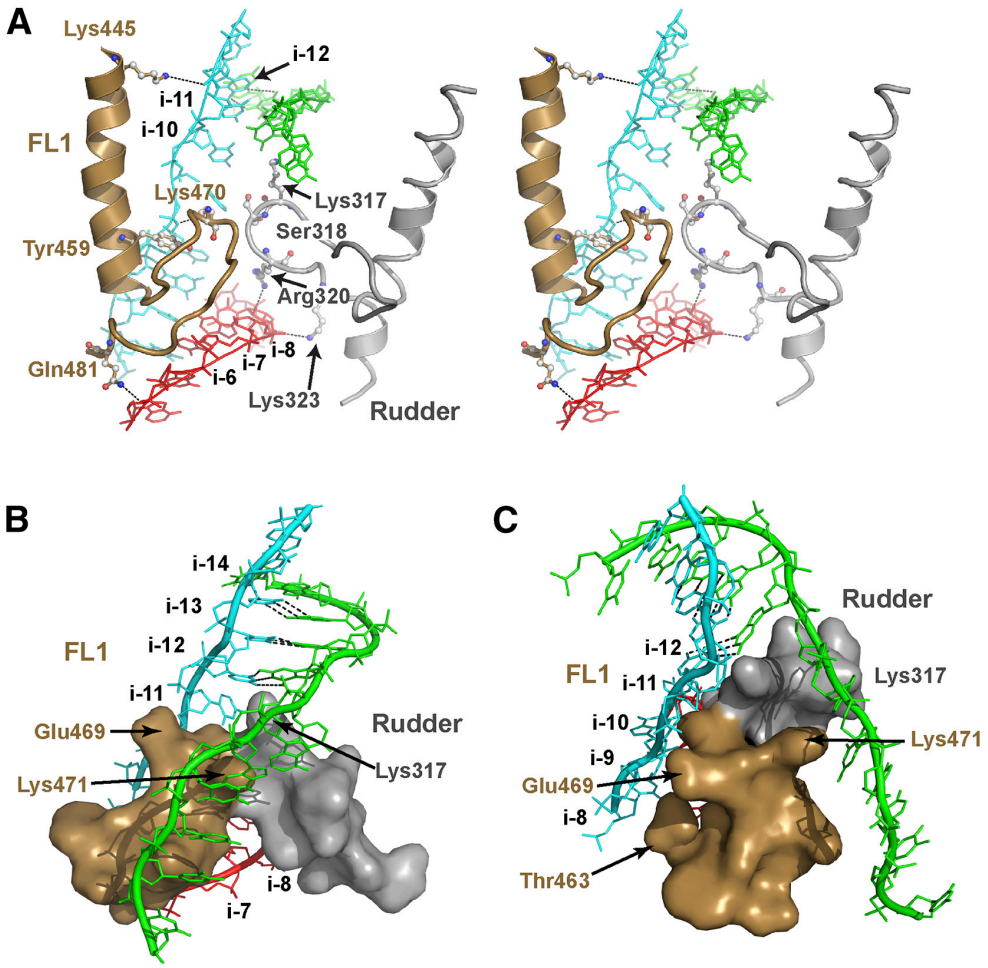
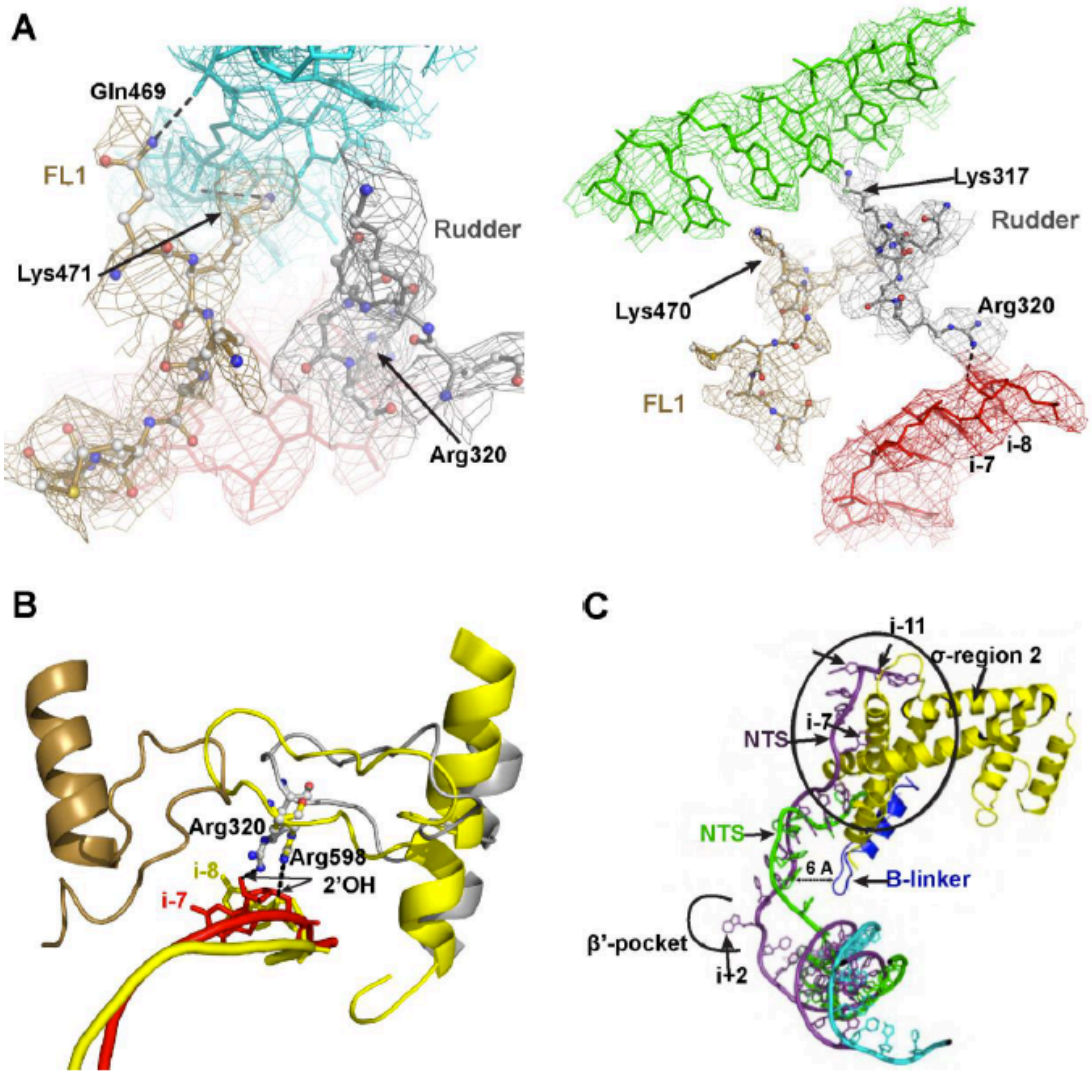


Figure A-7. Architecture of the Upstream Fork Junction.
(A) Stereo-view of the tripartite coordination of the three nucleic acid chains. Arch residues at the back, top, and bottom adopt unique conformations—with respect to apo- and elongation structures—that interact with template, non-template, and RNA strands, respectively. Rpb1 residue Arg^{320} forms a H-bond with the 2′-OH of the nascent transcript at position i−8. (B and C) Front view (B) and side view (C) of the upstream (closing) end of the bubble. Rudder (silver) and FL1 (sand) residues reach within 4 Å across the midline to form an arch that provides a scaffold for template and non-template strand annealing at i−12. Contacts include packing interactions between template strand i−8 and Tyr^{459} and potential H-bonds between Rpb2 residues Thr^{463} with template strand i−8, Glu^{469} with template strand i−10, and Lys^{471} with non-template strand i−9. Lys^{317} participates in contacts with non-template strand i−11 and i−12 (see also **Figure A-8**).

Figure A-8. Comparison of bacterial and eukaryotic RNA Polymerases.
A,B Architecture of the transcription bubble at the upstream (closing) end (see text). (A) template strand (blue) and arch residues, non-template strand (green), RNA (red) and arch residues. A refined 2Fobs-Fcalc electron density map (contoured at 1 σ) was calculated around both regions. (B) Interactions between rudder residues Arg320 (S. cerevisiae, silver) or Arg598 (T. thermophilus, yellow) with the 2´ hydroxyl of the nascent transcript (S. cerevisiae, red, T. thermophilus, yellow) are structurally conserved (observed also in transcribing complex PDB:ID 1Y1W). FL1 residues are not present in bacterial polymerases, instead a flexible β´ rudder (residues 583-603, yellow) (Vassylyev et al., 2007) reaches towards the mid-line forming a dome-like structure that spatially organize the junction of the two strands. (C) Structure overlay between sA factor region 2 (yellow) and open promoter complex (purple) from PDB:ID 4G7O and TFIIB bound to Pol II (PDB:ID 3K7A) (blue). Residues from sA region-2 and B-linker differ significantly. While sA can recognize specific bases at i-11 and i-7 (DNA scaffold in purple), the B-linker reaches within 6Å to interact with template and non-template strands at the opening end of the transcription bubble and could play a role in scaffold loading.


A
Gln469
FL1
Rudder
Lys471
Arg320
Lys317
Rudder
Lys470
Arg320
FL1
i-8
i-7
B
Arg320
Arg598
2'OH
i-8
i-7
C
i-11
σ-region 2
i-7
NTS
NTS
6 A
B-linker
β'-pocket
i+2

Strand Separation at the Downstream Transcriptional Fork

Strand separation in our structure appears to take place at i+5, where base-pair distance begins to increase progressively. However, it was not possible to define its exact location since our artificial bubble enforced non-complementarity precisely at i+5 (see below). However, our structure gives possible insight into the mechanism by which Pol II residues promote and sustain DNA strand separation. The template strand interacts with switch 1 and switch 2 residues, driving it toward the active site as originally described (Gnatt, Cramer et al. 2001), while the non-template strand interacts with two groups of Pol II residues. The first involves a positively charged cleft formed by Rpb1 residues Arg^{1386}-His^{1387}-Arg^{1391} (switch 1) and Lys^{1102}-Lys^{1109}-Asn^{1110} (located in a structurally conserved U-loop) in the vicinity of non-template strand i+5 and i+6 (**Figure A-9**A and **Figure A-10**A) (Kettenberger, Armache et al. 2004, Cheung, Sainsbury et al. 2011). The second involves interactions with Rpb2 fork loop 2 (FL2) residues 501–510 (**Figure A-9**A) providing packing contacts with the non-template strand i+3 to i+1 and Arg^{508} reaching within H-bond distance of the non-template phosphate chain (**Figure A-9**B and **Figure A-10**B). Structural overlay of FL2 residues from published crystal structures suggests they could be grouped in two major states (**Figure A-9**C). The first one is an "open" state (represented by PDB: 3PO2, 1Y1W, 3HOW), where FL2 residues interact with positions i+2 or i+3 on the non-template strand, respectively, allowing access to a non-specific nucleotide-binding pocket (Cheung and Cramer 2011). The second is a closed state, where FL2 residues appear to rotate about Pro^{501} and Pro^{510} (our structures, PDB: 3FKI, 3K7A), blocking access to the pocket. Positioning of FL2

residues in the latter conformation appears to guide the non-template strand under a four-strand β sheet dome (**Figure A-9**A) toward its junction above the arch. Other interactions include contacts with Rpb2 β strand 245–255 residues (**Figure A-10**B) and a patch of positive charges from the Rpb2 protrusion helix (**Figure A-4**C).

Fluorescence and Structural Experiments Suggest that the Downstream Fork Is Dynamic

Given that the precise location of strand separation was not able to be determined from the structure—and in light of previous structural studies showing a closed bubble at i+2 and i+3, respectively (Kettenberger, Armache et al. 2004, Cheung, Sainsbury et al. 2011), as well as fluorescence studies where the use of a 35-nucleotide scaffold with a partial (17 nucleotides) non-template strand showed complementarity at i+2 (Kashkina, Anikin et al. 2007)—we wished to ascertain whether it was possible to detect an open bubble in the presence of a 45-nucleotide scaffold bearing a fully complementary non-template strand (**Figure A-10**C). 2-aminopurine (2AP) is a fluorescent nucleotide analog that is significantly quenched upon base pairing to either T or C, as well as by stacking interactions with adjacent nucleotides (Stivers 1998, Liu and Martin 2001). Therefore, we placed 2AP in the template strand at positions i+2, i+3, i+5 (at the boundary of strand separation), and downstream at i+8, where the two DNA strands were likely to be paired (**Figure A-10**C).

Figure A-9. Architecture of the Downstream Fork Junction.
(A) Stereo-view of the architecture of the downstream fork. Relevant interactions include U-loop (UL) and switch 1 residues with non-template strand i+5 and i+6, FL2 residues with non-template strand i+2 to i+3, and Rpb2 residues 221–282 (forming a 5 strand β sheet, Dome) with non-template strand i+1 to i−2. (B) Unbiased *Fobs*− *Fcalc* electron density map contoured at 3 σ (FL2 residues 498–512 were not included in map calculation). (C) Structural overlay of published FL2 conformations during different stages of transcription. Δ4/7-TC (sand), PDB: 1Y1W (blue, elongation complex), PDB: 3HOW (cyan, backtracked complex), and PDB: 3PO2 (red, backtracked complex). (D) Representative spectra of 2AP probes at i+2 or i+3, bound to complementary non-template strand in the absence (open red circles and squares, respectively) and presence (blue circles and squares, respectively) of Pol II. The excitation wavelength was 315 nm and the fluorescence emission (shown in counts per second × 10^6) was collected from 340 to 400 nm. (E) Normalized fluorescence values for polymerase bound to ssDNA (primer-template) where 2AP is at the i+2, i+3, i+5, or i+8 position (red) and polymerase bound to dsDNA (primer-template annealed to fully complementary non-template strand) (blue). Error bars are SD of the mean (n = 3, see also **Figure A-10**).

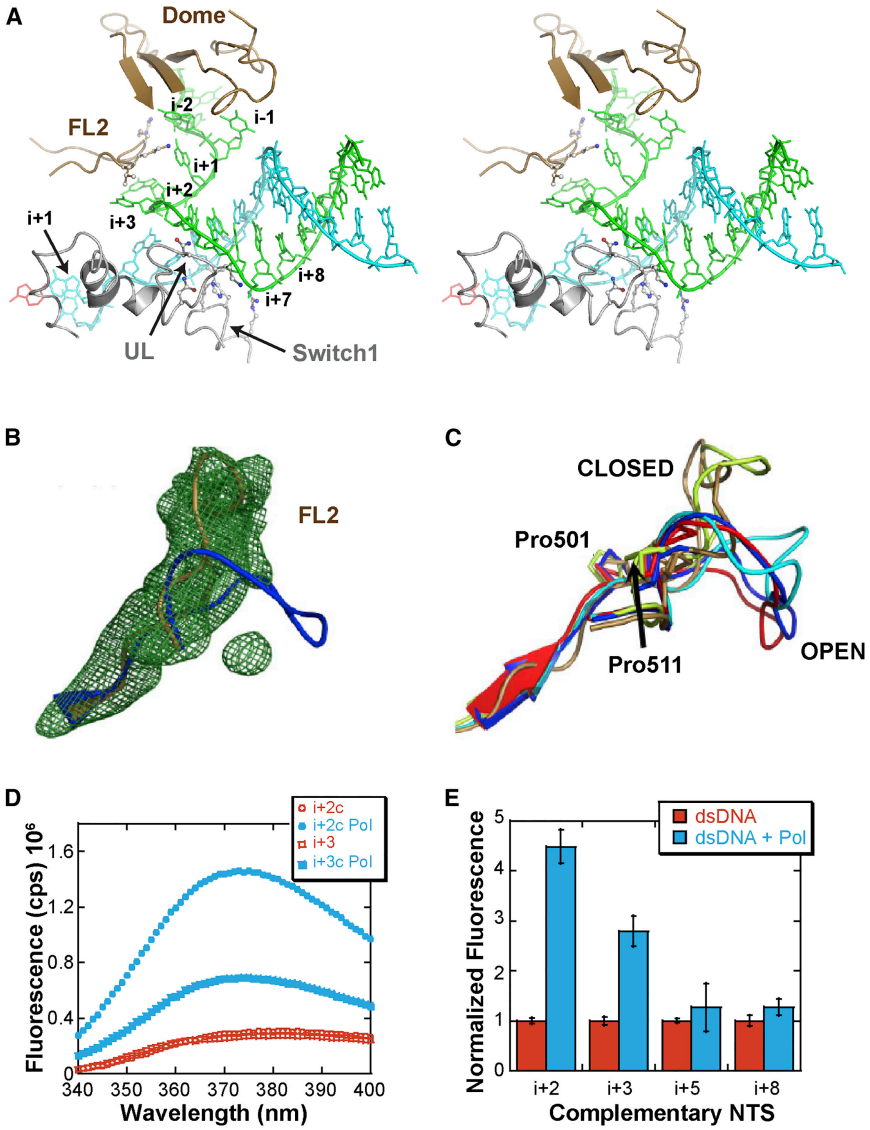
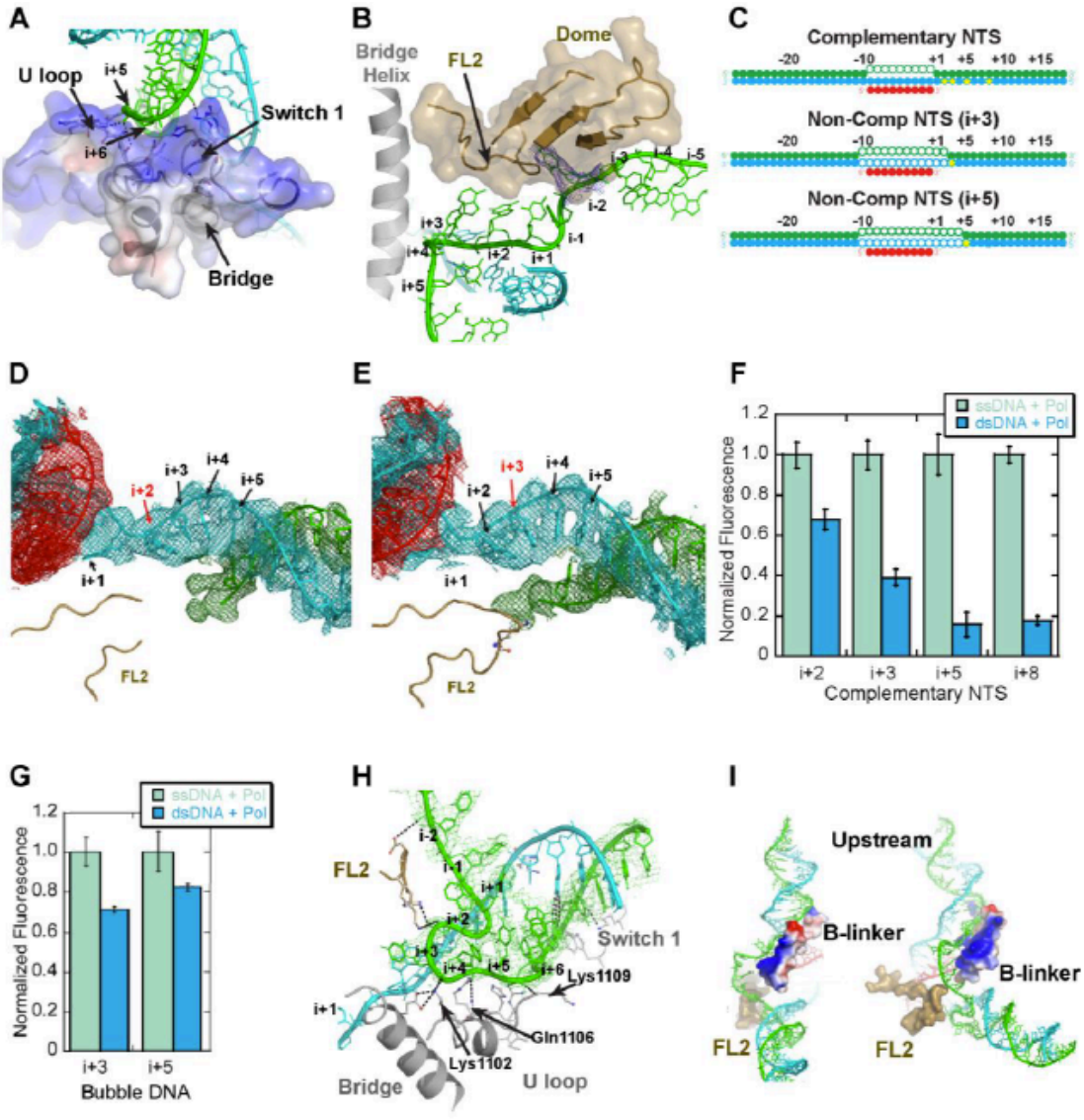

A
Dome
FL2
i-2
i-1
i+1
i+2
i+3
i+1
i+8
i+7
UL
Switch1
B
FL2
C
CLOSED
Pro501
Pro511
OPEN
D
i+2c
i+2c Pol
i+3
i+3c Pol
Fluorescence (cps) 10⁶
1.6
1.2
0.8
0.4
0
340
350
360
370
380
390
400
Wavelength (nm)
E
dsDNA
dsDNA + Pol
Normalized Fluorescence
5
4
3
2
1
0
i+2
i+3
i+5
i+8
Complementary NTS

Figure A-10. Investigating the dynamics of the downstream fork. (A) Solvent accessible surface electrostatic representation calculated using APBS (Baker et al., 2001) in PyMOL to illustrate how non-template strand elements (green) trail a crevice of positive charges (blue) at the opening end of the transcription bubble. Interactions between Pol II elements (U-loop and Switch 1) and the non-template strand phosphate backbone suggest that the region from i+3 to i+6 has the potential to be non-duplexed. (B) FL2 interactions with the non-template strand phosphate backbone lift the strand towards a short tunnel-like structure whose "dome" is a four-strand β-sheet from Rpb2. Electron density for non-template strand position i-2 is shown in blue. (C) Cartoon schematics of DNA represents the final annealed product used for each 2AP experiment, where yellow stars highlight the position of the 2AP probe, the template strand is in cyan, non-template strand in green, and RNA in red. (D,E) Electron density maps countered at 1s _of Pol II transcribing complexes composed of scaffold 2 + Tfg2 (D) and scaffold 3 alone (E) bearing non-complementary nucleotides at position i+2 or i+3 respectively. FL2 is disordered in i+2 but ordered in i+3 and is primed to engage with the non-template strand at the opening end of the transcription bubble. FL2 residue Arg504 reaches in close proximity to the nucleotide at i+3 position.
(F) Normalized fluorescence values for polymerase bound to ssDNA (template/RNA hybrid) where 2AP is at the i+2, i+3, i+5 or i+8 position before addition of the non-template (green) and after addition of non-template strand (blue). Error bars are standard deviation of the mean (n=3). (G) Comparison of 2AP fluorescence against non-complementary (bubble) DNA shows that when non-complementarity is extended to the i+3 or i+5 position, fluorescence measurements increase significantly compared to their complementary counterparts (panel F). Error bars are standard deviation of the mean (n=3). (H) Refined 2Fobs-Fcalc electron density map (green, contoured at 0.8s) calculated around the opening end of the non-template strand in the Pol II-TC. Evidence of a scrunched state –where template and non-template strand bases are compacted in space relative to relaxed conformations – has been proposed during the early stages of transcription initiation (Kapanidis et al., 2006; Revyakin et al., 2006). Similarly, other transcriptional events such as backtracking or interactions with elongation or termination factors might alter the number of bases (and location) in the mature transcription bubble. (I) Cartoon and surface representation of a front view (left) and front view rotated 60° counterclockwise (right) of the overlay between PDB:ID 3K7A (Liu et al., 2010) and Δ4/7-TC. The B-linker region of the general transcription factor TFIIB reaches within 6 Å at the downstream fork junction and could interact with the non-template strand and assist transcription bubble loading during initiation. The electrostatic surface representation calculated using APBS (Baker et al., 2001) in PyMOL illustrates that one face of the B-linker is positively charged and could possibly allow interactions with non-template strand phosphate chain. FL2 residues could potentially assist during bubble loading via interactions with the non-template strand.


A
U loop
i+5
i+6
Switch 1
Bridge
B
Bridge Helix
FL2
Dome
i-3
i-4
i-5
i-2
i-1
i+3
i+4
i+2
i+1
i+5
C
Complementary NTS
-20
-10
+1
+5
+10
+15
Non-Comp NTS (i+3)
Non-Comp NTS (i+5)
D
i+2
i+3
i+4
i+5
i+1
FL2
E
i+3
i+4
i+5
i+2
i+1
FL2
F
ssDNA + Pol
dsDNA + Pol
Normalized Fluorescence
1.2
1.0
0.8
0.6
0.4
0.2
0
i+2
i+3
i+5
i+8
Complementary NTS
G
ssDNA + Pol
dsDNA + Pol
Normalized Fluorescence
1.2
1.0
0.8
0.6
0.4
0.2
0
i+3
i+5
Bubble DNA
H
FL2
i-2
i-1
i+1
i+2
i+3
i+4
i+5
i+6
Switch 1
Lys1109
i+1
Gln1106
Lys1102
Bridge
U loop
I
Upstream
B-linker
B-linker
FL2
FL2

As expected when the primer-template is annealed to a complementary non-template strand, full quenching is observed, consistent with stable base pairing of 2AP (**Figure A-9**D and E). When polymerase was added to the scaffold with a fully complementary non-template strand, we see a 4.5-fold increase in fluorescence for the i+2 substrate and a 2.8-fold increase in fluorescence for the i+3 (**Figure A-9**D and E). This increase in fluorescence indicates a disruption in stable 2AP base pairing at both locations and could suggest that Pol II can unwind the non-template strand at i+3. There is 1.2-fold fluorescence enhancement for i+5 and none for i+8, suggesting 2AP is more stably base paired at those positions (**Figure A-9**E). Moreover, crystal structures of Pol II bound to scaffolds 2 and 3 bearing non-complementary transcription bubbles with strand separation at i+2 and i+3, respectively; both showed base-pairing at i+5, recapitulating the results of 2AP experiments (**Figure A-10**D and E).

To observe a full non-template strand, it was necessary to utilize a bubble with non-complementary DNA to i+5, since Pol II transcribing complexes bearing non-complementary bubbles at positions i+2 and i+3 showed minimal non-template strand density (scaffolds 2 and 3; **Figure A-3**A, **Figure A-10**D and E). Thus, we analyzed whether forced non-complementarity at i+5 in the transcribing complex contributed to increased mobility of the non-template strand and its subsequent capture in our scaffold 1 complex. Consistent with this possibility, 2AP experiments where complementary base pairing begins at positions i+3 and i+5 showed an increased fluorescence signal when compared to bases where complementarity begins at the i+1 position (**Figure A-10**F and G). This might indicate increased motion of the non-

template strand at the leading edge of the artificial transcription bubble bearing non-complementary bases. Thus, it is possible that the use of a scaffold with non-complementary base pairing at i+5 resulted in a kinked conformation of the non-template strand in Pol II-TC structure (**Figure A-10**H). It is also possible, while highly speculative, that the presence of Tfg2 (without Tfg1) in the cleft induced such conformation.

Interactions with the Non-template Strand Are Associated with the Off State of the Trigger Loop

A highly conserved loop comprising Rpb1 residues 1078–1097, the trigger loop (TL), has been shown to play a fundamental role in nucleotide selection and catalysis, while it is also proposed to govern translocation (Bar-Nahum, Epshtein et al. 2005, Wang, Bushnell et al. 2006, Feig and Burton 2010, Kaplan, Jin et al. 2012, Kireeva, Opron et al. 2012, Larson, Zhou et al. 2012, Kaplan 2013). Evidence from single-molecule studies on Pol II indicates that mutation of TL residues alters Pol II translocation properties, consistent with these models (Larson, Zhou et al. 2012). Structurally, the TL locates between Rpb1 helices 1064–1078 ($TL\alpha_1$) and 1097–1106 ($TL\alpha_2$), which in turn are part of a universally conserved five-helix bundle (hereafter known as TL bundle, TLB) that includes Rpb1 helices 826–846 (bridge helix), 1340–1357 ($TL\alpha_4$), and 1365–1379 ($TL\alpha_5$) supported by packing of hydrophobic residues at the bundle core (Figure A-11A).

Figure A-11. Conformational changes involved in DNA translocation.
(A) Cartoon and surface representation of the Rpb1 5-helical bundle comprising the trigger loop-bundle (TLB). a1, residues 1063-1077; a2, residues 1098-1106; Bridge helix, residues 810-845; a4, residue 1341-1357; and a5, residue 1365-1377. TL helices a1-a2 (hinges), and the U-loop of Pol II-TC are illustrated in yellow (Inset, also shows non-template strand i+5 and i+6). The core of the bundle is held together by hydrophobic interactions. (B) Conformational changes observed between "off" and "on" (structural overlay with PDB:ID 2NVZ, blue) states of the TL and Rpb1 funnel helices. All "on" state residues are indicated in blue, "off" state residues are indicated in grey for Pol II-TC. A 2Fobs-Fcalc map rendered at 1.0 s is contoured around TL and funnel residues. (C) Conformational changes observed between various "off" states of the TL: Δ4/7-TC, silver; Pol II-TC, yellow; PDB:ID 1Y1V, red; PDB:ID 3CQZ, magenta; PDB:ID 2NVX, green. Differences in secondary structure for TL residues 1093-1098 in Pol II transcribing structures, suggests that we were able to capture "off" state intermediates. (D) Cartoon representation of Rpb5 residues during "off" (illustrated in magenta) and "on" conformations, (illustrated in blue). During "on" state three loops of the jaw domain become disordered, suggesting that this region "moves" during this stage facilitating translocation of the nucleic acid scaffold. Additional conformational changes are observed in Rpb5 wedge residues. These changes could be related to motion in TLB residues. (E) Illustration of residues that potentially contribute to hydrogen bond stabilization of the nucleic acid scaffold. Rpb1, Rpb2 and Rpb5 residues are illustrated in silver, copper and magenta respectively. Over 40 potential H-bonds are disrupted during each translocation cycle (H-bond distances of 3.5 Å or less). (F) Surface representation of a transcribing Pol II complex. Cyan regions highlight interactions between Pol II and the nucleic acid scaffold, which may comprise the structural framework for high fidelity DNA tracking.

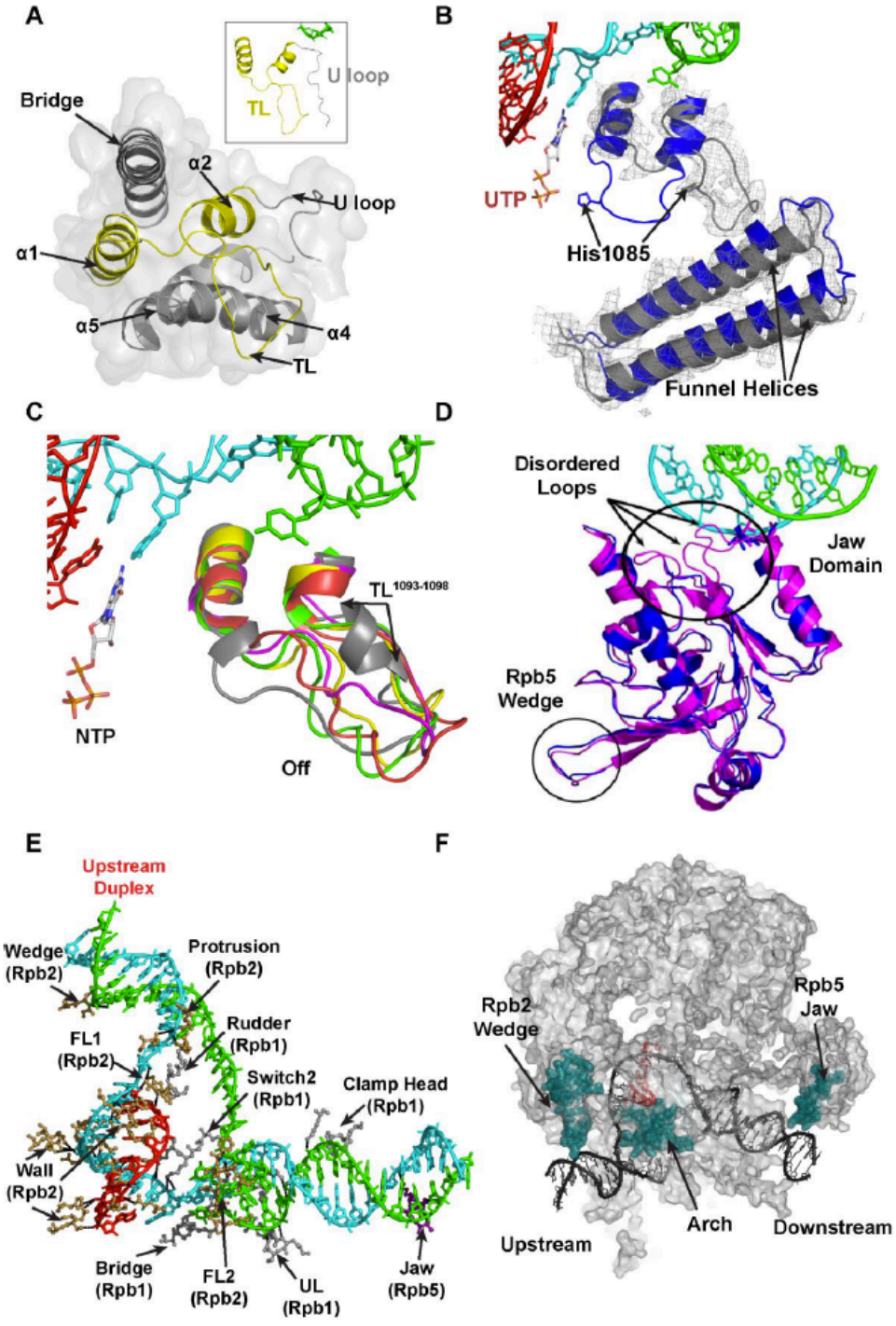
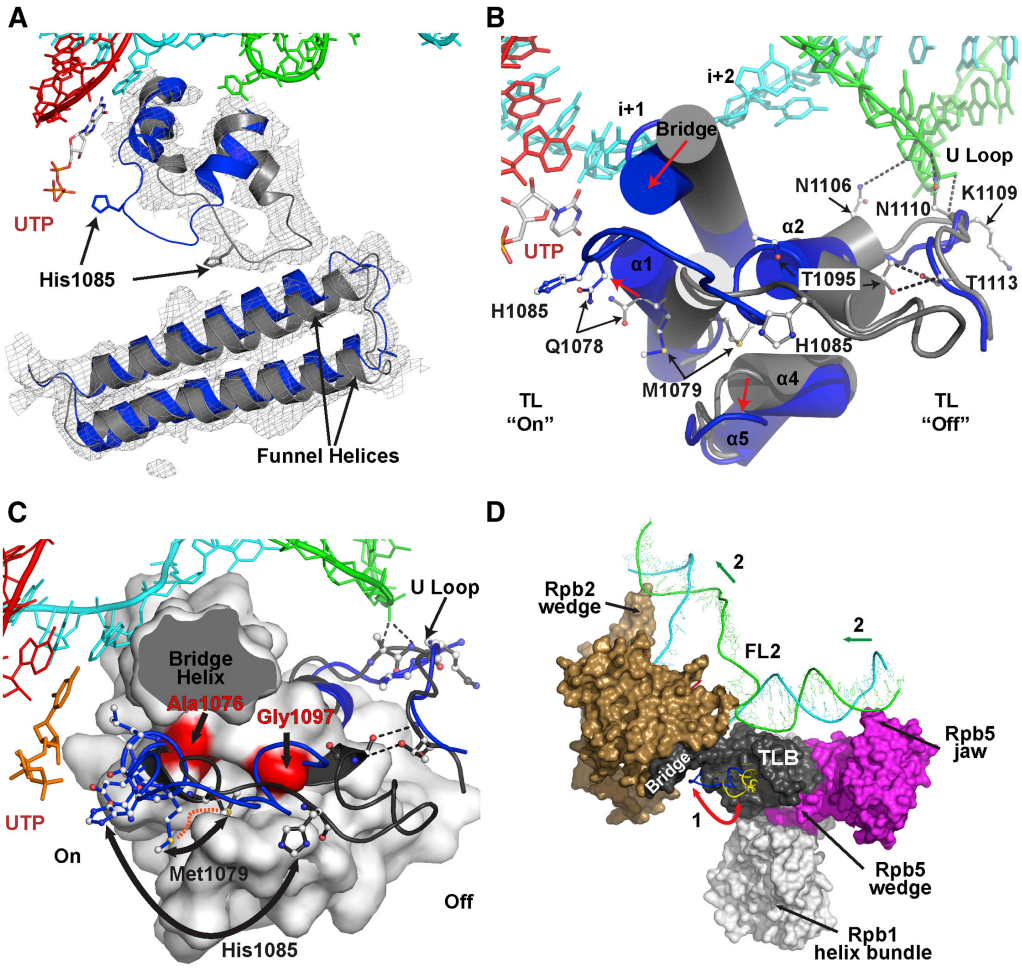

A
Bridge
α2
U loop
α1
α5
α4
TL
U loop
TL
B
UTP
His1085
Funnel Helices
C
NTP
TL 1093-1098
Off
D
Disordered Loops
Jaw Domain
Rpb5 Wedge
E
Upstream Duplex
Wedge (Rpb2)
Protrusion (Rpb2)
Rudder (Rpb1)
FL1 (Rpb2)
Switch2 (Rpb1)
Clamp Head (Rpb1)
Wall (Rpb2)
Bridge (Rpb1)
FL2 (Rpb2)
UL (Rpb1)
Jaw (Rpb5)
F
Rpb2 Wedge
Rpb5 Jaw
Arch
Upstream
Downstream

Figure A-12. Trigger Loop and Nucleic Acid Scaffold Interactions during Translocation. (A–D) On- and off-state residues will be indicated in blue and gray, respectively. A modeled UTP (PDB: 2NVZ) is indicated in light gray/orange.
(A) Conformational changes observed between off and on (structural overlay with PDB: 2NVZ) states of the TL and Rpb1 funnel helices. A *2Fobs− Fcalc* map rendered at 1.0 σ is contoured around TL and funnel residues. (B) Conformational changes observed between TL off and on states. Red arrows indicate motion. During off-/on-state conformational changes, most TL stabilizing interactions are disrupted, including: (1) release of Met^{1079} from its hydrophobic pocket, (2) disruption of α2-bridge helix H-bonds, resulting in bridge helix displacement, (3) disruption of Thr^{1095}-Thr^{1113} H-bonds allowing counterclockwise TL motion, (4) disruption of non-template strand/U-loop bonds, possibly leading to non-template strand release and translocation.
(C) Mutations of residues that disrupt Met^{1079} hydrophobic pocket result in gain-of-function phenotypes (Kaplan et al., 2012). TLB residues are represented as a solid silver surface. Motion of Met^{1079} might occur through a defined pathway on the protein surface (orange trace). Mutations of Ala^{1076} and Gly^{1097} (red surface) for bulkier residues can potentially disrupt the vestibule of the hydrophobic pocket (Kireeva et al., 2012).
(D) Possible coupling of the global translocation of the scaffold to local motion of the TL. Pol II regions in contact with upstream and downstream duplexes, Rpb2 (sand) and Rpb5 (magenta), respectively, are coupled through TLB residues (dark gray). TL off and on conformations are illustrated in yellow and blue, respectively (see also **Figure A-11**).


A
UTP
His1085
Funnel Helices
B
i+2
i+1
Bridge
U Loop
N1106
N1110
K1109
α2
UTP
α1
T1095
T1113
H1085
Q1078
H1085
M1079
α4
TL
"On"
α5
TL
"Off"
C
U Loop
Bridge
Helix
Ala1076
Gly1097
UTP
On
Met1079
Off
His1085
D
2
Rpb2
wedge
2
FL2
Rpb5
jaw
Bridge
TLB
1
Rpb5
wedge
Rpb1
helix bundle

The Pol II TL is intrinsically mobile and has only been detected in X-ray structures when bound to small molecules (such as a matched NTP or α-amanitin) or protein-ligands (TFIIS) that each stabilize a particular conformation (Kettenberger, Armache et al. 2003, Wang, Bushnell et al. 2006, Brueckner and Cramer 2008, Kaplan, Larsson et al. 2008). Observed conformations define a closed "on" state where TL residues interact with a matched nucleotide in the addition or "A" site (effectively isolating a reaction chamber and blocking access to additional substrates) (Feig and Burton 2010, Kireeva, Opron et al. 2012) or an open "off" state where TL residues move away from the A site. During advanced stages of refinement, electron density for the full TL backbone was clearly discernible (**Figure A-12**A and **Figure A-11**B), and partial residue placement was feasible with help of map sharpening in *Coot*. TL backbones were found in "off" states that differ from previously reported conformations (Kettenberger, Armache et al. 2003, Wang, Bushnell et al. 2006, Wang, Bushnell et al. 2009) such as the Pol II-TFIIS complex, where direct contacts with TFIIS displaced and stabilized TL residues (**Figure A-11**C). Interestingly, our structures show two distinct off-state conformations. The first one present in the Δ4/7-TC closely resembles the TL loop in the "on" conformation (PDB: 2E2H, 2NVZ) (Wang, Bushnell et al. 2006) but rotated counter-clockwise approximately 60° away from the addition site (**Figure A-12**A). The second one, observed in Pol II-TC, resembles a hairpin loop, and its shape is similar to previously reported off-state structures (**Figure A-11**A–C) (Kettenberger, Armache et al. 2004, Wang, Bushnell et al. 2006). Both conformations induce considerable changes in funnel and neighboring residues (**Figure A-12**A and **Figure A-**

11B). Despite these differences in the positions of the TL's off state, a common set of interactions can be observed among these structures and all structures crystallized with a DNA scaffold (exclusive of the on-state structures). These include: (1) interactions between TL hinge regions ($TL\alpha_1$ and $TL\alpha_2$ helices) with neighboring Rpb1 residues and (2) the burying of Met^{1079} inside a small hydrophobic pocket at the core of the TLB (**Figure A-12**B).

Structural overlay between off and on states suggests substantial rearrangements in TLB helices and TL residues during the nucleotide addition cycle (**Figure A-12**B). Transition to the on state involves TL residues swinging toward the addition site, where hinge contacts are disrupted and Met^{1079} moves out of its pocket at the core of the TLB (Figure 3-12C). Genetic interactions and gene expression profiling of substitutions of Rpb1 Ala^{1076}, Gly^{1097}, and Leu^{1101} with residues that disrupt the Rpb1 Met^{1079} hydrophobic pocket (Braberg, Jin et al. 2013) support a model where the integrity of the pocket is critical for stabilization of the off state in eukaryotic (Kaplan, Larsson et al. 2008, Kaplan, Jin et al. 2012) and archaeal (Fouqueau, Zeller et al. 2013) polymerases. Such substitutions are highly related to those that hamper off-state conformations by destabilization of the C-terminal TL hinge region; therefore, these substitutions are similarly predicted to alter translocation rate and catalysis (Malagon, Kireeva et al. 2006, Wang, Bushnell et al. 2006, Kaplan, Larsson et al. 2008, Kaplan, Jin et al. 2012, Kireeva, Opron et al. 2012), suggesting that the TL "off" state is specifically required for proper transcription.

Structurally Linked DNA-Interacting Domains from Rpb2, Rpb5, and TLB Residues Could Possibly Coordinate Translocation

Three regions of Pol II furnish residues that can potentially form H-bond contacts with the nucleic acid scaffold (**Figure A-11**E). These include (from upstream to downstream): (1) wedge, wall, and FL1 (Rpb2) with the upstream duplex, template strand, and the non-template strand (respectively); (2) TLB, rudder, switch (1 and 2), and clamp residues (Rpb1) with DNA-RNA hybrid, template strand, downstream fork, and downstream duplex (respectively); and (3) jaw residues (Rpb5) with the downstream duplex (**Figure A-12**D, **Figure A-11**E-F). Remarkably, these regions have substantial interactions among them: TLB helices ($TL\alpha_4$ and $TL\alpha_5$) and Rpb1 residues from a seven-helix bundle ($Rpb1^{846-1064}$) form a large pocket that buries a two-strand hairpin from $Rpb5^{193-214}$, which is in the immediate neighborhood of the jaw motif. Similarly, TLB contacts (via the bridge helix) with Rpb2 are extensive and involve "wall" residues (with the phosphate chain of the template strand) in the immediate neighborhood of Rpb2's wedge (**Figure A-5**B and **Figure A-12**D). Since these regions are coupled extensively, it is possible that they could play an important role during translocation.

Discussion

Molecular Basis for DNA-Tracking

Real-time microscopy experiments demonstrated the ability of RNA polymerases to rotate DNA by tracking with high fidelity its right-hand helix (Harada, Ohara et al. 2001). Our structures suggest that it is possible that engagement by wedge (upstream

duplex), arch (closing end of the bubble), and jaw and clamp (downstream duplex) residues (**Figure A-4, Figure A-3** and **Figure A-11**F) could comprise the structural framework that explains such tracking mechanisms. The interactions that Rpb1 (head clamp) and Rpb5 (jaw) residues make with the downstream duplex were described in several published crystal structures (Westover, Bushnell et al. 2004, Wang, Bushnell et al. 2006). However, since contacts with the upstream duplex and arch were not previously observed, the correlation between DNA tracking, and its structural underpinnings could not be established. Moreover, given that the crystal structures of Pol II transcribing complex and mtRNAP show interactions with upstream and downstream duplexes, it is possible that tracking mechanisms are conserved in transcription (**Figure A-5**C **and Figure A-6**).

We assessed multiple *rpb2* wedge alleles to ascertain the role of these elements in vivo. The strongest growth effects observed in vivo required removal of the loop (**Figure A-6**D), which may be considered an extensive perturbation to the Pol II structure. However, our interpretation of the direct or indirect functions of this loop is based on what is known about other Pol II mutants. We found that wedge alleles were MPA sensitive and shifted start sites downstream (**Figure A-5**H and **Figure A-6**D). This profile is relatively rare for Pol II alleles or known general transcription factor alleles (reviewed in (Kaplan 2013)), especially for mutants unrelated to the active center (Rpb1 N488D within the Pol II active site has this phenotypic profile; (Malagon, Kireeva et al. 2006)). Recent crystal structures of Pol II-TFIIB structures show that the wedge domain interacts with the TFIIB core N-terminal cyclin fold (Sainsbury, Niesser et al. 2013).

Since TFIIB is necessary for start site selection, it seems plausible that changes in Pol II-TFIIB interactions could affect this process. Deletion of the wedge loop confers a strong defect in occupancy of Pol II at a reporter gene, consistent with an initiation defect. However, the wedge mutants do not phenocopy TFIIB (*sua7*) alleles in relation to MPA sensitivity or *IMD2* expression (**Figure A-6**D), suggesting that some functions may be independent of TFIIB or relate to an initiation defect not observed in particular *sua7* alleles. The wedge domain may possibly have a role during transition from initiation to early elongation by providing interactions with the upstream duplex that could assist TFIIB ejection (**Figure A-6**F).

Downstream Fork Flexibility

The downstream fork is a dynamic region, and different conformations of the non-template strand could be allowed during different stages of transcription. Structural overlay of the non-template strand from our structures and published structures crystallized in the presence of a nucleic acid scaffold show small positional differences from i+5 to i+7, which are anchored by U-loop and switch 2 residues (Figure 3-10A). However, positions i+2 to i+4 differ in all crystal structures and can be found paired or un-paired to the template strand. These observations suggest that there may be a range within which the leading edge of the downstream fork can move; hence, fluidity within this region could be required to promote bubble opening and maintenance. Moreover, non-template strand flexibility correlates with FL2 flexibility (**Figure A-9**C), which adopts multiple conformations (Wang, Bushnell et al. 2006, Meyer, Ye et al. 2009, Liu, Bushnell et al. 2010, Cheung and Cramer 2011) to allow interactions with the non-

template strand during the various transcriptional stages. The size of the bubble itself, which appears to be dynamic, might not be so critical due to the fact that Pol II can accommodate different lengths and possibly different locations of the transcription bubble inside the cleft (Pal, Ponticelli et al. 2005).

The crystal structure of the ternary complex between *T. thermophilus* polymerase (RNAP), σ^A, and an open promoter complex (Zhang, Feng et al. 2012) shows that FL2 residues fold to form part of a 2′-deoxy-GTP "selectivity pocket" for non-template strand i+2 during initiation. Also, the crystal structure of the RNAP ternary complex revealed that promoter recognition, melting, and bubble loading are carried out by σ^A region-2 residues (**Figure A-8**C). Thus, architecture of the downstream fork in RNAP during initiation is determined by σ^A as well as FL2 residues.

The recent crystal structure of *E. coli* RNAP holoenzyme in complex with a 15-nucleotide bubble at 6 Å resolution revealed the architecture of the RNAP transcription bubble stabilized by σ elements (Zuo and Steitz 2015). However, overlay of RNAP bubble structure with our structures shows different trajectories of the non-template strand and also different position of the upstream duplex. Such differences could be explained structurally, since on the one hand FL1 residues (not conserved in RNAP) would clash with the non-template strand of RNAP, and on the other, the presence of σ, which itself interacts with the non-template strand, in the RNAP cleft shifts dramatically the position of the upstream duplex.

Interestingly, overlay of the co-crystal structures of Pol II in complex with TFIIB (Bushnell, Westover et al. 2004, Kostrewa, Zeller et al. 2009, Liu, Bushnell et al. 2010)

with our structure shows that TFIIB linker domain reaches within 6.5 Å of the fork junction and could assist bubble loading (Liu, Bushnell et al. 2010, Sainsbury, Niesser et al. 2013) and prevent re-annealing of downstream fork during initiation (**Figure A-10**I). Thus, σ^A and TFIIB could sustain an open fork (during initiation) using different mechanisms.

TL Allosteric Effects and DNA Translocation

It has been proposed by a number of groups that TL movement contributes to or controls translocation (Bar-Nahum, Epshtein et al. 2005, Brueckner and Cramer 2008, Feig and Burton 2010, Kaplan, Jin et al. 2012, Larson, Zhou et al. 2012). Our structures show Pol II in a post-translocated state, with a TL in the off state due to interactions with neighboring Rpb1 residues (**Figure A-12**A-B, and **Figure A-11**B). It is possible that a matched NTP at the i+1 position might disrupt these interactions, leading to an on-state confirmation. The on-state structure (Wang, Bushnell et al. 2006) shows Gln^{1078} as one of the key residues stabilizing a matched nucleotide through formation of H-bonds with the ribose. Structural overlay of TL residues between off and on conformations shows that Gln^{1078} moves approximately 3 Å to form H-bonds with the matched nucleotide (**Figure A-12**B; see also (Cheung, Sainsbury et al. 2011). The position of this residue might be critical, since on the one hand it could constitute part of the nucleotide selection mechanism (Yuzenkova and Zenkin 2010, Fouqueau, Zeller et al. 2013); on the other hand, displacement of Gln^{1078} could trigger extraction of the neighboring Met^{1079} from its hydrophobic pocket, initiating a cascade of events that would lead into the full on state. Along with our observations, genetic evidence has shown that mutations

of Gln^{1078} have a comparable effect on Pol II elongation activity as mutations on the catalytic His^{1085} (Kaplan, Jin et al. 2012). Importantly, Gln^{1078} and His^{1085} are genetically distinguishable, suggesting a multistep process in TL function, consistent with initial substrate-Gln^{1078} interactions and subsequent TL movement or folding. Moreover, Pol II activity is also exquisitely sensitive to substitutions around the Met^{1079} hydrophobic pocket (Figure A-12C) (Kaplan, Jin et al. 2012, Kireeva, Opron et al. 2012). These substitutions invariably lead to genetic phenotypes, genetic interaction, and gene expression profiling phenotypes consistent with increased Pol II activity, most likely due to destabilization of the TL off state (Braberg, Jin et al. 2013).

Critical questions for the Pol II mechanism are: how does translocation occur, and what are the molecular determinants of its linkage to the nucleotide addition cycle? Our structure reveals that a second pivotal role played by TLB residues could possibly include stabilization of the non-template strand. Crystal structures of Pol II bound to partial or full nucleic acid scaffolds (Kettenberger, Armache et al. 2004, Westover, Bushnell et al. 2004, Wang, Bushnell et al. 2006, Cheung, Sainsbury et al. 2011) show that $TL\alpha_2$ and UL residues Lys^{1102}, Asn^{1106}, Lys^{1109} (amide backbone), and Asn^{1110} locate within H-bond distance to the phosphate chain of non-template strand i+5 and i+6 (**Figure A-10**A). Overlay of the two states shows that the hinged motion of the TL during matched nucleotide binding results in conformational changes leading to increased U-loop/non-template strand distance and hence disruption of potential H-bond contacts (**Figure A-12**B and C). Although speculative, this could suggest that TL "off/on" transitions are allosterically coupled to "latch and release" (respectively) events

of the non-template strand phosphate chain by U-loop residues. Furthermore, comparisons between on- and off-state conformations show that on-state Rpb5 jaw residues (interacting with the downstream duplex) become disordered, hence decreasing the number of effective contacts with the duplex (**Figure A-11**D). Since TLB, Rpb5 (jaw), and Rpb2 (wall, wedge) regions are coupled through extensive interactions, it is possible that TL conformational changes are transferred allosterically to downstream and upstream duplexes to assist global translocation (**Figure A-12**D, **Figure A-11**E-F).

Our refined structure sheds light into four fundamental mechanistic aspects of transcription: (1) In addition to hybrid interactions (Gnatt, Cramer et al. 2001), Pol II has four major contact points with a nucleic acid scaffold: wedge, arch, clamp, and jaw domains (**Figure A-4**A and B). Engagement of the minor groove by Rpb2 wedge (upstream), arch (closing end of the bubble), and jaw and clamp (downstream) residues could comprise the structural framework that explains the mechanism for high-fidelity DNA "tracking" observed using real-time optical microscopy (Harada, Ohara et al. 2001). (2) Pol II residues define the architecture of the transcriptional fork. Arch residues coordinate annealing of template and non-template strands at the upstream fork; the downstream fork is a highly dynamic area where FL2 residues accommodate different conformations of the non-template strand. (3) The structure also suggests that Gln^{1078} is positioned to couple extraction of Met^{1079} out of the hydrophobic pocket to interactions with an incoming matched NTP, initiating a cascade of events leading to a full on state of the TL, followed by nucleotide incorporation and subsequent DNA/RNA translocation. (4) Finally, our structure shows that Pol II regions in contact with the

nucleic acid scaffold are connected as rigid bodies (from Rpb1, Rpb2, and Rpb5) and that TL on-/off-state conformational changes could possibly be tied to global translocation.

Experimental Procedures

Pol II Transcribing Complex Purification and Assembly

Saccharomyces cerevisiae Pol II and Tfg2 were purified as previously described (Chung, Craighead et al. 2003, Wang, Bushnell et al. 2006). To assemble a Pol II transcribing complex, equimolar concentrations of oligonucleotides containing a single stretch of non-complementary bases and a 9-mer RNA were annealed. The resulting nucleic acid scaffolds were mixed with Pol II (3:1 molar ratio), and excess scaffold was removed using size-exclusion chromatography (Superdex200, GE LifeSciences) against Buffer A (25 mM HEPES [pH 7.5], 100 mM KCl, 5 mM DTT, 0.5 mM EDTA, 10 μM $ZnCl_2$). Purification and assembly of Pol II-TC and Δ4/7-TC was achieved as previously described (Pullara, Guerrero-Santoro et al. 2013). An SDS-PAGE of the final complex is illustrated in **Figure A-1**E. Ethidium bromide staining confirmed the presence of the nucleic acid scaffold (**Figure A-1**E). Catalytic activities in the presence of TFIIF or Tfg2 were not determined.

In Vitro Elongation Assay

Pol II enzymes for in vitro assays were purified from yeast strains expressing wild-type or the mutant *rpb2* gene from a low copy plasmid, as described above. In vitro elongation assays were performed as described in (Kaplan, Larsson et al. 2008, Kaplan,

Jin et al. 2012) with minor modifications in the amount of nucleic acids and Pol II used for elongation complexes.

Chromatin Immunoprecipitation Assays

Epitope-tagged (*RPB3::3XFLAG::kanMX*) wild-type or mutant strain containing a galactose-inducible *YLR454w* reporter (*kanMX::GAL1p::YLR454w*) gene were used for ChIP assays. Chromatin immunoprecipitation experiments for in vivo elongation rate determination were performed as described previously (Hazelbaker, Marquardt et al. 2013), with slight modifications.

Northern Blotting and Primer Extension Analysis

Northern blotting was performed as previously described (Kaplan, Jin et al. 2012), essentially following the instructions of GeneScreen hybridization membranes (PerkinElmer) with minor modifications.

2AP Fluorescence Spectroscopy

Steady-state fluorescence measurements for Pol II-TCs were performed as previously described (Liu and Martin 2001, Kashkina, Anikin et al. 2007) with minor modifications using a Fluoromax-3 (HORIBA Scientific). Excitation wavelengths included both 280 and 315 nm, and fluorescence emission was collected from 340 to 400 nm. In addition to Pol II-TCs, spectra were collected for buffer, DNA, and Pol II alone to correct for background fluorescence. All measurements were performed at 25°C in triplicate (n = 3).

Crystallization and Refinement

Pol II transcribing complexes were concentrated to 8–10 mg/ml, and crystallization trials produced crystals in several conditions and were verified by SDS-PAGE analysis of crystals (Figure 3-1F). The structures were solved by molecular replacement using 12-subunit Pol II (PDB: 3FKI) (Meyer, Ye et al. 2009) for Pol II scaffolds in the presence or absence of Tfg2 and 10-subunit Pol II (PDB: 1SFO) (Westover, Bushnell et al. 2004) for Δ4/7-TC in Phaser (McCoy, Grosse-Kunstleve et al. 2007). Molecular replacement models were refined using the program Buster (Blanc, Roversi et al. 2004), Refmac (Murshudov, Vagin et al. 1997), and CNS (Brunger, Adams et al. 1998) followed by several cycles of manual building with B factor sharpening in *Coot* (Emsley, Lohkamp et al. 2010). Inclusion of weak-intensity, high-resolution data improved refinement behavior and stereochemistry (Karplus and Diederichs 2012). All figures were rendered using PyMOL (Version 1.5.0.4 Schrödinger, LLC).

Author Contributions

C.O.B. and M.C. contributed to this work equally. C.O.B. and G.C. were responsible for crystallization, data collection, structure refinement, and analysis. M.C. was responsible for mutagenesis studies, structure refinement, and analysis. H.S. and G.C. performed preliminary protein purification and crystallization. A.C., G.L., I.S.B., Q.Z., and F.P. performed crystallization trials and X-ray data collection. I.M. and C.D.K. performed genetic experiments and analysis of Rpb2 wedge mutants in yeast. C.D.K. contributed to analysis of the structures. B.W.G. and M.A.T. performed fluorescence

experiments and analyzed results. C.O.B., M.C., C.D.K., and G.C. wrote the manuscript. All authors commented and approved the manuscript.

Accession Numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes PDB: 5C4X, 5C4J for Pol II-TC and Δ4/7-TC, respectively. Additional codes are also deposited for supporting structures (PDB: 5C3E, 5C44, 5C4A).