UNIVERSITY OF CALIFORNIA

Los Angeles

Molecular Basis of Coupling 3'-end Processing to Transcription in Mammals

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biochemistry and Molecular Biology

by

Benson Ngo

2012

ABSTRACT OF THE DISSERTATION

Molecular Basis of Coupling 3'-end Processing to Transcription

in Mammals

by

Benson Ngo

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2012

Professor Harold G. Martinson, Chair

A major focus of my studies is to understand how mRNA transcription is functionally interconnected, or coupled, to 3'-end processing (cleavage and polyadenylation) of mRNAs . A product of this coupling effect is that transcription can affect and enhance 3'-end processing. To understand how transcription is able to affect 3'-end processing, I investigated a few mechanisms that have been suggested in the literature to be the primary modes of coupling 3'-end processing to transcription: recruitment of processing factors to the promoter, cotranscriptional recruitment of processing factors during gradual extrusion of the poly(A) signal from the polymerase, and phosphorylation of C-terminal domain (CTD) of RNA polymerase II (Chapter 3). Using an in vitro system that can support 3'-end processing that is

functionally coupled to transcription, I tested the extent to which each of these mechanisms contribute to coupling in our in vitro system. I demonstrate that recruitment of processing factors to the promoter or to the transcription elongation complex prior to transcription of the poly(A) signal is not necessary for coupled 3'-end processing in our in vitro system. I found that assembly of cleavage and polyadenylation factors during gradual extrusion of the poly(A) signal from the polymerase is not important for coupling. I also show that the phosphates put on the C-terminal domain (CTD) of the transcribing polymerase at or shortly after the time of transcription initiation are unimportant for coupling in our system. Surprisingly, nothing that happens at the promoter is necessary to direct coupling since transcription elongation complexes (TEC) initiated on oligo(dT) tailed templates can still engage in 3'-end processing that is an order of magnitude more efficient than processing uncoupled from transcription. Therefore, these results argue that the ability to couple 3'-end processing to transcription is intrinsic to the structure of the TEC. Consistent with this, I found that coupling is inhibited if the RNA tether which connects the RNA transcript to the polymerase is severed. Based on these results, we propose that coupling is somehow mediated by having the poly(A) signal held close to the CTD via an intact RNA tether in the context of the TEC.

Another aspect of this coupling phenomenon was demonstrated by Blencowe's group in which they reported that transcriptional activators can contribute to 3'-end cleavage in vivo. It is intriguing that transcriptional activators that are recruited to the promoter can affect a process that occurs at the other end of the gene. To understand this further, I investigated the effect of transcriptional activators on 3'-end processing in vitro (Chapter 5). I found that while transcriptional activators can dramatically enhance transcription, they did not alter 3'-end

processing levels significantly in a coupled in vitro system. However, I did observe that transcription initiates randomly in the absence of a transcriptional activator to direct promoter-specific transcription. At face value, these results suggest that the mechanism by which transcriptional activators affect 3'-end processing cannot be supported in our in vitro system. However, my results do suggest a possibility that transcriptional activators may affect cleavage levels by redirecting initiation from cryptic promoters to promoter-specific initiation.

It has been suggested that coupling of 3'-end processing to transcription may be in part mediated by speeding up the assembly of the cleavage and polyadenylation apparatus (CPA) at the poly(A) signal. One way that this is thought to be achieved is by early recruitment of processing factors to the promoter during initiation and its subsequent transfer to the elongating polymerase so that assembly can begin as soon as the poly(A) signal is transcribed. While this may be the case, very little is known regarding the actual assembly process of the CPA and the rate-limiting step that may be subject to regulation. Therefore, I sought to dissect the assembly pathway of the CPA in hopes of revealing how 3'-end processing may be regulated by transcription. I decided to begin by resolving whether ATP is a required cofactor for 3'-end processing since there have been conflicting reports in the literature. Interestingly, I found that ATP is indeed a required cofactor for 3'-end cleavage in our in vitro system since withholding ATP can inhibit cleavage (Chapter 4). I found that this inhibition is reversible since cleavage resumes upon re-addition of ATP. I also demonstrate that a partial CPA can form in the absence of ATP and this partial apparatus consists of at least CPSF and CstF, the core cleavage and polyadenylation complexes. Based on preliminary data, the rate-limiting step of the cleavage pathway occurs during or after the ATP-requiring step. These results suggest that ATP may play

an important role in regulation of 3'-end cleavage. The significance of an ATP requirement for 3'-end cleavage in our in vitro system will be discussed.

An interesting product of the interconnection between 3'-end processing to transcription is the possibility that there may be a coordinated surveillance mechanism that can assess the quality of the transcript being made and decides the fate of the transcript by directing it to process or to degrade. Here, we studied poly(A)-dependent pausing in vitro, which has been proposed as a surveillance checkpoint, and poly(A)-dependent degradation of unprocessed transcripts from weak poly(A) signals (Chapter 2). We confirm directly, by measuring the length of RNA within isolated transcription complexes, that a newly transcribed poly(A) signal reduces the rate of elongation by RNA polymerase II, resulting in the accumulation of these complexes downstream of the poly(A) signal. We then show that if the RNA in these elongation complexes contain a functional but unprocessed poly(A) signal, degradation of the transcripts ensues. We propose that during normal 3'-end processing, a decision is made whether to process or to degrade. In the case of weak poly(A) signals, where cleavage at the poly(A) site is slow, the default pathway to degradation predominates.

The dissertation of Benson Ngo is approved.

Arnold J. Berk

Guillaume Chanfreau

Joseph A. Loo

Michael F. Carey

Harold G. Martinson, Committee Chair

University of California, Los Angeles

2012

Table of Contents

Section		pages
List of Figures		ix
Acknowledgements		x
Vita		xii
Chapter 1	Introduction	1
	References	12
Chapter 2	Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro	16
	Abstract	17
	Introduction	17
	Results	18
	Discussion	26
	Materials and Methods	28
	References	29
Chapter 3	Transcription elongation complex directs coupling of 3'-end processing to transcription without assistance of events from the promoter	31
	Abstract	32
	Introduction	33
	Results	38
	Discussion	50
	Materials and Methods	56
	Figures	64
	References	82
Chapter 4	Preliminary studies on ATP requirement for 3'-end cleavage	88
	Abstract	89
	Introduction	90
	Results	93
	Discussion	101
	Materials and Methods	106
	Figures	111
	References	119
Chapter 5	Preliminary studies: Transcriptional activators may contribute to 3'-end processing by regulating promoter-specific and non-promoter-specific transcription	121
	Abstract	122
	Introduction	123
	Results	126
	Discussion	133
	Materials and Methods	138
	Figures	141

References 149

List of Figures

Figures (Chapter, Figure #) pages
2.1 19
2.2 20
2.3 22
2.4 23
2.5 24
2.6 25
3.1 64
3.2 66
3.3 69
3.4 72
3.S1....74
3.S2....76
3.S3....78
3.S4....80
4.1 111
4.2 114
4.3 117
5.1 141
5.2 143
5.3 145
5.4 147

Acknowledgements

First and foremost, I would like to thank my advisor Dr. Harold G. Martinson for helping me develop into the biochemist researcher I am today. His tremendous patience and excellent mentorship allowed me to achieve all of the work that I present in this dissertation.

I am extremely grateful to my committee members for their research advice, constructive criticisms, and rigor during my research presentations and in the evaluation of my dissertation thesis.

My work would not have been possible without the generous gifts from various labs. I would also like to thank Carol Eng from Arnold J. Berk's lab for her assistance in growing up HeLa cell cultures for nuclear extract preparations. I would like to thank Dr. Benjamin Blencowe for the activator-inducible dsx DNA plasmid template and pBS exon2 GAA6 plasmid for RNase protection analysis, Dr. James Goodrich for Gal4-p53 expression plasmid, Dr. Jay Gralla for Gal4-AH expression plasmid, Dr. Clint MacDonald for anti-CstF antibodies, and Dr. Christine Milcarek for anti-CPSF antibody.

I thank the RNA society for allowing me to use the reprint of the paper entitled "Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro" as Chapter 2 in my dissertation (RNA 2010, 16: 197-210 or doi:10.1261/rna.1622010). I would like to thank both Dr. Amir Kazerouninia, who contributed the majority of the experimental work in this paper (Figure 1-4, 6), and Dr. Harold G. Martinson for their research advice and direction that led me to make the important finding that transcripts bearing a weak poly(A) signal were being targeted by the poly(A) dependent discard pathway because they are too slow to engage in 3'-end cleavage (Figure 5).

There were many obstacles that I had to overcome during these six years of graduate research but I was able to get through them thanks to the support of my fellow colleagues, friends, and family. I want to thank Frank Rigo for giving me an inspiring talk on the importance of research in teaching. I would like to thank my colleague Huimin Zhang for our stimulating scientific discussions. I want to thank my friends Laura Ackerman, Theonalyn Aquino and Yahumara Cuellar for reminding me of the worth of becoming a professor and teacher. Finally, I thank my girlfriend, Nancy Wen, my mom, my brother, and friends for their support throughout my graduate career.

VITA

Education:

2002-2006: Bachelor of Arts in Biology-Chemistry with Honors
Claremont McKenna College, California (CA)
Graduated Cum Laude

Research Experience:

2006-present: In vitro studies on the mechanism of coupling RNA 3'-end processing to transcription
Advisor: Harold G. Martinson, Ph.D.
University of California, Los Angeles

2005-2006: Studies on the cellular localization of Kic1p in *S. pombe*
Advisor: Zhaohua Irene Tang, Ph.D.
Claremont McKenna College

Publications:

Tang, Z., Luca, M., Portillio, J., **Ngo, B**., Chang, C., Wen, T., Murray, J., Carr, A. "LAMMER Kinase Kic1 is involved in pre-mRNA processing. Experimental Cell Research 317: 2308-2320 (2011).

Kazerouninia, A., **Ngo, B.**, Martinson, H.G. "Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro." RNA. 16:197-210 (2010).

Tang, Z., Tsurumi, A., Alaei, S., Wilson, C., Chiu, C., Oya, J., **Ngo, B**. "Dsk1p kinase phosphorylates SR proteins and regulates their cellular localization in fission yeast." Biochem. J. 405: 21-30 (2007).

Poster Presentations at Scientific Meetings:

45th Annual Meeting of the American Society for Cell Biology (San Francisco, CA, USA. December 2005). Tang, Z., Alaei, S., Wilson C., Tsurumi, A., Chiu, C., Oya, J., Shin, J., Buitron, H., Hardy, H. "The cellular localizations of SR-related proteins in *Schizosaccharomyces pombe* fission yeast are affected by SR-protein specific kinase Dsk1."

Teaching Experience:

Served as a Teaching Assistant at University of California, Los Angeles

Fall 2009: **Chem 153A –** Introduction to Biochemistry: protein structure, enzymes and metabolism.

Summer B 2009: **Chem14CL –** Organic chemistry laboratory.

Summer A 2009: **Chem 153A –** Introduction to Biochemistry: protein structure, enzymes and metabolism.

Spring 2009: **Chem 14B** – General Chemistry: Introduction to thermodynamics, kinetics, electrochemistry, and organic chemistry.

Winter 2009: **Chem 14A** – General Chemistry: Introduction to atomic and molecular structure, equilibria, acids and bases.

Fall 2008: **Chem 153A –** Introduction to Biochemistry: protein structure, enzymes and metabolism.

Summer A 2008: **Chem 153A –** Introduction to Biochemistry: protein structure, enzymes and metabolism.

Spring 2008: **Chem 153B** – Biochemistry: DNA, RNA, and protein synthesis.

Spring 2007: **Chem 14B** – General Chemistry: Introduction to thermodynamics, kinetics, electrochemistry, and organic chemistry.

Winter 2007: **Chem 14A** – General Chemistry: Introduction to atomic and molecular structure, equilibria, acids and bases.

Awards, Fellowships and Scholarships:

Dissertation Year Fellowship Award (2011)

Radcliffe Dee Award (2010)

W.M. Keck Funding for Summer Research (2005)

Chapter 1

Introduction

Pre-mRNA 3'-end processing is an essential step in the maturation of eukaryotic mRNA. Defects in mRNA 3'-end processing have been associated with a variety of human diseases (Danckwardt et al. 2008). Therefore, it is important to understand how the pre-mRNA 3'-end processing machinery functions and how it is regulated.

Core elements of the poly(A) signal and the cleavage and polyadenylation apparatus

The core cis-elements and the majority of cleavage and polyadenylation factors required for 3'-end processing have been identified (Colgan et al. 1997; Zhao et al. 1999; Millevoi et al. 2010; Mandel et al. 2008). The core poly(A) signal consists of an AAUAAA hexamer and a variable GU-rich downstream element located 40-60 nucleotides downstream of the hexamer. These two elements determine the general location of the poly(A) cleavage site, which is located in between the poly(A) hexamer and the GU-rich element, approximately ~10-30 nucleotides downstream of the poly(A) hexamer. The AAUAAA hexamer is recognized by cleavage and polyadenylation specificity factor (CPSF) and the GU-rich downstream element is recognized by cleavage stimulatory factor (CstF) complex. In addition to these two complexes, cleavage factor Im (CFIm) also binds to a UGUAN motif. Besides these three complexes, recruitment of additional cleavage and polyadenylation factors including poly(A) polymerase, cleavage factor IIm (CFIIm), and C-terminal domain (CTD) of RNA polymerase II are required for the full assembly of the cleavage and polyadenylation apparatus (CPA). Following assembly, endonucleolytic cleavage of the RNA is believed to be carried out by the 73 kDa subunit of CPSF (Ryan et al. 2004; Mandel et al. 2006). Recently, purification of the functional human pre-mRNA 3'end processing complex has suggested that the cleavage and polyadenylation apparatus may consists up to 80 proteins (Shi et al. 2009). This is quite impressive since poly(A)

cleavage is a simple chemical reaction involving hydrolysis of a phosphodiester bond that could be carried out by a single protein. Also taking into account the polyadenylation reaction, the entire process of cleavage and polyadenylation could be carried out by a few proteins in theory. However, the fact that 3'-end processing requires the assembly of such a large apparatus suggest that 3'-end processing may serve as the center of an important surveillance activity in which the quality of the RNA transcript is "checked" before the decision to either cleave or discard the transcript is made.

Assembly of the cleavage and polyadenylation apparatus

While the core cis-elements and majority of the trans-acting factors required for 3'-end processing have been identified, less is known regarding the assembly of the cleavage and polyadenylation apparatus. However, there is evidence that assembly of the cleavage and polyadenylation apparatus on the poly(A) signal is a stepwise process (Chao et al. 1999). It has also been observed in vitro (Rigo et al. 2005) that when processing is coupled to transcription, a substantial lag precedes a dramatic burst of rapid cleavage, which is consistent with the idea that assembly of the apparatus may be a multi-step process.

More recently, an assembly pathway was proposed based on immunoprecipitation experiments that had identified discrete complexes involving CPSF, CstF and C-terminal domain (CTD)-less RNA polymerase II and CTD-intact RNA polymerase II (Nag et al. 2007). The pathway begins with CPSF riding on the polymerase body before the poly(A) signal is transcribed. CPSF scans the RNA from the body of the polymerase and upon transcription of the poly(A) hexamer, CPSF binds the AAUAAA sequence and directs the polymerase to pause (see "surveillance mechanisms" below). Once the GU-rich region is transcribed, CstF joins CPSF at the poly(A)

signal, and the two complexes are transferred to the CTD. Additional cleavage and polyadenylation factors join sometime later and cleavage and polyadenylation ensues.

In order to further investigate the assembly pathway of the CPA, we explored different methods to block 3'-end cleavage at discrete steps along the pathway and interestingly enough, we found that ATP is a required cofactor for 3'-end cleavage in our in vitro system (Chapter 4). By withholding ATP, we show that a discrete step in the assembly pathway is blocked and this inhibition can be reversed upon addition of ATP. We show that a partial cleavage and polyadenylation apparatus consisting of at least CPSF and CstF can be recruited in the absence of an ATP cofactor. We have previously reported that there is a conspicuous lag between transcription of the poly(A) signal and 3'-end cleavage in our in vitro system (Rigo et al. 2005). Interestingly, we find that the rate-limiting step of the cleavage pathway occurs at a step at or after the ATP-requiring step. We propose that ATP may be required for the stable recruitment of poly(A) polymerase to the CPA or it could reflect an ATP-induced structural change or remodeling of the CPA.

Surveillance mechanisms

One of the first discernable consequences following transcription through a poly(A) signal is the slowing down of the transcribing elongation complex. This reduction in rate of elongation, or pausing, has been found to be dependent on the interaction of CPSF with the poly(A) hexamer (Orozco et al. 2002; Park et al. 2004; Nag et al. 2006; 2007) and does not require CstF or the GU-rich region (Nag et al. 2007). This poly(A) signal-dependent pausing can be seen in the 3' flanks of genes across the genome (Gromak et al. 2006; Boireau et al. 2007; Glover-cutter et al. 2008; Lian et al. 2008) and has been proposed to coincide with a checkpoint activity that leads

to processing, degradation, or continued transcription elongation (Orozco et al. 2002; Rigo et al. 2005; Nag et al. 2007). This surveillance activity would be consistent with the recent finding that the cleavage and polyadenylation apparatus is surprisingly large (Shi et al. 2009), in the sense that additional factors or complexes beyond the core cleavage and polyadenylation factors would be required for "checking" the quality of the RNA transcript and to degrade the transcript if necessary.

One possible function of this surveillance activity may be to degrade inefficiently processed RNAs. It has been observed that transcripts with weak poly(A) signals are degraded when they fail to get cleaved and polyadenylated in vivo. For example, the relatively weak SV40 early poly(A) signal produces one-fifth as much mRNA as does the stronger SV40 late poly(A) signal, even when both are transcribed at the same rate (Carswell et al. 1989). However, although most transcripts carrying SV40 early poly(A) sites do not get processed at those sites, little or no RNA carrying uncleaved poly(A) sites can be found in steady-state nuclear RNA from such cells (Connelly et al. 1988; Chao et al. 1999; Park et al. 2004; Nag et al. 2006). These results suggest that these RNAs are extremely unstable. However, these RNAs are not unstable simply because it is unprocessed because inactivation of the weak poly(A) signal by mutation does lead to accumulation in the nuclear compartment (Connelly et al. 1988; Nag et al. 2006). Therefore, these results suggest the existence of a mechanism that degrades weak poly(A) signal transcripts in a poly(A)-dependent manner, which likely involves the cleavage and polyadenylation apparatus.

Consistent with this, we studied both poly(A) pausing and degradation of inefficiently processed RNAs from weak poly(A) signals using a coupled in vitro system (Chapter 2). We show directly,

by measuring length of RNA within isolated elongation complexes, that the elongation rate of transcription elongation complexes is reduced upon transcribing the poly(A) signal and causes the accumulation of elongation complexes downstream of the poly(A) signal. We show that if the RNA in these elongation complexes contains a functional but unprocessed poly(A) signal, degradation of the transcripts ensues. The degradation depends on the unprocessed poly(A) signal being functional, and does not occur if a mutant poly(A) signal is used. We propose that during normal 3'-end processing, a decision is made whether to process or to degrade. For weak poly(A) signals in which cleavage at the poly(A) site is slow, the default pathway to degradation predominates.

Mechanisms of how transcription affects 3'-end processing

Within the last two decades, it has become clear that RNA processing (capping, splicing, and cleavage and polyadenylation) are functionally interconnected, or coupled, to each other and to transcription. However, despite this, the mechanisms of how transcription affects 3'-end processing is still relatively unclear. Several mechanisms of how transcription can affect 3'-end processing have been proposed (Bentley 2005): 1) recruitment of processing factors to the elongation complex to increase local concentration of processing factors at the site of processing, 2) kinetic coupling, and 3) allosteric activation by the C-terminal domain (CTD) of RNA polymerase II.

Pre-loading of processing factors on the transcription elongation complex

The idea that processing factors are recruited early to the elongating polymerase to act at the site of processing originated from a report in which CPSF was found to co-purify with transcription initiation factor TFIID in an immunoprecipitation against TBP (TATA-binding

protein) (Dantonel et al. 1997), suggesting that CPSF may be recruited to the promoter during transcription initiation. During transcription, CPSF was found to associate with the transcribing polymerase and not TFIID in vitro, which is consistent with CPSF being transferred to the elongating polymerase from TFIID (Dantonel et al. 1997). Consistent with the idea of preloading processing factors to the promoter, chromatin immunoprecipitation (chIP) studies have localized CPSF and in some cases, CstF, to the promoter and/or body of protein-coding genes (Glover-cutter et al. 2008).

In addition to recruitment by TFIID, transcriptional activators have been demonstrated to enhance 3'-end cleavage in vivo (Rosonina et al. 2003). It was proposed that enhancement of 3'-end cleavage by transcriptional activators is mediated by recruitment of a processing factor called PSF (polypyrimidine tract binding protein-associated splicing factor) at the promoter (Rosonina et al. 2005). Consistent with this, they showed that various transcriptional activators enhanced 3'-end cleavage with different strengths and this strength correlated with their ability to bind to PSF in vitro. Recently, Manley and colleagues have reported that transcriptional activators can also enhance polyadenylation in vitro (Naigaike et al. 2011). They found that transcriptional activators associated with members of the Paf1c complex in in vitro binding assays. Paf1c complex is an elongation complex that was recently shown to associate with processing factors CPSF and CstF (Rozenblatt-Rosen et al. 2009) and it is believed that Paf1c helps mediate the pre-recruitment of CPSF and CstF to the transcription elongation complex for transfer to the poly(A) signal. Despite this growing body of evidence, it has not been directly demonstrated that pre-recruitment of processing factors and stimulation of 3'-end processing is a cause-and-effect relationship. Using an in vitro system that can support 3'-end processing

that is functionally coupled to transcription (Rigo et al. 2005), we found that transcriptional activators had very little effect on 3'-end cleavage in vitro (Chapter 5). Our studies suggest that the ability of transcriptional activators to enhance 3'-end cleavage may be limited in an in vitro system. However, our results also suggest the possibility of transcriptional activators in controlling 3'-end cleavage levels by altering the levels of nonpromoter-specific versus promoter-specific transcription.

Kinetic coupling

Transcription elongation rates can influence processing by affecting assembly of RNA-protein complexes (de la Mata et al. 2003; Howe et al. 2003; Pinto et al. 2011). In a situation where two processing sites can be utilized, slow transcription favors the use of proximal processing sites over distal ones if the appearance of the distal site is significantly delayed. This is due to the presumption that slow elongation enables more time for the processing apparatus to assemble and act on the proximal site before transcription of the second signal. Indeed, elongation rates have been seen to affect both alternative splicing (de la Mata et al. 2003; Howe et al. 2003) and alternative polyadenylation (Pinto et al. 2011). Although this concept of kinetic coupling is applied in situations where more than one processing signal is present, it is interesting to speculate whether transcription rates can also affect processing at individual sites by influencing order of assembly of processing factors.

Role of CTD in coupling transcription to 3'-end processing

The CTD of RNA polymerase II has been proposed to be important for coupling 3'-end processing to transcription in two ways. First, the CTD can act as a landing pad to recruit processing factors and to deliver them to the emerging transcript during transcription

(McCracken et al. 1997; Proudfoot 2004; Bentley 2005). Second, the CTD can act to allosterically activate the cleavage and polyadenylation apparatus (CPA) (Hirose et al. 1998). An intriquing property of the CTD is that it is made up of many repeats (up to 52 in mammals) of the heptameric sequence, YSPTSPS (Tyr-Ser-Pro-Thr-Ser-Pro-Ser), which is modified extensively in vivo through phosphorylation, glycosylation, and prolyl isomerization (Egloff et al. 2008). Moreover, CTD phosphorylation is a dynamic process that is regulated by a few CTD kinases during transcription. Of the various residues that is known to be phosphorylated, Ser2 and Ser5 phosphorylation is by far the most-well studied and therefore best illustrates how CTD modification can affect various stages of transcription (Buratowski 2003).

Phosphorylation of the CTD is an orchestrated event beginning with Ser5 phosphorylation by the Cdk7 subunit of TFIIH at the promoter. This phosphorylation leads to promoter escape and transition into early elongation mode. Ser5 phosphorylation also facilitates recruitment of capping factors for 5' capping (Komarnitsky et al. 2000; Schroeder et al. 2000). Soon after, the elongation complex encounters proximal pausing. This is overcome by phosphorylation of the CTD at Ser2 and associated elongation factors by Cdk9, subunit of P-TEFb, which transforms them into processive elongation complexes (Peterlin et al. 2006). The recruitment of P-TEFb is thought to occur generally through a bromodomain protein Brd4 (Jang et al. 2005; Yang et al. 2005), although DNA-sequence specific transcriptional activators that can bind to P-TEFb has also been suggested to recruit P-TEFb to specific promoters (Zhou et al. 2006). Besides Ser2 and Ser5 phosphorylation, not much is known about the role of Tyr1, Thr4, and Ser7 phosphorylation in transcription. Tyr1 can be phosphorylated by c-Abl in mammals but no homolog is present in yeast (Baskaran et al. 1999). More recently, Ser7 phosphorylation was

found to occur at 3' ends of protein-coding genes, although its current role in transcription is still unclear (Chapman et al. 2007). In addition, Thr4 of the CTD has also been reported to be phosphorylated and is required for 3'-end processing of histone pre-mRNAs (Hsin et al. 2011). Based on the growing body of evidence of a CTD code and the recently discovered but unknown functions of Thr4 and Ser7 phosphorylation, it is interesting to speculate whether the orchestration of phosphorylation during transcription is important for coupling transcription to 3'-end processing.

There is evidence to suggest that co-transcriptional CTD phosphorylation is essential for coupling of 3'-end processing to transcription. Bentley and colleagues utilized frog oocytes to study transcription-coupled and transcription-uncoupled processing by either injecting a reporter DNA template (transcription-coupled) or pre-made RNAs (transcription-uncoupled) into frog oocytes (Bird et al. 2004). They looked at the effect of blocking CTD phosphorylation with kinase inhibitors on transcription-coupled and uncoupled processing and found surprisingly that CTD-kinase inhibitors negatively affect 3'-end processing that is coupled to transcription but did not have any effect on 3'-end processing uncoupled from transcription (Bird et al. 2004).

Role of promoter in coupling

It has been thought that gene expression is regulated primarily by elements at the promoter and the availability of trans-acting factors. For example, Kornblihtt and colleagues have found that splicing of the fibronectin EDI exon is dependent on elements found at the promoter (Cramer et al. 1997). They found promoter swapping affected fibronectin EDI exon inclusion (Cramer et al. 1997) and they later proposed that the promoter was influencing exon inclusion

by assisting in the recruitment of SR protein SF2/ASF to the exon splicing enhancer (ESE) element (Cramer et al. 1999). These studies have generally led to the idea that the fate of the transcript is controlled by the promoter.

Although multiple mechanisms have been proposed to link 3'-end processing of the RNA to transcription, it is not known to what extent do each of these mechanisms contribute to this interconnection or coupling between transcription and 3'-end processing. Therefore, we investigated a few of these mechanisms using an in vitro system that can support the coupling between transcription and 3'-end processing (Chapter 3). We demonstrate that recruitment of processing factors to the transcription elongation complex is not necessary for coupling. We also show that co-transcriptional recruitment of cleavage and polyadenylation factors to the poly(A) signal is not required. We even show that the orchestration of CTD phosphorylation as directed by the promoter is not important since removal of CTD phosphates put on during initiation and early elongation does not inhibit 3'-end processing. Even when we had bypassed the events at the promoter altogether, by loading transcription elongation complexes on 3'-extended dT tails, the TECs still processed an order of magnitude more efficiently than processing uncoupled from transcription. Surprisingly, these results indicate that the transcription elongation complex can direct transcription-coupled 3'-end processing without assistance from the promoter. We propose that coupling is somehow mediated by having the poly(A) signal held close to the CTD in the context of the TEC.

References

Baskaran, R., Escobar, S.R., and Wang, J.Y.J. (1999). Nuclear c-Abl is a COOH-terminal repeated domain (CTD)-tyrosine kinase-specific for the mammalian RNA polymerase II: possible role in transcription elongation. Cell Growth Differ. *10*, 387-396.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr. Opin. Cell Biol. *17*, 251-256.

Bird, G., Zorio, D.A., and Bentley, D. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Mol. Cell Bio. *24*, 8963-8969.

Boireau, S., Maiuri, P., Basyuk, E., de la Mata, M., Knezevich, A., Pradet-Balade, B., Backer, V., Kornblihtt, A., Marcello, A. and Bertrand, E. (2007). The transcriptional cycle of HIV-1 in real-time and live cells. J. Cell Biol. *179*, 291-304.

Buratowski S. (2003). The CTD code. Nat. Struct. Biol. *10*, 679-680.

Carswell, S. and Alwine, J.C. (1989). The efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. Mol. Cell Biol. *9*, 4248-4258.

Chao, L.C., Jamil, A., Kim, S.J., Huang, L., and Martinson, H.G. (1999). Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. Mol. Cell. Biol. *19*, 5588-5600.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science. *318*, 1780-1782.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev. *11*, 2755-2766.

Connely, S. and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev *2*, 440-452.

Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. Proc. Natl. Acad. Sci. USA *94*, 11456-11460.

Cramer, P., Caceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E., and Kornblihtt, A.R. (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. Mol. Cell *4*, 251-258.

Danckwardt, S., Hentze, M.W., and Kulozik, A.E. (2008). 3'end mRNA processing: molecular mechanisms and implications for health and disease. EMBO J. *27*, 482-498.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3'-end of mRNA. Nature *389*, 399-402.

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

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. Trends in genetics *24*, 280-288.

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

Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol. Cell Biol. *26*, 3986-3996.

Hirose, Y., and Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

Howe K.J., Kane C.M., and Ares J. Jr. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA *9*, 993-1006.

Hsin, J., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3'-end processing. Science *334*, 80-87.

Jang, M. K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol. Cell *19*, 523-534.

Komarnitsky, P., E. J. Cho, and S. Buratowski. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. *14*, 2452-2460.

Lian, Z., Karpikov, A., Lian, J., Mahajan, M.C., Hartman, S., Gerstein, M., Snyder, M. and Weissman, S.M. (2008). A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3'end RNA polyadenylation. Genome Res. *18*, 1224-1237.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end processing endonuclease. Nature *444*, 953-956.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. Cell Mol. Life Sci. *65*, 1099-1122.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res. *38*, 2757-2774.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

Nag, A., Narsinh, K., Kazerouninia, A., and Martinson, H.G. (2006). The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. RNA *12*, 1534-1544.

Nag, A., Narsinh, K., and Martinson, H.G. (2007). The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat. Struct. Mol. Biol. *14*, 622-669.

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

Orozco, I.J., Kim, S.J., and Martinson, H.G. (2002). The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. J Biol. Chem. *277*, 42899-42911.

Park, N.J., Tsao, D.C. and Martinson, H.G. (2004). The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II CTD. Mol. Cell Biol. *24*, 4092-4103.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297-305.

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

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3'-end formation to transcription. Curr. Opin. Cell Biol. *16*, 272-278.

Rigo, F., Kazerouninia, A., Nag, A., and Martinson, H.G. (2005). The RNA tether from the poly(A) signal to the polymerase mediates coupling of transcription to cleavage and polyadenylation. Mol. Cell *20*, 733-745.

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

Rosonina, E., Ip, J. Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J., and Blencowe, B.J. (2005). Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. Mol. Cell. Biol. *25*, 6734-6746.

Rozenblatt-Rosen, O., Nagaike, T., Francis, J.M., Kaneko, S., Glatt, K.A., Hughes, C.M., LaFramboise, T., Manley, J.L., and Meyerson, M. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. Proc. Natl. Acad. Sci. USA *106*, 755-760.

Ryan, K., Calvo, O., and Manley, J.L. (2004). Evidence that polyadenylation factor CPSF-73 is the mmRNA 3' processing endonuclease. RNA *10*, 565-573.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev. *14*, 2435-2440.

Shi Y., Di Giammartino D.C., Taylor D., Sarkeshik A., Rice W.J., Yates J.R., Frank J., and Manley J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Mol. Cell *33*, 365-376.

Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol. Cell *19*, 535-545.

Zhao, J., Human, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol. Mol. Biol. Rev. *63*, 405-445.

Zhou, Q., and Yik, J.H. (2006). The yin and yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. Microbiol. Mol. Biol. Rev. *70*, 646-659.

Chapter 2

Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro

Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro

AMIR KAZEROUNINIA,[1] BENSON NGO,[1] and HAROLD G. MARTINSON[1,2]

[1]Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, California 90095-1569, USA
[2]Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90095-1569, USA

ABSTRACT

The poly(A) signal has long been known for its role in directing the cleavage and polyadenylation of eukaryotic mRNA. In recent years its additional coordinating role in multiple related aspects of gene expression has also become increasingly clear. Here we use HeLa nuclear extracts to study two of these activities, poly(A) signal-dependent transcriptional pausing, which was originally proposed as a surveillance checkpoint, and poly(A) signal-dependent degradation (PDD) of unprocessed transcripts from weak poly(A) signals. We confirm directly, by measuring the length of RNA within isolated transcription elongation complexes, that a newly transcribed poly(A) signal reduces the rate of elongation by RNA polymerase II and causes the accumulation of elongation complexes downstream from the poly(A) signal. We then show that if the RNA in these elongation complexes contains a functional but unprocessed poly(A) signal, degradation of the transcripts ensues. The degradation depends on the unprocessed poly(A) signal being functional, and does not occur if a mutant poly(A) signal is used. We suggest that during normal 3′-end processing the uncleaved poly(A) signal continuously samples competing reaction pathways for processing and for degradation, and that in the case of weak poly(A) signals, where poly(A) site cleavage is slow, the default pathway to degradation predominates.

Keywords: poly(A) signal; transcriptional coupling; transcriptional pausing; cleavage and polyadenylation; poly(A) signal-dependent RNA degradation; poly(A) signal-dependent surveillance

INTRODUCTION

The cleavage and polyadenylation reaction, which is responsible for defining the ends of most eukaryotic messenger RNAs, has emerged also as one of the central hubs for coordinating the transcription, splicing, release, transport, and surveillance of those same transcripts (Proudfoot et al. 2002; Nag et al. 2007; Perkins et al. 2008; Rigo and Martinson 2008; Rougemaille et al. 2008; Schmid and Jensen 2008). In keeping with this complexity of function, the apparatus designed to carry out cleavage and polyadenylation is also very complex—more than an order of magnitude larger and more complex, for example, than the capping apparatus (Pillutla et al. 1998; Shi et al. 2009), although cleavage/polyadenylation per se is chemically no more complicated than capping.

Reprint requests to: Harold G. Martinson, Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA 90095-1569, USA; e-mail: hgm@chem.ucla.edu; fax: (310) 206-4038.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1622010.

Assembly of the cleavage/polyadenylation apparatus appears to involve a delicately orchestrated sequence of events beginning, possibly, with the recruitment of some factors to the promoter (Dantonel et al. 1997; Calvo and Manley 2003; Glover-Cutter et al. 2008). The canonical mammalian poly(A) signal is characterized by two elements that flank the poly(A) cleavage site, the AAUAAA hexamer and a downstream G/U-rich region (Zhao et al. 1999). Almost immediately upon transcription of the poly(A) signal, the AAUAAA hexamer is bound by CPSF (Nag et al. 2007). Recruitment of CstF, which binds the G/U-rich region, appears to occur subsequently, and may be accompanied by substantial rearrangements among the components of the assembling apparatus (Nag et al. 2007). During this period, functional connections are established between the cleavage/polyadenylation apparatus and the upstream splicing apparatus (Rigo and Martinson 2008). Ultimately, the cleavage/polyadenylation apparatus becomes firmly attached to the polymerase as it commits to cleavage at the poly(A) site (Rigo et al. 2005; Rigo and Martinson 2008). It is possible that these culminating events coincide

with the establishment of gene looping by the cleavage/polyadenylation apparatus (Uhlmann et al. 2007; Perkins et al. 2008).

The first discernable consequence of transcription across the mammalian poly(A) signal is a reduction in the rate of elongation by the polymerase due to the interaction of CPSF with the poly(A) signal hexamer (Orozco et al. 2002; Park et al. 2004; Nag et al. 2006, 2007). This poly(A) signal-dependent pausing can be seen in the 3′ flanks of genes across the genome (Gromak et al. 2006; Boireau et al. 2007; Glover-Cutter et al. 2008; Lian et al. 2008) and has been proposed to coincide with checkpoint activity that leads either to cleavage at the poly(A) site or, alternatively, to continued transcription, or to degradation of the transcript (Orozco et al. 2002; Rigo et al. 2005; Nag et al. 2007). In yeast, surveillance during the polyadenylation phase of 3′-end processing is well documented for situations in which the polyadenylation has been compromised by defects in mRNP maturation factors or by the action of special regulatory elements (Schmid and Jensen 2008; Roth et al. 2009).

The efficiency of cleavage and polyadenylation is an important determinant of the overall level of gene expression in mammals, as in other eukaryotes. Thus, for similar rates of transcription, decreasing efficiencies of cleavage and polyadenylation give rise to decreasing final levels of gene expression (Edwalds-Gilbert et al. 1993; Chao et al. 1999; Wu and Alwine 2004). An unresolved question concerns the mechanism by which transcripts with weak poly(A) signals are discarded when they fail to get cleaved and polyadenylated. For example, the relatively weak SV40 early poly(A) signal supports mRNA production less than one-fifth as efficiently in mammalian cells as does the stronger SV40 late poly(A) signal, even when both are transcribed at the same rate (Carswell and Alwine 1989). Yet, although most transcripts carrying SV40 early poly(A) sites do not get processed at those sites, little or no RNA carrying uncleaved poly(A) sites can be found in the steady-state nuclear RNA from such cells (Connelly and Manley 1988; Chao et al. 1999; Park et al. 2004; Nag et al. 2006). Therefore, this unprocessed RNA is exceedingly unstable. Nevertheless, it is not merely the lack of processing at the poly(A) site that renders uncleaved RNA unstable, because full inactivation of the weak poly(A) signal, either mutationally or by using antisense, does lead to accumulation of uncleaved RNA in the nuclear compartment (Connelly and Manley 1988; Chao et al. 1999; Nag et al. 2006). Thus, paradoxically, the ability to discard unprocessed RNA appears to require the same weak poly(A) signal whose inefficient cleavage and polyadenylation was responsible for the lack of processing in the first place. Since weak poly(A) signals process more slowly than strong ones (Chao et al. 1999), this situation is suggestive of a surveillance function in which transcripts that are not processed quickly enter a poly(A) signal-dependent discard (PDD) pathway.

Here, we have sought to reproduce in vitro the poly(A) signal-dependent pause and the accompanying surveillance that may be responsible for the PDD of unprocessed transcripts bearing weak poly(A) signals. Although PDD is not evident in previous experiments involving weak poly(A) signals in vitro (Ryner et al. 1989; Prescott and Falck-Pedersen 1992), the processing in those experiments was not coupled to transcription. Therefore, since mRNA surveillance is thought to begin cotranscriptionally (Andrulis et al. 2002; Hieronymus et al. 2004), we decided to search for evidence of poly(A) signal-dependent transcript degradation using a transcription system that is capable of responding to the poly(A) signal in vitro (Tran et al. 2001). In this report we show that several aspects of the poly(A) signal-dependent pause are reproduced in vitro, and that coincident with or immediately following this transcriptional slowdown, PDD of nascent transcripts bearing uncleaved poly(A) signals becomes evident.

RESULTS

The poly(A) signal directs the polymerase to pause in vitro

To study poly(A) signal-dependent pausing in vitro (Fig. 1) we compared transcripts produced by polymerases that have crossed a functional poly(A) signal with transcripts produced by polymerases that have crossed an inactive, mutated poly(A) signal (Tran et al. 2001; Nag et al. 2006). Polymerases that have been slowed down by pausing will not transcribe as far during the time of the experiment, and can be identified by the shorter transcripts they carry. To ensure that any short transcripts detected in the analysis reflect slow polymerases, rather than terminated polymerases or poly(A) site-cleaved RNA, we analyzed only transcripts that were isolated from transcription elongation complexes (TECs) that had been separated from free RNA by size exclusion chromatography under strongly dissociating conditions (1 M NaCl, 1% Sarkosyl). In addition, to be sure that any short transcripts detected in the analysis reflect slow polymerases rather than partial degradation of the RNA, we used a pulse-chase format (Fig. 1A). This made sure that any shortening of the TEC RNA due to degradation would simultaneously remove [^{32}P] label, thereby reducing or eliminating the ability of the 5′ truncated RNA to contribute [^{32}P] signal to the analysis. Finally, to focus on the transcriptional effects of the intact unprocessed poly(A) signal, we used the SV40 early poly(A) signal core, which is so weak in a coupled transcription/processing system that hardly any processing occurs, though it nevertheless communicates well with the polymerase (Tran et al. 2001).

Figure 1B displays the results from such an experiment. In Figure 1B, lane 1, a region of intensity slightly above the position of the poly(A) site reveals an accumulation of

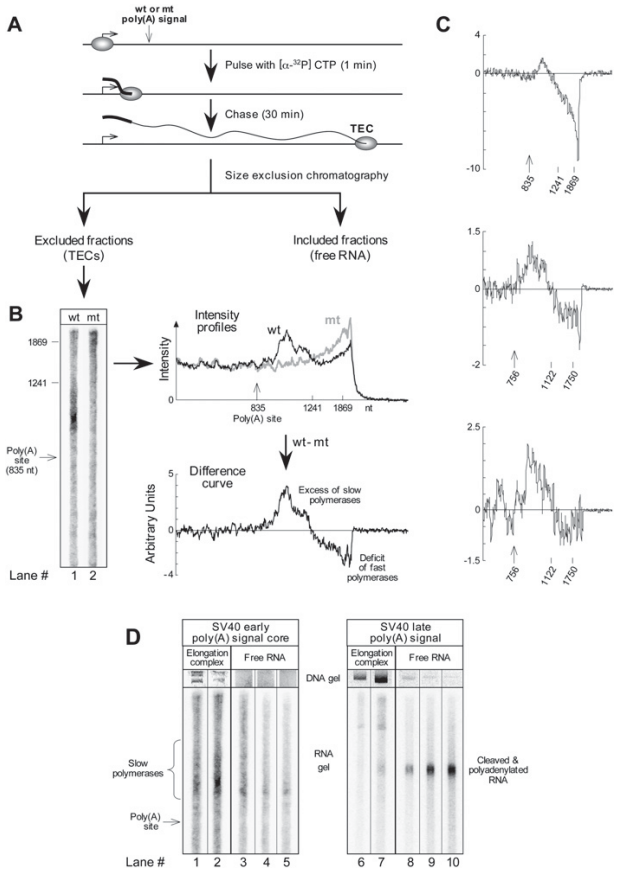


FIGURE 1. Survey of paused polymerases by transcript analysis. (*A*) Experimental protocol. DNA from Reporter 3 was transcribed in vitro using a pulse-chase format very similar to that of Rigo et al. (2005), and then TECs were isolated by size fractionation in the presence of 1 M NaCl and 1% Sarkosyl. A fourfold increase over the typical reaction size was used. The length of the pulse was chosen to be short enough to preclude any polymerases from reaching the poly(A) site before the start of the chase. (*B*) Comparison of TEC fractions for wild-type and mutant versions of the reporter. Size calibrations are based on RNA transcribed from the same or similar template in the presence of complementary DNA oligonucleotides that direct RNase H to cut the RNA at specified locations (Rigo et al. 2005). TECs (identified by the presence of plasmid DNA) eluted from the column in fractions 8 and 9 for the wild type and in fractions 7 and 8 for the mutant. Gel lanes *1* and *2* show column fractions 9 and 8, respectively. However, the intensity profiles to the *right* of the gel show the summed intensities for both TEC fractions eluting from the column. In preparing these profiles, the data were background corrected and normalized, as described below, to control for variations in the amount of template used in the different transcription reactions and for variations in sample recovery. First, the background in the gel above the region containing RNA was subtracted from the signal in all lanes, bringing the far *right* of the line graphs to zero. Then, the remaining signal in each lane was normalized at each position to the sum of the signal for TECs that had not reached the poly(A) signal. Thus, the average signal to the *left* of the poly(A) site for each graph is identical, by definition—as it should be, since the wild-type and mutant templates are identical upstream of the poly(A) signal. The difference curve was calculated by subtracting the mutant intensity profile from the wild-type one and expressing the result in arbitrary units related to the fraction of all RNA produced in the experiment. (*C*) Difference curves from three additional experiments. Reporters 3 and 2a were used for the *top* panel and for the *bottom* two panels, respectively. (*D*) Gel-electrophoretic analysis of RNA and DNA from size fractionated transcription reactions for poly(A) signals that either do not (lanes *1–5*) or do (lanes *6–10*) process efficiently in vitro. Fractions 8, 9, 11, 12, and 13 are shown for the SV40 early column, and fractions 7, 8, 10, 11, and 12 are shown for the SV40 late column. Transcription and processing of the SV40 late RNA was exactly as described in Rigo et al. (2005) with a 2-min pulse and a 13-min chase. Lane *2* in *D* is the same as lane *1* in *B*.

polymerases a short distance downstream from the wild-type poly(A) signal. In contrast, Figure 1B, lane 2, shows no such accumulation downstream from a mutant poly(A) signal. Instead, many polymerases on this template have proceeded much farther downstream, producing RNAs that migrate in the poorly resolved upper region of the gel (Fig. 1B, lane 2). This provides the first direct evidence for substantial slowing by polymerases once they cross a poly(A) signal. Although we have not explicitly ruled out the formal possibility that the poly(A) signal causes polymerases to arrest on the template, we will assume that this poly(A) signal-dependent pausing is the same poly(A) signal-dependent pausing that has been described in vivo. Note that the pulse-chase format of the experiment shows that this effect on transcription is imposed by the poly(A) signal on the polymerase in the absence of cleavage at the poly(A) site (which would remove all label from the elongation complexes) (see Fig. 1A).

In the upper graph to the right of the gel in Figure 1B the intensities for both lanes on the gel are plotted as a function of distance and superimposed on each other after background subtraction and normalization, as described in the figure legend. The result shows clearly that, after crossing a poly(A) signal, many polymerases linger behind that would otherwise proceed rapidly down the template. The graph at the bottom of Figure 1B summarizes both the wild-type and the mutant results in a single difference curve, obtained by subtracting the mutant from the wild-type curve in the upper graph. This difference curve illustrates clearly the generation of slow polymerases at the expense of rapid ones downstream from a functional poly(A) signal.

The results of Figure 1B are confirmed by difference curves from three additional independent experiments, as summarized in Figure 1C. Given the complexity of these experiments, it is not surprising that there is significant scatter in the relative sizes of the peaks and valleys of these curves. Nevertheless,

in all cases there is a poly(A) signal-dependent increase in the number of slow polymerases downstream from the poly(A) signal, followed by a poly(A) signal-dependent deficit in the number of faster polymerases farther downstream.

To confirm that the material identified as paused RNA in Figure 1B originated from elongation complexes and not from processed or partially processed RNA, we carried out an additional control. In Figure 1D the size exclusion data from the experiment that produced the SV40 early RNA of Figure 1B (lane 1) are compared directly with comparable data for RNA containing the strong SV40 late poly(A) signal, which processes efficiently in vitro (Rigo et al. 2005). The first two lanes of each panel in Figure 1D show the leading edge of the elongation complex fractions from the exclusion column (identified by the presence of plasmid DNA), and the remaining three lanes in each panel show the first few free RNA-containing fractions (see Nag et al. 2006 for further characterization of this column system). It can be seen that the SV40 early RNA (lanes 1–5) emerges predominantly in the elongation complex fractions with some trailing into the free RNA, while the processed SV40 late RNA (lanes 6–10) clearly fractionates like free RNA, with a little spreading into the proximal elongation complex fraction. Thus, the prominent region of RNA intensity centered at about 1 kb in the wild-type lane of Figure 1B does indeed consist predominantly of RNA from paused elongation complexes, and is not simply processed or partially processed RNA.

Further analysis of poly(A) signal-dependent pausing

The difference curves in Figure 1, B and C, are all in qualitative agreement that polymerases slow down after crossing a wild-type, but not a mutant poly(A) signal. However, the curves vary substantially in shape. We believe this most likely reflects sample-to-sample variations that are magnified when the intensities along one lane of a gel are subtracted from the intensities of another. We therefore decided to characterize poly(A) signal-dependent pausing also by a more quantitative approach that allowed the use of internal standards in all of the samples.

Figure 2A outlines such an experiment. Templates bearing wild-type or mutant poly(A) signals were transcribed in separate reactions (together with a reference plasmid), and TECs were separated from free RNA by size exclusion chromatography. G-less cassettes in the template, positioned both next to the poly(A) signal (pre-cassette) as well as 1-kb farther downstream (post-cassette), permitted polymerase progress to be monitored quantitatively. A reduction in transcription velocity results in less post-cassette being transcribed relative to pre-cassette, because fewer of the slow polymerases reach the post-cassette during the time allotted for transcription (Tran et al. 2001). To quantitate cassette transcription, the transcripts were digested with RNase T1 (which is G-specific) and the surviving G-less cassettes were resolved on a polyacrylamide gel (Fig. 2B).

The gel in Figure 2B shows that, relative to the 120-nucleotide (nt) pre-cassette, significantly less 261-nt post-cassette was transcribed from the template bearing a wild-type poly(A) signal (Fig. 2B, lane 1) than from the template with a mutant poly(A) signal (Fig. 2B, lane 2). This confirms that the polymerases slow down upon encountering the wild-type poly(A) signal, and are less likely to

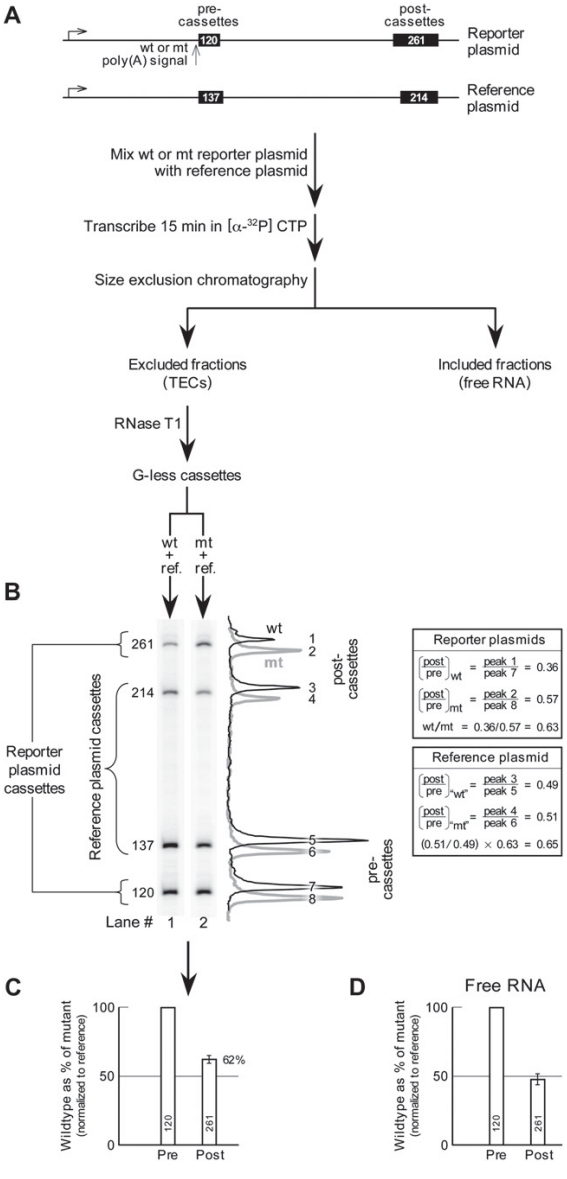


FIGURE 2. (Legend on next page)

reach the post-cassette during the time of the experiment. The visual comparison in the gel of Figure 2B was facilitated by adjusting the overall intensities in the two lanes to equalize the 120-nt pre-cassette bands. This operation is formally equivalent to numerically normalizing the intensity of the post-cassette to that of the pre-cassette in each lane prior to calculating the wild-type/mutant ratio for the post-cassettes, as is done in the upper table to the right of the gel. As shown in this table, only 63% as many polymerases reached the post-cassette in this experiment after transcribing a wild-type as compared with a mutant poly(A) signal.

The gel lanes in Figure 2B also contain bands from the reference plasmid transcripts. Since the post/pre cassette ratios for transcripts from the reference plasmid should theoretically be identical in all reactions, these ratios can be used to correct the reporter plasmid numbers for sample-to-sample variation, as shown in the lower table to the right of the gel. The averaged results, which incorporate this reference correction, for two independent experiments are illustrated in Figure 2C. Here, we define the number of polymerases crossing the pre-cassette as 100% and then show that, on average, only 62% as many of these reach the postcassette when they carry a wild-type poly(A) signal (''wild-type'' TECs) as when they carry a mutant poly(A) signal (''mutant'' TECs).

It is formally possible that the deficit of wild-type TECs carrying post-cassettes results, not from pausing before they reach the post-cassette, but from release after they pass it. However, this would predict a corresponding increase in post-cassettes in the free wild-type RNA from the included fractions of the size exclusion chromatography, and Figure 2D shows that this is not the case. In fact, Figure 2D resembles Figure 2C, suggesting that the free RNA may simply be a random sample of nascent cassette-containing RNA, which is released from the DNA by the poly(A)-independent transcript release factor TTF2 present in the extract (Liu et al. 1998; Hara et al. 1999).

We also examined poly(A) signal-dependent pausing by analysis of heparin and Sarkosyl-washed TECs isolated using immobilized templates rather than size exclusion chromatography (Fig. 3). Although the magnitude of the poly(A) signal-dependent pausing response, like cleavage and polyadenylation itself (Yonaha and Proudfoot 2000; Rigo et al. 2005), is reduced on immobilized templates, these experiments confirm that TECs proceed more slowly when they carry a wild-type poly(A) signal than when they carry a mutant poly(A) signal (Fig. 3B,C).

Pausing in vitro occurs independently of DNA sequence downstream from the poly(A) signal

Poly(A) signal-dependent termination can occur independently of sequence downstream from the poly(A) signal (Yeung et al. 1998). To establish whether poly(A) signal-dependent pausing also is independent of downstream sequence, the various pausing experiments in Figures 1 and 3 were carried out using two different reporters having completely unrelated sequences between their cassettes. Thus, in the case of the reporter used for Figure 1B and the top frame of Figure 1C, the slow TECs accumulate over an arbitrarily chosen eukaryotic sequence, whereas for the reporter used for the bottom two frames of Figure 1C they accumulate over prokaryotic DNA. Similarly, in the cassette analyses of Figure 3, wild-type TECs exhibit a clear delay in reaching the second cassette, regardless of whether the intercassette region consists of eukaryotic (Fig. 3B) or prokaryotic (Fig. 3C) DNA. The slightly reduced effect seen in Figure 3C presumably reflects the somewhat shorter distance separating the cassettes in the reporter used for those experiments (see Fig. 3A; Tran et al. 2001).

Involvement of the SV40 early poly(A) signal in a nascent transcript discard pathway

As noted in the Introduction, weak poly(A) signals, whose processing is slow, direct not only processing (at a low level), but also the degradation of those transcripts not destined to be processed. Since recapitulation of PDD by weak poly(A) signals in vitro may require the use of a transcriptionally coupled system (see Introduction), we turned our attention to the possibility that poly(A) signal-coupled surveillance may accompany the poly(A) signal-dependent transcriptional pause described here.

Studying specific degradation events in a transcriptionally coupled system requires the ability to detect and quantitate transcript loss from a highly heterogeneous population and to mitigate or control for changes in the transcript population that arise from new transcription, pausing, or processing itself. The approach described in Figures 2 and 3, in which the SV40 early poly(A) signal was used in

FIGURE 2. Poly(A) signal-dependent pausing in vitro on circular templates. (*A*) Experimental protocol with plasmid templates drawn to scale. In vitro transcriptions for column separation were carried out as described previously (Nag et al. 2006) using Reporter 1. (*B*) Cassette analysis of the principal TEC column fraction (fraction 8) for a wild-type and a mutant transcription reaction. The signal intensities of the cassette bands in the wild-type and mutant gel lanes are presented as line graphs, which are offset slightly to enhance clarity. Some key ratios are given in the table. The ''wt'' and ''mt'' designations in the case of the reference plasmid cassette ratios refer to transcription mixtures where the reporter was either wild type or mutant (although the same reference plasmid was used for both). (*C*) Overall cassette analysis for the TEC fractions. Data were summed over the two TEC fractions 7 and 8 for the experiment in *B*, and the resulting reporter wt/mt ratio was normalized to the reference wt/mt ratio. The experiment was repeated, and the average and range of values obtained is shown. There is no error bar for the 120-nt cassette because normalization was to this cassette so that this value, by definition, is 100%. (*D*) Cassette analysis of a representative free RNA fraction 11 for the experiment in *B*, normalized to the reference and then averaged across two experiments as in *C*.

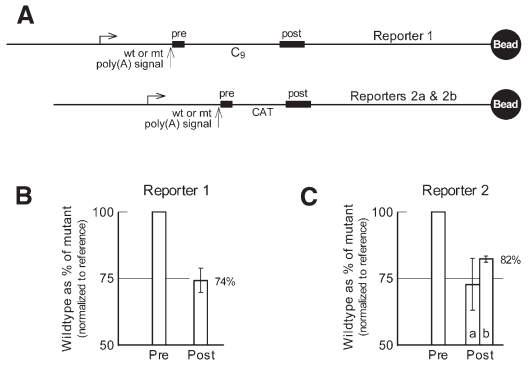


FIGURE 3. Poly(A) signal-dependent pausing in vitro on immobilized templates. (*A*) The immobilized reporter DNAs drawn to scale. Transcription was carried out as for Figure 2. (*B*) Cassette analysis for washed TECs isolated using Reporter 1. The average and standard deviation for three independent experiments is shown. (*C*) Cassette analysis for washed TECs isolated using Reporters 2a and 2b (independently cloned, presumably identical plasmids; see Materials and Methods). The average and range for two independent experiments is shown for reporter 2a, and the average and standard deviation for three independent experiments is shown for reporter 2b. Note that four out of the five results obtained clustered around the value of 82%. We consider the one low value (the 63% for reporter 2a) to be an outlier.

conjunction with G-less cassettes, appeared to be an appropriate point of departure for several reasons. First, the SV40 early poly(A) signal is weak in vivo (Carswell and Alwine 1989; Chao et al. 1999) and in vitro (Rigo et al. 2005). Second, the almost complete absence of processing under the conditions used here minimizes cleavage and polyadenylation as a competing process to be quantitated during the analysis. Third, the use of G-less cassettes allows the quantitation of discrete bands on a gel despite the heterogeneity of the transcript population. Finally, transcription and pausing are easily eliminated as variables by stopping transcription with α-amanitin and then measuring the amount of transcript remaining as a function of time.

To quantitate the amount of transcript present, we used the G-less cassettes carried by the reporter plasmid shown in Figure 4A. All transcripts, both short and long, contained a 76-nt cassette near the 5′-end. In addition, 120-nt and 261-nt cassettes appeared only in transcripts long enough to contain a poly(A) signal. However, transcription time was minimized so that most transcripts (about two-thirds) remained too short to contain a poly(A) signal. This allowed us to express the amount of poly(A) signal-containing transcripts (containing the 120-nt and the 261-nt cassettes) as a fraction of the total population of transcripts (the 76-nt cassettes), most of which lacked a poly(A) signal. Expressing the amount of poly(A) signal-containing transcripts as a ratio in this way eliminated various uncertainties in the data, such as the efficiency of sample recovery. To extract poly(A)-specific information we further normalized all data for the wild-type poly(A) signal to data obtained in parallel using a mutant poly(A) signal. PDD is indicated by a decrease in this ratio as a function of time.

The gel in Figure 4B shows the cassette bands from an experiment designed to detect PDD. Since the short transcription time allowed only 5% of the polymerases to reach the 261-nt cassette, its signal is comparatively weak, and an enhanced version of the 261-nt cassette panel is included in the figure. Note that while there is obvious poly(A) signal-dependent pausing (deficit of 261-nt cassettes in the wild-type lanes, compared with the mutant), this does not affect the analysis of PDD, because the pausing reflects events that occurred before transcription was stopped, prior to the zero-time reference point for PDD (see Fig. 4A). PDD is then revealed by the comparison between 0 and 19 min. For example, the intensity of the 120-nt cassette band in Figure 4B decreases more for the wild type than for the mutant between 0 and 19 min. Quantitative analysis (beneath the gel) shows that the wild-type/mutant ratios for both the 261-nt and the 120-nt cassettes decrease significantly in this time interval (from 0.40 to 0.29 and from 0.95 to 0.61, respectively), indicating the preferential degradation of one-quarter to one-third of the transcripts bearing wild-type versus mutant poly(A) signals. Note that this degradation is in addition to the obvious poly(A)-independent degradation displayed by the 76-nt cassette (Fig. 4B, cf. lanes 1,2 and lanes 3,4), which has been normalized out in the analysis below the gel. The results from several such experiments are summarized in Figure 4C.

Because of the superficial similarities between the G-less cassette assays for poly(A) signal-dependent pausing (Figs. 2, 3) and for PDD (Fig. 4), we want to emphasize the substantial differences between them. Thus, whereas the pausing assay measures primarily ongoing transcription, and reveals diminished transcription of the 261-nt cassette relative to the 120-nt cassette (Figs. 2, 3), the PDD assay measures only the degradation that occurs after transcription has been stopped and detects similar losses for both of the cassettes downstream from the poly(A) signal (Fig. 4C). The simultaneous loss of both of these cassettes from a population of unprocessed transcripts in a poly(A) signal-dependent fashion suggests the existence of a discard pathway in which the entire unprocessed transcript is eliminated.

Pausing alone does not lead to discard

We have previously shown that the poly(A) signal hexamer alone, like the SV40 early poly(A) signal (under the conditions used here), triggers pausing but cannot be processed (Nag et al. 2006). We wondered if it is simply the pausing, when not followed by processing, that triggers degradation of unprocessed transcripts. To answer this question we used a construct in which the SV40 early poly(A) signal had been replaced by a sequence containing

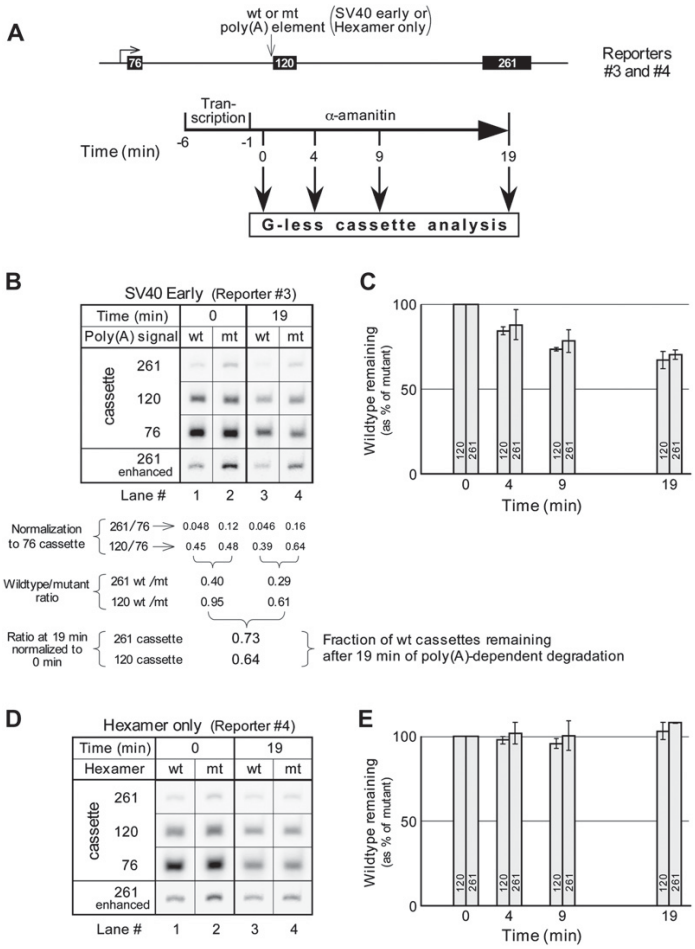


FIGURE 4. Poly(A) signal-dependent degradation in vitro. (*A*) Reporter construct drawn to scale, and the experimental protocol. Transcription was carried out according to Nag et al. (2006) with minor variations. The reaction was scaled up fivefold and transcription was stopped after 5 min by adding 5 μL of 0.5 μg/μL α-Amanitin. At the time intervals indicated, 13.5 μL aliquots were withdrawn and digested with RNase T1 in the presence of EDTA. (*B*) A gel showing the G-less cassette bands from the 0- and 19-min time points of an SV40 early PDD analysis. (*C*) A quantitative summary of several SV40 early PDD analyses, calculated as shown beneath the gel of part *B*. The averages and standard deviations for the 19-min time point are from six independent experiments, and for the 5- and 9-min time points from four independent experiments. The averages include data from both Reporters 3a and 3b (independently cloned, presumably identical plasmids; see Materials and Methods). (*D*) A gel showing the G-less cassette bands from the 0- and 19-min time points of a PDD analysis of Reporter 4, which contains two AATAAA hexamers in place of a poly(A) signal. (*E*) A quantitative summary showing the averages and ranges from two independent experiments. The range for the 261-nt cassette at 19 min was too small to be visible in the graph.

two AATAAA hexamers instead of a hexamer and a G/T-rich region (Nag et al. 2006). Figure 4D shows the relevant bands from an experiment, like that of Figure 4B, designed to detect hexamer-dependent degradation using this hexamer-only construct. Although the gel confirms hexamer-dependent pausing (less 261-nt cassette in the wild-type lanes than in the mutant lanes), there is no evidence of any time-dependent loss of cassettes in a hexamer-dependent (i.e., wild-type versus mutant) fashion. The averaged results from this and one other experiment are summarized in Figure 4E. The contrast between the hexamer-only element (Fig. 4D,E) and the SV40 early poly(A) signal (Fig. 4B,C) is clear—while both cause pausing (albeit less for the hexamer-only element in these experiments), only the poly(A) signal causes wild-type-specific degradation of transcripts. Therefore, pausing per se, not followed by processing, is insufficient to trigger the discard of unprocessed transcripts. Instead, it is the failure of some subsequent event in the processing pathway that leads to rejection of the transcript.

Canonical poly(A) site cleavage does not precede discard

There are at least three distinguishable events that follow pausing in the canonical 3′-end processing pathway, whose failure could conceivably lead to degradation. These are cleavage at the poly(A) site, initiation of poly(A) tail synthesis, and the elongation phase of poly(A) tail growth (Wahle and Kühn 1997). In yeast, pre-mRNA is shunted to a degradation pathway in the presence of certain mRNP maturation factor mutations that affect the elongation phase of polyadenylation. These mutations apparently have little effect on cleavage at the poly(A) site (Saguez et al. 2008). We decided to test whether impaired poly(A) tail growth might similarly account for the tendency of pre-mRNA with weak poly(A) signals to be degraded rather than processed when processing is coupled to transcription.

If differences in the efficiency of poly(A) tail elongation, not differences in the efficiency of poly(A) site cleavage, are what account for the predominance of processing for strong, but the predominance of discard for weak poly(A) signals, then complete elimination of the elongation phase of poly(A) tail growth for both weak and strong poly(A) signals should eliminate the distinction between them. The ATP analog, 3′-dATP, can be used to prevent the addition of all but the first adenine of the poly(A) tail. If impaired poly(A) tail growth is what leads to degradation of RNA with weak poly(A) signals, then one might expect a strong poly(A)

signal to similarly be degraded if poly(A) tail growth is blocked using 3′-dATP. In this case, both weak and strong poly(A) signals would produce low yields of cleaved RNA, bearing a single 3′-dATP at the 3′-end. Alternatively, if the appended 3′-dATP itself inhibits degradation in some way, perhaps because it prevents the residual synthesis of short oligo(A) tails that may be required for the degradation (Saguez et al. 2008), or by failing to support access to a 3′ → 5′ exonuclease, then both weak and strong poly(A) signals should produce high yields of cleaved RNA. Note that by either scenario the yield of cleaved but not polyadenylated RNA is expected to be the same for both weak and strong poly(A) signals.

Figure 5A shows the procedure, and Figure 5B the results for such an assay. The incubation time in α-amanitin (and 3′-dATP) was 50% longer than that used for Figure 4 to ensure that some processing would be detectable even for the weak SV40 early poly(A) signal. The transcript analysis in Figure 5B shows that the difference in processing strength (eight- to ninefold) between the weak SV40 early and the strong SV40 late poly(A) signals is not abrogated by the presence of 3′-dATP; cleavage at the SV40 early poly(A) site is barely detectable (Fig. 5B, lane 2), whereas cleavage at the SV40 late site is robust (Fig. 5B, lane 5). Therefore, lack of poly(A) tail growth does not shunt the SV40 late RNA into the discard pathway, nor does the lack of residual oligo(A) tail synthesis allow the SV40 early poly(A) signal to produce large amounts of RNA. This is consistent with our previous finding that the use of 3′-dATP does not interfere with coupled processing (Rigo et al. 2005; Rigo and Martinson 2008, 2009).

The results of Figure 5B show that the comparative weakness of the SV40 early poly(A) signal is not attributable to a deficiency in the elongation phase of poly(A) tail growth. Therefore, the deficiency that triggers degradation is either in the poly(A) site cleavage step or in the ability to add the first adenine to the new 3′-end generated by the cleavage. To distinguish between these possibilities we compared the degradation of downstream RNA using aliquots of the same samples used for Figure 5B. Canonical poly(A) site cleavage is followed by 5′ → 3′ exonucleolytic degradation of the downstream RNA by Xrn2 (Kaneko et al. 2007). If the reason for the weakness of the SV40 early poly(A) signal is a deficiency in the ability to add the first adenine to the upstream RNA following cleavage (rather than a deficiency in cleavage itself), then the poly(A) signal-dependent degradation of downstream RNA observed in Figure 4C would be attributable to the Xrn2 exonucleolysis that follows cleavage at the poly(A) site, just as for a strong poly(A) signal. According to this scenario, degradation of downstream RNA for a weak poly(A) signal would differ little from that for a strong poly(A) signal.

In contrast, as judged by recovery of the 120-nt cassette, the results shown in Figure 5C indicate that the RNA immediately downstream from the strong SV40 late poly(A) signal (Fig. 5C, lane 7) is degraded to a significantly greater extent than the RNA immediately downstream from the weak SV40 early poly(A) signal (Fig. 5C, lane 3). The averaged results from Figure 5C and a second independent

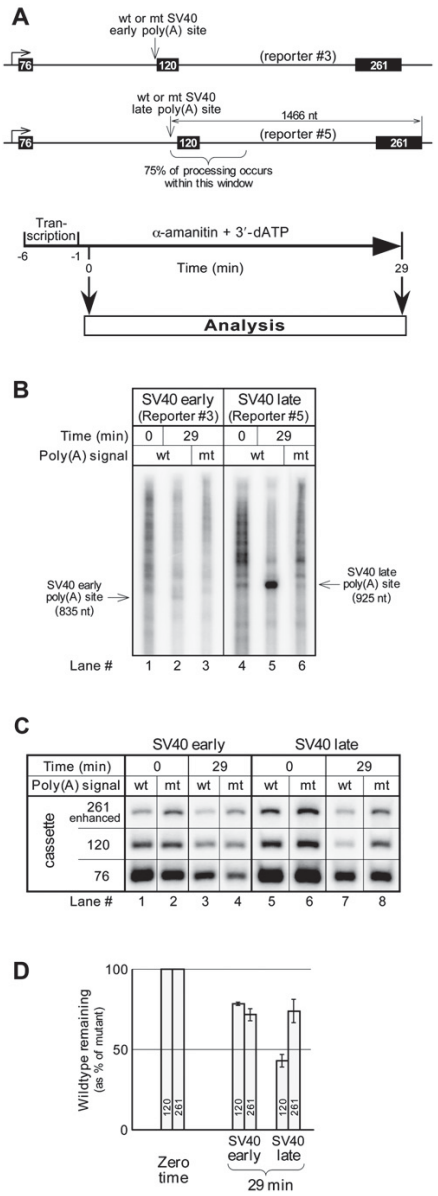


FIGURE 5. Poly(A) signal-dependent degradation in the presence and absence of processing. (*A*) Reporter constructs drawn to scale, and the experimental protocol. (*B*) Direct analysis of transcripts on a gel. (*C*) G-less cassettes from the transcripts. (*D*) A quantitative summary, calculated as for Figure 4C, showing the averages and ranges from two independent experiments. For purposes of normalization to the 76-nt cassette, we assume the stability of cleaved, but nonpolyadenylated RNA is similar to that of uncleaved RNA in our system.

experiment are summarized in Figure 5D, which was calculated as shown in Figure 4B. The quantitation confirms that substantially more RNA degradation occurs immediately downstream from the SV40 late poly(A) signal (as indicated by the 120-nt cassette) than occurs at any distance downstream from the SV40 early poly(A) signal. This more efficient degradation is correlated with more efficient cleavage at the SV40 late poly(A) site, and can therefore be attributed to Xrn2. Conversely, the much less-efficient degradation downstream from the SV40 early poly(A) site suggests strongly that cleavage is inefficient for this poly(A) signal, and that it is inefficient cleavage, rather than efficient $3' \rightarrow 5'$ degradation of efficiently cleaved RNA, which accounts for the low recovery of cleaved RNA in Figure 5B, lane 2.

We hypothesize that inefficient cleavage accounts for the inefficient processing at the SV40 early poly(A) site and that uncleaved RNA, waiting to be processed, is at risk of entering a discard pathway that does not depend on cleavage at the poly(A) site. Initial support for this idea comes, as discussed above, from a comparison of 120-nt cassette degradation rates downstream from the SV40 early and the SV40 late poly(A) sites. However, the idea is strengthened and extended by a consideration of the SV40 late poly(A) signal alone, and the differing fates of the 120-nt and 261-nt G-less cassettes in that construct (Fig. 5D). We have shown previously, in experiments like those of Figure 5B, that the probability of 3′-end processing progressively decreases for transcription complexes halted by α-amanitin at increasing distances beyond the poly(A) signal (Rigo et al. 2005). Nearly 75% of the transcription complexes that are destined to cleave their RNAs at the poly(A) site are found within the first 400 base pairs downstream (Rigo et al. 2005). Extrapolation of the data indicates that few if any of the transcription complexes that reach the end of the 261-nt G-less cassette before the α-amanitin is added engage in processing. Therefore, SV40 late transcripts containing the 261-nt G-less cassette, like all SV40 early transcripts, have little chance of entering the processing pathway and are expected to be subject to PDD. Accordingly, degradation of the long SV40 late transcripts (containing the 261-nt cassette) is the same as for all SV40 early transcripts (Fig. 5D), suggesting that all poly(A) signals are endowed with the same discard capability, but differ in their ability to process.

The discard pathway operates at the level of the transcription elongation complex

The assays of Figures 4 and 5 were conducted on unfractionated transcription reaction mixtures. If PDD is coupled to transcription and is directed at uncleaved RNA waiting to be processed, it should be detectable by analysis of purified TECs. We therefore repeated the transcript degradation analysis of Figure 4, B and C using size exclusion chromatography (see Fig. 6A). We also analyzed the degradation results using two completely independent methods of normalization (Fig. 6D). This more rigorous

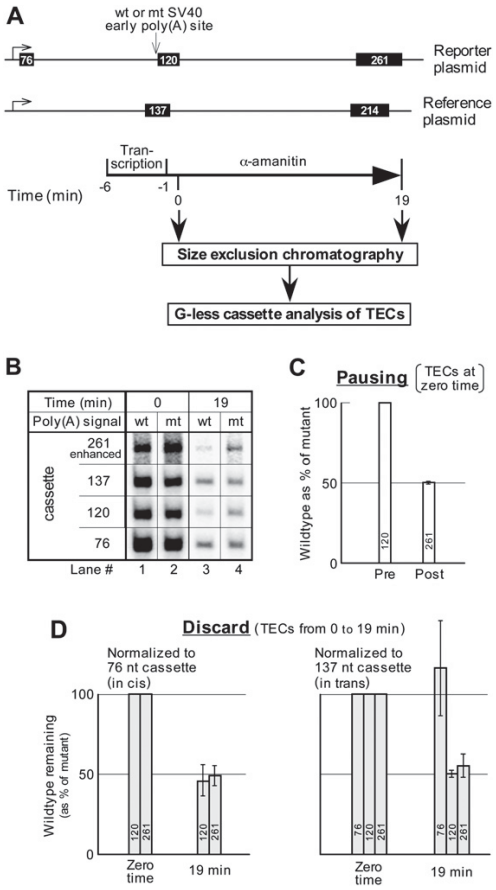


FIGURE 6. Poly(A) signal-dependent degradation is coupled to transcription. (*A*) Reporter 3 and the reference plasmid drawn to scale, together with the experimental protocol. (*B*) G-less cassettes from the transcripts in the peak fraction of the column excluded volume. (*C*) Pausing analysis at the zero time point of the TECs in the principal two fractions of the column excluded volume, showing the average and range from two independent experiments. Cassette analysis was as for Figure 2C but without normalization to the transcripts from the reference plasmid, because the 214-nt cassette band had insufficient counts for quantitation in one of the experiments. (*D*) PDD analysis of TECs at the 19-min time point. Two types of normalization are shown. Either the two downstream cassettes from the reporter transcripts were normalized to the 76-nt cassette as in Figure 4, or all three reporter cassettes were normalized to the 137-nt reference cassette. The average and range for two independent experiments is shown. Note that whereas the wild-type/mutant ratio for the 261-nt cassette is less than that for the 120-nt cassette in *C*, the two cassettes are shown as being the same at zero time (both 100%) in *D*. This is simply a consequence of our standard transcript degradation analysis, for which the 120-nt and the 261-nt cassette levels at all time points are normalized to the respective levels for these cassettes found at the 0 time point (see Fig. 4B). Thus, the equal zero time values in *D* simply reflect self-normalization at this time point.

elongation complex analysis confirmed the results of Figure 4B. First, there was strong poly(A) signal-dependent pausing at the zero time point. This is apparent from lanes 1 and 2 of Figure 6B, where the intensity of the 261-nt band, as compared with the 120-nt band, is less for the wild type than for the mutant (see quantitation in Fig. 6C). Second, between the zero time point and 19 min, there is substantial PDD. This can be seen in Figure 6B, where the intensities of the 120-nt and the 261-nt cassette bands decrease much more for the wild type than for the mutant (see quantitation in Fig. 6D). These results confirm that the PDD pathway operates on TECs, presumably through surveillance of the nascent transcripts. Significantly, appreciably more PDD is observed in assays on purified TECs (Fig. 6) than in assays of the total transcription reaction mixture (Figs. 4, 5). This may reflect the presence in the total reaction mixture of TTF2-released RNA (see above), which, being uncoupled from transcription, is no longer subject to the PDD pathway.

Two different methods of standardizing the data have been used in this study. For Figures 2 and 3 the standard was provided in *trans*, and was a reference plasmid included in the reaction mixture. For Figures 4 and 5 the standard existed in *cis*, being a 76-nt G-less cassette adjacent to the promoter of the reporter plasmid itself. Although the *cis* location of the 76-nt cassette was considered desirable, the proximity to the promoter led to abortive elongation (Marshall and Price 1992; Orozco et al. 2002; Park et al. 2004). This is reflected in the 261/76 and the 120/76 ratios for the mutant templates in Figure 4B, which both increased significantly between 0 and 19 min (importantly, these ratios did not increase for the wild-type template, on account of PDD of the 261-nt and 120-nt cassettes). Although the poly(A)-independent degradation of the 76-nt cassette is theoretically normalized out by the taking of wild-type/mutant ratios, we nevertheless sought to increase our confidence in the final results by designing the experiment of Figure 6 to allow the use of both normalization approaches in parallel. Figure 6D shows that the extent of PDD is statistically identical whether normalization is to the 76-nt cassette in *cis* or to the 137-nt cassette in *trans*. This not only strengthens our confidence in the PDD results overall, but also provides strong experimental validation of both normalization procedures. Moreover, normalization of the 76-nt cassette itself to the 137-nt reference cassette in Figure 6D, right, confirms that the 76-nt cassette undergoes little or no PDD—as expected for TECs that have mostly not reached the poly(A) signal.

Technical considerations

We emphasized earlier that transcriptional pausing cannot contribute in any way to the results of a poly(A) signal-dependent transcript degradation experiment, because the latter experiments were initiated only after blocking transcription with α-amanitin. However, we now evaluate the converse possibility that PDD may contribute to the results of a G-less cassette pausing experiment. Although, in the absence of transcription, PDD affects both the 120-nt and the 261-nt cassettes similarly, it is conceivable that during active transcription PDD might affect the 261-nt cassette preferentially, and thereby give results similar to poly(A) signal-dependent pausing. This possibility exists because, when transcription is ongoing, the poly(A) signal in transcripts long enough to carry a 261-nt cassette has spent more time at risk for degradation than a poly(A) signal in shorter transcripts carrying only the 120-nt cassette. The following analysis suggests that this possibility contributes, at most, minimally to our cassette-based measurements of pausing.

Figure 6D reveals about 50% PDD after 19 min of incubation. Yet, Figure 6C shows pausing of the same magnitude, resulting from only 5 min of transcription—which includes the time taken for the polymerases to reach the poly(A) signal in the first place (about 2 min) (Tran et al. 2001; Rigo et al. 2005). Thus, there is insufficient time for differential PDD to make more than a minor contribution to the 261-nt cassette deficit that we attribute to pausing. Moreover, if there is a lag before the start of PDD, as there is before cleavage and polyadenylation (Rigo et al. 2005), then the putative degradation may not contribute at all to pausing measurements after times of transcription as short as those in Figures 4–6. Finally, note that this analysis applies only to the quantitative pausing measurements carried out using G-less cassettes. The accumulation of slow polymerases seen by direct observation of intact transcripts (Fig. 1B) cannot be accounted for, or contributed to, in any way by PDD.

DISCUSSION

Almost immediately upon emerging from the mammalian RNA polymerase II in vivo, the poly(A) signal directs the polymerase to pause (Orozco et al. 2002; Nag et al. 2007). This pausing appears to be initiated well before the recruitment of the full cleavage and polyadenylation apparatus (Nag et al. 2007), probably during the lag that precedes 3′-end processing (Rigo et al. 2005). We have suggested that this pause serves as an important node for surveillance activity, determining the fate of the transcript (Orozco et al. 2002; Nag et al. 2007). One manifestation of surveillance appears to be the transcript degradation that accounts for the reduced mRNA output from pre-mRNAs bearing weak poly(A) signals. This degradation depends on the poly(A) signal itself (see Introduction) and is most easily imagined as being in competition with authentic cleavage and polyadenylation. Thus, strong poly(A) signals process quickly (Chao et al. 1999) and produce much mRNA, whereas weaker poly(A) signals process more slowly,

allowing many transcripts to be diverted to the discard pathway.

Here, we have studied pausing and discard using an in vitro system in which both of these activities are coupled to transcription. The studies were facilitated by use of a relatively weak poly(A) signal, which displays only moderate processing activity in vivo, and which is hardly processed at all under the in vitro conditions used here. This allowed us to visualize pausing and discard without significant interference by simultaneous processing. Figure 1B demonstrates directly that TECs pause after crossing a poly(A) signal, and Figure 6D shows that the transcripts associated with such complexes are then degraded in a poly(A) signal-dependent manner with a half-life of <20 min.

Poly(A) signal-dependent pausing of transcription

The experiments of Figures 1–3, all based on isolated TECs, confirm known properties of poly(A) signal-dependent pausing determined in vivo and provide new insights. First, poly(A) signal-dependent pausing depends on an intact AAUAAA hexamer in the poly(A) signal (Figs. 1–3), consistent with the central role of CPSF in pausing (Nag et al. 2007). Second, the poly(A) signal does not act like a pause site, or blockade, to polymerase progress. That is, polymerases do not pause at the poly(A) signal itself but, consistent with a mechanism in which the elongation properties of the polymerase have been modified, the polymerases exhibit reduced elongation rates over the DNA downstream. Third, although poly(A) signal-dependent pausing occurs over downstream DNA, the pausing requires only the poly(A) signal and does not depend on any sequence elements in the downstream DNA (Figs. 1B,C, 3). Fourth, pulse-chase experiments show that the paused elongation complexes carry RNA that is not cleaved at the poly(A) site (Fig. 1). Therefore, pausing does not require processing, consistent with the appearance of pausing in the lag that precedes processing in coupled transcription-processing reactions using strong poly(A) signals in vitro (e.g., early time points of Fig. 1A in Rigo et al. 2005).

Finally, the modified properties of the TEC responsible for pausing appear to be transient. This is evident in the difference curves of Figure 1, B and C, that reveal a concentration of slow polymerases immediately downstream from the poly(A) signal, which dissipates after several hundred base pairs under these in vitro conditions. The resumption of full-speed transcription may reflect the disruption of the CPSF–polymerase interaction (the initial cause of pausing) that has been proposed to occur during maturation of the cleavage and polyadenylation apparatus (Nag et al. 2007). The reduced number of paused elongation complexes at distal positions on the template in Figure 1 may also reflect the increasing effects of PDD on longer transcripts that have spent more time at risk. Interestingly, poly(A) signal-dependent pausing can extend considerably beyond a few hundred base pairs in vivo (e.g., Gromak et al. 2006; Kaneko et al. 2007; Glover-Cutter et al. 2008), especially if either the CTD or the poly(A) signal has been compromised by mutation (Park et al. 2004; Nag et al. 2006, 2007). Since the polymerase and the poly(A) signal are in continuous communication as elongation proceeds (Kim and Martinson 2003), various elements downstream from the poly(A) signal would be expected to modulate poly(A) signal-dependent pausing through effects on the CTD and/or on the assembling cleavage and polyadenylation apparatus.

Poly(A) signal-dependent degradation of nascent transcripts

The experiments of Figures 4 and 5 reveal a new activity of the mammalian poly(A) signal—PDD of nascent transcripts. This activity was anticipated because inactivation of weak poly(A) signals in vivo leads to accumulation of unprocessed RNA that would otherwise be degraded (see Introduction).

PDD is distinct from the recently described pre-mRNA surveillance pathway in yeast, where inefficient polyadenylation leads to degradation of RNA that has been cleaved at the poly(A) site (Saguez et al. 2008). The pathway described by Saguez et al. is triggered by defects in proteins responsible for maturation and transport of the mRNA, whereas PDD appears to be a normal activity, not triggered by defects, that is directed by the poly(A) signal itself. Moreover, the substrate for PDD appears to be the nascent uncleaved RNA (Figs. 5 and 6), not post-transcriptional RNA that has already been cleaved at the poly(A) site. This is consistent with the properties of PDD in vivo, which indicate that it is slow cleavage, not slow polyadenylation, that leads to degradation. In vivo, most RNA-bearing weak poly(A) signals gets degraded; yet, much of this RNA can be rescued by placing a strong poly(A) signal downstream from the weak one (e.g., Peterson and Perry 1989; Enriquez-Harris et al. 1991; Andrews and DiMaio 1993; Yang and Melera 1994). Since strong poly(A) signals process faster than weak ones (Chao et al. 1999), this observation is readily explained on the basis of a kinetic competition between the cleavage reactions. The rescue cannot be explained by competition between polyadenylation reactions on RNA that has already been cleaved at the weak site. Apparently, the PDD pathway is faster than cleavage at these weak sites but slower than cleavage at the strong sites, consistent with the results of Figure 5. Moreover, like cleavage itself, the PDD pathway operates on RNA that still contains an intact poly(A) signal.

Recently, an interesting example of experimentally induced PDD in vivo was reported (Fong et al. 2009). When cleavage at a strong poly(A) signal was suppressed by decapping the nascent transcript, the uncleaved RNA did

not accumulate, but instead was degraded. The degradation was blocked, and the uncleaved RNA was allowed to accumulate by a point mutation in the poly(A) signal hexamer, confirming that the degradation caused by decapping was due to PDD.

Three mechanisms potentially responsible for PDD can be imagined. The first invokes the newly discovered endonuclease activity of the exosome (Lebreton et al. 2008; Schaeffer et al. 2009). Since the exosome most likely engages the nascent RNA cotranscriptionally (Andrulis et al. 2002; Custódio et al. 2007; Schmid and Jensen 2008) and interacts directly or indirectly with poly(A) polymerase (Burkard and Butler 2000), it is possible that the exosome itself is positioned at or near the poly(A) site, so as to cleave (and then to degrade) when cleavage directed by the 3′-end processing apparatus of a poly(A) signal is slow.

A second mechanism that could conceivably account for PDD invokes the known multifunctionality of CPSF. For example, in mammalian histone RNA processing, cleavage by CPSF73 is not coupled to polyadenylation, and it is CPSF73 itself, not Xrn2, that exonucleolytically degrades the downstream RNA after cleavage (Yang et al. 2009). Perhaps, for poly(A) signals, one mode of CPSF73 cleavage is coupled to polyadenylation, whereas the other is coupled to degradation of the upstream RNA by the exosome. A kinetic competition between these two functionalities would predict high levels of degradation when authentic processing is slow.

A third mechanism that might account for PDD, and the one that we prefer, is suggested by the fact that some poly(A) signal-dependent (but cleavage-independent) release of downstream cassettes into the free RNA column fractions always occurs in our size exclusion chromatography experiments (data not shown). This observation is consistent with a discard mechanism in which transcripts that remain unprocessed are at risk of poly(A) signal-dependent release from the template, followed by $3' \rightarrow 5'$ exosomal degradation. Because there is no cleavage at the poly(A) site, this release could not be an Xrn2 exonuclease-dependent event. However, the release could be mediated by the poly(A) signal-dependent action of Pcf11 in the absence of cleavage (Sadowski et al. 2003; Qu et al. 2006). According to this mechanism, Xrn2-mediated termination would be coupled to processing, whereas Pcf11-mediated termination would be coupled to degradation. This mechanism would also explain how transcription could be subject to efficient poly(A) signal-dependent termination, even for poly(A) signals that cleave their transcripts very slowly and which also lack any special terminator elements (e.g., SV40 early in Orozco et al. 2002). This model is also consistent with the observation that knockdown of the nuclear exosome subunit PMScl100 in vivo increases the recovery of 3′-end processed RNA for weak poly(A) signals (West and Proudfoot 2009).

MATERIALS AND METHODS

Plasmid templates

Reporter 1 (drawn to scale in Figs. 2A, 3A) is closely related (and identical in the region shown in Fig. 2A) to pAP⟨C_9⟩ of Tran et al. (2001). The sequence between the cassettes, called C_9, is a ninefold repeated snippet of eukaryotic DNA. The poly(A) signal mutant was derived from the wild type by site-directed mutagenesis (AATAAA → AgTAct). The elongation reference plasmid was derived from this mutant by removing Gs to lengthen the pre-cassette (GGGCAAGCTG → aaaCAAtCTt) and by adding Gs to shorten the post-cassette (TCATTAT → TCgagAT). Reporter 2a (drawn to scale in Fig. 3A) is closely related (and identical from the post-cassette leftward) to the pAP⟨117cat⟩ construct in Tran et al. (2001). The sequence between the cassettes is a piece of the prokaryotic chloramphenicol acetyl transferase gene. Reporter 2b is presumed to be identical to 2a, having been derived, as a control, from the poly(A) signal mutant of 2a by back-mutagenesis. Reporter 3a (drawn to scale in Fig. 4A) was generated from Reporter 1 by cutting with StuI next to the promoter and inserting the following G-less cassette sequence (5′ → 3′): TGCCATACCCT TCCTCCATCTATACCACCCACTCTCCTTTCCTCATTATTCCT CCTATTATCTTCTCCTCTTCTCTCAGG. Reporter 3b, presumably identical to Reporter 3a, is a control derived from the mutant of 3a by back-mutagenesis. Reporter 4 is identical to Reporter 3 except that the poly(A) signal was replaced by an identical length of DNA containing two AATAAA hexamers instead of a hexamer and a G/T-rich region (insert iii of Nag et al. 2006). The mutant version of this reporter, in which both hexamers are mutated, contains insert v from Nag et al. Reporter 5 is identical to Reporter 3 except that a SmaI–BamHI fragment containing the SV40 late poly(A) signal (Tran et al. 2001) replaces the HpaI–BamHI fragment of Reporter 3 that contains the SV40 early poly(A) signal. The mutant form of Reporter 5 contained AAgtAc in place of the wild-type poly(A) signal hexamer, AATAAA.

In vitro transcription

HeLa nuclear extract preparation and in vitro transcription were based on the procedures described by Tran et al. (2001). The concentrations of magnesium and citrate used in the assays were individually optimized for each extract preparation. Typically, 5–7 mM magnesium and 6–8 mM citrate were used, except for Figure 1D, right (see legend). Where applicable, α-amanitin and 3′-dATP were added as described in Rigo et al. (2005).

Size exclusion chromatography

Size exclusion chromatography of transcription reactions was done using a 3.1 mL (0.6 cm × 11 cm) column as described by Nag et al. (2006), except that 1% rather than 0.5% Sarkosyl was used for the experiments of Figure 1, B and C. Wild-type and mutant samples were transcribed in parallel, stopped as described by Nag et al. (2006), and then fractionated in succession on a size exclusion column. Since fractionating the wild-type and mutant samples in succession required that one of the samples be stored on ice for about 25 min before the fractionation, we alternated the column load order whenever experiments were repeated. We

found that the initial extraction of column fractions should be with phenol rather than with TRIzol to avoid size bias in the recovery of the RNA. For Figure 2 the recovered RNA was then digested with RNase T1 to release the G-less cassettes, TRIzol extracted, and then loaded on the gel. For Figure 1 the column fractions were treated briefly with proteinase K before, and then with DNase I (rather than RNase T1) after the phenol extraction. Figure 6 was as for Figure 2 except that brief, successive RNase T1 and proteinase K treatments of the column fractions preceded the phenol extraction.

Immobilized templates

End-filling was used to generate biotinylated DNA for attaching to streptavidin-coated magnetic beads. Wild-type:mutant comparisons were made only between preparations made in parallel. All plasmid DNA was cut with EagI, and the end was filled using Klenow polymerase, dGTP, and Biotin-14-dCTP (Invitrogen). All DNAs were then cut a second time (Reporter 1 and the reference plasmid with AseI, and Reporter 2 with BsaI), and the resulting large biotinylated fragment from each plasmid was resolved using agarose gel electrophoresis and isolated using a Qiagen Gel Extraction kit. The Qiagen eluate for each reporter DNA was then mixed with eluate for reference DNA and added to an equal volume of magnetic beads suspended in kilobaseBINDER solution (1 mg beads per 10 μg DNA). After rotation at room temperature for ≥10 h, the beads were washed twice with Dynal B&W Buffer, four times with 10 mM Tris, 0.1 mM EDTA (pH 8), and resuspended in 10 mM Tris, 0.1 mM EDTA (pH 8). About 30% of the input DNA became attached to the beads as estimated by restriction digestion and agarose gel analysis of a small aliquot of the preparation.

Following transcription, beads were removed by a 2-min magnetic selection, washed twice with nuclear extract buffer D (Tran et al. 2001) containing 0.5 mg/mL heparin and 0.5% Sarkosyl, and then the beads were treated with RNase T1 followed by TRIzol to recover the cassettes (Nag et al. 2006). RNA was precipitated with 20 μg of Glycogen (Ambion) and 350 μL of isopropanol, and the cassettes were separated on an 8% polyacrylamide gel.

Since the comparison of wild-type to mutant transcription for immobilized templates necessarily required comparing different immobilized template preparations to each other, we used transcription of the reference DNA (included on all the beads) to screen for equivalent performance of the immobilized templates as a condition for using any of the data from those preparations. Ideally, the elongation efficiencies, as indicated by the post/pre cassette ratios (see Fig. 2B; see the text), should be identical for the reference DNAs of all of the immobilized template preparations transcribed in parallel. But, in actuality, there was considerable variation among the preparations. Therefore, we included in our wild-type/mutant comparisons only immobilized template preparations that differed from each other by no more than 10% in their reference post/precassette ratios (for both pellet and supernatant). For the template preparations that survived this screen, the wild-type/mutant cassette ratio obtained for the reporter plasmids was then normalized to the wild-type/mutant ratio for their accompanying reference, as illustrated by the example in Figure 2B. Thus, the normalized wt/mt ratio is given by the following expression: $\left[(\text{post}/\text{pre})_{\text{wt}}/(\text{post}/\text{pre})_{\text{mt}}\right]_{\text{reporter}}/\left[(\text{post}/\text{pre})_{\text{wt}}/(\text{post}/\text{pre})_{\text{mt}}\right]_{\text{reference}}$.

ACKNOWLEDGMENTS

We thank Frank Rigo and Anita Nag for important discussions during the course of this work, and the NIH and the Jonsson Cancer Center Foundation for support.

Received March 1, 2009; accepted September 22, 2009.

REFERENCES

Andrews EM, DiMaio D. 1993. Hierarchy of polyadenylation site usage by bovine papillomavirus in transformed mouse cells. *J Virol* **67:** 7705–7710.

Andrulis ED, Werner J, Nazarian A, Erdjument-Bromage H, Tempst P, Lis JT. 2002. The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature* **420:** 837–841.

Boireau S, Maiuri P, Basyuk E, de la Mata M, Knezevich A, Pradet-Balade B, Backer V, Kornblihtt A, Marcello A, Bertrand E. 2007. The transcriptional cycle of HIV-1 in real-time and live cells. *J Cell Biol* **179:** 291–304.

Burkard KT, Butler JS. 2000. A nuclear 3′–5′ exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein npl3p. *Mol Cell Biol* **20:** 604–616.

Calvo O, Manley JL. 2003. Strange bedfellows: Polyadenylation factors at the promoter. *Genes & Dev* **17:** 1321–1327.

Carswell S, Alwine JC. 1989. Efficiency of utilization of the simian virus 40 late polyadenylation site: Effects of upstream sequences. *Mol Cell Biol* **9:** 4248–4258.

Chao LC, Jamil A, Kim SJ, Huang L, Martinson HG. 1999. Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. *Mol Cell Biol* **19:** 5588–5600.

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

Custódio N, Vivo M, Antoniou M, Carmo-Fonseca M. 2007. Splicing- and cleavage-independent requirement of RNA polymerase II CTD for mRNA release from the transcription site. *J Cell Biol* **179:** 199–207.

Dantonel JC, Murthy KG, Manley JL, Tora L. 1997. Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. *Nature* **389:** 399–402.

Edwalds-Gilbert G, Prescott J, Falck-Pedersen E. 1993. 3′ RNA processing efficiency plays a primary role in generating termination-competent RNA polymerase II elongation complexes. *Mol Cell Biol* **13:** 3472–3480.

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

Fong N, Ohman M, Bentley DL. 2009. Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. *Nat Struct Mol Biol* **16:** 916–922.

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

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

Hara R, Selby CP, Liu M, Price DH, Sancar A. 1999. Human transcription release factor 2 dissociates RNA polymerases I and II stalled at a cyclobutane thymine dimer. *J Biol Chem* **274:** 24779–24786.

Hieronymus H, Yu MC, Silver PA. 2004. Genome-wide mRNA surveillance is coupled to mRNA export. *Genes & Dev* **18:** 2652–2662.

Kaneko S, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2007. The multifunctional protein p54nrb/PSF recruits the exonuclease

XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. *Genes & Dev* **21:** 1779–1789.

Kim SJ, Martinson HG. 2003. Poly(A)-dependent transcription termination: Continued communication of the poly(A) signal with the polymerase is required long after extrusion in vivo. *J Biol Chem* **278:** 41691–41701.

Lebreton A, Tomecki R, Dziembowski A, Seraphin B. 2008. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* **456:** 993–996.

Lian Z, Karpikov A, Lian J, Mahajan MC, Hartman S, Gerstein M, Snyder M, Weissman SM. 2008. A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3′ end RNA polyadenylation. *Genome Res* **18:** 1224–1237.

Liu M, Xie Z, Price DH. 1998. A human RNA polymerase II transcription termination factor is a SWI2/SNF2 family member. *J Biol Chem* **273:** 25541–25544.

Marshall NF, Price DH. 1992. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol* **12:** 2078–2090.

Nag A, Narsinh K, Kazerouninia A, Martinson HG. 2006. The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. *RNA* **12:** 1534–1544.

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

Orozco IJ, Kim SJ, Martinson HG. 2002. The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. *J Biol Chem* **277:** 42899–42911.

Park NJ, Tsao DC, Martinson HG. 2004. The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II CTD. *Mol Cell Biol* **24:** 4092–4103.

Perkins KJ, Lusic M, Mitar I, Giacca M, Proudfoot NJ. 2008. Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. *Mol Cell* **29:** 56–68.

Peterson ML, Perry RP. 1989. The regulated production of μ_m and μ_s mRNA is dependent on the relative efficiencies of μ_s poly(A) site usage and the Cμ4-to-M1 splice. *Mol Cell Biol* **9:** 726–738.

Pillutla RC, Yue Z, Maldonado E, Shatkin AJ. 1998. Recombinant human mRNA cap methyltransferase binds capping enzyme/RNA polymerase IIo complexes. *J Biol Chem* **273:** 21443–21446.

Prescott JC, Falck-Pedersen E. 1992. Varied poly(A) site efficiency in the adenovirus major late transcription unit. *J Biol Chem* **267:** 8175–8181.

Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. *Cell* **108:** 501–512.

Qu X, Perez-Canadillas JM, Agrawal S, De Baecke J, Cheng H, Varani G, Moore C. 2006. The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3′-end processing. *J Biol Chem* **282:** 2101–2115.

Rigo F, Martinson HG. 2008. Functional coupling of last intron splicing and 3′-end processing to transcription in vitro: The poly(A) signal couples to splicing before committing to cleavage. *Mol Cell Biol* **28:** 849–862.

Rigo F, Martinson HG. 2009. Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. *RNA* **15:** 823–836.

Rigo F, Kazerouninia A, Nag A, Martinson HG. 2005. The RNA tether from the Poly(A) signal to the polymerase mediates coupling of transcription to cleavage and polyadenylation. *Mol Cell* **20:** 733–745.

Roth KM, Byam J, Fang F, Butler JS. 2009. Regulation of NAB2 mRNA 3′-end formation requires the core exosome and the Trf4p component of the TRAMP complex. *RNA* **15:** 1045–1058.

Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J, Jensen TH, Stutz F, Devaux F, et al. 2008. THO/Sub2p functions to coordinate 3′-end processing with gene-nuclear pore association. *Cell* **135:** 308–321.

Ryner LC, Takagaki Y, Manley JL. 1989. Sequences downstream of AAUAAA signals affect pre-mRNA cleavage and polyadenylation in vitro both directly and indirectly. *Mol Cell Biol* **9:** 1759–1771.

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

Saguez C, Schmid M, Olesen JR, Ghazy MA, Qu X, Poulsen MB, Nasser T, Moore C, Jensen TH. 2008. Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. *Mol Cell* **31:** 91–103.

Schaeffer D, Tsanova B, Barbas A, Reis FP, Dastidar EG, Sanchez-Rotunno M, Arraiano CM, van Hoof A. 2009. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat Struct Mol Biol* **16:** 56–62.

Schmid M, Jensen TH. 2008. Quality control of mRNP in the nucleus. *Chromosoma* **117:** 419–429.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol Cell* **33:** 365–376.

Tran DP, Kim SJ, Park NJ, Jew TM, Martinson HG. 2001. Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. *Mol Cell Biol* **21:** 7495–7508.

Uhlmann T, Boeing S, Lehmbacher M, Meisterernst M. 2007. The VP16 activation domain establishes an active mediator lacking CDK8 in vivo. *J Biol Chem* **282:** 2163–2173.

Wahle E, Kühn U. 1997. The mechanism of 3′ cleavage and polyadenylation of eukaryotic pre-mRNA. *Prog Nucleic Acid Res Mol Biol* **57:** 41–71.

West S, Proudfoot NJ. 2009. Transcriptional termination enhances protein expression in human cells. *Mol Cell* **33:** 354–364.

Wu C, Alwine JC. 2004. Secondary structure as a functional feature in the downstream region of Mammalian polyadenylation signals. *Mol Cell Biol* **24:** 2789–2796.

Yang H, Melera PW. 1994. A genetic polymorphism within the third poly(A) signal of the DHFR gene alters the polyadenylation pattern of DHFR transcripts in CHL cells. *Nucleic Acids Res* **22:** 2694–2702.

Yang XC, Sullivan KD, Marzluff WF, Dominski Z. 2009. Studies on the 5′ exonuclease and endonuclease activity of CPSF-73 in histone pre-mRNA processing. *Mol Cell Biol* **29:** 31–42.

Yeung G, Choi LM, Chao LC, Park NJ, Liu D, Jamil A, Martinson HG. 1998. Poly(A)-driven and poly(A)-assisted termination: Two different modes of poly(A)-dependent transcription termination. *Mol Cell Biol* **18:** 276–289.

Yonaha M, Proudfoot NJ. 2000. Transcriptional termination and coupled polyadenylation in vitro. *EMBO J* **19:** 3770–3777.

Zhao J, Hyman L, Moore C. 1999. Formation of mRNA 3′ends in eukaryotes: Mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63:** 405–445.

Chapter 3

Transcription elongation complex directs coupling of 3’-end processing to transcription without assistance of events from the promoter

Abstract

It has been known for more than two decades now that 3'-end processing is coupled to transcription. Several mechanisms by which transcription can affect 3'-end processing have been proposed. These mechanisms include pre-recruitment of processing factors to the promoter, cotranscriptional assembly of cleavage and polyadenylation factors during gradual extrusion of the poly(A) signal, and phosphorylation of the C-terminal domain (CTD) of RNA polymerase II. In order to understand the contributions from each of these distinct mechanisms, we investigated these mechanisms in a coupled in vitro system. We found that recruitment of processing factors to the promoter is not required for coupled 3'-end processing in our in vitro system. We demonstrate that co-transcriptional recruitment of cleavage and polyadenylation factors during gradual extrusion of the poly(A) signal is not important. We also found that CTD phosphorylation history is unimportant since removal of CTD phosphates put on during initiation and early elongation does not abrogate 3'-end processing, as long as elongation complexes are given opportunity to rephosphorylate. Surprisingly, even when events at the promoter were bypassed altogether, by loading elongation complexes onto a promoterless template via a 3'-extended oligo(dT) tail, the transcription elongation complexes (TECs) were still able to process an order of magnitude more efficiently than processing that is uncoupled from transcription. These results argue that the ability to couple 3'-end processing to transcription is intrinsic to the structure of the TEC. Consistent with this, I found that coupling is inhibited if the RNA tether which connects the RNA transcript to the polymerase is severed. Based on these results, we propose that coupling is somehow mediated by having the poly(A) signal held close to the CTD via an intact RNA tether in the context of the TEC.

Introduction

Within the last decade, it has become increasingly clear that capping, splicing and, 3'-end formation is functionally interconnected to each other and to transcription (Maniatis et al. 2002). Despite this, it is still not clear how transcription affects 3'-end processing. In a review by Bentley (Bentley 2005), three possible mechanisms have been proposed: 1) recruitment of processing factors to the elongation complex to increase local concentration of processing factors at the site of processing, 2) kinetic coupling, and 3) allosteric activation by the C-terminal domain (CTD) of RNA polymerase II.

The idea that processing factors are recruited early to the elongating polymerase to act at the site of processing originated from a report in which a core cleavage and polyadenylation specificity factor (CPSF) was found to co-purify with transcription initiation factor TFIID in an immunoprecipitation against TBP (TATA-binding protein), suggesting that CPSF may be recruited to the promoter during transcription initiation (Dantonel et al. 1997). During transcription, CPSF was found to associate with the transcribing polymerase and not TFIID in vitro, which is consistent with CPSF being transferred to the elongating polymerase from TFIID (Dantonel et al. 1997). Consistent with the idea of preloading processing factors to the promoter, chromatin immunoprecipitation (chIP) studies showed that specific subunits of CPSF and the other core processing factor, cleavage stimulatory factor (CstF), are found at the promoter and in the body of protein-coding genes (Glover-cutter et al. 2008). In addition to recruitment by TFIID, transcriptional activators can stimulate 3'-end processing in vivo and in vitro and the extent of activation correlates with the ability of the activator to bind/recruit processing factors to the promoter either directly or indirectly (Rosonina et al. 2003; Rosonina

et al. 2005; Nagaike et al. 2011). Moreover, processing factors have been reported to associate with the elongating polymerase through interactions with the elongation Paf1c complex (Rozenblatt-Rosen et al. 2009) and transcription factor ELL2 (Martincic et al. 2009). Despite this growing body of evidence, it has not been directly demonstrated that these pre-recruited processing factors are functional and participate later in the 3'-end processing reaction.

Transcription rates can influence processing by affecting assembly of RNA-protein complexes (de la Mata et al. 2003; Howe et al. 2003; Pinto et al. 2011). In a situation where two processing sites can be utilized, slow transcription favors the use of proximal processing sites over distal ones if the appearance of the distal site is significantly delayed. This is due to the presumption that slow elongation enables more time for the processing apparatus to assemble and act on the proximal site before transcription of the second signal. Indeed, elongation rates have been seen to affect both alternative splicing (de la Mata et al. 2003; Howe et al. 2003) and alternative polyadenylation (Pinto et al. 2011). Although this concept of kinetic coupling is applied in situations where more than one processing signal is present, it is interesting to speculate whether transcription rates can also affect processing at individual sites by influencing order of assembly of processing factors.

The CTD of RNA polymerase II has been proposed to be important for coupling 3'-end processing to transcription in two ways. First, the CTD can act as a landing pad to recruit processing factors and to deliver them to the emerging transcript during transcription (McCracken et al. 1997; Proudfoot 2004; Bentley 2005). Second, the CTD can act to allosterically activate the cleavage and polyadenylation apparatus (CPA) (Hirose et al. 1998). An intriquing property of the CTD is that it is made up of many repeats (up to 52 in mammals) of

the heptameric sequence, YSPTSPS (Tyr-Ser-Pro-Thr-Ser-Pro-Ser), which is modified extensively in vivo through phosphorylation, glycosylation, and prolyl isomerization (Egloff et al. 2008). The realization that there are numerous possible combinations of CTD modification eventually led to proposal of the CTD code (Buratowski 2003). CTD code postulates that the modification state of the CTD can be "read, written, and erased" by the transcriptional machinery to dictate the association or dissociation of complexes important to carry out all stages of transcription and RNA processing. An outcome of this model is that a specific CTD code is generated to direct efficient 3'-end processing. Of the various CTD modifications, Ser2 and Ser5 phosphorylation is by far the most-well studied and therefore best illustrates how CTD modification can affect various stages of transcription (Buratowski 2003).

Phosphorylation of the CTD is an orchestrated event beginning with Ser5 phosphorylation by the Cdk7 subunit of TFIIH at the promoter. This phosphorylation leads to promoter escape and transition into early elongation mode. Ser5 phosphorylation also facilitates recruitment of capping factors for 5' capping (Komarnitsky et al. 2000; Schroeder et al. 2000). Soon after, the elongation complex encounters proximal pausing. This is overcome by phosphorylation of the CTD at Ser2 and associated elongation factors by Cdk9, subunit of P-TEFb, which transforms them into processive elongation complexes (Peterlin et al. 2006). The recruitment of P-TEFb is thought to occur generally through a bromodomain protein Brd4 (Jang et al. 2005; Yang et al. 2005), although DNA-sequence specific transcriptional activators that can bind to P-TEFb has also been suggested to recruit P-TEFb to specific promoters (Zhou et al. 2006). Besides Ser2 and Ser5 phosphorylation, not much is known about the role of Tyr1, Thr4, and Ser7 phosphorylation in transcription. Tyr1 can be phosphorylated by c-Abl in mammals but no

homolog is present in yeast (Baskaran et al. 1999). More recently, Ser7 phosphorylation was found to occur at 3' ends of protein-coding genes, although its current role in transcription is still unclear (Chapman et al. 2007). In addition, Thr4 of the CTD has also been reported to be phosphorylated and is required for 3'-end processing of histone pre-mRNAs (Hsin et al. 2011). Based on the growing body of evidence of a CTD code and the recently discovered but unknown functions of Thr4 and Ser7 phosphorylation, it is interesting to speculate whether the orchestration of phosphorylation during transcription is important for coupling transcription to 3'-end processing.

The core poly(A) signal is made up of an AAUAAA hexamer sequence and a downstream G/U rich region, which is recognized by core processing factors CPSF and CstF, respectively (Zhao et al. 1999). In addition to these two complexes, the full assembly of the 3'-end processing complex may involve greater than 80 proteins (Shi et al. 2009). Since the only chemical reaction demanded of this apparatus is hydrolysis of a single phosphodiester bond in the RNA, the requirement for such a large apparatus suggests that 3'-end formation may be the center of a large regulatory network to ensure the quality of the RNA being produced.

To investigate the mechanistic details of how transcription is coupled to 3'-end formation, we utilized a previously reported in vitro transcription system (Rigo et al. 2005). This HeLa nuclear extract-based system has been shown to support both functional splicing and 3'-end processing that is coupled to each other and to transcription (Rigo et al. 2005; Rigo et al. 2008; Rigo et al. 2009). To our surprise, we found that coupling did not depend on pre-recruitment of processing factors to the promoter, co-transcriptional assembly of cleavage and polyadenylation factors during gradual extrusion of the poly(A) signal, or orchestration of CTD

phosphorylation in our in vitro system. In fact, coupling still persisted even when transcription was initiated on a template without a promoter. Instead, we found that the transcription elongation complex can direct efficient 3'-end processing as long as the processing signal remains close with the elongating polymerase via an intact RNA tether.

Results

3'-end processing of dsxSV40L template is coupled to transcription

Previously, we have demonstrated that our in vitro system can support 3'-end processing that is functionally coupled to transcription (Rigo et al. 2005). Our definition of coupling is not simply that 3'-end processing and transcription is carried out concurrently (or at the same time). Rather, the process of transcription is functionally connected to 3'-end processing and a consequence of this is that 3'-end processing occurs more efficiently than processing that is uncoupled from transcription (Rigo et al. 2005, Figure 2C-E). Moreover, we have demonstrated that this transcription-coupled 3'-end processing is inhibited if the integrity of the transcription elongation complex is compromised by severing the RNA tether and dissociating the RNA transcript from the transcribing polymerase (Rigo et al. 2005, Figure 3B). Therefore, we sought to understand how 3'-end processing is coupled to transcription using this coupled in vitro system.

To facilitate studies on the mechanism of coupling, we immobilized the dsxSV40L template (also known as Gal5-HIV2dsxΔInt(+ESE); Rosonina et al. 2003) on magnetic streptavidin beads (Fig. 1A, template diagram). The dsxSV40L template contains an SV40 late poly(A) signal, whose expression is placed under the control of a HIV2 promoter and five Gal4 binding sites. The Gal 4 sites allow us to control transcription levels by the addition of an exogenous transcriptional activator such as Gal4-p53. To confirm that processing is efficient with this construct, we carried out an in vitro transcription-processing assay (Rigo et al. 2005). Figure 1A shows 3'-end processing of dsxSV40L transcripts is efficient in our coupled in vitro system. After forming the preinitiation complex on the beaded templates, we initiated transcription with a pulse of [α-

^{32}P]-CTP and then chased with a high concentration of unlabeled CTP (Figure 1A, experimental timeline diagram). Transcription was stopped with the addition of α-amanitin and 3'dATP (to block polyadenylation) and poly(A) site cleaved RNA was allowed to accumulate for an hour. Processing efficiency, expressed as a percentage, is given below the gel of Figure 1A. To correct for transcript accumulation due to pausing, we subtracted the processing efficiency at 60 minutes of α-amanitin and 3'dATP incubation (Fig.1A, lane 2) from the 0 minute timepoint (Fig.1A, lane 1). The difference is given as corrected processing (%). Moreover, we verified that the dsxSV40L transcripts are accurately cleaved via an RNase Protection Assay (see Supplementary Figure 1).

Recruitment of processing factors to the TEC and gradual extrusion of the poly(A) signal is not important for coupled processing

Next, we began to address the mechanism of this coupling activity by first asking whether the coupling occurs via the recruitment model. If the basis for the coupling activity is the recruitment of processing factors to the promoter followed by subsequent transfer onto the transcribing polymerase, then the removal of those recruited factors prior to the extrusion of the poly(A) signal should abolish processing. To do this, we generated transcription elongation complexes (TECs) with a brief pulse. The TECs were then isolated by magnetic selection, washed with low-salt transcription buffer (rinse) or with a high salt, high detergent, 1M KCl, 1% Sarkosyl buffer (strip) to remove all initiation and elongation factors from the elongation complex (Adamson et al 2003). The TECs were then put back into extract and chased past the poly(A) site in the absence of radiolabelled nucleotides or Gal4-p53. Transcription was halted by α-amanitin and 3'dATP and cleaved RNA was allowed to accumulate for an hour. Figure 1B

shows that processing remained efficient despite stripping (compare lane 2 with lane 1). We conclude that the recruitment of processing factors to the promoter does not contribute to coupled 3'-end processing in our in vitro system.

Transcription elongation rates have been known to influence choice of processing sites by affecting assembly of RNA-protein complexes in alternative splicing (de la Mata et al. 2003; Howe et al. 2003). Since transcription is directional, assembly of RNA-processing factor complexes may be dictated by the order of appearance of the elements in the poly(A) signal. Therefore, we wondered whether slow extrusion of the poly(A) signal promotes important RNA-protein interactions necessary for efficient processing. We reasoned that if cotranscriptional assembly of processing factors during gradual extrusion of the poly(A) signal is important then stripping the TEC of those interactions would eliminate coupling. To do so, TECs were generated with a pulse and chased past the poly(A) site before subjecting them to a strip (1M KCl, 1% Sarkosyl wash). These elongation complexes with the poly(A) signal already fully transcribed (devoid of any processing factors) were plunged directly back into extract with α-amanitin and 3'dATP already present in the processing reaction to prevent any elongation. Figure 1C shows that the processing efficiencies are almost identical despite stripping the TECs and removing any cotranscriptionally assembled cleavage and polyadenylation factors. Based on this result, cotranscriptional assembly of processing factors during extrusion of the poly(A) signal is not important for coupling 3'-end processing to transcription in our in vitro system. Interestingly, this result also indicates that pre-recruitment of processing factors to the TEC prior to extrusion of the poly(A) signal is not required to undergo coupled processing.

History of CTD phosphorylation is unimportant for coupling.

CTD phosphorylation is an orchestrated event that begins with serine 5 and 7 phosphorylation by CDK7/TFIIH during transcription initiation and serine 2 phosphorylation by CDK9/PTEFb during transcription elongation (Buratowski et al. 2009; Egloff et al. 2008; Akhtar et al. 2009). What is currently not known is whether this orchestration of CTD phosphorylation as directed by the promoter is required for efficient 3'-end processing. Therefore we investigated the importance of the history of CTD phosphorylation in a phosphatase experiment. We formed early TECs with a pulse, stripped the TECs of all transcription factors and treated the TECs with calf intestinal phosphatase. A western blot of the phosphatase-treated TECs and the SDS gel of a parallel gamma ^{32}P-ATP labeling experiment (Fig. 2A Western and SDS gel, lanes 1 and 2) both show that the phosphates put on the polymerase CTD during initiation are completely removed following the phosphatase treatment. Despite that, the TECs can still process efficiently after re-addition of nuclear extract and allowing the elongation complexes to rephosphorylate and elongate past the poly(A) signal (Fig. 2A RNA gel, lanes 1 and 2). Figure 2A, lanes 3 and 4 shows the same experiment carried out with a different extract. Under these conditions, processing efficiency was modestly reduced by phosphatase treatment, presumably because of incomplete rephosphorylation. Indeed, processing was completely restored when the phosphatase-treated TECs were rephosphorylated in extract for 15 minutes with 3'dATP prior to elongation past the poly(A) signal (Fig. 2A, lanes 3 and 6). This effect cannot be explained simply as a result of excess CTD phosphorylation due to 3'dATP because mock phosphatase-treated TECs do not process more efficiently after being subjected to the same 3'dATP treatment (compare Fig. 2A, lanes 3 and 5). Therefore, it is the rephosphorylation of the CTD that had restored coupling,

not simply *more* phosphorylation. Interestingly, the fact that processing can be restored with an extended incubation in extract even despite having all of the initial phosphates removed from the CTD argues that CTD phosphorylation that occurs at the promoter is not important. Despite the history of CTD phosphorylation being unimportant, these results are consistent with the observation that CTD phosphorylation is essential for coupling.

It is peculiar that extended treatment of phosphatase-treated TECs in extract with 3'dATP was needed to restore efficient processing despite the fact that the TECs have plenty of opportunity for rephosphorylation during the chase and an hour long incubation in 3'dATP and α-amanitin (Figure 2A experimental timeline diagram). One possibility is that extensive phosphorylation of the CTD has to occur before the poly(A) signal is extruded from the polymerase in order for cleavage to take place. To test this hypothesis, TECs were formed with a pulse and chased past the poly(A) signal. The TECs were then treated with phosphatase before they were returned to an equivalent reaction mixture with nuclear extract for processing. Figure 2B shows that processing was reduced dramatically when TECs with the poly(A) signal already fully extruded were treated with phosphatase (compare lanes 2 and 4). If the observed reduction in processing efficiency is simply because the phosphatase-treated TECs have not had enough time to be rephosphorylated, then extending the incubation period in 3'dATP for an additional 15 minutes (based on the length of 3'dATP incubation in experiment shown in Figure 2A, lane 6) should restore processing to the same extent as mock-treated TECs. The comparison of Figure 2B, lanes 4 and 1 shows that phosphatase-treated TECs do not process as well as mock-treated TECs even though the phosphatase-treated TECs were incubated for 15 minutes longer in 3'dATP. Therefore, these results show that TECs that are dephosphorylated after crossing the

poly(A) signal cannot process even though the TECs are presumably rephosphorylated during the hour-long incubation period (see Figure 2B, experimental flow chart diagram). Apparently, the phosphates put on afterwards are too late. This result not only further emphasizes the importance of CTD phosphorylation in coupling but also suggests that the TECs have to be sufficiently phosphorylated prior to the extrusion of the poly(A) signal in order for cleavage to occur. Remarkably, this finding is consistent with a quality control checkpoint early in the 3'-end processing pathway, which evaluates the phosphorylation status of the CTD as a consideration in deciding the fate of the transcript (see discussion).

Since CTD phosphorylation is crucial for coupling, we wanted to investigate whether Pin1, a CTD-modifying prolyl isomerase and a regulator of phosphorylation signaling, is important for coupling. We depleted Pin1 from our extracts using a commercial rabbit α-Pin1 antibody and carried out in vitro coupled processing assay on a dsxSV40L plasmid template. Figure 2C RNA gel, lane 2 show that processing remains efficient despite extensive depletion of Pin1 (compare Fig. 2C western, lanes 1 and 2). We conclude that Pin1 is not required for coupled processing.

A promoter is not required for basic coupling

Despite evidence in the literature suggesting a role for the promoter in mRNA processing, we eliminated two possible roles for the promoter in coupling 3'-end processing in our in vitro system: pre-recruitment of processing factors to the TEC, and directing CTD phosphorylation. This begged the question: is the promoter necessary for coupling in our in vitro system? To address this, we designed a promoterless template. It is known that purified RNA polymerase II (devoid of transcription initiation factors) is capable of initiating on a 3' extended single strand tail protruding from the duplex DNA. Therefore, we constructed a dsxSV40L template with a 3'

oligo(dT) tail (oligo-dTtaildsx) by first PCRing up the dsxSV40L from the +4 transcription start site down to 464nt past the poly(A) cleavage site (Fig. 3A). The first three bases of the sense strand were replaced with GGG, as previously studies by Chamberlin's group have shown that this substitution reduces the tendency of formation of extended RNA:DNA hybrids (Dedrick et al. 1985). The reverse primer was biotinylated to allow immobilization onto streptavidin-coated magnetic beads. Following attachment, the dT tail was added via a terminal transferase reaction. The presence of the bead on the downstream end of the template forces the oligo dT tail to be added only on the upstream end, presumably because of steric interference. The oligo(dT) tail was chosen over an oligo(dC) tail because dC tails are able to hybridize to and promote RNaseH-cutting of the GU-rich region of SV40 late poly(A) signal, which interferes with processing.

A common problem with transcription on oligo tailed templates is the formation of RNA:DNA hybrids (Kadesch et al. 1982; Dedrick et al. 1985). To confirm that the RNA generated from the oligo-dTtaildsx template is single-stranded, we initiated transcription on oligo-dTtaildsx templates using purified RNA Pol II with a pulse and a chase, and treated the resultant TECs with various RNases including RNaseH and RNase A/T1 (Supplementary Figure 2). Results show that the RNA transcripts generated from this template are extremely sensitive to RNase A/T1 cocktail but not to RNaseH unless a DNA oligo complementary to the RNA transcript is added. We conclude that the RNA transcripts generated from the oligo-dTtaildsx template are primarily single-stranded. To confirm that only RNA Pol II is transcribing in this system, we carried out an α-amanitin titration experiment. It is known that only RNA pol II transcription is inhibited at 0.1 μg/mL of α-amanitin (Weinmann et al. 1974). Figure S3, lanes 1 and 2 show

that transcription is completely inhibited at 0.1 μg/mL of α-amanitin. Therefore, we conclude that the complexes that elongate past the poly(A) site are exclusively RNA polymerase II.

To see whether coupling occurs in the absence of a promoter, we carried out in vitro transcription-processing assay on oligo-dTtaildsx templates. The template was incubated with purified RNA polymerase II briefly and transcription was initiated with a pulse with no ATP. It has been previously shown that transcription elongation complexes initiated on tailed templates pause strongly between 13-17 nt (Sluder et al. 1988). We found that withholding ATP exacerbated this effect, allowing us to form early elongation complexes with uniform transcript length (data not shown). These paused complexes were then stripped with 1M KCl and 1% Sarkosyl to remove unbound proteins, resuspended in NE, chased past the poly(A) site and transcription was halted by α-amanitin and 3'dATP. Figure 3B, lane 1 shows that processing is still efficient despite having initiated on a template without the assistance of a promoter or transcription initiation factors.

To assess whether processing on oligo-dTtaildsx templates is coupled, we employed a tether cutting assay (Rigo et al. 2005). We have previously shown that the integrity of the RNA tether is vital for coupling since severing the tether by oligo-directed RNaseH cutting blocked 3'-end processing that is coupled to transcription but not uncoupled processing (Rigo et al. 2005, Figure 3B-C). We therefore reasoned that if processing is coupled on promoterless oligo-dTtaildsx templates, then severing the RNA tether will inhibit processing. TECs with 13-17 nt-long transcripts were generated with a brief pulse without ATP, stripped, put back in extract and chased in the presence of a short DNA oligo that is complementary to sequences ~79nt downstream of the SV40 late poly(A) signal. The resultant hybrid formation by the oligo leads

to cutting by RNaseH endogenous to the extract. Moreover, addition of the oligo with the chase enables RNaseH cutting to take place as soon as the tether is extruded from the polymerase. Following the chase, transcription was then halted by the addition of α-amanitin and 3'dATP and processed transcripts were allowed to accumulate for an hour. Figure 3B lane 3 shows that severing the tether reduces processing dramatically compared to a sample in which an irrelevant oligo is added (Fig. 3B, lane 2). We conclude that initiation at a promoter is not required for coupling.

Because the results were so striking, we devised an alternative method to verify coupling by comparing the processing efficiency of transcripts released from elongation complexes with identical transcripts that remained within elongation complexes. In one sample, chased TECs initiated on an oligo-dTtaildsx were directly subjected to processing conditions (Fig. 3C. lane 1). In another sample, the chased TECs were first treated with chymotrypsin to remove the CTD, and residual chymotrypsin was removed by stripping the TECs. The transcripts were then released in buffer by cutting the tether with recombinant RNaseH and a DNA oligo. We found that removing the CTD prior to RNaseH digestion of the transcripts promoted release of transcripts (Kaneko et al. 2005; data not shown). Following an hour of digestion, the beads were removed and the released transcripts were added back to nuclear extract and subjected to identical processing conditions. Figure 3C, lane 2 shows that cleavage of the free RNA is inefficient. To confirm that these RNA transcripts were not damaged in some way (for example from trace amounts of chymotrypsin that remained after washing), an identical reaction was carried out and the isolated RNA was put back into identical processing conditions but with the addition of polyvinyl alcohol (PVA). We have previously shown that PVA can stimulate 3'-end

processing of transcripts even when it is uncoupled from transcription (Rigo et al. 2005, Figure 2E). Comparison of correcting processing (%) of Figure 3C, lane 3 and lane 1 shows that in the presence of PVA, cleavage of the free RNA is restored. Taken together, these results demonstrate that the transcription elongation complex can direct coupling of 3'-end processing to transcription without assistance of events at the promoter.

To rule out the possibility that RNaseH-released transcripts processed inefficiently because they were generated from a promoterless template, we carried out a similar experiment using the dsxSV40L template (Figure 4A). Figure 4A shows a time course processing experiment in which intact TECs gave rise to efficient processing (lanes 1-6) while the RNaseH-released transcripts were inefficiently processed even though they were incubated under identical processing conditions (lanes 7-12). We conclude that coupling requires an intact TEC.

The association of the RNA transcript with RNA polymerase II in the TEC promotes coupling

Despite having tested prominent models in the field regarding the mechanism of coupling 3'-end processing to transcription, the essential question still remains unanswered: what is the nature of coupling? The results of Figure 3B, 4A, and 2B suggest a model in which the close proximity between the transcribing polymerase with a phosphorylated CTD and the poly(A) signal promotes important contacts made between the TEC and the processing factors, which allows for better recruitment of processing factors or allosteric activation. This model requires that the integrity of the TEC remain preserved until the coupling event has taken place. To evaluate the idea that the association of the RNA to the polymerase is crucial for coupling, we generated TECs, severed the RNA tether via endogenous RNaseH and asked if the released RNAs can process as efficiently as the TEC-associated transcripts. To provide the released RNAs

with the greatest chance to process efficiently, the RNAs are released gently via endogenous RNaseH-directed cutting under processing conditions and not in buffer with recombinant RNaseH and secondly, the TECs are given opportunity for partial assembly of the cleavage and polyadenylation apparatus (CPA) prior to cutting the tether. Figure 4B top diagram shows the outline of the experiment. TECs were formed with a pulse and a chase on immobilized dsxSV40L. The TECs were then stripped and put back in extract briefly to allow opportunity for partial assembly of the cleavage and polyadenylation apparatus before the addition of the tether-directed DNA oligo. After allowing 5 minutes for tether-cutting, the beads were separated from the supernatant and both fractions were incubated under identical processing conditions. We found that the transcripts that remain associated with the TECs processed efficiently (Fig. 4B, lanes 1 and 2) while those that were released into the supernatant do not (Fig. 4B, lanes 3 and 4). To ensure that the released transcripts were not defective, we generated released transcripts in an identical manner and put them back under identical processing conditions in the presence of PVA. Figure 4B, lanes 5 and 6 show that in the presence of PVA, the released RNA can process. The results of Figure 4B and Figure 3B, taken together, demonstrate that the association of the RNA transcript with the polymerase is crucial for coupling.

We noticed that the majority of the tether-cut RNAs remain associated with the beaded template after separation of the supernatant from the beads (compare Fig. 4B, lanes 1 and 3). We were curious whether these transcripts were functionally distinct from those that were released. For example, do these tether-cut RNAs have some special property that may allow them to engage in efficient 3'-end processing? To address this possibility, we digested off the

CTD using chymotrypsin prior to tether-cutting by endogenous RNaseH to release the majority of the tether-cut RNA into the supernatant and then asked whether these released RNAs can process efficiently. Note that residual chymotrypsin was again removed by washing the TECs prior to resuspension in nuclear extract under processing conditions as in Figure 3C. The results show that even after releasing nearly 80% of the tether-cut RNA, these RNA transcripts cannot process without addition of a crowding agent such as PVA (Supplementary Figure S4). These results (Fig. 4B and Fig S4) remain consistent with the notion that coupling requires the RNA transcript to remain associated with the elongation complex.

Discussion

Using an in vitro system that can support 3'-end processing that is functionally coupled to transcription (Rigo et al. 2005), we have demonstrated that coupling does not require pre-recruitment of processing factors to the TEC (Figure 1B), gradual extrusion of the poly(A) signal (Figure 1C), or orchestration of CTD phosphorylation beginning with serine 5 phosphorylation by TFIIH at the promoter (Figure 2A, 2B). Even when we had bypassed the events at the promoter altogether, by loading the TECs onto the DNA template via a single stranded oligo-dT tail, the TECs were still able to process an order of magnitude more efficiently than free RNAs (Figure 3C). Apparently, the transcription elongation complex can direct the necessary modifications (such as CTD phosphorylation and capping) required for coupled 3'-end processing of the RNA transcript without any assistance from the promoter.

Taken together, these results suggest a model in which coupling at the fundamental level is mediated by having the poly(A) signal held close to the polymerase CTD in the context of a TEC. The close proximity of the poly(A) signal and the polymerase stabilizes and directs efficient recruitment of the cleavage and polyadenylation apparatus and/or activation of the processing apparatus. This model is supported by results which show that the association of the transcript with the elongating polymerase is important, since disruption of this association by severing the tether via RNaseH-mediated cutting inhibits coupling (Figure 3B, 3C, 4A, 4B, and Rigo et al. 2005). We were able to recently confirm that it is the act of severing the tether that disrupts coupling and not simply due to antisense-oligo binding to the tether per se because equal-molar addition of 2'-o'methoxyethyl (MOE)-modified oligonucleotides directed to bind at the

same location of the RNA tether (but cannot direct RNaseH to cut) did not have any effect on 3'-end processing (unpublished results).

Moreover, we have demonstrated directly by treatment of stripped TECs with calf intestinal phosphatase (CIP) that CTD phosphorylation is critical for efficient processing (see Figure 2A and 2B). Our approach avoids the use of kinase inhibitors such as DRB, H8, or flavopiridol which may indirectly affect coupling by targeting other classes of kinases that phosphorylate components of the 3'-end processing machinery such as CFIm or CFIIm (Ryan et al. 2007). The importance of CTD phosphorylation in processing is consistent with the proposed model (above) that the CTD in close proximity to the poly(A) signal (by virtue of the TEC), participates in either assembly of the CPA or activation.

Interestingly, there seems to be a strict requirement that the CTD has to be sufficiently phosphorylated before crossing the poly(A) site in order for cleavage to occur (Figure 2B). When phosphatase-treated TECs were granted ample time to rephosphorylate prior to crossing the poly(A) site, processing was rescued (Fig. 2A, lanes 3-6). But if the TECs are dephosphorylated after crossing the poly(A) site, the transcripts were inefficiently processed even though the TECs have an hour of opportunity to rephosphorylate in extract (Figure 2B). This stringent requirement for adequate CTD phosphorylation to be in place before extrusion of the poly(A) signal is consistent with a previously mentioned checkpoint model (Orozco et al. 2002; Nag et al. 2007). Previously, we found that almost immediately upon crossing the poly(A) signal, the polymerase is directed to pause and we have suggested that this pause serves as an important intersection point for surveillance activity to determine the fate of the transcript (Orozco et al. 2002; Nag et al. 2007). Based on the CIP studies, we believe that proper CTD

phosphorylation may serve as one of the many inputs in deciding whether the TEC should proceed with efficient 3'-end processing or to abort the process for degradation of the transcript, perhaps via the poly(A)-dependent degradation pathway (Chapter 2 or Kazerouninia et al. 2010). The CTD is an obvious participant in this checkpoint pathway because phosphorylation of the CTD is thought to be a dynamic, reversible process that changes throughout all stages of the transcription cycle to assist in transcription elongation as well as capping, splicing, and 3'-end processing of transcripts (Egloff et al. 2008).

How does CTD phosphorylation act as an input in the checkpoint pathway to decide the fate of the transcript? We envision that immediately following the extrusion of the poly(A) signal, assembly of the CPA begins and the CTD participates in this process. If the CTD is not properly phosphorylated by the time assembly begins, the recruitment of the CPA occurs too slowly or activation does not occur and dead-end complexes form. Once dead-end complexes form, the CTD has lost its opportunity to influence the fate of the transcript, even if the CTD becomes rephosphorylated at a later point. Because CPA assembly is incomplete or inactive, the transcripts are directed to a degradation pathway and the polymerase becomes terminated and recycled. In support for the role of CTD in assembly of the CPA, Pcf11 is a processing factor that is known to bind more efficiently to a phosphorylated CTD in yeast (Licatalosi et al. 2002; Meinhart et al. 2004). There has also been a demonstration that Yhh1p, yeast homologue of 160-kDa subunit of CPSF, also binds to phospho-CTD (Dichtl et al. 2002). However, in vitro studies have not revealed an effect of CTD phosphorylation on binding of 3'-end processing factors CPSF or CstF in humans (Fong et al. 2001; McCracken et al. 1997). It remains a possibility that these in vitro studies may not have revealed an effect because of the

requirement for a phosphorylated CTD in the context of a transcribing elongating polymerase. Another possible, non-mutually exclusive role for CTD phosphorylation is that it can serve to allosterically activate the CPA. It has been shown that a phosphorylated CTD can stimulate uncoupled 3'-end processing in vitro up to 5-fold (Hirose et al. 1998).

Interestingly, although CTD phosphorylation is necessary for efficient 3'-end processing, we found that Pin1, a regulator of CTD phosphorylation signaling does not appear to play an important role in 3'-end processing in general since depletion of >90% of Pin1 had no effect (Figure 2C). Although we think it is unlikely, it's possible that the small pool of Pin1 that survived depletion is sufficient to direct efficient processing. It is also possible that other prolyl isomerase enzymes may substitute for Pin1 although Pin1 is the only one that is phosphorylation-dependent, with unique substrate specificity for peptides with phospho-serine and phospho-threonine residues preceding the proline (Yaffe et al. 1997).

Significantly, our results demonstrate that coupling does not require pre-recruitment of processing factors to the TEC (Figure 2A and 2B), which is consistent with an earlier report (Adamson et al. 2005). This is surprising to us given the number of reports in the literature suggesting that processing factors are found at the promoter both in vivo and in vitro (Dantonel et al. 1997; Glover-cutter et al. 2008; Rosonina et al. 2005). There have also been reports suggesting that these processing factors were found riding with the elongating polymerase (Dantonel et al. 1997; Rozenblatt-Rosen et al. 2009; Martincic et al. 2009). A possible explanation is that pre-recruitment of processing factors may be more important for weak poly(A) signals. On a strong poly(A) signal, the cis-acting elements, together with the CTD, are robust enough to recruit processing factors in a timely manner. Conversely, weak poly(A)

signals are weak because the elements are suboptimal and therefore assembly of CPA occurs slowly (Chao et al. 1999), making it susceptible for negative factors to inhibit processing by competing with processing factors for sites on the poly(A) signal. Consistent with the possibility that usage of weaker poly(A) sites is regulated by recruitment of processing factors to the TEC, overexpression of CstF in vivo results in the preferred use of proximal, weaker μs poly(A) signal over the distal but stronger μm poly(A) signal (Takagaki et al. 1996) and it was later proposed that this may be due to CstF being loaded onto the elongating polymerase by the elongating factor ELL2 (Martincic et al. 2009). It is therefore conceivable that preloading of processing factors to the elongating complex to increase the local concentration of processing factors at the processing site may be more relevant for weaker poly(A) signals.

In terms of the role of transcription elongation in coupling, we found that there was little to no effect for co-transcriptional assembly of cleavage and polyadenylation apparatus on coupling (Figure 1C). Apparently, the RNA-protein complexes that form during extrusion of the poly(A) signal can be reassembled efficiently even when the entire poly(A) signal is pre-transcribed and plunged back into extract, as long as it is part of a TEC (Figure 1C, lanes 3-4). Based on our understanding that the SV40 late poly(A) signal is strong because it can direct efficient assembly of the CPA on a TEC, it's possible that the importance of co-transcriptional assembly of the CPA may be masked for strong poly(A) signals. It will be interesting therefore to see whether the result may be different for a weak poly(A) signal. Given that transcription is directional, it's plausible that sequential assembly beginning with CPSF to the hexamer acts to promote recruitment of CstF to the G/U-rich region, which leads to efficient processing by reducing the

likelihood for negative factors to compete with CstF or other downstream processing factors for their binding sites such as hnRNP F in regulating μs poly(A) usage (Veraldi et al. 2001).

On another note, the possibility that elongation rates can affect 3'-end processing efficiency by affecting RNA folding remains an interesting avenue of investigation that we have not yet addressed. There has been demonstration that elongation rates can affect time-sensitive RNA folding in RNA genes (Lubkowska et al. 2011; Lewicki et al. 1993; Chao et al. 1995). This may be important since there have also been suggestions that the SV40 late poly(A) signal folds into a functional secondary structure in order to direct efficient processing (Hans et al. 2000; Wu et al. 2004). Therefore, future studies on weak poly(A) signals may help us better understand the significance of various mechanisms of coupling and their relative importance for weak and strong poly(A) signals.

Materials and Methods

Oligos used:

1: 5'-GTTGGACTCAAGACGATAGTTACC-3'

2: 5'-biotin-CACATTTCCCCGAAAAGTGCCACC-3'

3: 5'-GGGTCTGCGGAGAGGCTGGCAGATTG-3'

4: 5'-CCAAGCTACCGAGCTCTTTTTTGGTACCCCTTGGGAGC-3'

5: 5'-GCTCCCAAGGGGTACCAAAAAAGAGCTCGGTAGCTTGG-3'

6: 5'-CCACTCATGTATGTGATTTTTTAGGACGCAGACGCCAAC-3'

7: 5'-GTTGGCGTCTGCGTCCTAAAAAATCACATACATGAGTGG-3'

79.ds oligo: 5'-GGTGACACTATAGAACTCG-3'

Control oligo: 5'-CCAATACGCAAACCGCCTCTCC-3'

Immobilized templates

DsxSV40L immobilized templates (1.7 kb) were generated by PCRing up the Gal5-HIV2dsxΔInt(+ESE) template (a gift from Rosonina et al. 2003) from -781 bp upstream of the transcription start site down 464 bp past the SV40 Late poly(A) cleavage site using oligos number 1 and 2 as primers. Oligo number 2 is a reverse biotinylated primer to allow for bead attachment on the downstream end of the template. The PCR product of correct length was gel-extracted and purified using the QIAquick Gel Extraction Kit. Binding of the template to magnetic Dynal M-280 Streptavidin beads (Invitrogen) was carried out according to manufacturer's protocol. The final concentration of DNA on streptavidin beads was 0.8-3.1 pmol of DNA template per mg of beads. We found that a DNA density greater than 3.1 pmol of 1.7 kb-long DNA per mg of beads gave reduced processing efficiency in a coupled processing

assay, presumably due to steric interference. Reducing DNA density on the beads to less than 3.1 pmol of DNA per mg of beads did not have any further effect on processing efficiency but did increase transcription.

Promoterless oligo-dTtaildsx template was synthesized by PCRing up the dsxSV40L template from the +4 position and replacing the first three basepairs with GGG using the forward primer (oligo number 3) and a biotinylated reverse primer (oligo number 2) to amplify up to 464 bp downstream of the SV40 late cleavage site. The templates are then bound to Dynal M-280 Streptavidin beads according to manufacturer's protocol. Final concentration of DNA concentration on beads used was 3.1 pmol of DNA template per mg of beads. The oligo-dT tail was then added by treatment with 40 units of terminal transferase (New England Biolabs) and 7.5 μM TTP for 30 minutes at 37^{o}C. The presence of the magnetic beads on the downstream end of the template forced the oligo dT tail to be added only to the upstream end of the template. To discourage RNA pol III transcription, two patches of 6 consecutive thymidine residues (one present beginning at position 174 bp and another at 218 bp downstream of the transcription start site) were inserted by 2 rounds of site-directed mutagenesis (Strategene) using oligos 4 and 5 as primers for position 174 and oligos 6 and 7 as primers for position 218. It has been reported that RNA polymerase III terminates at sites containing a stretch of 4-6 thymidine residues (Bogenhagen et al. 1981; Cozzarelli et al. 1983). Immobilized beaded templates were stored at 4^{o}C for no more than a month.

Purification of Gal4-p53 activator

G4-p53 construct (gift from James Goodrich) was expressed in the E. coli BL21 strain. T7 promoter expression was induced in these cells with 0.4 mM IPTG for 2 hours. Cells were

collected, resuspended in TGED buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 20% glycerol, 0.1 mM EDTA, 1 mM DTT, 1mM PMSF) containing 5 mM $MgCl_2$, and lysed via sonication. Cell lysates were incubated with ANTI-FLAG M2 affinity gel (Sigma, A2220) for 3 hours at 4^{o}C. The affinity gel was washed with TGED buffer and TGED buffer containing 1M NaCl. The affinity gel was transferred to a Poly-Prep Chromatography Column (Biorad) and Gal4-p53 proteins were released with TGED buffer containing 0.2 mg/mL FLAG peptide at room temperature. The fractions were collected and immediately stored at -80^{o}C. Peak fractions containing Gal4-p53, as determined by SDS-PAGE and sypro ruby staining, were later pooled together, re-aliquoted and stored at -80^{o}C.

Purification of RNA Polymerase II from HeLa nuclear extract

8WG16 ascites fluid (Covance) was purified using Protein-A Agarose Beads (Santa Cruz Biotechnology, sc-2001) as described (Thompson et al. 1990). Purified 8WG16 antibodies were conjugated to CNBr-activated Sepharose 4B beads (GE Healthcare Life Sciences) as described (Thompson et al. 1990), except a final concentration of 2.2 mg of 8WG16 antibodies per mL of beads was used for antibody conjugation and 0.1 M Tris-HCl buffer pH 8.0 was used to block excess unreacted groups on the beads. Nuclear extract was incubated with 8WG16 sepharose beads at a ratio of 3.5 mL of nuclear extract per mL of beads for 3 hours. The beads were washed extensively with 150 bed volumes of TE buffer (50 mM Tris-HCl pH 7.9 and 0.1 mM EDTA) containing 200 mM ammonium sulfate. Polymerase II was eluted with 4 bed volumes of EB buffer containing 0.75 M ammonium sulfate and 40% ethylene glycol at room temperature and dialyzed for 5-6 hours at 4^{o}C against TE buffer containing 1 mM DTT, 150 mM ammonium sulfate and 50% glycerol. Purified RNA Polymerase II was stored in liquid nitrogen.

In vitro transcription (Coupled processing assay)

Hela nuclear extract was prepared as described (Tran et al. 2001).

A typical pulse-chase assay began with 3-9 µL of nuclear extract that was mixed with Gal4-p53, anti-RNase (Ambion), DTT, $MgCl_2$, sodium citrate, DNA plasmid or beaded DNA template, and water up to 8.7 µL. Volume of Gal4-p53 addition was selected based on a titration experiment determining the minimum volume required to achieve maximum transcription in a typical coupled processing experiment. The mixture was preincubated at 30^{o}C for 30 min and then pulsed with 3 µL containing 20 µCi of [α-^{32}P] CTP (Perkin Elmer, 800 Ci/mmol), nucleotide triphosphates and creatine phosphate. Then, 2.6 µL of chase mix was added containing a high concentration of nonradiolabeled CTP. Next, 1 µL of α-amanitin and 3'dATP were added. Final concentrations in a standard pulse-chase assay were as follows: 8.7% glycerol, 8.5 mM HEPES (pH 7.9), 42 mM KCl, 86 µM EDTA, 1.3 mM DTT, 58 µM PMSF, 10 U Anti-RNase, 2.9 mM $MgCl_2$, 2.3 mM Citrate (pH 6.7), 0.16 pmol DNA template, 230 µM ATP, 117 µM UTP and GTP, 12 mM creatine phosphate, and 1.2 mM CTP, 23 ng/µL α-amanitin and 250 µM 3'dATP. PVA, when used, was added with α-amanitin and 3'dATP at a final concentration of 2.1%. Oligos, when used, was added with the chase at a final concentration of 9 ng/µL.

Following α-amanitin and 3'dATP incubation, the reaction was terminated by the addition of a stop solution: 65 µL of 10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS, and 100 µg proteinase K (Ambion). RNA was extracted with 350 µL TRIzol (Invitrogen) and 70 µL chloroform and then precipitated with 4 µL of 5 mg/mL glycogen (Ambion) and 350 µL isopropanol (30 min, room temperature), and finally ran on a 5% polyacrylamide gel. After electrophoresis, results were

recorded and analyzed by using a PhosphorImager with ImageQuant software (Molecular Dynamics).

When TECs were washed, beads were collected on magnet (Dynal MPC-S) for approximately 30-60 seconds and washed with either 200 µL of high salt, high detergent strip buffer (20 mM HEPES pH 7.9, 1 M KCl, 1% Sarkosyl) or 200 µL of low-salt transcription buffer (3.9 mM HEPES, 20 mM KCl, 39 µM EDTA, 3.9% glycerol, 0.1 mM DTT, 20 µM PMSF). Depending on the experiment, the TECs were then washed with either CIP buffer for phosphatase treatment (see CIP treatment), chymotrypsin buffer for chymotrypsin treatment (see Chymotrypsin treatment), or transcription buffer to remove any residual contamination before returning them to extract for further transcription or processing. Washed TECs were placed back into the original reaction conditions prior to wash unless otherwise noted.

Promoterless template transcription

TECs were generated from promoterless oligo-dT tailed templates by incubating 2.4 µL of purified RNA polymerase II with anti-RNase (Ambion), DTT, $MgCl_2$, sodium citrate, beaded DNA template, and water up to 7.8 µL. Volume of RNA polymerase II was selected based on a titration experiment determining the minimum volume required to achieve maximum transcription from 0.16 pmol of oligo-dTtaildsx template in a typical coupled processing experiment. The mixture was preincubated at 30^{o}C for 5 min and then pulsed with 3 µL containing 20 µCi of [α-^{32}P]-CTP and nucleotide triphosphates without ATP. These complexes were then stripped, washed and resuspended in transcription buffer with nuclear extract to proceed with chase according to steps outlined in coupled processing assay (above). The final

concentrations of every component is the same as stated under coupled processing assay, except the final concentrations of ATP, UTP, and GTP is 260 μM.

Chymotrypsin treatment

Stripped TECs were washed with and resuspended in chymotrypsin buffer (20 mM HEPES, 60 mM KCl, 7 mM $MgCl_2$, 200 μg/mL BSA). Chymotrypsin was added in the same buffer at a final concentration of 5 ng/μL. Proteolysis was allowed to occur for 2 minutes at 37^{o}C. The reaction was terminated and enzyme removed by stripping the complexes twice. Western blot analysis confirmed complete removal of CTD under these conditions (data not shown). Stock solutions of 0.5 mg/mL chymotrypsin (Sigma, C4129) was prepared in 1 mM HCl, 2 mM $CaCl_2$ and stored in -20^{o}C for no more than a month.

CIP treatment

Stripped TECs were washed with and resuspended in CIP buffer (1X NEBuffer 3). Either 2 μL of 10 unit/μL CIP enzyme (New England Biolabs) or 2 μL CIP storage buffer (10 mM Tris-HCl pH 8.2, 50 mM KCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 50% glycerol) was added in the same buffer. Phosphatase reaction was allowed to occur for an hour at 37^{o}C. The reaction was terminated and enzyme removed by stripping the TECs twice.

Recombinant RNaseH treatment

Stripped TECs were washed with transcription buffer and resuspended in a buffer containing anti-RNase (Ambion), DTT, $MgCl_2$, sodium citrate, and water up to 4.2 μL. 5 units of recombinant RNaseH (New England Biolabs) and 79.ds oligo was added to the reaction to a final volume of 5.7 μL. The final concentration of oligo was 15 ng/μL. RNaseH cutting was allowed to occur for an hour at 30^{o}C. Released transcripts were separated from the beads and added

back to processing conditions with nuclear extract (same recipe as stated under coupled processing assay).

Western Blot Analysis

To observe the phosphorylation state of TECs following CIP treatment, mock- and phosphatase-treated TECs were separated on 4% SDS-PAGE gel and transferred to a nitrocellulose membrane. The membrane was blocked with Western Blocking Reagent (Roche, cat # 11921681001) and blotted for α-CTD antibody (8WG16, Covance, 1:1,000) and visualized using the Super Signal West Pico Chemiluminescent Substrate (Thermo Scientific, cat# 34080) using secondary antibodies from Calbiochem (anti-mouse IgG), and the blot was exposed to Hyper film (Amersham). Pin1-depleted extracts were separated on a 10-20% gradient SDS-Tricine gel (Biorad), transferred to nitrocellulose membrane, and the blot was probed with rabbit α-Pin1 antibody (PC270, Oncogene, 1:1,000) and mouse α-CTD antibody (8WG16, Covance, 1:1,000).

SDS gel γ-P^{32} experiment

To evaluate the effectiveness of CIP treatment, transcription elongation complexes were generated as described in "coupled processing assay" except that the pulse consisted of 20 µCi γ-^{32}P-ATP (Perkin Elmer, 3000mCi/mmol), 2 µM cold ATP, and 230 µM each of CTP, UTP, and GTP. After the pulse, the TECs were stripped and then subjected to CIP treatment (see above). Following treatment, the beads were collected and the proteins were recovered by boiling in SDS sample buffer and separated on 4% SDS-PAGE gel. After electrophoresis, the results were recorded using a PhosphoImager with ImageQuant software.

Depletion of Pin1

Mock- and Pin1-depleted extracts were generated by subjecting the extract through two rounds of depletion. Each round of depletion involved incubation of 20 µL extract with either normal rabbit IgG antibody-bound Protein A magnetic beads or rabbit α-Pin1-bound Protein A magnetic beads for 1-2 hours. These beads were prepared by incubating either 4 µg of normal rabbit IgG (Santa Cruz Biotechnology, sc-2027) or 3.2 µg of rabbit α-Pin1 antibody (PC270, Oncogene, 1:1,000) with 0.6 mg Dynabeads Protein-A magnetic beads (Invitrogen) overnight at 4^{o}C. These beads were rinsed with PBS before incubation with nuclear extract. Following both rounds of depletion, the extract was incubated for another hour with BSA-blocked Protein A magnetic beads (generated by incubating 100 µg BSA with 0.6 mg Protein A beads overnight) to remove residual antibodies from the extract. To evaluate the extent of depletion, 5 µL of extract was loaded for western blot analysis. For coupled processing assay, 3 µL of depleted extracts were used.

Figures

Figure 1. Pre-recruitment of cleavage factors to the TEC and slow extrusion of the poly(A) signal are not required for efficient 3'-end processing dsxSV40L transcripts.

(**A**) Immobilized dsxSV40L linear template was transcribed for 4 minutes before transcription was stopped by the addition of α-amanitin and 3'dATP. Poly(A) site cleaved RNAs were allowed to accumulate for 60 min. Processed (%) is the fraction of poly(A) site cleaved RNAs over the total amount of RNA extending past the poly(A) site. Corrected processing (%) refers to the difference of RNAs processed (%) at time x of α-amanitin, 3'dATP incubation from the 0 min timepoint to correct for transcriptional pausing around the poly(A) site region.

(**B**) Transcription complexes were generated from 2-fold scaled-up reactions, collected by magnetic selection, and washed with either low-salt transcription buffer (Rinse) or with 1M KCl, 1% sarkosyl buffer (Strip). The TECs were then washed in low-salt transcription buffer to remove any residual contamination before returning them to extract, chasing them past the poly(A) signal and allowing processed RNAs to accumulate for an hour. Rinses contained 200 µg/mL BSA to facilitate bead collection on magnet. Final ATP concentration was 117 µM in this experiment.

(**C**) Transcription complexes were generated as in part B except the TECs were chased past the poly(A) signal prior to collecting them by magnetic selection and washing them with a Rinse or a Strip. The TECs were then washed again in low-salt transcription buffer, returned to extract under the same processing conditions as in part B for RNA processing. Rinses contained 200 µg/mL BSA to facilitate bead collection on magnet.

Figure 1

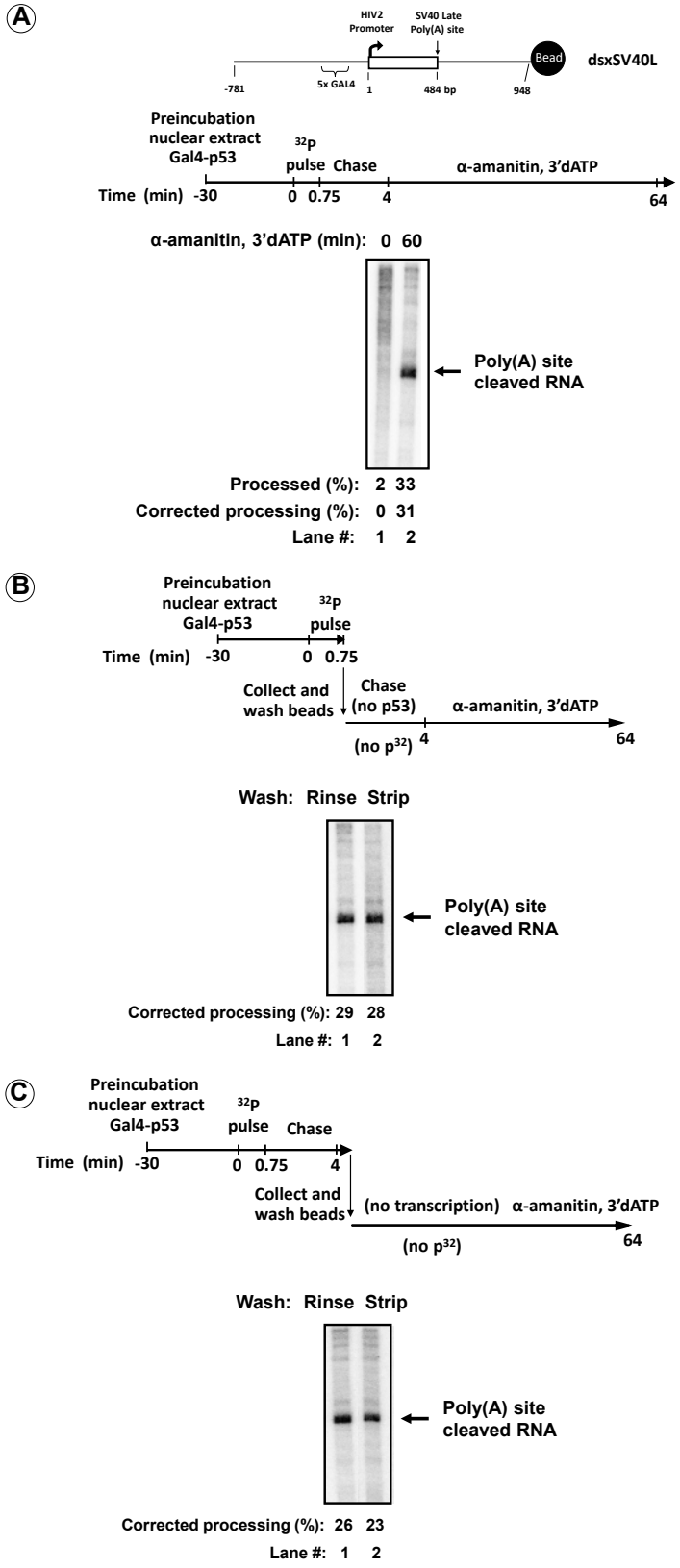

A
HIV2 Promoter
SV40 Late Poly(A) site
Bead
dsxSV40L
-781
5x GAL4
1
484 bp
948
Preincubation nuclear extract Gal4-p53
32P pulse
Chase
α-amanitin, 3'dATP
Time (min) -30
0
0.75
4
64
α-amanitin, 3'dATP (min): 0 60
Poly(A) site cleaved RNA
Processed (%): 2 33
Corrected processing (%): 0 31
Lane #: 1 2
B
Preincubation nuclear extract Gal4-p53
32P pulse
Time (min) -30
0 0.75
Collect and wash beads
Chase (no p53)
α-amanitin, 3'dATP
(no p32)
4
64
Wash: Rinse Strip
Poly(A) site cleaved RNA
Corrected processing (%): 29 28
Lane #: 1 2
C
Preincubation nuclear extract Gal4-p53
32P pulse
Chase
Time (min) -30
0 0.75
4
Collect and wash beads
(no transcription)
α-amanitin, 3'dATP
(no p32)
64
Wash: Rinse Strip
Poly(A) site cleaved RNA
Corrected processing (%): 26 23
Lane #: 1 2

Figure 2. Orchestration of CTD phosphorylation as directed by the promoter is unimportant for coupled 3-end processing.

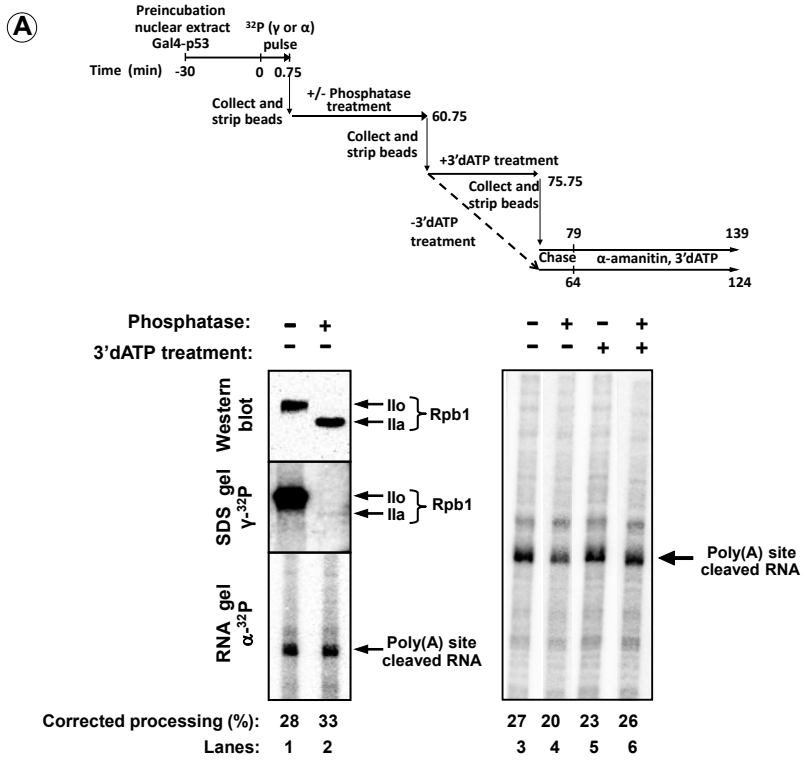
(**A**) TECs were generated from 7-fold scaled-up reactions. Phosphatase treatment of elongation complexes was carried out by washing the TECs in phosphatase buffer (1X NEBuffer 3) after stripping and then incubating the TECs in fresh phosphatase buffer with 1U/µL of calf intestinal phosphatase (New England Biolabs) for 60 min at 37^{o}C. The western blot and RNA gel in lanes 1-2 are from the same experiment in which an aliquot (5-fold reaction equivalence) was taken following phosphatase treatment to analyze the phosphorylation status of the CTD. The membrane was blotted with α-CTD antibody (8WG16, Covance, 1:1,000). The SDS gel in lanes 1-2 is from a parallel experiment in which γ-^{32}P-ATP was included in the pulse of a 5-fold scaled-up reaction to verify the effectiveness of the phosphatase treatment. The experiment shown in lanes 3-6 was carried out using a different extract. 10 mM EDTA was used to stop transcription after pulse and was eliminated in the washes prior to the chase.

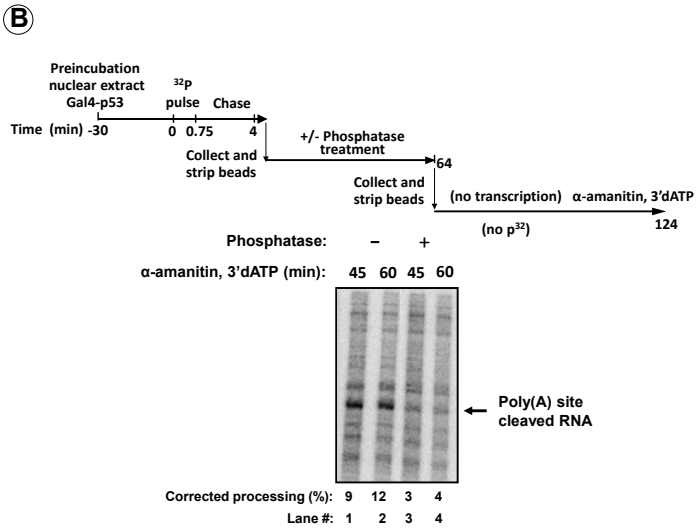
(**B**) The experiment is similar to part A except that the TECs (generated from 2-fold scaled-up reactions) were treated with phosphatase after they had crossed the poly(A) site. 10 mM EDTA was used to stop transcription after chase and was eliminated in the washes prior to resuspension and processing in fresh extract.

(**C**) Depletion of Pin1 does not inhibit coupling. Mock- and Pin1-depleted extracts were generated by subjecting the extract through two rounds of depletion. Each round of depletion involved incubation of 20 µL extract with either normal rabbit IgG antibody-bound Protein A magnetic beads (20 µL) or rabbit α-Pin1-bound Protein A magnetic beads for 1-2 hours. Following depletion, the extract was incubated for another hour with BSA-blocked Protein A

magnetic beads to remove residual antibodies from the extract. 5 µL of extract was loaded for western blot analysis. The blot was probed with rabbit α-Pin1 antibody (PC270, Oncogene, 1:1,000) and mouse α-CTD antibody (8WG16, Covance, 1:1,000).

Figure 2


A
Preincubation nuclear extract Gal4-p53
32P (γ or α) pulse
Time (min) -30 0 0.75
Collect and strip beads
+/- Phosphatase treatment
60.75
Collect and strip beads
+3'dATP treatment
75.75
Collect and strip beads
-3'dATP treatment
79
139
Chase
α-amanitin, 3'dATP
64
124
Phosphatase: - + - + - + - +
3'dATP treatment: - - - - + +
Western blot
IIo
IIa
Rpb1
SDS gel γ-32P
IIo
IIa
Rpb1
RNA gel α-32P
Poly(A) site cleaved RNA
Poly(A) site cleaved RNA
Corrected processing (%): 28 33 27 20 23 26
Lanes: 1 2 3 4 5 6


B
Preincubation nuclear extract Gal4-p53
32P pulse Chase
Time (min) -30 0 0.75 4
Collect and strip beads
+/- Phosphatase treatment
64
Collect and strip beads
(no transcription) α-amanitin, 3'dATP
(no p32)
124
Phosphatase: - +
α-amanitin, 3'dATP (min): 45 60 45 60
Poly(A) site cleaved RNA
Corrected processing (%): 9 12 3 4
Lane #: 1 2 3 4

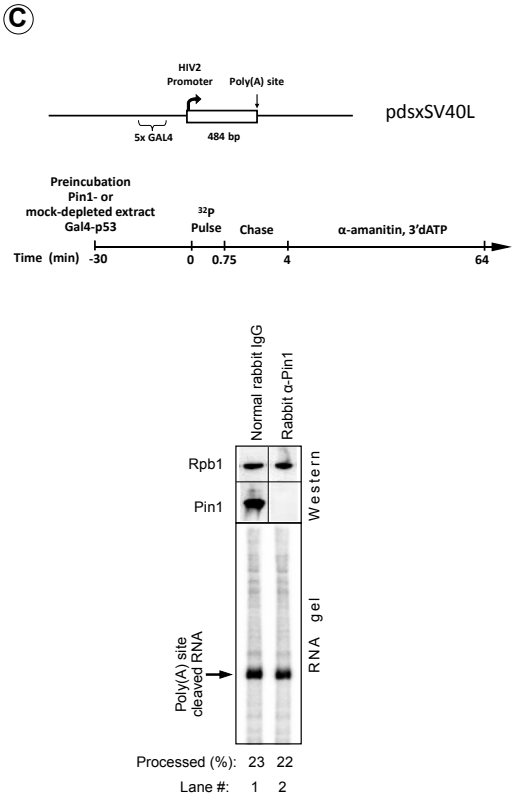

C
HIV2 Promoter
Poly(A) site
pdsxSV40L
5x GAL4
484 bp
Preincubation Pin1- or mock-depleted extract Gal4-p53
32P Pulse Chase
α-amanitin, 3'dATP
Time (min) -30 0 0.75 4 64
Normal rabbit IgG
Rabbit α-Pin1
Rpb1
Pin1
Western
Poly(A) site cleaved RNA
RNA gel
Processed (%): 23 22
Lane #: 1 2

Figure 3. The transcription elongation complex can direct coupling of 3'-end processing to transcription without assistance of events at the promoter.

(**A**) Design of the promoterless oligo-dTtaildsx template. Oligo-dTtaildsx template was synthesized by PCRing up the dsxSV40L template from the +4 position and replacing the first three basepairs with GGG using the forward primer and a biotinylated reverse primer to amplify up to 464 nt downstream of the SV40L cleavage site. The templates are then bound to streptavidin beads prior to treatment with terminal transferase and 7.5 μM TTP for 30 minutes at 37^{o}C to ensure that the oligo dT tail is added only to the upstream end of the template. To eliminate potential RNA pol III transcription, two patches of 6 consecutive thymidine residues (one present beginning at position 174 bp and another at 218 bp) were inserted by site-directed mutagenesis. It has been reported that RNA polymerase III terminates at sites containing a stretch of 4-6 thymidine residues (Bogenhagen et al. 1981; Cozzarelli et al. 1983).

(**B**) Initiation on a promoter is not required for efficient 3'-end processing. Transcription on Oligo-dTtaildsx templates was initiated with a pulse with no ATP to generate 13-17 nt long paused transcription complexes in an 8-fold scaled-up reaction. These complexes are then stripped and chased past the poly(A) site either with no oligo, ctrl oligo (non-complementary DNA oligo whose sequence is derived from 273 nt upstream of the transcription start site) or 79.ds oligo (complementary DNA oligo that directs RNaseH to cut ~79nt downsteam of the cleavage site). After the chase, TECs were processed in α-amanitin, 3'dATP for an hour. The resultant cut RNAs directed by the 79.ds oligo in lane 3 are referred to as "tether-cut RNA" on the RNA gel. The oligonucleotide name (79.ds oligo) refers to the distance from the SV40 late cleavage site to the predominant RNaseH cutting site downstream (Wu et al. 1999).

(**C**) Processing of free RNA is inefficient compared to processing of RNA transcripts in elongation complexes. Chased TECs were formed from 8-fold scaled-up reactions, incubated in 5 ng/µL chymotrypsin for 2 minutes at 30^{o}C to digest off the CTD, washed, and incubated in 5 units of recombinant RNaseH (New England Biolabs) with 79.ds oligo (15 ng/µL) in transcription buffer for an hour at 30^{o}C to release the RNA transcripts. The beads were then removed and the released RNAs were processed under typical processing conditions with or without polyvinyl alcohol (PVA). In another sample, chased TECs were formed and then directly subjected to similar processing conditions (with the addition of recombinant RNaseH and a ctrl oligo).

Figure 3

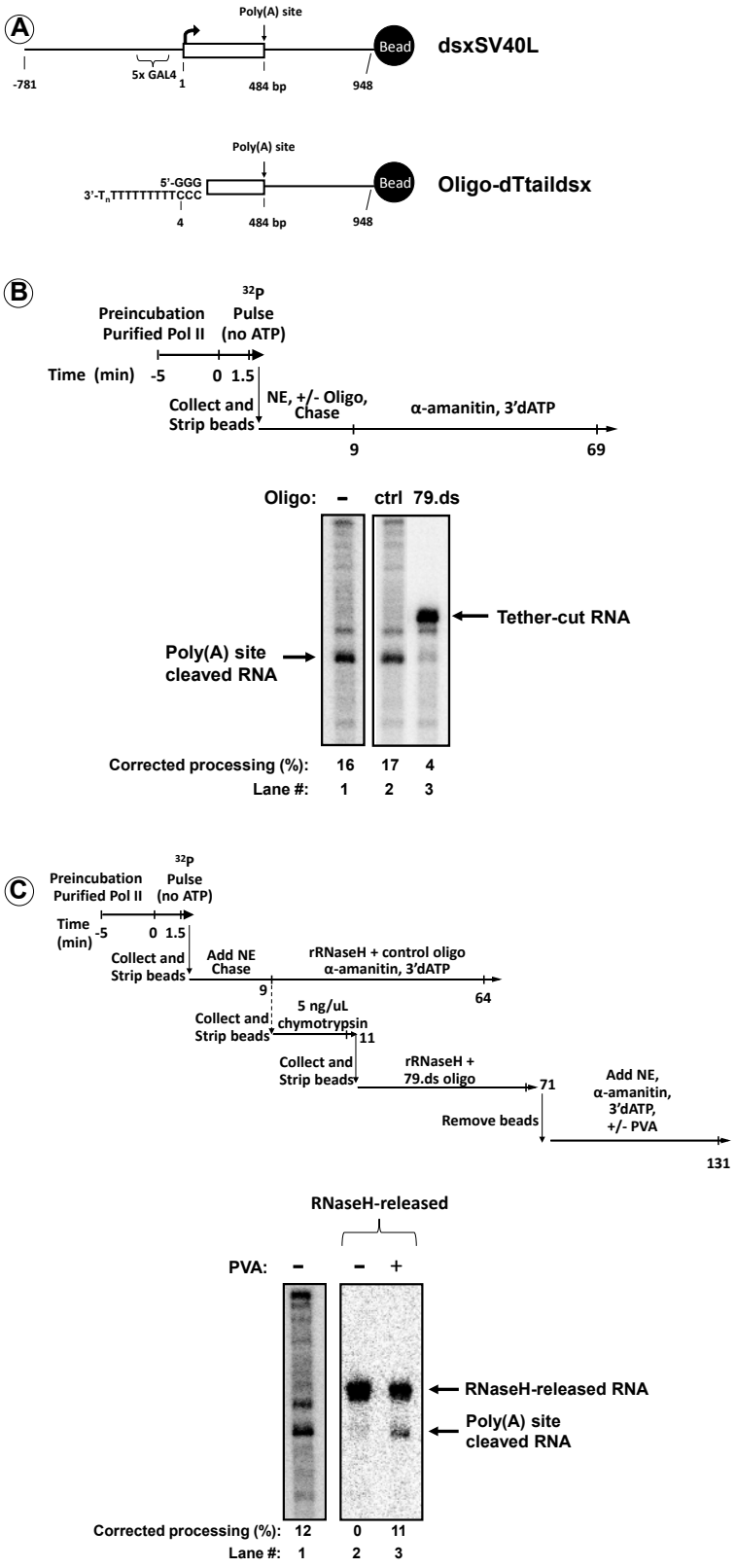

A
Poly(A) site
Bead
dsxSV40L
-781
5x GAL4
1
484 bp
948
Poly(A) site
5'-GGG
3'-T$_n$TTTTTTTTTCCC
Bead
Oligo-dTtaildsx
4
484 bp
948
B
^{32}p
Preincubation
Purified Pol II
Pulse
(no ATP)
Time (min) -5 0 1.5
Collect and
Strip beads
NE, +/- Oligo,
Chase
α-amanitin, 3'dATP
9
69
Oligo: – ctrl 79.ds
Tether-cut RNA
Poly(A) site
cleaved RNA
Corrected processing (%): 16 17 4
Lane #: 1 2 3
C
^{32}p
Preincubation
Purified Pol II
Pulse
(no ATP)
Time
(min) -5 0 1.5
Collect and
Strip beads
Add NE
Chase
rRNaseH + control oligo
α-amanitin, 3'dATP
9
64
Collect and
Strip beads
5 ng/uL
chymotrypsin
11
Collect and
Strip beads
rRNaseH +
79.ds oligo
71
Remove beads
Add NE,
α-amanitin,
3'dATP,
+/- PVA
131
RNaseH-released
PVA: – – +
RNaseH-released RNA
Poly(A) site
cleaved RNA
Corrected processing (%): 12 0 11
Lane #: 1 2 3

Figure 4. Released transcripts process less efficiently than TEC-associated transcripts.

(**A**) RNaseH-released transcripts cannot process efficiently even when they are transcribed from a promoter template. Chased TECs were formed from a 12-fold scaled-up reaction on dsxSV40L templates, stripped, treated with chymotrypsin and the RNA transcripts were recovered as in Figure 3C. The transcripts were then added back to a processing reaction with nuclear extract under typical processing conditions. In another sample, chased TECs were formed, stripped, and put back into nuclear extract under similar processing conditions (with the addition of recombinant RNaseH and a ctrl oligo.

(**B**) TEC-associated transcripts process more efficiently than RNA released from elongation complexes. Chased TECs were formed from a 2-fold scaled-up reaction, stripped and added back to nuclear extract for 1 minute prior to addition of 79.ds oligo under processing conditions to allow opportunity for CPA assembly. Following 5 minute incubation with the 79.ds oligo, the beads were separated from the supernatant and both were processed under identical conditions for an additional 55 minutes. In lanes 5-6, PVA was added with the oligo to a final concentration of 2.1%.

Figure 4

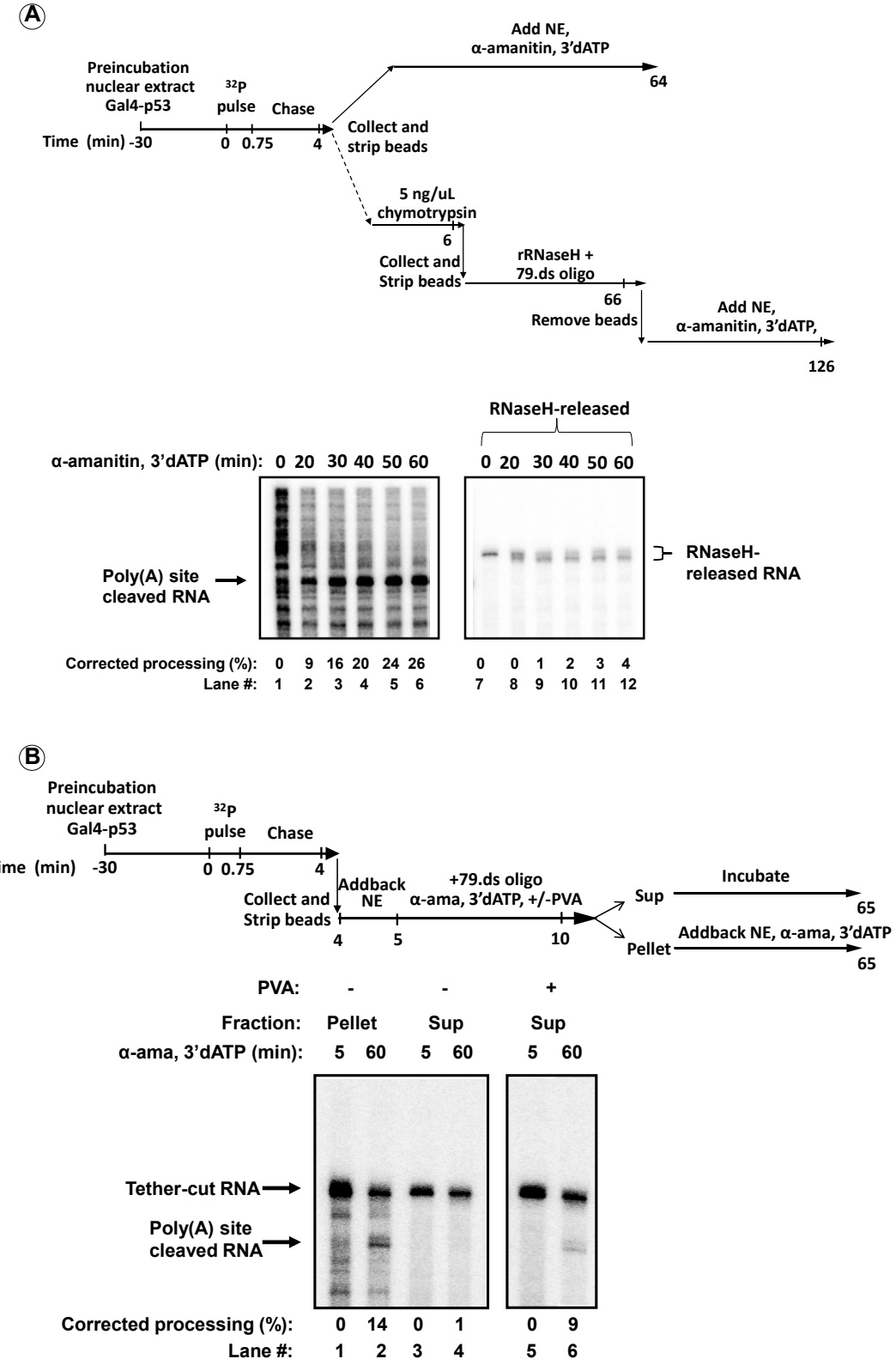

A
Add NE,
α-amanitin, 3'dATP
64
Preincubation
nuclear extract
Gal4-p53
32p
pulse
Chase
Time (min) -30
0
0.75
4
Collect and
strip beads
5 ng/uL
chymotrypsin
6
Collect and
Strip beads
rRNaseH +
79.ds oligo
66
Remove beads
Add NE,
α-amanitin, 3'dATP,
126
RNaseH-released
α-amanitin, 3'dATP (min): 0 20 30 40 50 60
0 20 30 40 50 60
Poly(A) site
cleaved RNA
RNaseH-
released RNA
Corrected processing (%): 0 9 16 20 24 26 0 0 1 2 3 4
Lane #: 1 2 3 4 5 6 7 8 9 10 11 12
B
Preincubation
nuclear extract
Gal4-p53
32p
pulse
Chase
Time (min) -30
0
0.75
4
Collect and
Strip beads
Addback
NE
+79.ds oligo
α-ama, 3'dATP, +/-PVA
4
5
10
Sup
Incubate
65
Pellet
Addback NE, α-ama, 3'dATP
65
PVA: - - +
Fraction: Pellet Sup Sup
α-ama, 3'dATP (min): 5 60 5 60 5 60
Tether-cut RNA
Poly(A) site
cleaved RNA
Corrected processing (%): 0 14 0 1 0 9
Lane #: 1 2 3 4 5 6

Figure S1. dsxSV40L transcripts are accurately processed.

Coupled processing assay was carried out on dsxSV40L template with cold nucleotides. The RNAs were extracted, digested with DNaseI (Roche), hybridized at 65^{o}C, and subjected to RNase Protection Assay with RNase T1/A cocktail (Ambion). Probe was generated from pBluescript dsx exon2 $(GAA)_6$ (Rosonina et al. 2003). Digesting the resulting plasmid with NotI, and transcribing with T7 RNA polymerase generase an RNase protection probe (323 nt) that will result in protected fragments from dsxSV40L transcripts with sizes of 270 nt for uncleaved and 196 nt for cleaved RNA. The presence of another strong uncleaved band (~246 nt) most likely reflects pausing on 6 consecutive thymidines present in the SV40 late poly(A) signal. Corrected processing (%) is calculated as the fraction of RNAs that were cleaved over total RNA (uncleaved + cleaved RNAs) and corrected for the 0 min timepoint.

Figure S1

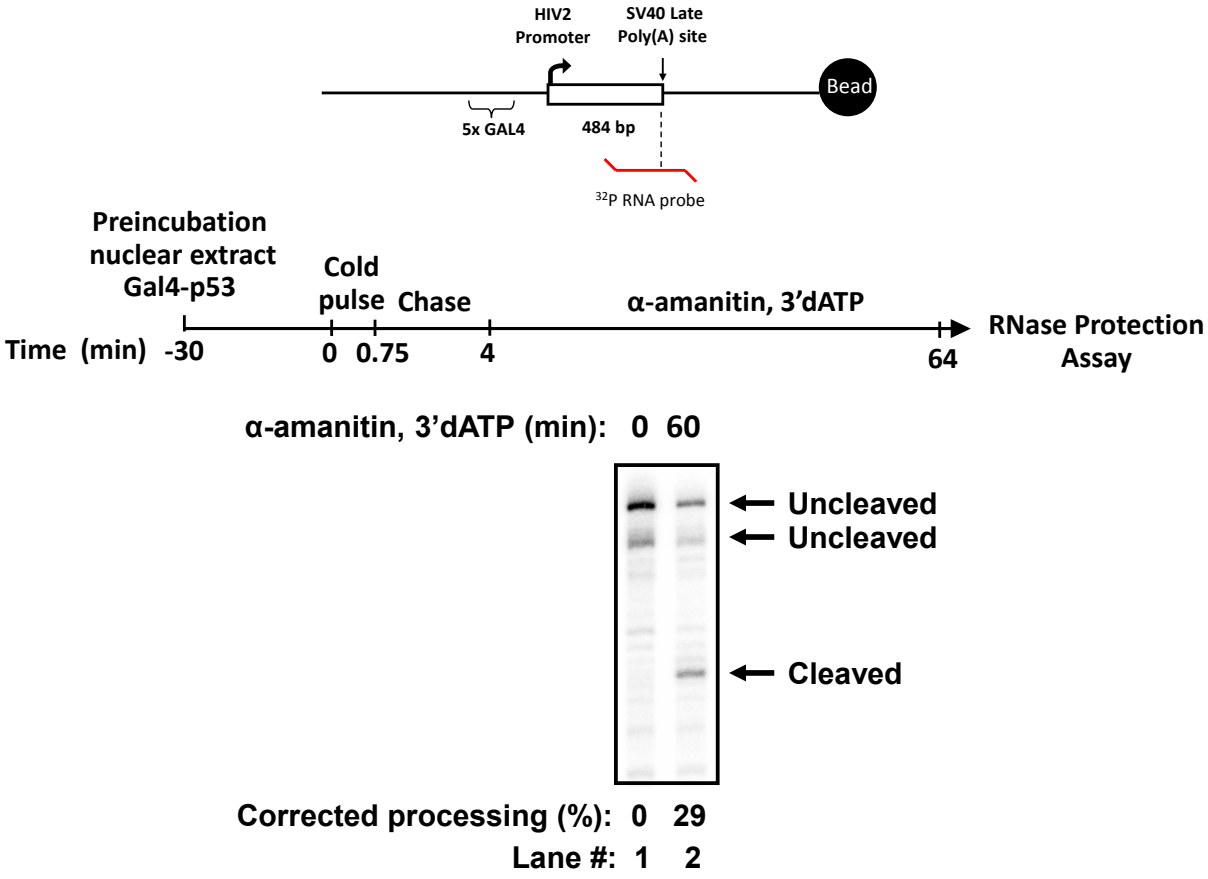

HIV2
Promoter
SV40 Late
Poly(A) site
Bead
5x GAL4
484 bp
32P RNA probe
Preincubation
nuclear extract
Gal4-p53
Cold
pulse
Chase
α-amanitin, 3'dATP
Time (min) -30
0
0.75
4
64
RNase Protection
Assay
α-amanitin, 3'dATP (min): 0 60
Uncleaved
Uncleaved
Cleaved
Corrected processing (%): 0 29
Lane #: 1 2

Figure S2. RNA transcripts synthesized from oligo-dTtaildsx templates are single-stranded. Transcription was carried out on a precursor of the oligo-dTtaildsx template, which do not carry the inserted thymidine patches (see figure legend of Figure 3A). TECs were generated, collected, rinsed with transcription buffer and resuspended in fresh transcription buffer with either no enzyme, RNaseH, RNase T1/A cocktail, and/ or 79.ds oligo. Treatment with either no enzyme or RNaseH +/- 79.ds oligo proceeded for an hour at 30^{o}C (lanes 2-3) while treatment with RNase T1/A cocktail proceeded for 30 minutes at 30^{o}C (lane 4).

Figure S2

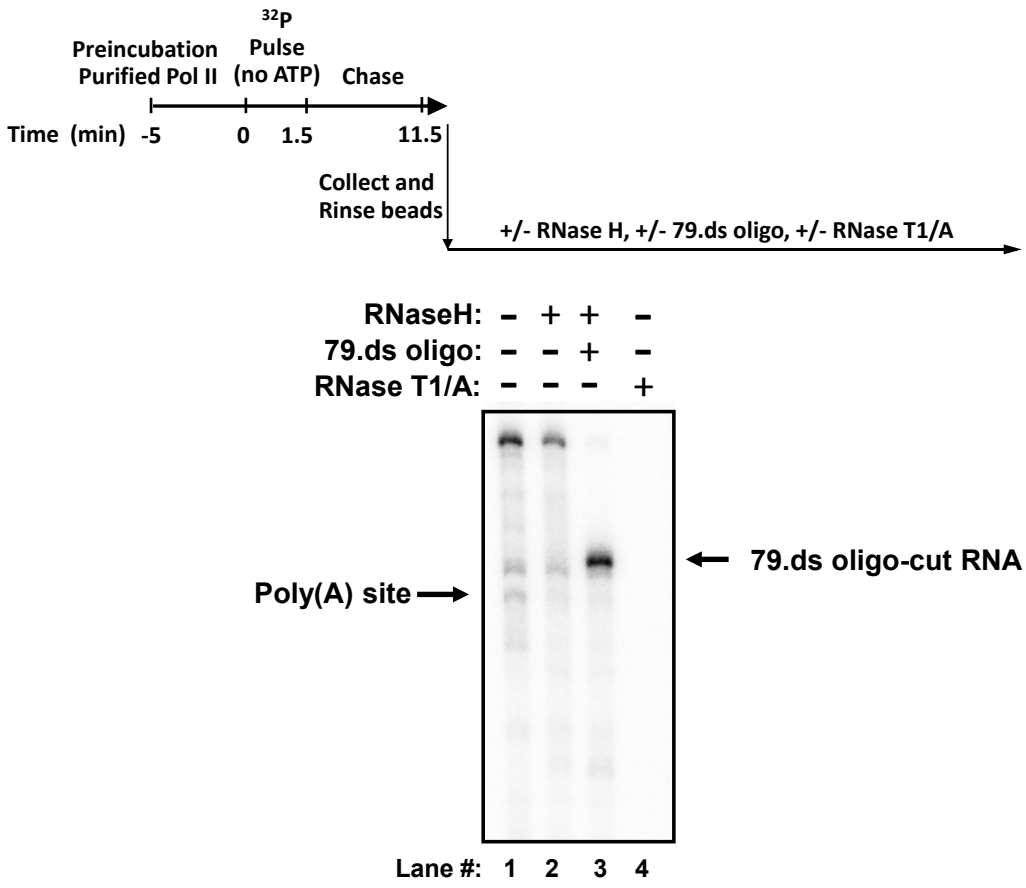

32P
Pulse
(no ATP)
Preincubation
Purified Pol II
Chase
Time (min) -5 0 1.5 11.5
Collect and
Rinse beads
+/- RNase H, +/- 79.ds oligo, +/- RNase T1/A
RNaseH: – + + –
79.ds oligo: – – + –
RNase T1/A: – – – +
Poly(A) site
79.ds oligo-cut RNA
Lane #: 1 2 3 4

Figure S3. RNA transcripts generated using oligo-dTtaildsx templates are synthesized by RNA polymerase II.

Oligo-dTtaildsx templates were transcribed in the presence of 0, 0.1 or 100µg/mL α-amanitin. Majority of transcription was inhibited at a concentration of 0.1 µg/mL α-amanitin, indicating that these transcripts are primarily synthesized by RNA polymerase II. It is known that only RNA pol II transcription is inhibited at a concentration of 0.1 µg/mL of α-amanitin (Weinmann et al. 1974).

Figure S3

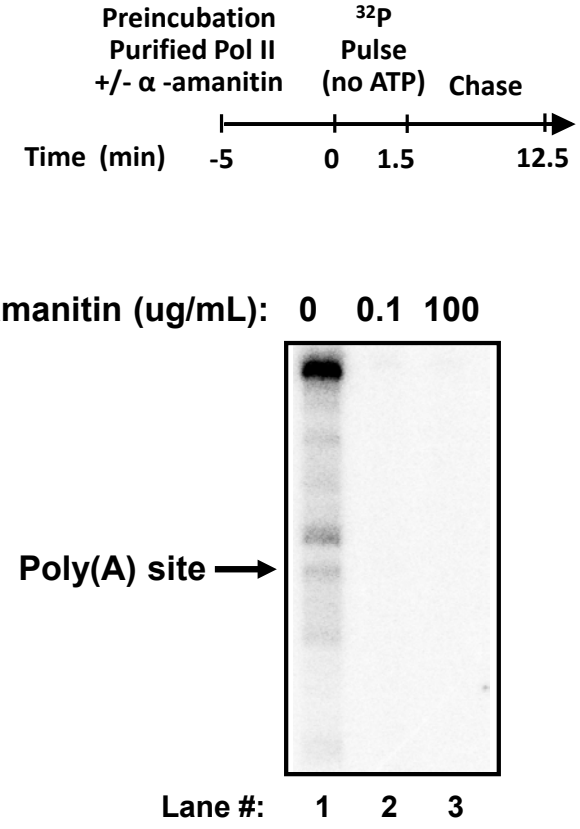
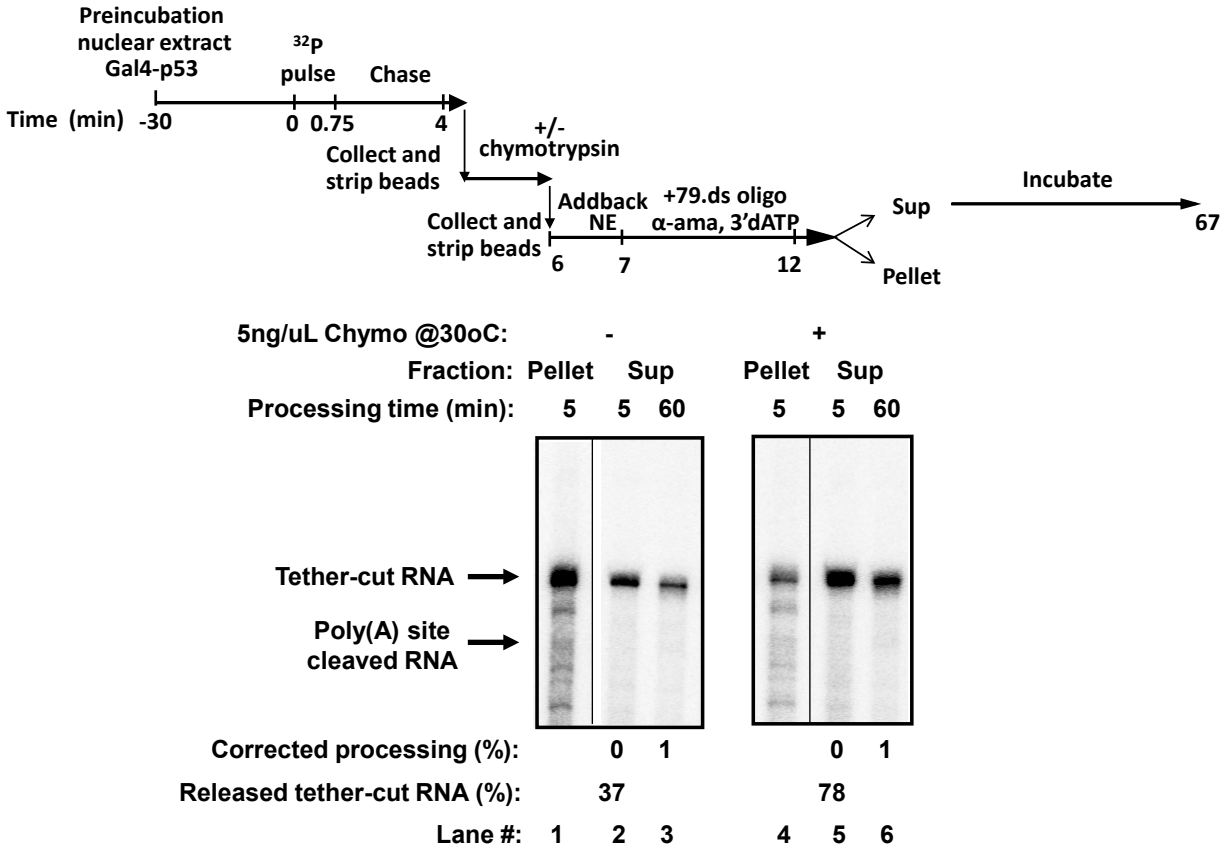

Preincubation
Purified Pol II
+/- α -amanitin
32P
Pulse
(no ATP)
Chase
Time (min)
-5
0
1.5
12.5
α-amanitin (ug/mL): 0 0.1 100
Poly(A) site
Lane #: 1 2 3

Figure S4. Released tether-cut RNAs do not process efficiently.

TECs were formed and treated with buffer with no enzyme or with 5 ng/µL of chymotrypsin for 2 min at 37^{o}C to remove the CTD before adding the TECs back to extract for release of RNAs by endogenous RNaseH. The released RNAs (supernatant) were separated from the beads and allowed to process for an additional 55 minutes.

Figure S4


Preincubation
nuclear extract
Gal4-p53
32p
pulse
Chase
Time (min) -30
0 0.75
4
Collect and
strip beads
+/-
chymotrypsin
Collect and
strip beads
Addback
NE
+79.ds oligo
α-ama, 3'dATP
6
7
12
Sup
Pellet
Incubate
67
5ng/uL Chymo @30oC:
-
+
Fraction: Pellet Sup Pellet Sup
Processing time (min): 5 5 60 5 5 60
Tether-cut RNA
Poly(A) site
cleaved RNA
Corrected processing (%): 0 1 0 1
Released tether-cut RNA (%): 37 78
Lane #: 1 2 3 4 5 6

References

Adamson, T.E., Shore, S.M., and Price, D.H. (2003). Analysis of RNA polymerase II elongation in vitro. Methods Enzymol. *371*, 264-275.

Adamson, T.E., Shutt, D.C., and Price, D.H. (2005). Functional coupling of cleavage and polyadenylation with transcription of mRNA. J. Biol. Chem. *280*, 32262-32271.

Akhtar, M.S., Heidemann, M., Tietjen, J., Zhang, D., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH kinase places bivalent marks on the carboxyl-terminal domain of RNA polymerase II. Mol. Cell *34*, 387-393.

Baskaran, R., Escobar, S.R., and Wang, J.Y.J. (1999). Nuclear c-Abl is a COOH-terminal repeated domain (CTD)-tyrosine kinase-specific for the mammalian RNA polymerase II: possible role in transcription elongation. Cell Growth Differ. *10*, 387-396.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr. Opin. Cell Biol. *17*, 251-256.

Bogenhagen, D.F., and Brown, D.D. (1981). Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell *25*, 261-170.

Buratowski S. (2003). The CTD code. Nat. Struct. Biol. *10*, 679-680.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541-546.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science. *318*, 1780-1782.

Chao, M.Y., Kan, M.C., and Lin-Chao, S. (1995). RNAII transcribed by IPTG-induced T7 RNA polymerase is non-functional as a replication primer for ColEl-type plasmids in Escherichia coli. Nucleic Acids Res. *23*, 1691-1695.

Chao, L.C., Jamil, A., Kim, S.J., Huang, L., and Martinson, H.G. (1999). Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. Mol. Cell. Biol. *19*, 5588-5600.

Cozzarelli, N.R., Gerrald, S.P., Schlissel, M., Brown, D.D., and Bogenhagen, D.F. (1983). Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. Cell *34*, 829-835.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3'-end of mRNA. Nature *389*, 399-402.

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

Dedrick, R.L., and Chamberlin, M.J. (1985). Studies on transcription of 3'-extended templates by mammalian RNA polymerase II. Parameters that affect the initiaton and elongation reactions. Biochemistry *24*, 2245-2253.

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

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. Trends in genetics *24*, 280-288.

Fong, N., and Bentley, D. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev. *15*, 1783-1795.

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

Hans, H., and Alwine, J.C. (2000). Functionally significant secondary structure of the simian virus 40 late polyadenylation signal. Mol. Cell. Biol. *20*, 2926-2932.

Hirose, Y., and Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

Howe K.J., Kane C.M., and Ares J. Jr. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA *9*, 993-1006.

Hsin, J., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3'-end processing. Science *334*, 80-87.

Jang, M. K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol. Cell *19*, 523-534.

Kadesch, T.R., and Chamberlin, M.J. (1982). Studies of in vitro transcription by calf thymus RNA polymerase II using a novel duplex DNA template. J. Biol. Chem. *257*, 5286-5295.

Kaneko, S., and Manley, J.L. (2005). The mammalian RNA polymerase II C-terminal domain interacts with RNA to suppress transcription-coupled 3'-end formation. Mol Cell *20*, 91-103.

Kazerouninia, A., Ngo, B., and Martinson, H. (2011). Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro. RNA *16*, 197-210.

Komarnitsky, P., E. J. Cho, and S. Buratowski. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. *14*, 2452-2460.

Lewicki, B., Margus, T., Remme, J., and Nierhaus, K. (1993). Coupling of rRNA transcription and ribosomal assembly in vivo – formation of active ribosomal-subunits in Escherichia coli requires transcription of RNA genes by host RNA polymerase which cannot be replaced by T7 RNA polymerase. J. Mol. Biol. *231*, 581-593.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3'-end processing factors with RNA polymerase II. Mol. Cell *9*, 1101-1111.

Lubkowska, L., Maharjan, A.S., and Komissarova, N. (2011). RNA folding in transcription elongation complex: implication for transcription termination. J. Biol. Chem. *286*, 31576-31585.

Maniatis T., and Reed R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.

Martincic, K., Alkan, S.A., Cheatle, A., Borghesi, L., and Milcarek, C. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nat. Immunol. *10*, 1102-1109.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

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

Nag, A., Narsinh, K., and Martinson, H.G. (2007). The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat. Struct. Mol. Biol. *14*, 622-669.

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

Orozco, I.J., Kim, S.J., and Martinson, H.G. (2002). The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. J Biol. Chem. *277*, 42899-42911.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297-305.

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

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3'-end formation to transcription. Curr. Opin. Cell Biol. *16*, 272-278.

Rigo, F., Kazerouninia, A., Nag, A., and Martinson, H.G. (2005). The RNA tether from the poly(A) signal to the polymerase mediates coupling of transcription to cleavage and polyadenylation. Mol. Cell *20*, 733-745.

Rigo, F., and Martinson, H.G. (2008). Functional coupling of last intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. Mol. Cell Biol. *28*, 849-862.

Rigo, F., and Martinson, H.G. (2009). Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. RNA *15*, 823-836.

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

Rosonina, E., Ip, J. Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J., and Blencowe, B.J. (2005). Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. Mol. Cell. Biol. *25*, 6734-6746.

Rozenblatt-Rosen, O., Nagaike, T., Francis, J.M., Kaneko, S., Glatt, K.A., Hughes, C.M., LaFramboise, T., Manley, J.L., and Meyerson, M. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. Proc. Natl. Acad. Sci. USA *106*, 755-760.

Ryan, K. (2007). Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation. RNA Biol. *4*, 26-33.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev. *14*, 2435-2440.

Shi Y., Di Giammartino D.C., Taylor D., Sarkeshik A., Rice W.J., Yates J.R., Frank J., and Manley J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Mol. Cell *33*, 365-376.

Sluder, A.E., Price, D.H., and Greenleaf, A.L. (1988). Elongation by Drosophila RNA polymerase II. Transcription of 3'-extended DNA templates. J. Biol. Chem. 263, 9917-9925.

Takagaki Y., Seipelt R.L., Peterson M.L., and Manley J.L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell *87*, 941-952.

Thompson, N.E., Aronson, D.B., and Burgess, R.R. (1990). Purification of eukaryotic RNA polymerase II by immunoaffinity chromatography. Elution of active enzyme with protein stabilizing agents from a polyol-responsive monoclonal antibody. J. Biol. Chem. *265*, 7069-7077.

Tran, D.P., Kim, S.J., Park, J.J., Jew, T.M., and Martinson, H.G. (2001). Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. Mol. Cell. Biol. *21*, 7495-7508.

Veraldi, K.L., Arhin, G.K., Martincic, K., Chung-Ganster, L.H., Wilusz, J., and Milcarek, C. (2001). hnRNPF influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. Mol. Cell Biol. *21*, 1228-1238.

Weinmann, R., and Roeder, R.G. (1974). Role of DNA-dependent RNA polymerase III in the transcription of the tRNA and 5S RNA genes. Proc. Natl. Acad. Sci. USA *71*, 1790-1794.

Wu, H., Liima, W.F., and Crooke, S.T. (1999). Properties of cloned and expressed human RNase H1. J. Biol. Chem. 274, 28270-28278.

Wu, C., and Alwine, J.C. (2004). Secondary structure as a functional feature in the downstream region of mammalian polyadenylation signals. Mol. Cell Biol. 24, 2789-2796.

Yaffe, M.B., Schutkowski, M., Shen, M., Zhou, X.Z., Stukenberg, P.T., Rahfeld, J.U., Xu, J., Kuang, J., Kirschner, M.W., Fischer, G., Cantley, L.C., and Lu, K.P. (1997). Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. Science *278*, 1957-1960.

Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol. Cell *19*, 535-545.

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol. Mol. Biol. Rev. *63*, 405-445.

Zhou, Q., and Yik, J.H. (2006). The yin and yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. Microbiol. Mol. Biol. Rev. *70*, 646-659.

Chapter 4

Preliminary studies on ATP requirement for 3’-end cleavage

Abstract

3’-end processing (cleavage and polyadenylation) of mRNA transcripts requires the assembly of a very large apparatus that may involve greater than 80 proteins (Shi et al. 2009). While the majority of the trans-acting factors involved in the assembly of the cleavage and polyadenylation apparatus have been studied extensively, not much is known regarding the actual assembly process. However, there have been suggestions that the assembly of this apparatus may be a step-wise process. To study the assembly pathway, we began by resolving whether ATP is a required cofactor for 3’-end processing since there have been conflicting reports in the literature. Interestingly, we found that ATP is indeed a required cofactor for 3’-end cleavage in our in vitro system since withholding ATP can inhibit cleavage. We found that this inhibition is reversible since cleavage resumes upon re-addition of ATP. We show that withholding ATP blocks a discrete step in the assembly pathway since only a partial cleavage and polyadenylation apparatus can form in the absence of ATP cofactor. We provide evidence that this partial apparatus consists of at least CPSF and CstF and this apparatus can resume assembly following supplementation with ATP cofactor and nuclear extract. We also demonstrate that the assembly of this partial apparatus is not the rate-limiting step in the assembly pathway. Instead, the rate-limiting step occurs at or after the ATP-requiring step. These results suggest that ATP may play an important role in regulation of 3’-end cleavage. The significance of an ATP requirement for 3’-end cleavage in our in vitro system will be discussed.

Introduction

The core poly(A) signal is made up of an AAUAAA hexamer sequence and a downstream GU-rich region, which are recognized by core processing factors cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulatory factor (CstF), respectively (Zhao et al. 1999). Besides these two complexes, the full assembly of the cleavage and polyadenylation apparatus (CPA) may involve greater than 80 proteins (Shi et al. 2009). While much is known regarding the identity of the processing factors involved in the cleavage and polyadenylation reaction, very little is known regarding the assembly process of the CPA. However, there is evidence that assembly of the CPA on the poly(A) signal is a stepwise process in vivo (Chao et al. 1999). It has also been observed in vitro that when processing is coupled to transcription, a substantial lag precedes a dramatic burst of rapid cleavage, which further suggests that the assembly of the apparatus may be a multi-step process.

More recently, an assembly pathway was proposed based on immunoprecipitation experiments that had identified discrete complexes involving CPSF, CstF and C-terminal domain (CTD)-less RNA polymerase II and CTD-intact RNA polymerase II (Nag et al. 2007). The pathway begins with CPSF riding on the polymerase body before the poly(A) signal is transcribed. CPSF scans the RNA and upon extrusion of the poly(A) hexamer, CPSF binds the AAUAAA sequence and directs the polymerase to pause. Once the GU-rich region is transcribed, CstF joins CPSF at the poly(A) signal, and the two complexes are transferred to the CTD. Additional cleavage and polyadenylation factors such as cleavage factor Im (CFIm), cleavage factor IIm (CFIIm), and poly(A) polymerase (PAP) join sometime later and cleavage and polyadenylation ensues.

In addition to the recruitment of cleavage and polyadenylation factors, an early report suggested that ATP may be a required cofactor for 3'-end cleavage (Moore et al. 1985). Interestingly, the requirement for ATP is unusual in that hydrolysis of the β-γ phosphate bond is not required (Moore et al. 1985), suggesting that it is the binding of ATP, and perhaps triggering of a conformational change, that is required. However, the proposal that ATP is a required cofactor for cleavage was conflicted by later reports showing that cleavage can take place in absence of divalent cation and without the addition of exogenous ATP (Zarkower et al. 1986; Takagaki et al. 1988). Furthermore, it was later demonstrated that creatine phosphate is the required cofactor and not ATP (Hirose et al. 1997), since cleavage can take place without ATP, provided that creatine phosphate is at a high concentration. Manley and colleagues also ruled out the possibility that creatine phosphate was supporting cleavage by regenerating ATP in a reaction involving the enzyme creatine phosphokinase in their partially purified system (Hirose et al. 1997). Consistent with the idea that ATP is not required for 3'-end cleavage, Price and colleagues reported that under conditions that lack free divalent cations (in presence of EDTA), 3'-end cleavage can occur in their coupled in vitro system (Adamson et al. 2005). Because of these studies, it has been widely accepted that ATP is not required for 3'-end cleavage. However, it remains unclear whether the difference in ATP requirement may reflect a difference between the in vitro systems and conditions used. It is also not known what the exact role of creatine phosphate is in cleavage although it has been suggested that creatine phosphate may be mimicking some function of the CTD of RNA polymerase II (Hirose et al. 1998).

The goal of our initial studies was to devise a method to inhibit various steps of the assembly pathway in order to study the assembly process of the cleavage and polyadenylation apparatus. We decided to begin by resolving whether ATP is a required cofactor for 3'-end processing since there have been conflicting reports in the literature. We were surprised to find that ATP is indeed a required cofactor for efficient 3'-end cleavage in our coupled in vitro system. With this knowledge, we sought to investigate the role of ATP in the assembly pathway. We found that an incomplete cleavage and polyadenylation apparatus consisting of at least CPSF and CstF can assemble in the absence of ATP cofactor. We also demonstrate that the assembly of this partial apparatus is not rate-limiting but instead, the rate-limiting step of the cleavage pathway occurs at or after the ATP-requiring step. These results suggest that ATP may play an important role in regulation of 3'-end cleavage. The significance of an ATP requirement for 3'-end cleavage in our in vitro system will be discussed.

Results

In an effort to study the assembly pathway of the 3'-end processing apparatus, we searched for methods to block 3'-end processing at different steps along the assembly pathway. We decided to begin by resolving whether ATP is required for 3'-end cleavage to occur in a coupled in vitro system. To do this, we utilized a dsxSV40L DNA template immobilized on magnetic beads carrying an SV40 late poly(A) signal whose expression is under the control of five Gal4 sites (see Figure 1 DNA template diagram). Transcription was controlled by the addition of Gal4-p53 activator (two p53 activation domains in tandem fused to Gal4 DNA binding domain). Pulse-chase assays using this template have been described (Chapter 3). To rule out the possibility that our nuclear extract contains residual ATP, we asked if our nuclear extracts can support polyadenylation in the absence of exogenously added ATP. To do this, we formed transcription elongation complexes (TECs) and elongated them past the poly(A) site. We then isolated the TECs by pulling down the DNA templates using a magnet, washed off all transcription factors and residual ATP with a high detergent, high salt buffer (1M KCl, 1% sarkosyl) and resuspended the TECs in nuclear extract under processing conditions with or without exogenously added ATP and incubated for an hour to allow processing to occur. The RNA transcripts were subjected to oligodT selection after RNA extraction to look for evidence of polyadenylated transcripts. We have previously demonstrated that a naked elongation complex with the poly(A) signal fully transcribed can still undergo efficient coupled 3'-end cleavage when returned back to fresh nuclear extract under processing conditions (Chapter 3). Figure 1A, lane 1 shows that when TECs were processed in the presence of ATP, the transcripts were polyadenylated, as evident from the pull-down by oligo-dT beads. However, when no ATP was added, oligo-dT beads did

not pull down any polyadenylated transcripts (Figure 1A, lane 2). These results suggest that our nuclear extract does not contain sufficient amounts of residual ATP to support polyadenylation. In addition, our in vitro processing assay utilizes creatine phosphate, which is known to participate in an ATP-generating reaction catalyzed by creatine phosphokinase. Interestingly enough, these above results also suggest that at least under our experimental conditions, the creatine phosphate present during processing does not regenerate sufficient amounts of ATP, or if any, to support polyadenylation.

Now that we can be sure that there are insufficient amounts of ATP to support polyadenylation under our processing conditions, we wanted to ask if ATP is required for 3'-end cleavage in our coupled in vitro system. To test this, we repeated the experiment as in Figure 1A except 3'dATP was used and the extracted RNA was directly loaded on acrylamide gel (instead of being subjected to oligo-dT selection) to look for evidence of 3'-end cleaved transcripts. Figure 1B lane 2 shows when TECs were returned to nuclear extract under processing conditions with 3'dATP, transcripts were cleaved, giving rise to a strong poly(A) cleaved band. However, no poly(A) site cleaved RNA was observed when TECs were returned to fresh nuclear extract under similar processing conditions without 3'dATP (Figure 1B, lane 4). Therefore, these results suggest that ATP is indeed required for 3'-end cleavage in our in vitro system. This is surprising given that many studies have suggested that ATP is not required for 3'-end cleavage since 3'-end cleavage can occur in the absence of free magnesium (in presence of EDTA) or in the absence of exogenously added ATP (Hirose et al 1997; Zarkower et al. 1986; Takagaki et al. 1988; Adamson et al. 2005).

Although it is unlikely, another possibility is that the transcripts are in fact being polyadenylated by residual ATP in the extract but the poly(A) tails are insufficiently-long to be pulled down by oligo-dT beads. To address this possibility, we repeated the experiment in Figure 1A using cold nucleotides and subjected the RNA transcripts to RNase protection Assay. Figure 1C, lane 1 is a control lane showing that no cleaved RNA has accumulated yet in TECs that had just been resuspended in fresh nuclear extract (0 min timepoint of the hour-long processing step). Figure 1C, lane 2 shows that when TECs were resuspended in fresh extract and incubated for an hour under processing conditions in the absence of added ATP, no cleavage was observed. However, when ATP was present during the hour-long incubation, cleavage of the transcripts was strong and robust (Figure 1C, lane 3). Interestingly, cleavage was partially rescued when AMPPCP, an ATP analog with a non-hydrolyzable β-γ phosphate bond, was added (Figure 1C, lane 4). This is consistent with a report demonstrating that while ATP is required, hydrolysis of the β-γ phosphate is not required for 3'-end cleavage (Moore et al. 1985). These results demonstrate that ATP is indeed required for 3'-end cleavage in our coupled in vitro system.

Based on the above results, 3'-end processing is blocked when ATP is withheld. An important question is whether this inhibition is reversible. For example, one possibility is that ATP participates in the efficient recruitment of the cleavage and polyadenylation apparatus (CPA). When ATP is not present, assembly of the CPA is inefficient, which leads to the formation of irreversible dead-end complexes. Another possibility is that a particular step in the cleavage pathway is blocked when ATP is withheld but the cleavage pathway can resume once ATP is present. To address this experimentally, we asked whether TECs that had been incubating an hour-long in extract in the absence of added ATP (recall from Fig. 1B, lane 4, no poly(A) cleaved

RNA is generated) can resume processing following re-addition of 3'dATP. To do this, we formed TECs, stripped them, and incubated them in nuclear extract in the absence of ATP. After an hour, we added 3'dATP to the reaction and incubated for another hour. Figure 2A, lane 1 shows that TECs that were incubating for an hour in absence of added ATP can resume processing when 3'dATP was added. For comparison, Figure 2A, lane 2 shows again that 3'-end cleavage is efficient when stripped TECs were put directly back under processing conditions with 3'dATP. Figure 2A, lane 3 confirms the results of Fig. 1B, lane 4 that no poly(A) site-cleaved RNA is generated when TECs were put back under processing conditions without ATP. Together, these results demonstrate that the inhibition of cleavage can be reversed by the re-addition of ATP. The results also suggest that a discrete step in the assembly pathway is blocked.

Next, we wanted to ask if withholding ATP prevents initiation of CPA assembly or whether it inhibits an intermediate step in CPA assembly. A possible interpretation of the previous result, for example, is that no assembly of CPA was supported at all during incubation in ATP-less extract, and that assembly of CPA occurred only after 3'dATP was added. To address this, we needed to probe whether any assembly of the cleavage and polyadenylation apparatus can take place in ATP-less extract. To do this experimentally, we asked if TECs preincubated in ATP-less extract for an hour can still process if they are returned to an extract that had been depleted of CPSF and CstF. Notice that this approach presumes that at least CPSF and CstF are recruited to the poly(A) signal in the absence of ATP. This assumption makes sense given that CPSF and CstF are the core processing factors that recognize the hexamer and GU-rich region of the poly(A) signal and are believed to be among the first few proteins to be recruited to the

poly(A) signal. To do this, we pre-assembled the CPA in the absence of ATP, rinsed the complexes with low-salt transcription buffer and returned the complexes back into an extract that had been depleted of CstF and CPSF using a mixture of two-monoclonal antibodies directed to the 64 kDa subunit of CstF. Figure 2B is a western blot confirming that CPSF and CstF are both quantitatively depleted in the α-CstF64-depleted extracts (lane 1) while the largest subunit of RNA polymerase II, Rpb1, remain essentially the same compared to normal mouse-depleted extracts (lane 2) or non-depleted extracts (lane 3). Figure 2C, lane 1 shows that when TECs were incubated in nuclear extract in absence of ATP to allow pre-assembly of the CPA, the resulting complex was able to generate poly(A) cleaved RNAs even though it was resuspended and processed in α-CstF64-depleted extracts. Figure 2C, lane 2 confirms that α-CstF64-depleted extracts cannot support 3'-end processing on its own while normal mouse-depleted extracts can (Figure 2C, lane 3). Together these results demonstrate that there is at least partial assembly of the CPA during incubation of the TEC in nuclear extract in absence of ATP. In addition, this partial CPA complex consists of at least CPSF and CstF.

Since we were able to demonstrate that CPSF and CstF can assemble onto the poly(A) signal under conditions that lack ATP, the next obvious question was whether the entire CPA can assemble in the absence of ATP. Perhaps ATP is only required to stimulate or activate the CPA following assembly of the complete CPA. To address this, we generated TECs, allowed the CPA to pre-assemble in the absence of ATP, rinsed the complexes with low-salt transcription buffer, and then put back the complexes to similar processing conditions containing 3'dATP but no nuclear extract and asked if these TECs can process. Figure 2D, lane 2 shows that no poly(A) cleaved RNAs were generated under these conditions. These results suggest that the CPA

assembled in the absence of ATP is not a complete apparatus and that ATP may be required for the recruitment of one or more processing factors. However, we cannot rule out the possibility that the low-salt transcription buffer may have washed off a very weakly binding cleavage and polyadenylation factor required for 3'-end processing.

Next, we wanted to study the kinetics of 3'-end cleavage by exploiting the ATP requirement. We have previously observed that there is a several minute lag between the transcription of the full poly(A) signal and the accumulation of cleaved and polyadenylated RNAs (Rigo et al. 2005). We wanted to ask whether the partial assembly of the CPA in absence of ATP is rate-limiting for the 3'-end processing pathway. We reasoned if assembly of cleavage and polyadenylation factors that occur in absence of ATP is rate-limiting, then these pre-assembled complexes should process at a faster initial rate compared to the default stripped TECs. To do this, we formed stripped TECs, pre-assembled the partial CPA in absence of ATP, then added 3'dATP and incubated for another hour. For comparison, we assembled a separate batch of stripped TECs and directly incubated them in nuclear extract under processing conditions with 3'dATP. Here, Figure 3A shows the percentage of transcripts cleaved at the poly(A) site as a function of time between TECs with pre-assembled CPA (lanes 1-3) and naked TECs (lanes 4-6). Previously, we have found in our in vitro system that 3'-end cleavage begins roughly 10 minutes after extrusion of the poly(A) signal (Rigo et al. 2005) and is complete after about 40 minutes of incubation (Chapter 3, Figure 4A). If pre-assembly of CPA on TECs in absence of ATP is rate-limiting, then one would expect a dramatic initial burst of processing that would be near completion within 20 minutes. However, this was not the case (Figure 3A lanes 1-3). Assuming that the maximum processing efficiency is reached after 60 minutes for both types of

transcription complexes (Figure 3A, lanes 3 and 6), TECs with pre-assembled CPA was 20% complete after 20 minutes while naked TECs were 27% complete. This shows that pre-assembling the CPA on the TEC in absence of ATP did not confer any kinetic advantage for 3'-end cleavage of the associated transcripts. Therefore, these results suggest that the early steps of recruitment before the ATP-dependent step occur quickly and that the rate-limiting step in the 3'-end processing pathway occurs at or after the ATP-requiring step.

Our previous results suggested that the requirement for ATP in the 3'-end cleavage reaction is not dependent on ATP hydrolysis (recall Figure 1C, lane 4) since AMPPCP can support 3'-end cleavage. Therefore, we wondered whether ATP is required for binding to one or more processing factors. Perhaps, its role is to induce a slow, rate-limiting conformational change in one or more cleavage and polyadenylation factors to activate the cleavage apparatus. To test this hypothesis, we formed stripped TECs, pre-assembled a partial CPA in absence of ATP, rinsed off unbound proteins, and then treated these TECs with 3'dATP in buffer (no extract) to allow ATP binding and the hypothesized conformational change. We then put back these complexes in nuclear extract for processing to ask if these complexes will process at a faster rate. Figure 3B, lanes 1-3 show the processing efficiency of these TECs as a function of time. However, when compared to control TECs that had not been pre-treated with 3'dATP (Fig. 3B, lane 4-6), the rate of 3'-end processing is not significantly different between the two preparations of TECs. Therefore, these results are not consistent with this conformational change hypothesis. Alternatively, this experiment may not have been able to adequately address this hypothesis if ATP is also required for the recruitment of one or more processing factors to the CPA before induction of a conformational change. Therefore, these results do

not completely rule out the possibility for an ATP-induced conformational change. Interestingly, the results of Figure 3A and 3B together are consistent with the idea that the rate-limiting step of the cleavage and polyadenylation pathway occurs during or after the ATP-requiring step.

Discussion

We have demonstrated that ATP is a required cofactor for efficient 3'-end processing in our in vitro system. By withholding ATP, we were able to block the 3'-end cleavage reaction (Figure 1C, lane 2). We found that this inhibition was reversible since processing resumed upon re-addition of an ATP analog such as 3'dATP (Figure 2A, lane 1). Moreover, our results suggest that in the absence of ATP, an incomplete cleavage and polyadenylation apparatus is assembled onto the poly(A) signal (Figure 2D). This partial apparatus appears to be made up of at least CPSF and CstF (Figure 2C) and can remain poised up to an hour in nuclear extract and still engage in relatively efficient 3'-end cleavage of the transcript upon re-addition of ATP (Figure 2A). Importantly, this indicates that assembly of the CPA is not diverted into a non-productive pathway and result in dead-end complexes when ATP is not present.

What is the role of ATP in 3'-end cleavage? ATP can be envisioned to have two possible roles. First, ATP may be required for phosphorylation of one or more cleavage and polyadenylation factors. However, this is unlikely since we observed that an ATP analog with a non-hydrolyzable β-γ phosphate bond, AMPPCP, can support 3'-end cleavage, albeit inefficiently (Figure 1C). Second, ATP may be required to induce a conformational change by binding to the active or allosteric site in one or more proteins. One possible function for this ATP-binding is to directly induce a conformational change in the affected protein (e.g. poly(A) polymerase, PAP) to allow it to be recruited to the cleavage and polyadenylation apparatus. Another possible function is to induce a remodeling or structural change to the partially assembled CPA before further assembly can occur, eventually leading to the final cleavage step. However, one would expect that such a large-scale structural change would be a rate-limiting step. If that is the case, our

kinetic results would argue against this possibility since pre-incubation of a partially assembled CPA with 3'dATP did not speed up the rate of the cleavage reaction (Figure 3B).

While our experiments do not offer direct evidence for ATP binding-mediated recruitment of a cleavage factor, there is a report that is at least consistent with this idea. Jacob's lab found that ATP binding by poly(A) polymerase (PAP) is essential for 3'-end cleavage (Ghoshal et al. 1991). First, they demonstrated that an ATP analog called Ara-ATP (9-β-D-arabinofuranosyladenine triphosphate) can inhibit poly(A) cleavage in vitro. Next, they demonstrated that the inhibition by Ara-ATP can be partly overcome by supplementing with fresh poly(A) polymerase fraction in the presence of ATP (Ghoshal et al. 1991). This result suggests that Ara-ATP may be inhibiting cleavage by occupying the ATP binding site of PAP and that the supplementation with fresh poly(A) polymerase reduces the effective concentration of Ara-ATP and increases the chance of PAP binding to ATP. Therefore, Jacob's report implicate a role for ATP binding by PAP in 3'-end cleavage and is at least consistent with the idea that the role of ATP is to mediate the recruitment of PAP to the CPA.

Whether ATP is necessary for direct recruitment of PAP or "prepares" or induces structural changes in a partially assembled CPA apparatus, our results support the idea that ATP in one way or another leads to further assembly since the CPA that assembles in absence of ATP seems to be incomplete, as it cannot carry out 3'-end cleavage with re-addition of ATP alone but rather it has to be accompanied by addition of nuclear extract as well (Figure 2D and Figure 3B).

In addition to studying the role of ATP in the assembly pathway, we also discovered that the rate-limiting step of the 3'-end cleavage pathway appears to take place during or after the ATP-

dependent step. This is based on the observation that giving TECs an opportunity to assemble a partial CPA apparatus (in absence of ATP) did not help speed up the rate of 3'-end cleavage in our assays (Figure 3A and Figure 3B).

Perhaps the most surprising finding is that ATP is a required cofactor for 3'-end cleavage in our coupled transcription-processing assays. This is surprising since it has been widely accepted that 3'-end cleavage does not require ATP in an uncoupled processing assay. One obvious explanation for this discrepancy is that the requirement for ATP in our assays reflect a role of ATP in coupling transcription to 3'-end processing. However, there has been a previous report that demonstrated a requirement for ATP in an uncoupled 3'-end cleavage reaction (Moore et al. 1985). Therefore, it is unlikely that the ATP requirement reflects a unique role in coupling. Another possibility is that ATP is required under certain conditions but not others. For example, Manley and colleagues demonstrated using a partially purified system that while ATP is not a required cofactor for 3'-cleavage for SV40 late poly(A) signal, they did find that ATP can stimulate 3'-end cleavage on PAP-dependent poly(A) signal such as Adenovirus L3 poly(A) signal (Hirose et al. 1997). They suggested that the ability of ATP to stimulate 3'-end cleavage reaction for L3 may be mediated by poly(A) polymerase. This would be consistent with the fact that ATP did not stimulate 3'-end cleavage for SV40 late poly(A) signal since their processing reactions were devoid of poly(A) polymerase (SV40 late poly(A) signal is unusual in that cleavage does not require PAP under uncoupled processing conditions).

Given that the requirement for ATP may be different with different poly(A) signals and under different processing reaction conditions, how do we make sense of this? The most important factor to consider is what is most physiologically relevant. Although Manley and colleagues

were able to demonstrate that 3'-end cleavage of L3 can take place in the absence of ATP, they had to utilize very high concentrations (40-60 mM) of creatine phosphate (Hirose et al. 1997). And even under these conditions, 3'-end cleavage was very inefficient even for an uncoupled 3'-end cleavage assay. Interestingly, at lower concentrations of creatine phosphate (~20 mM), ATP is required for 3'-end cleavage of L3 (Hirose et al. 1997). We believe that the 40-60 mM creatine phosphate necessary for 3'-end cleavage of L3 in absence of ATP is not physiologically relevant given that in most tissues creatine phosphate is in the 5-10 mM range (Iyengar 1984). The only tissue with this high level of creatine phosphate (~ 40mM) is skeletal muscle (Beis et al. 1975; Kushmerick et al. 1992). Therefore, in most tissues, ATP would be required for optimal 3'-end cleavage for most pre-mRNAs (PAP-dependent) under physiological concentrations of creatine phosphate.

How do the experimental conditions alter the ATP requirement for cleavage? Based on these reports (Hirose et al. 1997; Moore et al. 1985) and our findings, one possibility is that the ATP requirement can be bypassed when high concentrations of creatine phosphate are used under uncoupled processing conditions. Uncoupled processing assays typically utilize a crowding agent such as polyvinyl alcohol (PVA) and while it is not clear what the role of PVA is, it is known that PVA can stimulate uncoupled 3'-end processing. It is possible that PVA, in concert with creatine phosphate, can bypass the requirement of ATP by stabilizing or mediating interactions between cleavage and polyadenylation factors on the CPA, which would've normally been carried out by poly(A) polymerase and ATP. Therefore, a requirement for ATP would only be observable under low creatine phosphate concentrations in the absence of PVA. Consistent with this idea, the studies conducted by Moore (Moore et al. 1985) and our studies do not

utilize PVA while studies by Manley (Hirose et al. 1997) and Price (Adamson et al. 2005) do. However, additional experiments will have to be conducted in order to better understand the role of ATP in 3'-end cleavage and how creatine phosphate and PVA may substitute for the role of ATP in vitro.

Materials and Methods

Oligos used:

1: 5'-GTTGGACTCAAGACGATAGTTACC-3'

2: 5'-biotin-CACATTTCCCCGAAAAGTGCCACC-3'

Immobilized templates

DsxSV40L immobilized templates (1.7 kb) were generated by PCRing up the Gal5-HIV2dsxΔInt(+ESE) template (a gift from Rosonina et al. 2003) from -781 bp upstream of the transcription start site down 464 bp past the SV40 Late poly(A) cleavage site using oligos number 1 and 2 as primers. Oligo number 2 is a reverse biotinylated primer to allow for bead attachment on the downstream end of the template. The PCR product of correct length was gel-extracted and purified using the QIAquick Gel Extraction Kit. Binding of the template to Dynal M-280 Streptavidin beads (Invitrogen) was carried out according to manufacturer's protocol. The final concentration of DNA on streptavidin beads was 0.8-3.1 pmol of DNA template per mg of beads. We found that a DNA density greater than 3.1 pmol of 1.7kb-long DNA per mg of beads gave reduced processing efficiency in a coupled processing assay, presumably due to steric interference. Reducing DNA density on the beads to less than 3.1 pmol of DNA per mg of beads did not have any further effect on processing efficiency but did increase transcription.

Purification of Gal4-p53 activator

G4-p53 construct (gift from James Goodrich) was expressed in the E. coli BL21 strain. T7 promoter expression was induced in these cells with 0.4 mM IPTG for 2 hours. Cells were collected, resuspended in TGED buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 20% glycerol, 0.1

mM EDTA, 1 mM DTT, 1mM PMSF) containing 5 mM $MgCl_2$, and lysed via sonication. Cell lysates were incubated with ANTI-FLAG M2 affinity gel (Sigma, A2220) for 3 hours at 4^{o}C. The affinity gel was washed with TGED buffer and TGED buffer containing 1M NaCl. The affinity gel was transferred to a Poly-Prep Chromatography Column (Biorad) and Gal4-p53 proteins were released with TGED buffer containing 0.2 mg/mL FLAG peptide at room temperature. The fractions were collected and immediately stored at -80^{o}C. Peak fractions containing Gal4-p53, as determined by SDS-PAGE and sypro ruby staining, were later pooled together, re-aliquoted and stored at -80^{o}C.

In vitro transcription (Coupled processing assay)

Hela nuclear extract was prepared as described (Tran et al. 2001). HeLa nuclear extracts were dialyzed against Buffer D (20% glycerol, 20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, and 0.1 mM PMSF).

A typical coupled processing assay began with 5 µL of nuclear extract that was mixed with Gal4-p53, anti-RNase (Ambion), DTT, $MgCl_2$, sodium citrate, immobilized DsxSV40L template, and water up to 9.9 µL. Volume of transcriptional activator addition was selected based on a titration experiment determining the minimum volume required to achieve maximum % poly(A) cleavage in a typical coupled processing experiment. The mixture was incubated at 30^{o}C for 30 min to form preinitiation complexes. Transcription was then initiated with a 3 µL pulse of 20 µCi of [α-^{32}P] CTP (Perkin Elmer, 800 Ci/mmol), nucleotide triphosphates and creatine phosphate. Then, 2.6 µL of chase mix was added containing a high concentration of nonradiolabeled CTP. Concentrations of reagents at this point are as follows: 7% glycerol, 6 mM HEPES (pH 7.9), 2.6 mM Tris-HCl (pH 7.4), 7.8 mM NaCl, 32 mM KCl, 69 µM EDTA, 1.8 mM

DTT, 83 µM PMSF, 10 U Anti-RNase, 4 mM $MgCl_2$, 3.2 mM Citrate (pH 6.7), 0.16 pmol DNA template, 320 µM ATP, 161 µM UTP and GTP, 1.6 mM CTP, and 16 mM creatine phosphate. Transcription was halted with addition of 11 mM EDTA and transcription complexes were isolated with magnet (Dynal MPC-S) for approximately 30-60 seconds and washed with 200 µL of high salt, high detergent buffer (20 mM HEPES pH 7.9, 1 M KCl, 1% Sarkosyl) to remove all transcription factors and then with 200 µL of low-salt transcription buffer (3.9 mM HEPES, 20 mM KCl, 39 µM EDTA, 3.9% glycerol, 0.1 mM DTT, 20 µM PMSF). For processing directly, washed TECs were then resuspended and incubated for an hour in a 16.5 µL processing reaction (see below for final concentrations of reagents) which includes 5 µL of nuclear extract with either ATP, 3'dATP, or AMPPCP. For pre-assembly of CPA on TEC, TECs were resuspended in the same 16.5 µL processing reaction except in the absence of any added ATP. In Figure 2D and 3B in which TECs with pre-assembled CPA were incubated with 3'dATP in buffer, TECs were resuspended in 16.5 µL processing reaction in the presence of 5 µL of Buffer D instead of nuclear extract. Final concentrations of reagents in a standard processing reaction (16.5 µL) were as follows: 7 % glycerol, 6 mM HEPES (pH 7.9), 2.4 mM Tris-HCl (pH 7.4), 7.3 mM NaCl, 30 mM KCl, 65 µM EDTA, 1.7 mM DTT, 78 µM PMSF, 10 U Anti-RNase, 3.8 mM $MgCl_2$, 3.0 mM Citrate (pH 6.7), 0.16 pmol DNA template, 15 mM creatine phosphate, 30 µg/mL α-amanitin and 325 µM ATP, 3'dATP or AMPPCP.

The reaction was terminated by the addition of a stop solution: 65 µL of 10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS, and 100 µg proteinase K (Ambion). RNA was extracted with 350 µL TRIzol (Invitrogen) and 70 µL chloroform and then precipitated with 4 µL of 5 mg/mL glycogen (Ambion) and 350 µL isopropanol (30 min, room temperature), and finally ran on a 5%

polyacrylamide gel. After electrophoresis, results were recorded and analyzed by using a PhosphorImager with ImageQuant software (Molecular Dynamics).

Depletion of CstF64 and Western blot analysis

Normal mouse- and CstF64-depleted extracts were generated by subjecting the extract through two rounds of depletion. Each round of depletion involved incubation of 20 µL extract with either normal mouse IgG antibody-bound Protein G magnetic beads or mouse α-CstF64 (3A7 & 6A9)-bound Protein A magnetic beads for 2 hours. These beads were prepared by incubating either 80 µL of normal mouse IgG (Santa Cruz Biotechnology, sc-2025) or 40 µL each of mouse α-CstF64 antibodies 3A7 and 6A9 (both are gifts from MacDonald) with 1.2 mg Dynabeads Protein-G magnetic beads (Invitrogen) overnight at 4^{o}C. These beads were rinsed with PBS before incubation with nuclear extract. Following both rounds of depletion, the extract was incubated for another hour with naked Protein G magnetic beads to remove residual antibodies from the extract. Naked Protein G magnetic beads were generated by incubating 1.2 mg Protein G beads in 200 µL binding buffer (Na-phosphate buffer + 0.01% Tween-20) with no antibodies overnight at 4^{o}C.

To confirm the depletion of CstF64, 5 µL of normal mouse- and CstF64-depleted extracts were separated on a 10-20% gradient SDS-Tricine gel (Biorad). The proteins were transferred over to a nitrocellulose membrane and the membrane was blocked with Western Blocking Reagent (Roche, cat # 11921681001) and blotted for CstF64 using mouse monoclonal α-CstF64 3A7 antibody (3A7, gift from MacDonald, 1:333), for CPSF160 using rabbit α-CPSF160 (gift from Milcarek, 1:1,000), and for RNA polymerase II using mouse α-CTD (8WG16, Covance, 1:1,000). Immunodetection was performed with LI-COR Odyssey (Li-Cor Biosciences) using an

IRDye800CW coupled goat anti-mouse IgG secondary antibody (LI-Cor Biosciences, Cat# 926-32210, 1:10,000) for CstF64, IRDye800CW coupled anti-rabbit IgG secondary antibody (LI-Cor Biosciences, Cat# 926-32211, 1:10,000) for CPSF160, and IRDye680 coupled anti-mouse IgG secondary antibody (LI-Cor Biosciences, Cat# 926-32220, 1:10,000) for CTD of RNA Pol II.

Figures

Figure 1. ATP is required for efficient 3'-end cleavage.

(**A**) Nuclear extracts under processing conditions cannot support polyadenylation without addition of exogenous ATP. Immobilized dsxSV40L template was transcribed for 4 minutes, and the TECs were isolated with a magnet and washed with 1M KCl, 1% sarkosyl buffer (Strip) to remove all nucleotides and all transcription factors associated with the elongation complexes (Adamson et al. 2003). The TECs were then washed in low-salt transcription buffer to remove any residual salt and detergent from the previous wash. These TECs were then resuspended under processing conditions with or without exogenously added ATP and incubated for an hour. The RNA transcripts were then recovered by RNA extraction and subjected to oligo(dT) selection using a Poly(A) Purist MAG kit (Ambion). After one round of selection, "Bound" RNA and "Unbound" RNA were electrophoresed on a 5% urea-polyacrylamide gel and visualized by autoradioagraphy.

(**B**) 3'-end cleavage requires ATP. Washed TECs were generated and resuspended under processing conditions with or without 3'dATP and processed RNAs were allowed to accumulate for an hour. Processing (%) refers to the fraction of poly(A) site cleaved RNAs over the total amount of RNA extending past the poly(A) site. Corrected processing (%) refers to the difference of processing (%) at time x of α-amanitin incubation from the 0 min timepoint to correct for transcriptional pausing around the poly(A) site region.

(**C**) 3'-end cleavage cannot occur without addition of ATP or an ATP analog, AMPPCP. Experiment was carried out similar to A except transcription was carried out with cold nucleotides and TECs were processed in the presence of ATP, AMPPCP or no ATP cofactor. RNA

was extracted by TRIzol/chloroform, digested with DNaseI (Roche), hybridized at 65^{o}C, and subjected to RNase protection with RNase T1/A cocktail (Ambion). Probe was generated from pBluescript dsx exon2 $(GAA)_6$ (Rosonina et al. 2003). Digesting the resulting plasmid with NotI, and transcribing with T7 RNA polymerase generates a labeled protection probe (323 nt) that will result in protected fragments of 270 nt for uncleaved and 196 nt for cleaved RNA. The asterisk indicates a strong uncleaved band (~246 nt) that most likely reflects pausing of elongation complexes on 6 consecutive thymidines just right downstream of the GU-rich region of the SV40 late poly(A) signal. This "uncleaved" band also was included in the calculation of corrected processing (%).

Figure 1

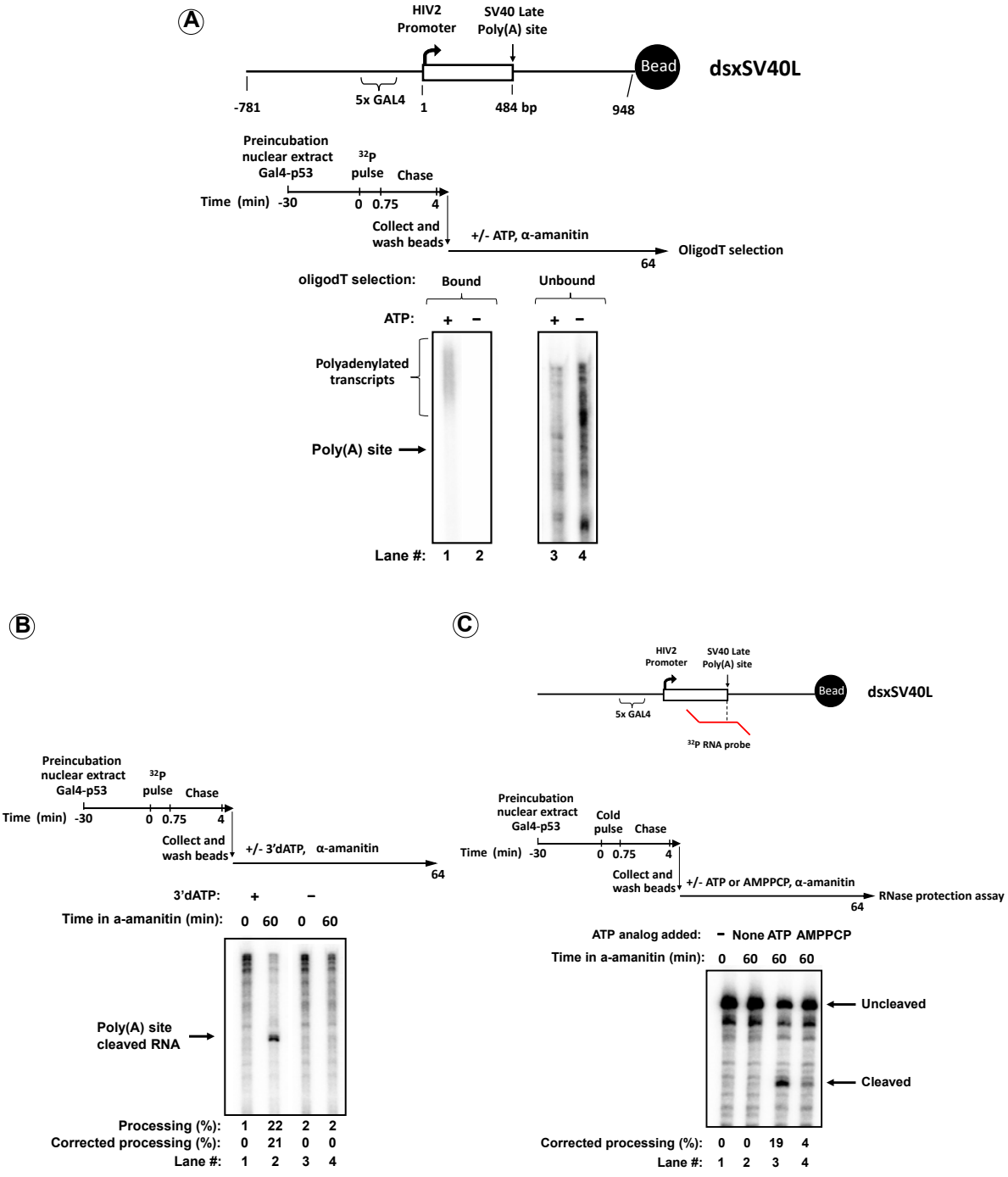

A
HIV2 Promoter
SV40 Late Poly(A) site
Bead
dsxSV40L
-781
5x GAL4
1
484 bp
948
Preincubation nuclear extract Gal4-p53
32P pulse
Chase
Time (min) -30 0 0.75 4
Collect and wash beads
+/- ATP, α-amanitin
64
OligodT selection
oligodT selection:
Bound
Unbound
ATP: + − + −
Polyadenylated transcripts
Poly(A) site
Lane #: 1 2 3 4
B
Preincubation nuclear extract Gal4-p53
32P pulse
Chase
Time (min) -30 0 0.75 4
Collect and wash beads
+/- 3'dATP, α-amanitin
64
3'dATP: + −
Time in a-amanitin (min): 0 60 0 60
Poly(A) site cleaved RNA
Processing (%): 1 22 2 2
Corrected processing (%): 0 21 0 0
Lane #: 1 2 3 4
C
HIV2 Promoter
SV40 Late Poly(A) site
Bead
dsxSV40L
5x GAL4
32P RNA probe
Preincubation nuclear extract Gal4-p53
Cold pulse
Chase
Time (min) -30 0 0.75 4
Collect and wash beads
+/- ATP or AMPPCP, α-amanitin
64
RNase protection assay
ATP analog added: − None ATP AMPPCP
Time in a-amanitin (min): 0 60 60 60
Uncleaved
Cleaved
Corrected processing (%): 0 0 19 4
Lane #: 1 2 3 4

Figure 2. A partial CPA assembles on the TEC during incubation in extract in absence of ATP and can participate in 3'-end cleavage upon addition of ATP.

(**A**) TECs incubated under processing conditions in absence of ATP can still process upon addition of ATP. In lane 1, TECs were formed, washed, and resuspended under processing conditions without ATP for an hour. 3'dATP was then added and the reaction was incubated for another hour. For comparison, separate batch of washed TECs were made and resuspended directly under processing conditions with (lane 2) or without ATP (lane 3).

(**B**) Majority of CPSF and CstF are depleted using antibodies directed to the 64 kDa subunit of CstF. CstF- and normal mouse- depleted extracts were generated by subjecting the extract through two rounds of depletion. Each round of depletion involved incubation of 20 µL of extract with either normal mouse IgG antibody-bound Protein G magnetic beads (1.2 mg) or mouse α-CstF64 3A7 and 6A9-bound Protein G magnetic beads for 2 hours. Following depletion, the extraction was incubated for another hour with naked Protein G beads to remove residual antibodies. 5 µL of each extract was loaded for western blot analysis. The blot was probed for 64 kDa subunit of CstF using mouse monoclonal α-CstF64 3A7 antibody (3A7, gift from MacDonald, 1:333), 160 kDa subunit of CPSF using rabbit polyclonal α-CPSF160 (Gift from Milcarek, 1:1,000), and for RNA polymerase II largest subunit using mouse α-CTD (8WG16, Covance, 1:1,000).

(**C**) CPSF and CstF are recruited as part of the CPA in absence of ATP cofactor and can functionally participate in 3'-end cleavage upon re-addition of ATP. In lane 1, TECs were formed and cleavage and polyadenylation factors were allowed to assemble in absence of ATP. TECs were then rinsed and returned to CstF-depleted extracts. To confirm that CstF-depleted

extracts cannot support 3’-end cleavage on its own, another preparation of washed TECs were resuspended in CstF- (lane 2) or normal mouse-depleted (lane 3) extracts under processing conditions with 3’dATP.

(**D**) CPA assembled in absence of ATP cofactor is not a complete apparatus. TECs with pre-assembled CPA were formed, rinsed with low-salt transcription buffer and returned under processing conditions with 3’dATP but no extract.

Figure 2

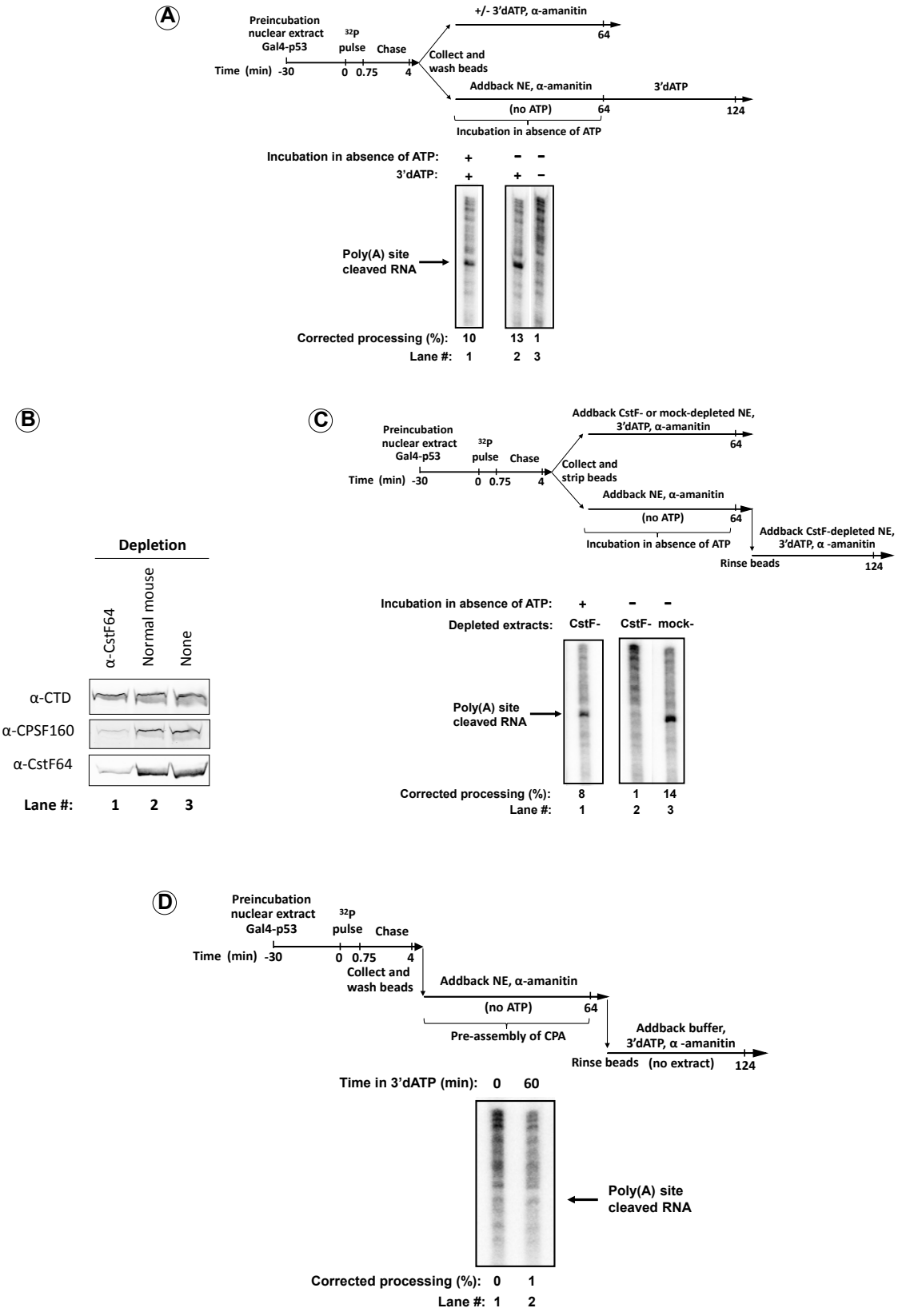
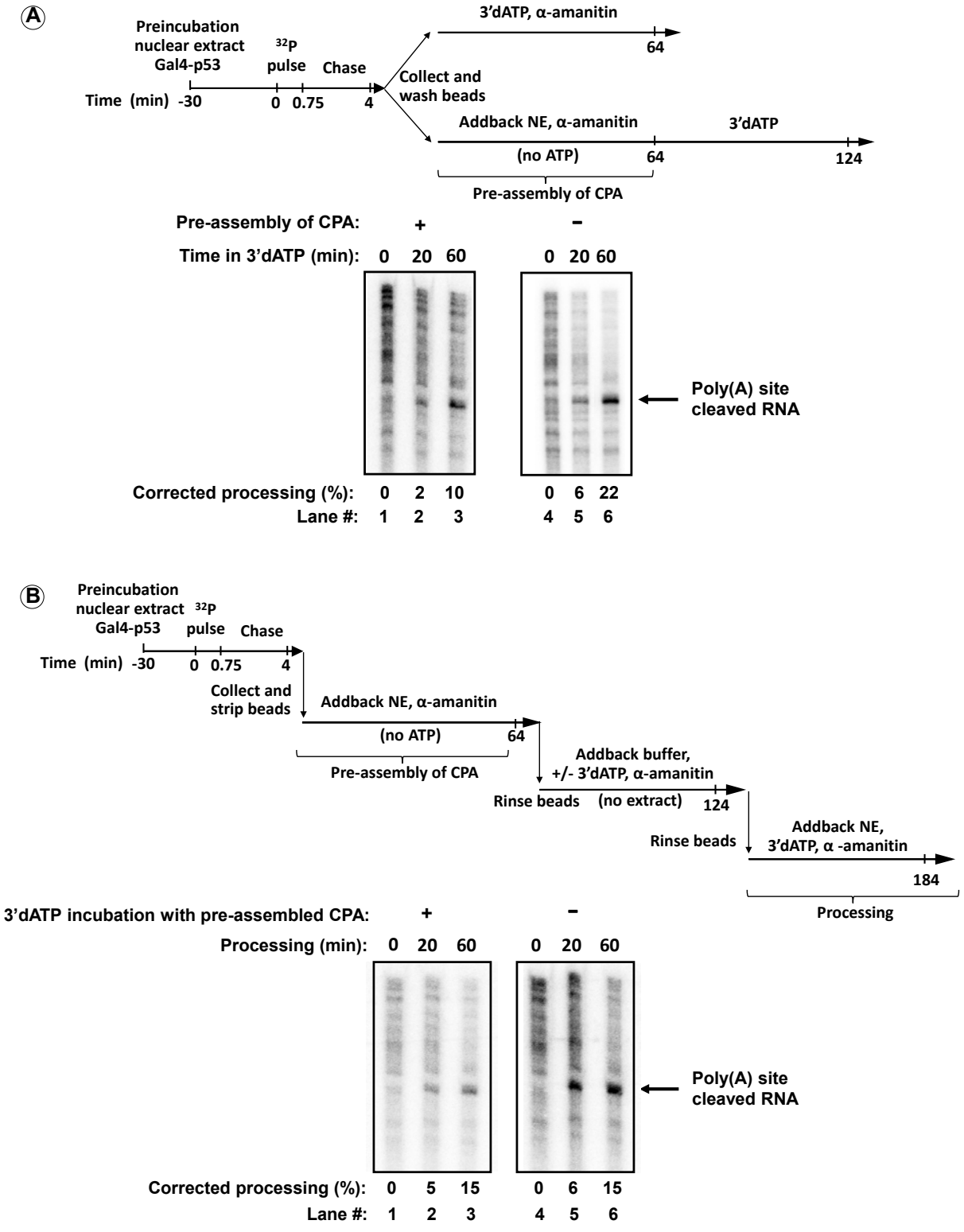

A
Preincubation nuclear extract Gal4-p53
32P pulse
Chase
+/- 3'dATP, α-amanitin
64
Time (min) -30
0
0.75
4
Collect and wash beads
Addback NE, α-amanitin
3'dATP
(no ATP)
64
124
Incubation in absence of ATP
Incubation in absence of ATP: + - -
3'dATP: + + -
Poly(A) site cleaved RNA
Corrected processing (%): 10 13 1
Lane #: 1 2 3
B
Depletion
α-CstF64
Normal mouse
None
α-CTD
α-CPSF160
α-CstF64
Lane #: 1 2 3
C
Preincubation nuclear extract Gal4-p53
32P pulse
Chase
Addback CstF- or mock-depleted NE, 3'dATP, α-amanitin
64
Time (min) -30
0
0.75
4
Collect and strip beads
Addback NE, α-amanitin
(no ATP)
64
Incubation in absence of ATP
Addback CstF-depleted NE, 3'dATP, α -amanitin
Rinse beads
124
Incubation in absence of ATP: + - -
Depleted extracts: CstF- CstF- mock-
Poly(A) site cleaved RNA
Corrected processing (%): 8 1 14
Lane #: 1 2 3
D
Preincubation nuclear extract Gal4-p53
32P pulse
Chase
Time (min) -30
0
0.75
4
Collect and wash beads
Addback NE, α-amanitin
(no ATP)
64
Pre-assembly of CPA
Addback buffer, 3'dATP, α -amanitin
Rinse beads (no extract)
124
Time in 3'dATP (min): 0 60
Poly(A) site cleaved RNA
Corrected processing (%): 0 1
Lane #: 1 2

Figure 3. The rate-limiting step of the 3'-end cleavage pathway occurs after the ATP-requiring step.

(**A**) TECs with pre-assembled partial CPA do not process more quickly than naked TECs when both are resuspended under processing conditions in presence of 3'dATP. TECs with pre-assembled partial CPA and naked TECs were generated and both were processed under identical conditions. Time points at 0 min, 20 min, and 60 min following addition of 3'dATP were taken to monitor how quickly processing occurs.

(**B**) Preincubation of TECs with pre-assembled partial CPA in 3'dATP in buffer does not increase rate of 3'-end cleavage. TECs with pre-assembled partial CPA were pre-incubated under processing conditions without nuclear extract with (lanes 1-3) or without 3'dATP (lanes 4-6). After an hour, the TECs were rinsed and resuspended under processing conditions with nuclear extract and 3'dATP.

Figure 3


A
Preincubation
nuclear extract
Gal4-p53
32P
pulse
Chase
Time (min) -30
0
0.75
4
Collect and
wash beads
3'dATP, α-amanitin
64
Addback NE, α-amanitin
(no ATP)
64
3'dATP
124
Pre-assembly of CPA
Pre-assembly of CPA: + -
Time in 3'dATP (min): 0 20 60 0 20 60
Poly(A) site
cleaved RNA
Corrected processing (%): 0 2 10 0 6 22
Lane #: 1 2 3 4 5 6
B
Preincubation
nuclear extract
Gal4-p53
32P
pulse
Chase
Time (min) -30
0
0.75
4
Collect and
strip beads
Addback NE, α-amanitin
(no ATP)
64
Pre-assembly of CPA
Addback buffer,
+/- 3'dATP, α-amanitin
Rinse beads (no extract)
124
Rinse beads
Addback NE,
3'dATP, α -amanitin
184
Processing
3'dATP incubation with pre-assembled CPA: + -
Processing (min): 0 20 60 0 20 60
Poly(A) site
cleaved RNA
Corrected processing (%): 0 5 15 0 6 15
Lane #: 1 2 3 4 5 6

References

Adamson, T.E., Shore, S.M., and Price, D.H. (2003). Analysis of RNA polymerase II elongation in vitro. Methods Enzymol. *371*, 264-275.

Adamson, T.E., Shutt, D.C., and Price, D.H. (2005). Functional coupling of cleavage and polyadenylation with transcription of mRNA. J. Biol. Chem. *280*, 32262-32271.

Beis, I. and Newsholme, E.A. (1975). The contents of adenine nucleotides, phosphagens, and some glycolytic intermediates in resting muscles from vertebrates and invertebrates. Biochem. J. *152*, 23-32.

Chao, L.C., Jamil, A., Kim, S.J., Huang, L., and Martinson, H.G. (1999). Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. Mol. Cell. Biol. *19*, 5588-5600.

Ghoshal, K. and Jacob, S.T. (1991). Ara-ATP impairs 3'-end processing of pre-mRNAs by inhibiting both cleavage and polyadenylation. Nuc. Acids Res. *19*, 5871-5875.

Hirose, Y. and Manley, J.L. (1997). Creatine phosphate, not ATP, is required for 3' end cleavage of mammalian pre-mRNA in vitro. J. Biol. Chem. *272*, 29636-29642.

Hirose, Y., and Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

Iyengar, M.R. (1984). Creatine kinase as an intracellular regulator. J. Muscle Res. Cell Motil. *5*, 527-534.

Kushmerick, M.J., Moerland, T.S., and Wiseman, R.W. (1992). Mammalian skeletal muscle fibers distinguished by contents of phosphocreatine, ATP, and Pi. Proc. Natl. Acad. Sci. USA *89*, 7521-7525.

Moore, C.L. and Sharp, P.A. (1985). Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell *41*, 845-855.

Nag, A., Narsinh, K., and Martinson, H.G. (2007). The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat. Struct. Mol. Biol. *14*, 622-669.

Rigo, F., Kazerouninia, A., Nag, A., and Martinson, H.G. (2005). The RNA tether from the poly(A) signal to the polymerase mediates coupling of transcription to cleavage and polyadenylation. Mol. Cell *20*, 733-745.

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

Shi Y., Di Giammartino D.C., Taylor D., Sarkeshik A., Rice W.J., Yates J.R., Frank J., and Manley J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Mol. Cell *33*, 365-376.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell *52*, 731-742.

Tran, D.P., Kim, S.J., Park, J.J., Jew, T.M., and Martinson, H.G. (2001). Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. Mol. Cell. Biol. *21*, 7495-7508.

Zarkower, D., Stephenson, P., Sheets, M., and Wickens, M. (1986). The AAUAAA sequence is required both for cleavage and polyadenylation of simian virus 40 pre-mRNA in vitro. Mol. Cell Biol. *6*, 2317-2323.

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol. Mol. Biol. Rev. *63*, 405-445.

Chapter 5

Preliminary study: Transcriptional activators may contribute to 3'-end processing by regulating promoter-specific and nonpromoter-specific transcription

Abstract

Transcription factors that are recruited at the promoter are known to participate in coupling 3'-end processing to transcription. How do events that occur at the promoter at the upstream end of a gene affect 3'-end processing of transcripts that occurs at the other end? To address this, we sought to investigate the mechanism by which transcriptional activators control 3'-end cleavage levels using a previously reported in vitro system that can support 3'-end processing that is functionally coupled to transcription. We demonstrated that while transcriptional activators can dramatically enhance transcription, they did not alter 3'-end processing levels significantly in vitro. At face value, these results suggest that the mechanism by which transcriptional activators affect 3'-end processing cannot be supported in an in vitro system. Interestingly, we observed that transcription initiates randomly in the absence of transcriptional activator to direct promoter-specific transcription. We discuss the possibility that transcriptional activators may modulate cleavage levels by redirecting initiation from cryptic promoters to promoter-specific sites.

Introduction

During transcription, mRNA is synthesized and has to be accurately processed (capped, spliced and, cleaved and polyadenylated) in order for it to be exported out into the cytoplasm for translation. There is considerable evidence that these steps of mRNA processing occur cotranscriptionally, and these processes are interrelated through a series of mutually synergistic interactions (Maniatis et al. 2002). Here, we focus on how transcription can affect 3'end processing (cleavage and polyadenylation).

Within the past two decades, multiple mechanisms have been proposed to explain how transcription can affect 3'-end processing (Bentley 2005). One of the first mechanisms came from a report demonstrating that the CTD (carboxy terminal domain) of RNA polymerase II is an integral component in linking transcription to 3'-end processing (McCracken et al. 1997). It was suggested that the CTD may act as a loading platform to recruit the processing machinery to act at the site of processing, the poly(A) signal (see Chapter 1 for review on poly(A) signal). However, the proposed recruitment function of the CTD is made complicated by the fact that the CTD is a required participant of and can stimulate a cleavage reaction uncoupled from transcription (Hirose et al. 1998). Therefore, it is not entirely clear how the CTD may function to couple 3'-end processing to transcription.

In addition to the CTD, there have been a number of reports that transcription factors can assist in the recruitment of processing factors to the elongating polymerase to act on the processing signal immediately after extrusion. The first report that introduced this mechanism proposed that CPSF was recruited to the promoter by TFIID and is later transferred to the elongating polymerase (Dantonel et al. 1997). Consistent with this idea, chromatin immunoprecipitation

studies have localized 3'-end processing factors to the promoter regions of yeast and mammals (Licatalosi et al. 2002, Calvo et al. 2005, Rozenblatt-Rosen et al. 2009, Glover-cutter et al. 2008). Additional transcription factors have been reported to play a role in recruitment of 3'-end processing factors. Transcriptional activators have been found to increase not only transcription but also splicing and 3'-end processing in vivo (Rosonina et al. 2003). Blencowe and colleagues were able to demonstrate a correlation between the ability of transcriptional activators to stimulate 3'-end processing and the strength of these activators to bind PSF, polypyrimidine tract binding protein-associated splicing factor (Rosonina et al. 2005). Moreover, overexpression of PSF bypasses the requirement for transcriptional activator in enhancing 3'-end cleavage levels (Rosonina et al. 2003). Therefore, these studies support the idea that transcriptional activators may directly or indirectly recruit PSF to facilitate pre-mRNA processing.

More recently, transcriptional activator was shown to enhance 3'-end processing in vitro (Nagaike et al. 2011) and this effect is consistent with the ability of transcriptional activator Gal4-VP16 to interact directly with the PAF1 complex (PAF1c). PAF1c is an elongation complex and has been previously reported to associate with components of the polyadenylation machinery, including CPSF and CstF (Rozenblatt-Rosen et al. 2009). Therefore, transcriptional activators may also enhance 3'-end processing by recruiting the PAF1 complex to the elongating polymerase which facilitates recruitment of CPSF and CstF to the transcript.

In this study, we sought to investigate the mechanism by which transcriptional activators control 3'-end cleavage levels using a previously reported in vitro system that can support 3'-end processing that is functionally coupled to transcription. However, we were surprised to

find that the majority of transcription in the absence of transcriptional activator was nonpromoter-specific transcription. This prevented us from drawing any conclusion based on pulse-chase assays or any assay that relied only on RNA length to determine processing efficiency. Using RNase protection assay, we found that transcriptional activator Gal4-p53 activated transcription robustly (~47-fold) but only enhanced 3'-end processing slightly at best. Based on the above findings, it appears that the mechanism by which transcriptional activators enhance 3'-end cleavage levels in vivo may not be supported in our in vitro system. We later discuss the possibility that transcriptional activators may modulate cleavage levels by redirecting initiation from cryptic promoters to promoter-specific sites.

Results

To study the mechanism of how transcriptional activators enhance 3'-end processing, we tethered various transactivation domains to the HIV2 promoter on the DNA template Gal5-HIV2dsxΔInt(+ESE) (see Figure 1, top diagram) and asked if they enhanced cleavage at the poly(A) site. This template has been previously used to study activator-dependent enhancement of 3'-end processing in vivo (Rosonina et al. 2003, Rosonina et al. 2005). This template contains exon 3 and exon 4 of doublesex with a strong SV40 late poly(A) signal, whose expression is placed under the control of an HIV2 promoter and five upstream binding sites for the yeast Gal4-DNA binding domain (DBD). The expression of this pre-mRNA can therefore be activated through the addition of Gal4-DBD fusion proteins containing different activation domains.

To verify that the effect of transcriptional activators on 3'-end processing can be reproduced in our in vitro system, we carried out an in vitro transcription/coupled processing assay using different Gal4-DBD fusion proteins. The DNA template was pre-incubated with nuclear extract with either no activator, Gal4-p53, Gal4-VP16, or Gal4-AH to form preinitiation complexes (see Figure 1 experimental outline). Transcription was initiated with a pulse of [α-^{32}P]-CTP and then chased with a high concentration of unlabeled CTP. Transcription was stopped with the addition of α-amanitin and 3'dATP (to block polyadenylation) and poly(A) site cleaved RNA was allowed to accumulate for an hour. The percentage poly(A) cleaved and transcriptional fold activation are given below the gel in Figure 1. Figure 1 shows that Gal4-p53 (lane 2) increased transcription by ~7-fold (compare "transcriptional activation" between lanes 1 and 2) and enhanced 3'-end cleavage efficiency by ~20 fold compared to the no activator control (lane 1

enhanced). This effect is not specific to Gal4-p53 because Gal4-VP16, and Gal4-AH also gave similar results (Fig. 1, compare lanes 4 and 5 to lane 3 enhanced, respectively). We concluded, at the time, that the transcriptional activators appeared to be enhancing 3'-end processing in our in vitro system. However, additional controls were needed to make sure that our interpretations were correct.

For one, we needed to validate that the transcripts are initiating from the HIV2-promoter. To do this, we carried out an oligo-cutting assay. By adding short DNA oligos that are complementary to the RNA transcripts, RNaseH endogenous to the nuclear extract will cut the resulting DNA:RNA hybrid to generate transcripts of discrete length. Therefore, the position of transcription initiation can be mapped accurately based on the length of the cut transcripts. To do this, transcription was carried out exactly as before except an oligo directed to cut 376 nt downstream of the presumed transcription start site was added during the preincubation (Figure 2, experimental flow chart diagram). The percentage oligo-cut is listed given below the gel in Figure 2. Figure 2 shows that the majority of transcripts generated in the presence of Gal4-p53 were cut and reduced to the predicted transcript length of 376 nt when the 376 oligo was added to the reaction (compare lane 2 to lane 1). In contrast, very little cutting was observed when transcription was carried out in the absence of activator (Fig. 2, compare lane 5 enhanced to lane 4 enhanced). In fact, most of the long transcripts beyond the oligo-cut site, are still present, and are unaffected by the oligo (Fig.2, lane 5 enhanced). An obvious interpretation of these results is that transcription is not initiating from the presumed HIV2 promoter, but is instead initiating elsewhere on the template. Since we do not see discrete cut bands beyond the 376 nt band but instead it is a smear of RNA transcripts (Fig.2 lane 5

enhanced, RNA longer than 376 nt), this could mean that the majority of nonpromoter-specific transcription may be initiating and transcribing in the reverse direction (which would be not be cut by RNaseH since the oligo would not be complementary) or it can be initiating in the correct direction but initiating either very far upstream such that it has not transcribed down to oligo-cutting site or it's initiating randomly everywhere and therefore even when cut by RNaseH, the RNAs are not giving a discrete oligo-directed cut band on the gel.

The presence of nonpromoter-specific transcription is a problem as it interferes with our calculation of processing efficiency for unactivated transcription and therefore our interpretations. If nonpromoter specific transcription is occurring, long transcripts may be generated without ever having transcribed a poly(A) signal and since our processing efficiency calculations assume that all transcripts longer than 484 nt (length of poly(A) cleaved RNAs initiated from HIV2 promoter; see Figure 1 top diagram) carries a poly(A) signal and has not been processed, the calculated processing efficiency for unactivated transcription would be a gross underestimate. Because our originally conclusion that transcriptional activators enhance 3'-end processing is derived from the difference in processing efficiencies between unactivated and activated transcription, the existence of nonpromoter specific transcription means we cannot draw any conclusions using this pulse-labeling method of RNA transcripts and we cannot determine if RNA transcripts are cleaved at the poly(A) site simply based on transcript length. Therefore, we investigated further to confirm our suspicion that nonpromoter-specific transcription occurs in absence of a transcriptional activator to direct promoter-specific initiation.

The fact that nonpromoter-specific transcription is so robust (Figure 2, lane 5 enhanced, long transcripts above oligo-cut RNA) is surprising to us given that RNA polymerase II transcription has been generally viewed to be quite specific. To rule out the possibility that we were observing transcription by another RNA polymerase enzyme, we carried out an α-amanitin sensitivity experiment. We found that the transcripts observed are indeed generated from RNA polymerase II since addition of 1μg/mL of α -amanitin completely eliminated transcription in both activated and unactivated transcription reactions (Fig. 2, lane 3 and lane 6 enhanced, respectively). It has been reported that a concentration of 1 μg/mL α-amanitin only inhibits RNA polymerase II transcription but not RNA polymerase I or III (Weinmann et al. 1974). Figure 2 lanes 7 and 8 shows the same oligo-cutting experiment carried out with a different extract. However, the results are the same; transcripts synthesized in absence of added activator are less prone to oligo-directed RNaseH cutting.

Since the oligo-cutting assay is a measure of sensitivity to oligo-directed RNaseH cutting, it is possible that the position where the oligo was directed to bind may be blocked by proteins (that are apparently removed by addition of activator). Another possibility is that the RNA generated in the absence of activator may fold into a secondary structure not suited to oligo-binding at that particular site. Because of these uncertainties, we used a different approach to examine whether transcription is initiating from the HIV2 promoter in the absence of activator. Our approach involved linearizing the Gal5-HIV2dsxΔInt(+ESE) plasmid template by EcoR1 digestion so that transcription initiating from the HIV2 promoter will generate runoff transcripts of 177 nt in length from these templates. Therefore, we carried out in vitro transcription on EcoR1-digested and undigested plasmid templates for comparison. Figure 3, lane 2 shows that

the majority of elongation complexes generated in the presence of activator yielded transcripts of the predicted run-off length, indicating that transcription is occurring mainly from the HIV2 promoter. However, Figure 3 lane 4 shows that elongation complexes generated in absence of an activator gave very little runoff transcripts. In fact, the majority of unactivated transcription (transcripts longer than 177 nt) were unaffected by EcoR1 digestion of the template (compare Figure 3 lane 4 with lane 3). Therefore, the EcoR1 digestion experiment (Fig. 3, lane 4) supports the idea that most of the transcription occurring in the absence of an activator is initiating from cryptic promoters. Interestingly, we noticed that most transcripts longer than the EcoR1 runoffs run as a smear (Fig 3, lane 4) rather than transcripts of discrete length, with the exception of one small population of RNA transcripts (Fig. 3, lane 4, labeled by asterisk). This suggests that either nonpromoter-specific transcription initiates randomly all over the template or that the bulk of nonpromoter-specific transcription is initiating so far upstream of the EcoR1-cut site that it is unaffected by the linearization of the plasmid template by EcoR1. While the former explanation is the simplest explanation, and by far the most likely, we cannot rule out the latter. We believe that some transcription may be also occurring in the reverse direction since a group of transcripts approximately ~100 nt longer than EcoR1-runoff transcripts appear on EcoR1-linearized templates (Figure 3, lane 4 labeled by asterisk) but is not identified in the oligo-cutting experiment as a discrete band (Figure 2, lane 5 enhanced). It is unlikely that these transcripts are initiating ~100 bp upstream of the HIV2 promoter and transcribing in the right direction because the oligo-cutting experiments did not reveal a band of transcripts ~100 nt longer than the oligo-cut transcripts of 376 nt (see Figure 2, lane 5 enhanced and lane 8). These

results (Figure 2 and 3) establish that nonpromoter-specific transcription occurs in absence of activator.

As a control, we were able to confirm that this non-HIV2 promoter specific transcription is not a result of transcription of the genomic DNA in our crude nuclear extract, since an equivalent transcription reaction carried out in the absence of a reporter template gave almost no transcription (Fig. 3, lane 5). It is interesting to note that non-HIV2 transcription was reduced when Gal4-p53 was present (compare transcripts longer than 177 nt in Figure 3 lane 4 with lane 2). This suggests that Gal4-p53 is reducing nonpromoter-specific transcription while activating promoter-specific transcription. Note that the oligo-cutting experiments (compare transcripts longer than 376 nt in Fig. 2 lanes 7 and 8) are also consistent with this observation that transcriptional activator reduced transcription from cryptic promoters. These results are consistent with the idea that transcriptional activators are activating transcription by directing the recruitment of transcription initiation factors to the HIV2 promoter and reducing the availability of initiation factors for use at cryptic promoters. In any case, because of the high contaminating transcription signal from cryptic promoters, we concluded that a pulse-chase assay cannot accurately measure processing efficiency of unactivated transcription.

Therefore, we proceeded to analyze the effect of transcriptional activators on 3'-end processing via RNase protection instead. We designed the labeled RNA probe to bind complementary to the region of the SV40 late poly(A) signal (Figure 4, top diagram). This allows us to examine specifically the effect of activators on RNA transcripts that contain a poly(A) signal. Figure 4 shows that Gal4-p53 activates transcription very robustly, by 47-fold (compare uncleaved band in lanes 1 and 3). This is in stark contrast to the 7 fold transcriptional activation by Gal4-p53

reported in Figure 1 (compare “transcriptional activation” between lane 1 and 2). This difference most likely reflects the fact that a lot of the nonpromoter-specific transcription has not transcribed the poly(A) signal (which the probe selective protects) and is therefore not picked up by the RNase protection assay. This quantitation (47-fold transcriptional activation) takes into account the fact that the “minus Gal4-p53” samples (Figure 4 lanes 3 and 4) are the results of an 8-fold scaled-up reaction. Surprisingly, under conditions in which Gal4-p53 strongly activates transcription, we found that Gal4-p53 did not enhance poly(A) cleavage very dramatically (compare Figure 4 lanes 2 and 4). Based on quantitation, the effect of Gal4-p53 addition increased poly(A) cleavage efficiency only slightly at best. From this result, we conclude that transcriptional activators do not enhance 3’-end processing in our in vitro system.

Discussion

Using an in vitro system that can support 3'-end processing that is functionally coupled to splicing and to transcription (Rigo et al. 2005; Rigo et al. 2008; Rigo et al. 2009), we found that the transcriptional activator Gal4-p53 activated transcription very strongly (~47-fold, see Figure 4) but enhanced 3'-end processing efficiency only slightly at best (see Figure 4). These results suggest that transcriptional activators do not enhance 3'-end processing strongly in vitro. What do these results mean in light of Blencowe's study showing that transcriptional activators, including Gal4-p53, enhance 3'-end processing in vivo (Rosonina et al. 2003)? One possibility is that the mechanism by which transcriptional activators control 3'-end processing in vivo is not supported in vitro. This may be because our in vitro system lacks certain components that are required for transcriptional activators to enhance processing. One obvious possibility is that transcriptional activators may require higher order DNA structure or chromatin templates in order to stimulate 3'-end cleavage. This would suggest a role for histone-modifying proteins in transcriptional activator-dependent stimulation of 3'-end cleavage. Another possibility stems from a report by Manley's group (Nagaike et al. 2011), in which they use a crowding agent in their in vitro system called polyethylene glycol (PEG). They were able to demonstrate under their in vitro conditions that transcriptional activators can enhance 3'-end processing by few-to-several fold (Nagaike et al. 2011). It's possible that the crowding agent PEG used in their studies (which is not used in our assays) may stabilize or promote important protein-protein or protein-RNA interactions that are necessary for the ability of transcriptional activators to enhance 3'-end processing. Further investigation will help us to determine precisely the role of

transcriptional activators on 3'-end processing and the ability of the in vitro system to study these effects.

While we did not observe a dramatic effect of transcriptional activators on 3'-end processing, we did find surprisingly however, that Pol II transcription initiates promiscuously at cryptic promoters on the template in the absence of a transcriptional activator. This is based on the fact that the unactivated transcripts were not sensitive to oligo-directed RNaseH cutting at 376 nt downstream of the HIV2 transcription start site (Figure 2 lane 5 enhanced; Figure 2 lane 8) and did not generate very much run-off transcripts when the DNA template was digested by EcoR1 177 basepairs downstream of the transcription start site (Figure 3, lane 4). Moreover, nonpromoter-specific transcription appears to be initiating randomly. First of all, our oligo-cutting experiments did not reveal the presence of multiple RNA bands of discrete length in response to addition of an oligo directing RNaseH to cut the transcripts at 376 nt downstream of the HIV2 transcription starts site (Fig. 2, lane 5 enhanced and Fig. 2, lane 8). If cryptic transcription was occurring at specific sites, one or two sites upstream of the HIV2 promoter for example, then we would expect to see discrete RNA bands longer than 376 nucleotides. Instead, the long RNA transcripts run as a smear. Second, transcription on EcoR1-cut templates also did not reveal multiple nonpromoter-specific run-off transcripts of discrete length (Figure 3, lane 4) with the exception of one band (Figure 3, lane 4 asterisk). If cryptic promoter-specific transcription was initiating primarily at a few sites, we would expect to see the complexes run off the EcoR1 site and yield transcripts of specific length whether or not it is transcribing the forward direction or the reverse direction. Rather, most of the long RNA runs as a smear (Figure 3, lane 4 transcripts longer than 177 nt). Although another possibility is that transcripts

may be initiating very far upstream such that the complexes have not transcribe far enough down to run off the template, random transcription is more likely since it requires less assumptions. It is also interesting to note that the addition of transcriptional activator Gal4-p53 reduced nonpromoter-specific transcription and activated HIV2 promoter-specific transcription (see Figure 2, compare transcripts longer than 376 nt in lanes 7 and 8; see Figure 3 compare transcripts longer than 177 nt in lanes 2 and 4). These results highlight the important role of transcriptional activators to direct proper transcription.

Consistent with our results showing that Pol II transcription initiates promiscuously at cryptic promoters in absence of transcription activator, Manley and colleagues also observed that the bulk of transcription in their assays had persisted even when the TATA box was mutated, indicating robust promoter non-specific transcription (Nagaike et al. 2011, Figure S1). Also in agreement with our results, they observed that addition of transcriptional activator Gal4-VP16 reduced promoter non-specific transcription and this effect was dependent on presence of Gal4 sites (Nagaike et al. 2011, Figure S1).

In consideration of our results on nonpromoter-specific transcription, our RNase protection results raise a few interesting questions or ideas. If unactivated transcription initiates randomly, then our RNase protection results, which demonstrated that unactivated transcripts are processed at the 3'-end as efficiently as activated transcription (Figure 4, compare lane 4 with lane 2), would suggest that transcripts generated from cryptic promoters process almost as efficiently as promoter-specific transcription. This is surprising given that promoters generally contain specific cis-acting elements (while cryptic promoters lack them) that are important for directing the necessary recruitment of transcription factors, processing factors,

and CTD-modifying kinases that may be required for efficient processing of the RNA transcript. For example, Kornblihtt and colleagues proposed the idea that elements at the promoter can dictate the fate of the transcript by demonstrating that promoter swapping affects the efficiency fibronectin EDI exon inclusion (Cramer et al. 1997). They later proposed that the promoter was influencing exon inclusion by assisting in the recruitment of SR protein SF2/ASF to the exon splicing enhancer (ESE) element (Cramer et al. 1999). Therefore, if these interpretations are correct, then our results raise an interesting question regarding the extent of the promoter in dictating the fate of the transcript in terms of cleavage and polyadenylation. However, further experiments would be required to draw concrete conclusions.

Another interesting possibility is if unactivated transcription is not random but if instead, most of the nonpromoter-specific transcription was derived from the wrong strand or that nonpromoter-specific transcription initiated very far upstream such that it was unable to transcribe the SV40 late poly(A) signal. In that case, our RNase protection assay results, which examined processing efficiencies of poly(A) signal-containing transcripts would not have detected non-promoter specific transcription. One indication that this may be true is the fact that our RNase protection assay showed that Gal4-p53 activated transcription by 47 fold (Figure 4, compare "transcriptional activation" in lanes 1 and 3) while Gal4-p53 appeared to activate transcription only by 7-fold in pulse-chase assays (Figure 1, compare "transcriptional activation" between lane 1 and 2). This difference can be explained if our RNase protection assay did not detect much of the nonpromoter-specific transcription that is detected by our pulse-chase assays. In other words, transcriptional activation in the RNase protection assay reflects primarily an increase in promoter-specific transcription while "transcriptional activation" our

pulse-chase assays refers to increase in overall transcription (regardless of whether transcription is promoter-specific or nonpromoter-specific). In that case, our results would not conflict with the notion that cryptic promoter transcription leads to inefficiently processed transcripts. Given the robustness of nonpromoter-specific transcription, there is a possibility that transcriptional activators may contribute to 3'-end processing through the control of nonpromoter-specific and promoter-specific transcription, by reducing inefficiently-processed non-promoter specific transcription and increasing efficiently-processed promoter-specific transcription. This could potentially account, at least in part, for Manley's finding that transcriptional activators can enhance 3'-end processing by few-to-several fold in their in vitro system (Nagaike et al 2011) if we assume that under their experimental conditions that cryptic promoter transcription was contributing a lot of unactivated transcription in their RNase protection assay. However, at this point, it is purely conjecture.

In any case, our results indicate that caution must be taken when studying transcriptional activators in vitro. Typical in vitro transcription experiments involve labeling the transcript with radioactive nucleotides and extracting information on processing efficiency based on length of processed and unprocessed transcripts. However, this is made nearly impossible because of the contamination from promoter non-specific transcription (Figure 2, lane 5 enhanced; Figure 2, lane 8; Figure 3 lane 4). Instead, we believe the use of RNase protection assays that can give specific information and differentiate between promoter- and nonpromoter-specific transcripts is necessary to obtain interpretable data.

Materials and Methods

Templates

Gal5-HIV2dsxΔInt(+ESE) was a gift from Benjamin Blencowe (Rosonina et al. 2003).

To generate EcoR1-cut Gal5-HIV2dsxΔInt(+ESE) templates for in vitro transcription, 18 µg of Gal5-HIV2dsxΔInt(+ESE) plasmid was digested with 20 units of EcoR1 (New England Biolabs) for 3 hours at 37^{o}C. The enzyme was inactivated by heating the reaction at 65^{o}C for 20 minutes. The linearized template was then separated on a 0.7% agarose gel and gel-extracted using the QIAquick Gel Extraction Kit. The linearized product was verified by gel-electrophoresis on 0.7% agarose gel to be free of un-cut plasmids, as judged by the presence of a fast migrating supercoiled band (data not shown).

Purification of Gal4-p53 activator

G4-p53 construct (gift from James Goodrich) was expressed in the E. coli BL21 strain. T7 promoter expression was induced in these cells with 0.4 mM IPTG for 2 hours. Cells were collected, resuspended in TGED buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 20% glycerol, 0.1 mM EDTA, 1 mM DTT, 1mM PMSF) containing 5 mM $MgCl_2$, and lysed via sonication. Cell lysates were incubated with ANTI-FLAG M2 affinity gel (Sigma, A2220) for 3 hours at 4^{o}C. The affinity gel was washed with TGED buffer and TGED buffer containing 1M NaCl. The affinity gel was transferred to a Poly-Prep Chromatography Column (Biorad) and Gal4-p53 proteins were released with TGED buffer containing 0.2 mg/mL FLAG peptide at room temperature. The fractions were collected and immediately stored at -80^{o}C. Peak fractions containing Gal4-p53, as determined by SDS-PAGE and sypro ruby staining, were later pooled together, re-aliquoted and stored at -80^{o}C.

Purification of Gal4-AH activator

pGEX3X-Gal4-AH (gift from Professor Jay Gralla, UCLA) was expressed in the E. coli BL21 strain. T7 promoter expression was induced in these cells with 1 mM IPTG for 3 hours. Cells were collected, washed with PBS and resuspended in lysis buffer (1X PBS, 1 mM DTT, 1 mM of PMSF) containing 1.2% sarkosyl, and lysed via sonication. Cell lysates were diluted 8-fold with lysis buffer to reduce sarkosyl concentration before incubation with glutathione-sepharose beads (GE Healthcare Lifesciences) for 3 hours at 4^{o}C. The beads were first washed with 10 bed volumes of lysis buffer 3 times and then 10 bed volumes of wash buffer (1X PBS, 1 mM DTT) 7 times. Gal4-AH was then recovered by incubating the beads in 10 U/mL human thrombin (GE Healthcare Lifesciences, Cat# 27-0846-01) for 1 hour at 4^{o}C. The supernatant was then dialyzed into Buffer D (20 mM HEPES pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, and 0.5 mM PMSF) overnight at 4^{o}C.

In vitro transcription (Coupled processing assay)

Hela nuclear extract was prepared as described (Tran et al. 2001).

A typical pulse-chase assay began with 3 µL of nuclear extract that was mixed with Gal4-VP16 (Protein One, Cat# P1019), Gal4-AH, Gal4-p53 or appropriate activator storage buffer (see above), anti-RNase (Ambion), DTT, $MgCl_2$, sodium citrate, DNA plasmid or linearized DNA template, and water up to 7.9 µL. Volume of transcriptional activator addition was selected based on a titration experiment determining the minimum volume required to achieve maximum % poly(A) cleavage in a typical coupled processing experiment. The mixture was preincubated at 30^{o}C for 30 min and then pulsed with 3 µL containing 20 µCi of [α-^{32}P] CTP (Perkin Elmer, 800 Ci/mmol), nucleotide triphosphates and creatine phosphate. Then, 2.6 µL of

chase mix was added containing a high concentration of nonradiolabeled CTP. Next, 1 µL of α-amanitin and 3'dATP were added. Final concentrations in a standard pulse-chase assay were as follows: 5.2 % glycerol, 5.2 mM HEPES (pH 7.9), 2.8 mM Tris-HCl (pH 7.4), 8.3 mM NaCl, 21 mM KCl, 47 µM EDTA, 1.9 mM DTT, 76 µM PMSF, 10 U Anti-RNase, 4.3 mM $MgCl_2$, 3.4 mM Citrate (pH 6.7), 0.16 pmol DNA template, 345 µM ATP, 172 µM UTP and GTP, 17 mM creatine phosphate, and 1.7 mM CTP, 34 µg/mL α-amanitin and 372 µM 3'dATP. Oligos, when used, was added during the preincubation at a final concentration of 3 ng/µL. For α-amanitin sensitivity assay (Figure 2, lane 3 and 6), α-amanitin was added during preincubation such that the final concentration would be 1 µg/mL during the pulse step.

The reaction was terminated by the addition of a stop solution: 65 µL of 10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS, and 100 µg proteinase K (Ambion). RNA was extracted with 350 µL TRIzol (Invitrogen) and 70 µL chloroform and then precipitated with 4 µL of 5 mg/mL glycogen (Ambion) and 350 µL isopropanol (30 min, room temperature), and finally ran on a 5% polyacrylamide gel. After electrophoresis, results were recorded and analyzed by using a PhosphorImager with ImageQuant software (Molecular Dynamics).

Figures

Figure 1. Transcriptional activators appears to enhance transcription and 3'-end processing. Circular plasmid DNA was transcribed for 4 minutes before transcription was halted by the addition of α-amanitin and 3'dATP, to allow poly(A) site cleaved RNAs to accumulate for 60 min. % Poly(A) cleaved is calculated as the fraction of poly(A) site cleaved RNA over the total amount of RNA extending past the poly(A) site. Transcriptional activation is calculated as the fold difference in total transcripts carrying a poly(A) signal and longer (poly(A) site cleaved RNA and longer). The experiment in lanes 3-5 was carried out in a final reaction volume of 13.5 µL instead of 14.5 µL (difference of 1 µL of H_2O).

Figure 1

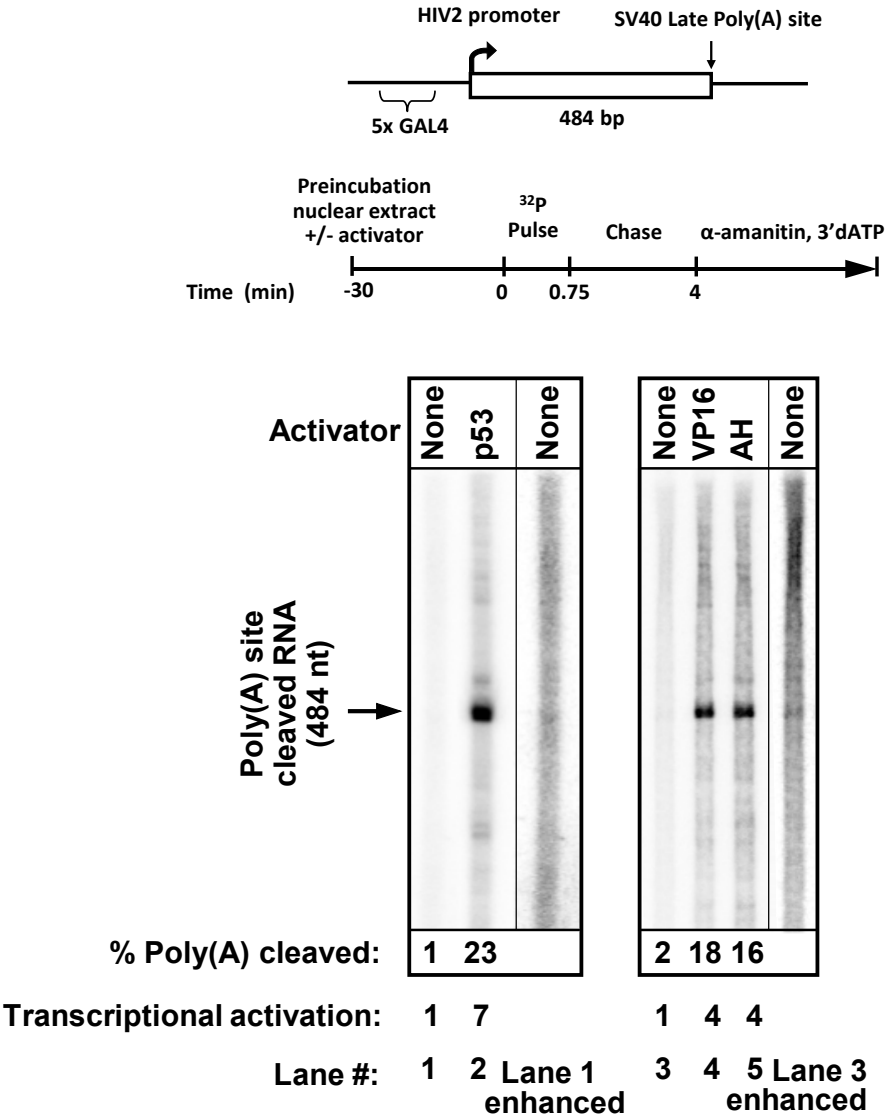

HIV2 promoter
SV40 Late Poly(A) site
5x GAL4
484 bp
Preincubation
nuclear extract
+/- activator
32P
Pulse
Chase
α-amanitin, 3'dATP
Time (min)
-30
0
0.75
4
Activator
None
p53
None
None
VP16
AH
None
Poly(A) site
cleaved RNA
(484 nt)
% Poly(A) cleaved:
1 23
2 18 16
Transcriptional activation:
1 7
1 4 4
Lane #:
1 2 Lane 1
enhanced
3 4 5 Lane 3
enhanced

Figure 2. Majority of transcripts synthesized in absence of activator is insensitive to oligo-directed RNaseH cutting.

In lanes 3 and 6, α-amanitin was added to the reaction during preincubation to examine the sensitivity of transcription to the drug. Following 3.25 min of chase, transcription was terminated immediately with stop solution instead of incubating for an additional hour. Lanes 7 and 8 were carried out exactly as lanes 2 and 5, respectively, except that a different preparation of nuclear extract was used. % oligo cut is calculated as the fraction of oligo cut RNA (376 nt) over the total amount of RNA at or longer than 376 nt. The oligonucleotide name (376 oligo) refers to the distance from the HIV2 transcription start site to the predominant RNaseH cutting site (Wu et al. 1999).

Figure 2

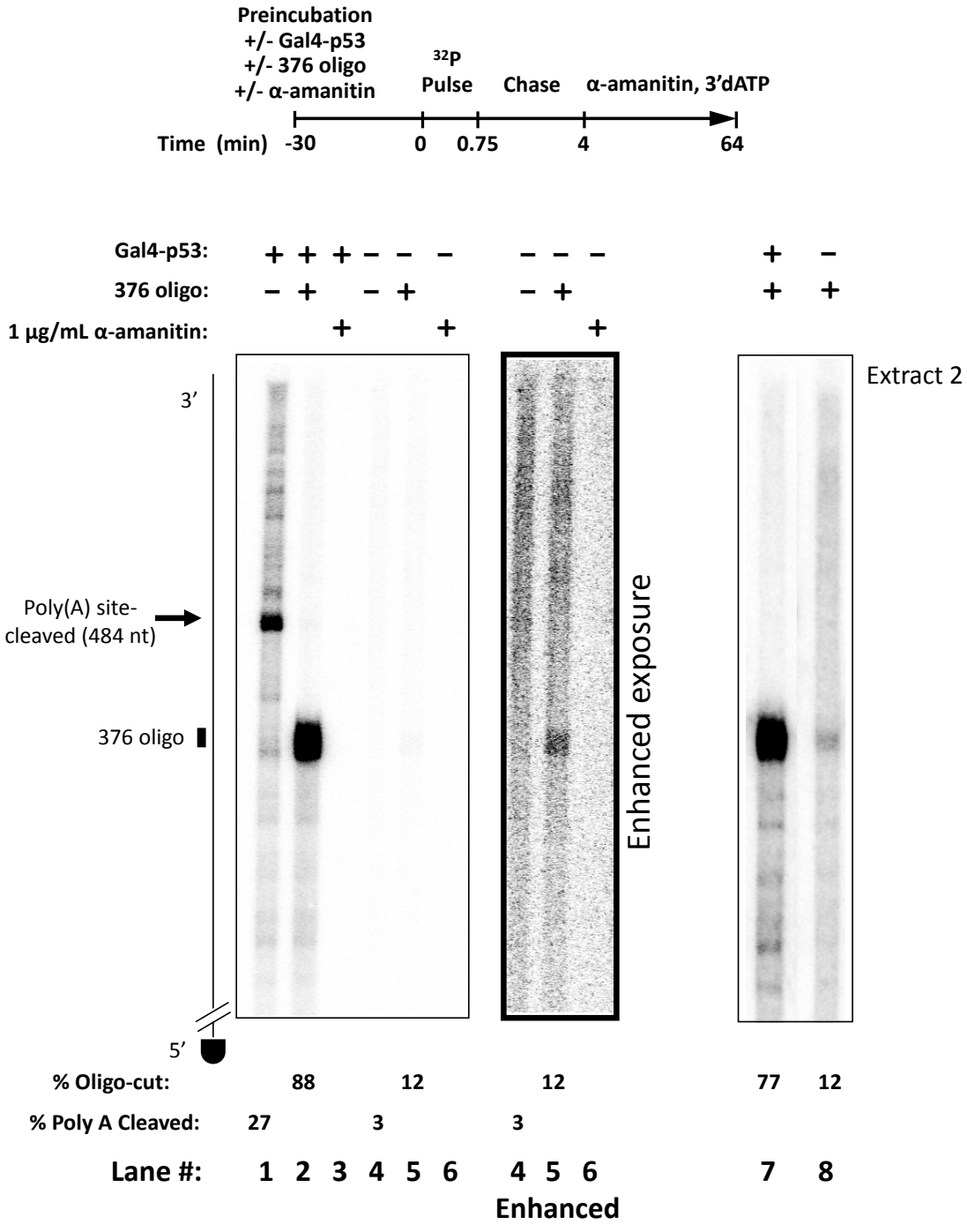

Preincubation
+/- Gal4-p53
+/- 376 oligo
+/- α-amanitin
32P
Pulse
Chase
α-amanitin, 3'dATP
Time (min) -30 0 0.75 4 64
Gal4-p53:
376 oligo:
1 μg/mL α-amanitin:
Extract 2
3'
Poly(A) site-cleaved (484 nt)
376 oligo
5'
Enhanced exposure
% Oligo-cut: 88 12 12 77 12
% Poly A Cleaved: 27 3 3
Lane #: 1 2 3 4 5 6 4 5 6 7 8
Enhanced

Figure 3. In the absence of transcriptional activator, transcription initiates at multiple sites on the template.

Gal5-HIV2dsxΔInt(+ESE) was linearized by EcoR1 and purified (see methods). EcoR1-cut or un-cut Gal5-HIV2dsxΔInt(+ESE) template was transcribed for 4 minutes and the RNA was visualized on the gel. Lane 5, no template was added to the reaction. % EcoR1 runoffs is calculated as the fraction of EcoR1 runoff transcripts (177 nt) over the total amount of RNA at or longer than 177 nt. The asterisk indicates ~315 nt-long transcripts that appear to be initiating somewhere downstream of the EcoR1 cut site and transcribing in the reverse direction and running off the EcoR1 cut site. This is deduced based on the fact that digestion with 376 oligo (Figure 2) did not reveal the existence of transcripts longer than the expected 377 nt band, which would indicate initiation occurring upstream of the HIV2 promoter.

Figure 3

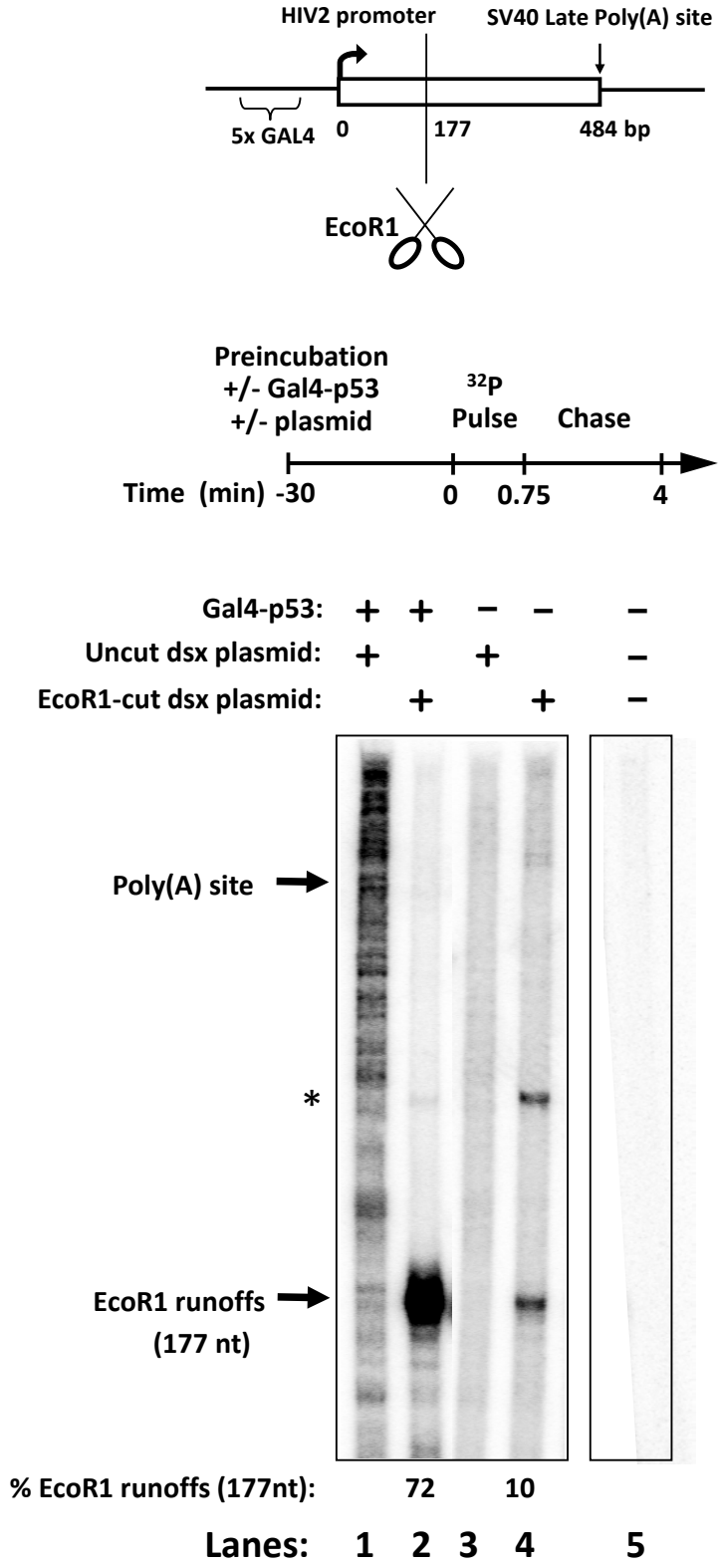
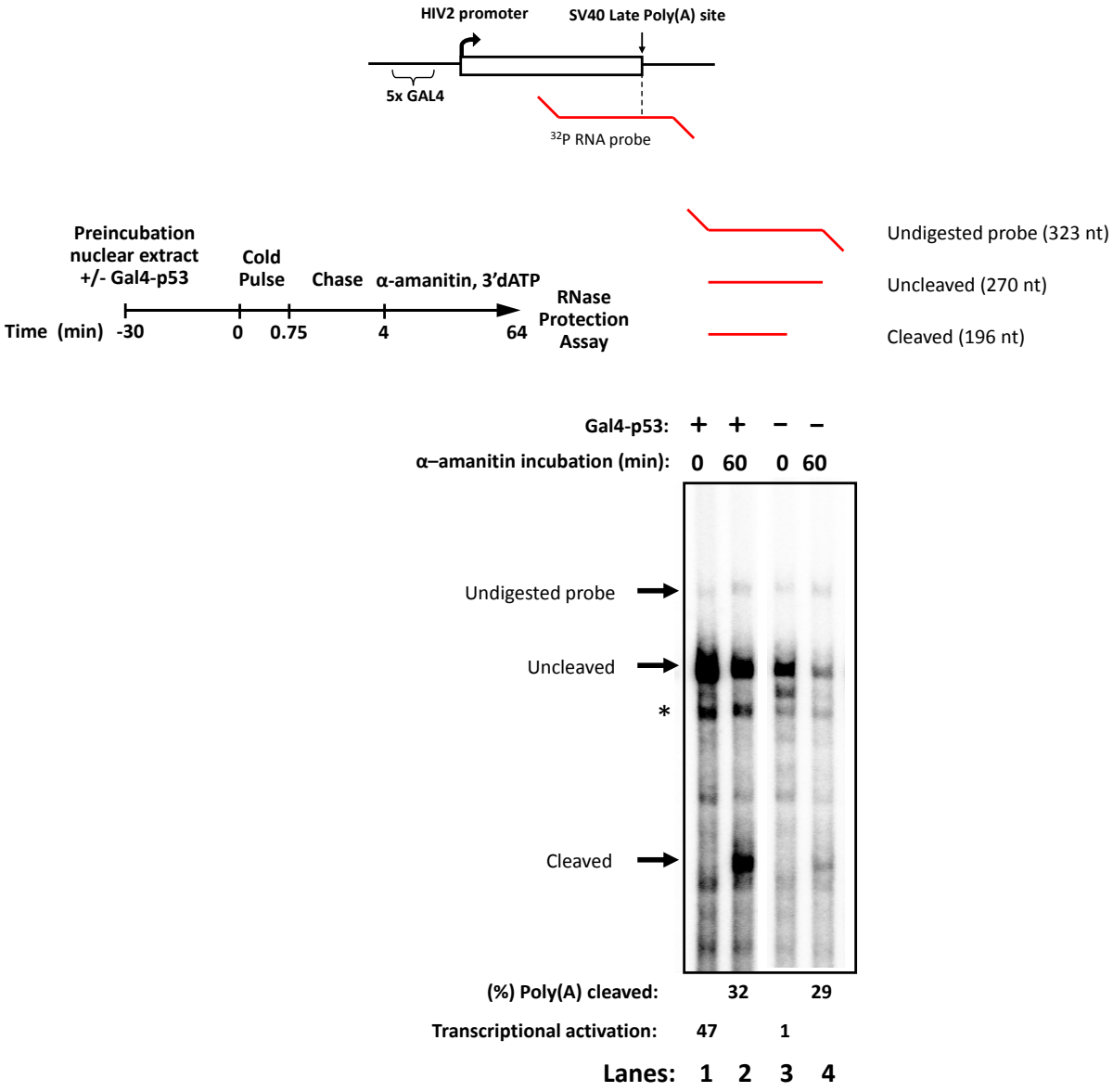

HIV2 promoter
SV40 Late Poly(A) site
5x GAL4
0
177
484 bp
EcoR1
Preincubation
+/- Gal4-p53
+/- plasmid
32P
Pulse
Chase
Time (min) -30
0
0.75
4
Gal4-p53: + + − − −
Uncut dsx plasmid: + + −
EcoR1-cut dsx plasmid: + + −
Poly(A) site
*
EcoR1 runoffs
(177 nt)
% EcoR1 runoffs (177nt): 72 10
Lanes: 1 2 3 4 5

Figure 4. Gal4-p53 activates transcription robustly but does not enhance 3'-end cleavage in a coupled extract.

Lanes 3 and 4 are the results of an 8-fold scaled-up reaction. Coupled processing assay was carried out on Gal5-HIV2dsxΔInt(+ESE) template with cold nucleotides. The RNAs were extracted by TRIzol/chloroform (see above), digested with DNaseI (Roche), hybridized at 65^{o}C, and subjected to RNase protection with RNase T1 (Ambion) according to Chao et al. 1999. Probe was generated from pBluescript dsx exon2 $(GAA)_6$ (Rosonina et al 2003). Digesting the resulting plasmid with NotI, and transcribing with T7 RNA polymerase generates a labeled protection probe (323 nt) that will result in protected fragments of 270 nt for uncleaved and 196 nt for cleaved RNA. The asterisk indicates a strong uncleaved band (~246 nt) that most likely reflects pausing of elongation complexes on 6 consecutive thymidines just right downstream of the GU-rich region of the SV40 late poly(A) signal. This "uncleaved" band also was included in the calculation of percentage poly(A) cleaved. Transcriptional activation was calculated as the fold difference of total transcripts (cleaved and uncleaved) compared to "minus Gal4-p53" (lane 3).

Figure 4


HIV2 promoter
SV40 Late Poly(A) site
5x GAL4
32P RNA probe
Preincubation nuclear extract +/- Gal4-p53
Cold Pulse
Chase
α-amanitin, 3'dATP
RNase Protection Assay
Time (min) -30 0 0.75 4 64
Undigested probe (323 nt)
Uncleaved (270 nt)
Cleaved (196 nt)
Gal4-p53: + + – –
α–amanitin incubation (min): 0 60 0 60
Undigested probe
Uncleaved
*
Cleaved
(%) Poly(A) cleaved: 32 29
Transcriptional activation: 47 1
Lanes: 1 2 3 4

References

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr. Opin. Cell Biol. *17*, 251-256.

Calvo, O., and Manley, J.L. (2005). The transcriptional coactivator PC4/Sub1 has multiple functions in RNA polymerase II transcription. EMBO J. *24*, 1009-1020.

Chao, L.C., Jamil, A., Kim, S.J., Huang, L., and Martinson, H.G. (1999). Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. Mol. Cell. Biol. *19*, 5588-5600.

Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. Proc. Natl. Acad. Sci. USA *94*, 11456-11460.

Cramer, P., Caceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E., and Kornblihtt, A.R. (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. Mol. Cell *4*, 251-258.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3'-end of mRNA. Nature *389*, 399-402.

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

Hirose, Y., and Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3'-end processing factors with RNA polymerase II. Mol. Cell *9*, 1101-1111.

Maniatis T., and Reed R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

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

Rigo, F., Kazerouninia, A., Nag, A., and Martinson, H.G. (2005). The RNA tether from the poly(A) signal to the polymerase mediates coupling of transcription to cleavage and polyadenylation. Mol. Cell *20*, 733-745.

Rigo, F., and Martinson, H.G. (2008). Functional coupling of last intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. Mol. Cell Biol. *28*, 849-862.

Rigo, F., and Martinson, H.G. (2009). Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. RNA *15*, 823-836.

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

Rosonina, E., Ip, J. Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J., and Blencowe, B.J. (2005). Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. Mol. Cell. Biol. *25*, 6734-6746.

Rozenblatt-Rosen, O., Nagaike, T., Francis, J.M., Kaneko, S., Glatt, K.A., Hughes, C.M., LaFramboise, T., Manley, J.L., and Meyerson, M. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. Proc. Natl. Acad. Sci. USA *106*, 755-760.

Tran, D.P., Kim, S.J., Park, J.J., Jew, T.M., and Martinson, H.G. (2001). Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. Mol. Cell. Biol. *21*, 7495-7508.

Weinmann, R., and Roeder, R.G. (1974). Role of DNA-dependent RNA polymerase III in the transcription of the tRNA and 5S RNA genes. Proc. Natl. Acad. Sci. USA *71*, 1790-1794.

Wu, H., Liima, W.F., and Crooke, S.T. (1999). Properties of cloned and expressed human RNase H1. J. Biol. Chem. 274, 28270-28278.