STUDIES OF POLYADENYLATION REGULATION OF U1A mRNA BY AN RNP COMPLEX CONTAINING U1A AND U1 snRNP

By

ROSE MARIE CARATOZZOLO

A dissertation submitted to the Graduate School – New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Program in Biochemistry

Written under the direction of

Samuel I. Gunderson, Ph.D.,

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2011

ABSTRACT OF THE DISSERTATION

STUDIES OF POLYADENYLATION REGULATION OF U1A mRNA BY AN RNP COMPLEX CONTAINING U1A AND U1 snRNP

By Rose Marie Caratozzolo

Dissertation Director:

Samuel I. Gunderson, Ph.D.

The 3'-end processing of nearly all eukaryotic pre-mRNAs comprises multiple steps which culminate in the addition of a poly(A) tail, which is essential for mRNA stability, translation, and export. Consequently, polyadenylation regulation is an important component of gene expression. One way to regulate polyadenylation is to inhibit the activity of a single poly(A) site, as exemplified by the U1A protein that negatively autoregulates itself by binding to a Polyadenylation Inhibitory Element (PIE) site within the 3' UTR of its own pre-mRNA. U1 snRNP, which is primarily involved in splice site recognition, inhibits poly(A) site activity in papillomaviruses by binding to 5' splice site-like sequences, which have recently been named "U1-sites". Here, a recently identified U1-site in the human U1A 3'UTR is examined and shown to synergize with the adjacent PIE site to inhibit polyadenylation. However, unlike the sites found in papillomaviruses, the U1A U1-site has no inhibitory activity on its own and is dependent on a wild-type PIE. This lack of activity is due to the site being masked within a phylogenetically conserved stem structure (U1-STEM). The secondary RNA structure of

this region was confirmed by RNase digestion analysis. Mutation of the U1-STEM, thereby opening up the U1-site, greatly increases U1-site mediated inhibition. The region between the U1-STEM and PIE (referred to as Region C) was also revealed to be required for synergy. Since biotin pulldown assays indicated that U1 snRNP binding to the U1-site was not affected by the presence of the U1-STEM, a model was proposed suggesting that U1 snRNP binds to the U1-STEM, but remains trapped in an inactive conformation until PIE is bound by two U1A molecules. However, further experiments showed that U1 snRNP binding did actually increase when the U1-STEM was mutated, but no corresponding change to the U1-STEM structure was detected. The discrepancies within these data suggest there is still much to be determined regarding the binding of U1 snRNP to the U1-Site. A more refined model is then presented which involves remodeling of Region C and part of the U1-STEM.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Samuel Gunderson, for all his guidance, support and encouragement throughout my graduate study at Rutgers. I am especially grateful for the endless patience he has shown in guiding me towards the completion of this thesis.

I would also like to thank current and former members of the Gunderson lab- Dr. Catherine Phillips, Dr. Rafal Goraczniak, Dr. Eric Ho, Steve Jung, Dr. Jianglin Ma, Dr. Fei Guan, Chris Jakubowski and Nick Zaphiros for their support throughout the years. I would especially like to thank Dr. Phillips for her advice and support in my graduate studies and beyond.

In addition, I would like to express my utmost gratitude to my friends, especially Dr. Alpa Gautam, who as a fellow gradate student with me at Rutgers has helped me through all stages of my graduate career.

Finally, and most importantly, I am exceedingly grateful to my parents, Roy and Rosemarie, for their endless love and encouragement. This dissertation would not have been possible without their sacrifice, patience and support.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter I: Introduction

In Eukaryotes, the 3'-end processing of nearly all pre-mRNAs is comprised of two steps: 1) endonucleolytic cleavage of the transcript and 2) synthesis of a poly(A) tail at the 3'-end of the upstream cleavage product (Figure 1.1, Zhao *et al.,* 1999). This 3'-end modification is essential for mRNA maturation. Interference with this process can lead to errors in cell development and function, as well as effect cell viability. For example, 3'-end processing promotes transport of the mRNA from the nucleus to the cytoplasm (Vinciguerra and Stutz 2004). Additionally, since mRNAs are degraded 3'→ 5', the addition of the poly (A) tail (with associated proteins) serves to increase transcript stability (Sachs and Wahle 1993; Ford *et al.,* 1997; Wickens *et al.,* 1997). Translation of mRNA into proteins is also greatly enhanced by 3'-end processing. The poly(A) tail and associated proteins form a pseudo-circular structure with the 5'-end to promote translation (Sachs *et al.,* 1997; Wilusz *et al.,* 2001). Other processes, such as transcription and splicing have also been linked to 3'-end processing (Hirose and Manley 2000; Proudfoot *et al.,* 2002). The study of the mechanisms of 3'-end processing and its regulation are essential in order to further understand their wide-ranging roles in cellular processes.

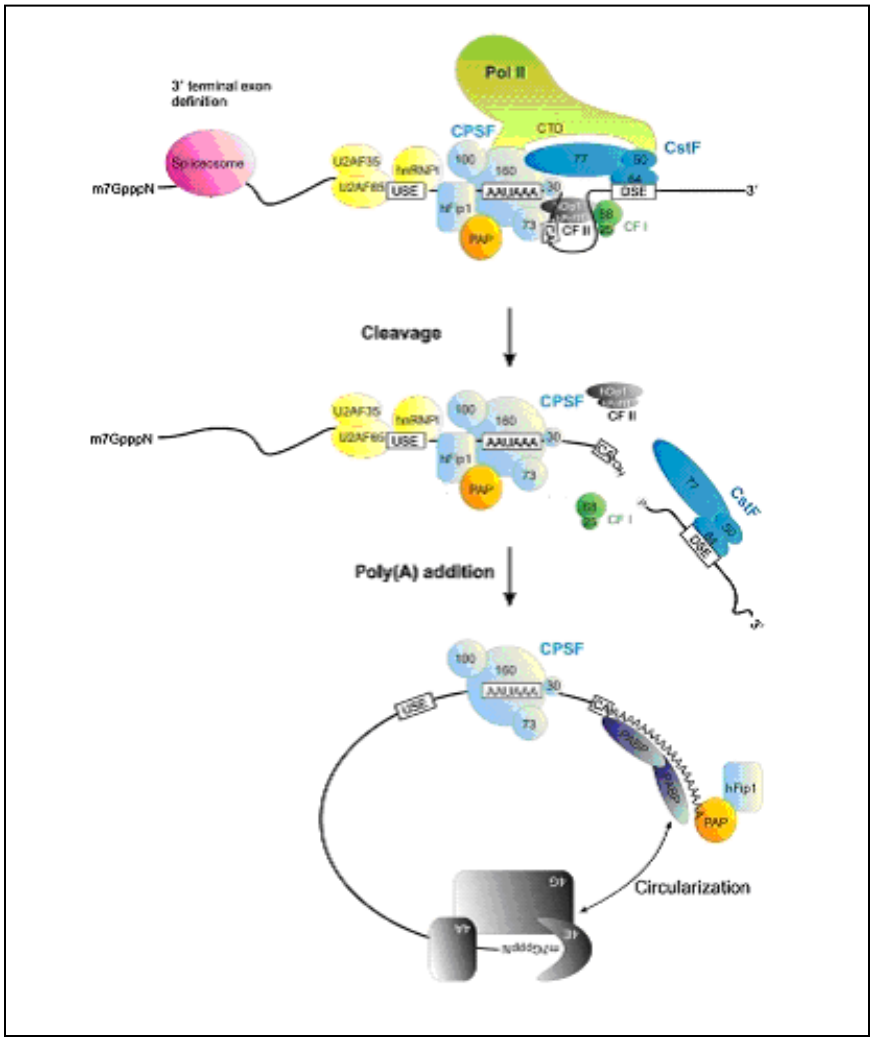

Figure 1.1 Mammalian 3'end Processing: Cleavage and Polyadenylation

The mammalian mRNA 3'end processing reaction comprises two steps: 1) endonucleolytic cleavage and 2) polyadenylation. The processing machinery includes the pre-mRNA, 3' end processing sequences (boxed sequences) and multiple protein factors. Cleavage results in an upstream product that is polyadenylated and a downstream product that is degraded. After cleavage, most protein factors are released from the complex and a full-length poly(A) tail is added. This modified mRNA can then circularize via interaction with the 5'-cap to provide stability and aid in translation. See text below and Figure 1.2 for more detailed description of sequences and protein factors. (Modified from Danckwardt *et al.*, 2008)

A. 3'-End Processing of Mammalian Pre-mRNAs

RNA Cis-Acting Sequence Elements: The cleavage and polyadenylation machinery is directed to the pre-mRNA via conserved sequence elements located within the 3' untranslated region (3'-UTR). Almost all Eukaryotic pre-mRNA that is processed contains these sequences. One notable exception are certain histone pre-mRNAs which are cleaved but do not undergo polyadenylation (Gilmartin 2005). These sequence elements comprise a core cleavage and polyadenylation signal and contain three primary sequences that define the polyadenylation site. There are also two auxiliary sequences, located both upstream and downstream of the primary elements, which serve to increase the efficiency of cleavage and polyadenylation (Figure 1.2, Zhao *et al.*, 1999).

• **Polyadenylation Signal (PAS):** This sequence element was the first element identified to be involved in 3'-end processing (Proudfoot and Brownlee 1976). It is highly conserved and consists of the hexamer sequence AAUAAA, located between 10-30 bases upstream of the cleavage site (Figure 1.2). Recent studies of human and mouse expressed sequence tags (ESTs) determined that approximately 70% contained this canonical sequence, while 15% contained the single nucleotide variant- AUUAAA, and 4% had no discernable PAS (Beaudoing *et al.*, 2000, Tian *et al.*, 2005). The remaining 11% contained single and double nucleotide variants of the canonical AAUAAA. Mutational studies in *Xenopus laevis* showed that point mutations in the PAS resulted in reduced polyadenylation and increased levels of unprocessed pre-mRNA transcripts (Wickens and

Stephenson 1984). Therefore, the canonical signal (and close variants) is necessary for efficient 3'-end processing.

• **Downstream Sequence Element (DSE):** This sequence element is less conserved than the PAS (Zhao *et al.,* 1999), and is located downstream of the cleavage site (Figure 1.2). The DSE has two parts: 1) a U-rich sequence, located 15-25 bases from the cleavage site, and 2) a GU-rich sequence, located approximately 5-10 bases from the cleavage site, with mRNAs containing one, none or both of these sequences (McLauchlan *et al.,* 1985; Gil and Proudfoot 1987; Sittler *et al.,* 1994; Salisbury *et al.,* 2006). These sequences lack sequence conservation, which means that point mutations have little effect on DSE functionality, while deletions/insertions are less tolerated (Zhao *et al.,* 1999; McDevitt *et al.,* 1986; Zarkower and Wickens 1988). Deletions/insertions that alter the proximity of the DSE to the cleavage site can result in alternate cleavage site choice and decreased cleavage efficiency (Zhao *et al.,* 1999).

• **Cleavage Site (pA):** The site of endonucleolytic cleavage, also called the poly(A) site (pA), is positioned between the PAS and the DSE (Figure 1.2, Chen *et al.,* 1995; Shatkin and Manley 2000), usually 10-30 nucleotides downstream of the PAS and 20-30 nucleotides upstream of the DSE (Zhao *et al.,* 1999). Approximately 70% of cleavage sites in vertebrates are adenosine (A) residues, with other bases preferred A>U>C>>G (Sheets *et al.,* 1990). While the sequence surrounding the cleavage site is not highly conserved, 59% of the sequences studied by Sheets *et al.* found that a cytosine residue

preceded the poly(A) site, indicating that CA is the optimal cleavage site (Sheets *et al.*, 1990).

• **Auxiliary Elements:** The majority of known auxiliary elements are located upstream of the PAS (e.g. upstream sequence element, USE). There is no consensus sequence, but usually these elements are short U-rich sequences- UUUU, UGUA or UAUA (Figure 1.2, Zhao *et al.*, 1999; Hu *et al.*, 2005). USEs have been found in many cellular genes and, although the accumulated evidence is limited, they are thought to help promote the binding of polyadenylation factors to the cleavage site (Brackenridge *et al.*, 1997; Moreira *et al.*, 1995 and 1998; Natalizio *et al.*, 2002). Downstream auxiliary elements are less documented, and while they lack a conserved sequence or distance from the poly(A) site, they are generally G-rich (Figure 1.2, Bagga *et al.*, 1995; Zhao *et al.*, 1999; Oberg *et al.*, 2005).

Cleavage and Polyadenylation Protein Factors: The cleavage and polyadenylation machinery in mammals requires over 14 protein factors, which combine to form the core complex required for 3'-end processing (Figures 1.1 and 1.2). The core complex encompasses the cleavage/polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), mammalian cleavage factors I and II (CF I_m and CF II_m), poly(A) polymerase (PAP), poly(A) binding protein II (PABII) and symplekin. The C-terminal domain (CTD) of RNA Polymerase II (Pol II) is also part of the overall complex. All of these proteins are involved in the cleavage reaction, except for PABII, while only CPSF, PABII and PAP are required for polyadenylation (Zhao *et al.*, 1999).

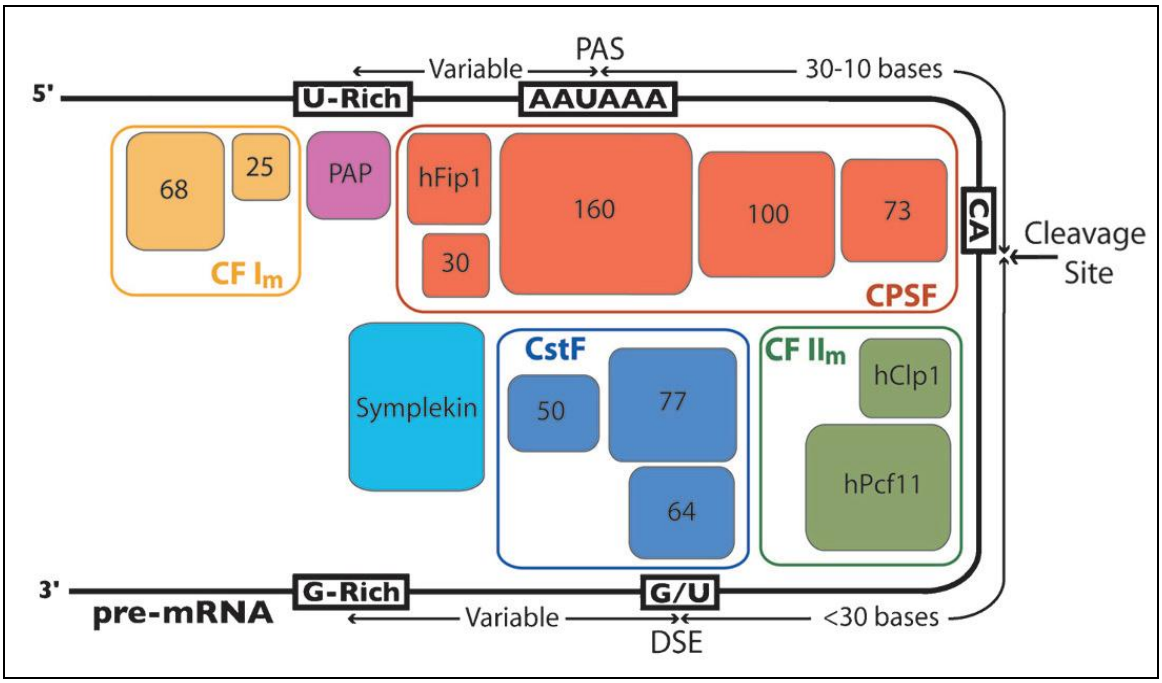

Figure 1.2 Cleavage and Polyadenylation Sequence Elements and Protein Factors

Schematic of the mammalian 3'-end processing machinery and the sequences recognized by individual factors. Cis-acting RNA sequences elements are boxed and approximate distances between elements are indicated. CPSF recognizes the PAS, while CstF recognizes the DSE, and together they help to define the Cleavage Site, CA. CF I_m, CF II_m and symplekin act as scaffolding proteins to strengthen protein-protein and protein-RNA interactions. (Modified from Mandel *et al.*, 2008)

• **Cleavage/Polyadenylation Specificity Factor (CPSF):** Mammalian CPSF consists of five subunits: CPSF-30, CPSF-73, CPSF-100, CPSF-160 and hFip1, all of which have homologs in yeast (Zhao *et al.*, 1999). The largest subunit, CPSF-160, binds directly to the PAS (Figure 1.2). It has the highest affinity for the canonical AAUAAA signal, with reduced affinity for PAS variants (Bienroth *et al.*,1991; Keller *et al.*, 1991; Zhao *et al.*, 1999). This binding is dependent on the cooperative binding of a CstF subunit, CstF-77, to the DSE (Murthy and Manley 1995). UV-crosslinking and biochemical studies have provided strong evidence that CPSF-73 is the subunit responsible for the actual

endonucleolytic cleavage reaction at the cleavage site (Figures 1.1 and 1.2, Ryan *et al.* 2004; Mandel *et al.,* 2006). This activity is not sequence specific, which agrees with the current theory that CPSF-160 and CstF-64 define the cleavage site by binding to the upstream and downstream elements flanking the CA nucleotides. The CPSF-30 subunit has five zinc finger motifs that preferentially bind poly(U) sequences that flank the cleavage site (Barabino *et al.,* 1997 and 2000). These zinc finger domains could also allow CPSF-30 to coordinate/recruit other proteins for cleavage and polyadenylation via protein-protein interactions. This subunit may also be involved in coupling 3'-end processing with splicing by association with the spliceosome (Li *et al.*, 2001; Rappsilber *et al.,* 2002). The primary function of the hFip1 subunit is most likely to help recruit and stimulate PAP along with CPSF-160 (Kaufmann *et al.*, 2004). The function of the final subunit CPSF-100 is still unknown, but its homolog is necessary for cell viability in yeast (Preker *et al.,* 1997) indicating it has an essential function.

• **Cleavage Stimulatory Factor (CstF):** This factor has three subunits: CstF-50, CstF-64 and CstF-77 (Zhao *et al.*, 1999). CstF-64 binds to the GU-rich DSE (Figure 1.2, MacDonald *et al.,* 1994), and together with CPSF-160 helps to define the cleavage site. This subunit has also been shown to effect expression of genes by inducing alternative poly(A) site choice (Shell *et al.,* 2005). CstF-77 interacts with CPSF-160 (see above) and PAP (Murthy and Manley 1995; Takagaki and Manley 2000). The interaction between CPSF-160 and CstF-77 is theorized to stabilize their interactions with the pre-mRNA (Wilusz *et al.,* 1990; MacDonald *et al.,* 1994; Murthy and Manley 1995) and more recent biochemical studies showed that CPSF and CstF associate prior to poly(A) signal

recognition (Takagaki and Manley 2000). Examination of the *Drosophila* homolog Su(f) has suggested that CstF-77 can self-associate and may act as a dimer during 3'-end processing (Simonelig *et al.,* 1996; Benoit *et al.,* 2002). While CstF-77 can also bind to the CTD of Pol II, CstF-50 binds with higher affinity (McCracken *et al.,* 1997). This subunit also interacts with the splicing factor SRm160 (McCracken *et al.,* 2003), providing more evidence of the link between 3'-end processing and transcription.

• **Mammalian Cleavage Factor I (CF I_m):** This factor functions as a heterodimer featuring a CF I_m-25 subunit paired with one of three larger subunits: CF I_m-59, CF I_m-68 or CF I_m-72. While all of these polypeptides co-purify from HeLa nuclear extract, CF I_m activity can be reconstituted *in vitro* with CF I_m-25 and CF I_m-68 (Figure 1.2, Rüegsegger *et al.,* 1996 and 1998). The subunits CF I_m-25, CF I_m-59 and CF I_m-68 preferentially bind to sequences (UGUAA) usually located close to the PAS (Brown and Gilmartin 2003). This binding enhances recognition of both the canonical and non-canonical poly(A) signals (Venkataraman *et al.,* 2005). This canonical signal enhancement is further aided by interactions between CF I_m and hFip1 and PAP. In addition to this augmentation of cleavage and polyadenylation, CF I_m can also inhibit cleavage for certain genes. Studies *in vitro* have shown that CF I_m , when bound to UGUAA sequences within the pre-mRNA of the CF I_m-68 subunit can suppress cleavage (Brown and Gilmartin 2003). This suggests that CF I_m may control its own pre-mRNA processing *in vivo* through autoregulation. This factor can also interact with U1 snRNP via CF I_m-25 interaction, indicating that it aids to couple 3'-end processing and splicing (Awasthi and Alwine 2003).

• **Mammalian Cleavage Factor II (CF II_m):** This factor is less characterized than CF Im. Purification from HeLa cells has allowed its activity to be separated into two parts: CF IIA_m and CF IIB_m (de Vries *et al.,* 2000). CF IIA_m is essential for cleavage and is made up of two polypeptides: hClp1 and hPcf11. The hPcf11 polypeptide interacts with the CTD of Pol II via a Pol II interacting domain (CID, Sadowski *et al.,* 2003; Zhang and Gilmour 2006). Mutations to this CID result in cell death due to errors in transcriptional termination (Sadowski *et al.,* 2003; Noble *et al.,* 2005; Zhang *et al.,* 2005), which is further evidence of the coupling of transcription to 3'-end processing. In addition to the CID, this subunit has CstF and RNA binding domains, allowing it to act as a scaffold for the 3'-end processing machinery (Zhang *et al.,* 2005). The second polypeptide, hClp1, also acts as a scaffold to tether CPSF and CF I_m (de Vries *et al.,* 2000). CF IIB_m is a non-essential, yet stimulatory component of CF II_m that did not co-purify with any known processing factors (de Vries *et al.,* 2000), indicating it may be a new factor involved in stimulating 3'-end processing.

• **Symplekin:** This protein was originally identified as a component of tight junctions (Figure 1.2, Koen *et al.,* 1996) and has high similarity to a yeast polyadenylation machinery factor- Pta1p (Zhao *et al.,* 1999; Takagaki and Manley 2000). It was also found to form a stable complex with CPSF and CstF in the nucleus (Takagaki and Manley 2000), and likely acts as a scaffold for the assembly of the cleavage and polyadenylation machinery. More recently, symplekin was found to co-localize with CPSF-100 during oocyte maturation in *Xenopus laevis* and was required for cytoplasmic polyadenylation in those oocytes (Hofmann *et al.,* 2002). This data implies that this

protein is involved in both 3'-end processing in the nucleus and regulation of polyadenylation in the cytoplasm. This model was supported by studies which found that symplekin interacts with cytoplasmic polyadenylation element binding protein (CPEB) and is required for CPEB-mediated cytoplasmic polyadenylation (Barnard *et al.*, 2004). Symplekin may also be a target for polyadenylation regulation through phosphorylation, as increased amounts of phosphorylated Pta1p resulted in shortened poly(A) tails (He and Moore 2005).

• **Poly(A) Binding Protein II (PABPII):** This protein is not required for polyadenylation, but it is necessary to regulate poly(A) tail length (Bienroth *et al.*, 1993; Amrani *et al.*, 1997; Wahle and Rüegsegger 1999). PABPII binds to short stretches of 11-14 adenosine residues as they are produced by PAP, and continue to bind until the desired tail length is reached (Figure 1.1, Keller *et al.*, 2000; Meyer *et al.*, 2002). It recognizes the adenosine bases via its RNA-recognition motifs (RRMs, Deo *et al.*, 1999; Wahle and Rüegsegger 1999). This protein also has a stimulatory effect on PAP by binding directly to the pre-mRNA, adjacent to PAP (Kerwitz *et al.*, 2000). This interaction increases polyadenylation efficiency 80-fold by increasing PAP processivity.

• **The CTD of RNA Polymerase II:** Evidence for the coupling of transcription and pre-mRNA processing (5'-end capping, splicing and 3'-end cleavage and polyadenylation) has steadily accumulated over the last decade. The majority of this evidence points to RNA Polymerase II (Pol II), specifically it's CTD, as the link between these processing steps. The PAS and DSE are known regulators of Pol II mediated transcription

termination (Whitelaw and Proudfoot 1986; Logan *et al.*, 1987; Connelly and Manley 1988). Also, in the absence of transcription, the CTD of Pol II is still required for cleavage *in vitro* (Hirose and Manley 1998), while *in vivo*, the CTD interacts with CPSF and CstF and is required for efficient splicing and 3'-end processing (McCracken *et al.*, 1997; Fong and Bentley 2001; Ryan *et al.*, 2002). Different segments of the CTD can independently stimulate the various steps of 3'-end processing (Fong and Bentley 2001). Consequently, the CTD of Pol II is proposed to function as a scaffold or recruitment platform for RNA processing factors (Figure 1.3, Bentley 2005).

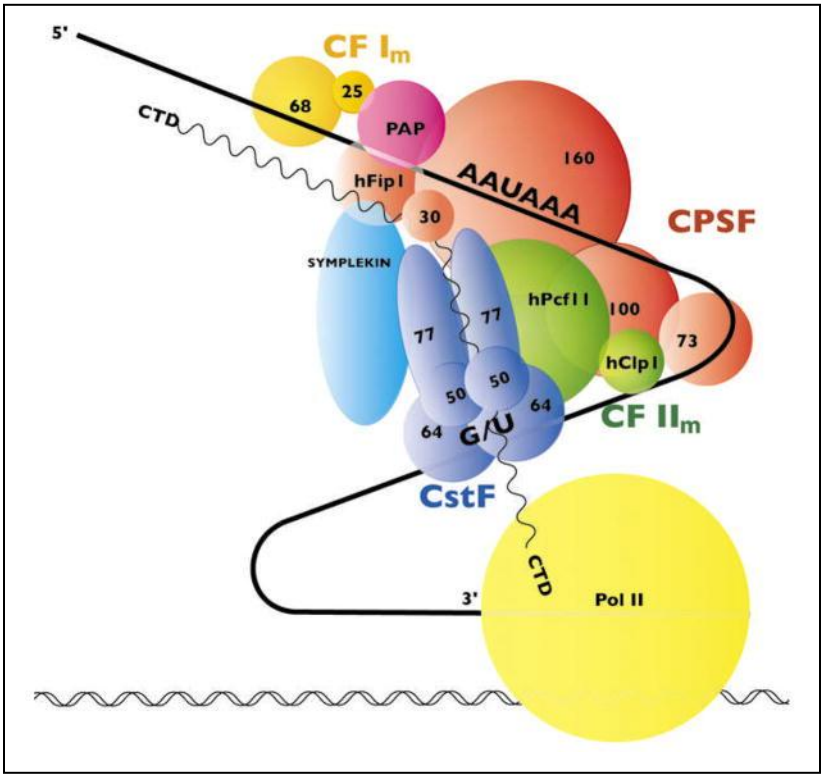

Figure 1.3 The CTD of Pol II Acts As A Scaffold For The 3'-end Processing Machinery

A graphic representation of the role of the CTD of Pol II during 3'-end processing. The CTD interacts with subunits of CPSF and CstF (shown here as a dimer) and provides a platform for other processing factors. (Modified from Mandel *et al.*, 2008)

• **Poly(A) Polymerase (PAP):** In mammals, PAP is required for cleavage and polyadenylation. It interacts with many factors of the 3'-end processing machinery, and while these interactions are important, they are not necessary for polyadenylation *in vitro* (Edmonds 2002). Multiple isoforms of PAP have been discovered from human, bovine, mouse and frog (Ryner *et al.,* 1989; Aström *et al.,* 1991; Wahle *et al.,* 1991; Raabe *et al.,* 1994; Thuresson *et al.,* 1994; Ballantyne *et al.,* 1995; Zhao and Manley 1996). Three human isoforms have been found in HeLa cells- 90, 100 and 106 kDa (Thuresson *et al.,* 1994). The 106 kDa isoform is the phosphorylated form of the 100 kDa isoforms, and together they are named PAPII. It is this protein that the majority of PAP function has been derived from. The 90 kDa isoform was more recently characterized and named PAPγ (Kyriakopoulou *et al.,* 2001; Topalian *et al.,* 2001). PAPII localizes to both the nucleus and cytoplasm, while PAPγ is strictly nuclear (Thuresson *et al.,* 1994; Kyriakopoulou *et al.,* 2001; Topalian *et al.,* 2001). The structures of the PAP catalytic core (about 400aa) from human, bovine and yeast have been determined (Bard *et al.,* 2000; Martin *et al.,* 2000 and 2004) and show that PAP is arranged into three domains- N-terminal (NTD), middle and C-terminal (CTD). The NTD contains the catalytic site and coordinates two metal ions (Mg^{2+} or Mn^{2+}) required for catalysis (Martin and Keller 1996; Martin *et al.,* 2000 and 2004). The active site is located in the large cleft between the NTD and CTD, while the CTD binds hFip1p, CPSF-160 and CF I_m-25 (Martin and Keller 1996; Kim and Lee 2001).

Cleavage and Polyadenylation Reactions: In summary, the cleavage and polyadenylation reactions occur as follows (see Figure 1.1): CPSF-160 recognizes and binds to the PAS (Bienroth *et al.,* 1991; Keller *et al.,* 1991), and CstF-77 recognizes and binds to the DSE (MacDonald et al. 1994). Together, they help to define the cleavage site and the interactions between the subunits stabilize their binding to RNA (Wilusz *et al.,* 1990, Macdonald *et al.,* 1994, Murthy and Manley 1995). Next, CF I_m binds to the pre-mRNA and CPSF, which enhances the assembly of the processing machinery and the cleavage reaction (Rüegsegger *et al.,* 1996 and 1998). The CTD of Pol II also helps to recruit CPSF, CstF and other processing factors to the pre-mRNA (McCracken *et al.,* 1997; Fong and Bentley 2001; Ryan *et al.,* 2002; Bentley 2005). PAP is recruited through interactions with CPSF-160, CstF-77 and CF I_m-25 (Murthy and Manley 1995, Kim and Lee 2001). CF II_m and symplekin join the forming complex and once the entire cleavage complex is assembled, CPSF-73 works as the endonuclease to cleave the pre-mRNA at the cleavage site (Ryan *et al.* 2004; Mandel *et al.,* 2006).

Following cleavage, CstF, CF I_m, CF II_m, symplekin and Pol II dissociate from the complex, leaving only CPSF and PAP. The downstream cleavage fragment is rapidly degraded while PAP adds approximately 11-14 adenosine residues at the PolyA site (Bienroth *et al.,* 1993; Zhao *et al.,* 1999). The process of adding these initial adenosines is very slow until PABPII binds, upon which PAP, along with the help of CPSF, rapidly increases the length of the tail to approximately 150-250 residues in mammals (Wahle 1991 and 1995) and 55-90 residues in yeast (Brown and Sachs, 1998).

B. Regulation of Polyadenylation

Significance of Polyadenylation: The extensive coupling of 3'-end processing (capping, splicing and polyadenylation) and transcription has recently emerged as vital for the regulation of cellular processes and gene expression. For example, transcription termination requires functional polyadenylation signals and factors (Rosonina *et al.,* 2006; Kaneko *et al.,* 2007; Danckwardt *et al.,* 2008). Mutations in these factors or signals have been linked to multiple human diseases including cancer and thalassemia (Danckwardt *et al.,* 2008). Polyadenylation factors interact with 3'-terminal intron splicing factors to promote splicing, cleavage and polyadenylation (Niwa *et al.,* 1990; Lutz *et al.,* 1996; Gunderson *et al.,* 1997; Vagner *et al.,* 2000b; Millevoi *et al.,* 2000 and 2006; Kyburz *et al.,* 2006; Danckwardt *et al.,* 2007). The poly(A) tail itself protects the mRNA from nuclear degradation, conferring stability to the transcript (Bousquet-Antonelli *et al.,* 2000). In addition, the poly(A) tail is needed for mRNA export and localization outside of the nucleus (Bousquet-Antonelli *et al.,* 2000; Proudfoot 2001). Finally, PABP, while attached to the poly(A) tail, interacts with the 5'-end cap binding protein eIF4E, which causes circularization of the mRNA (Sachs *et al.,* 1997; Wells *et al.,* 1998; Eldad *et al.,* 2008). This interaction facilitates translation initiation and increases its efficiency (Gallie 1991; Preiss and Hentze 1998; Borman *et al.,* 2000; Sachs *et al.,* 2000; Kahvejian *et al.,* 2005; Eldad *et al.,* 2008). The mRNA can become physically compacted through circularization mediated by protein-protein bridging, forming a closed-loop structure that can promote reloading of terminating ribosomes back onto the 5'-end of the transcript (Wells *et al.,* 1998; Sachs *et al.,* 2000; Eldad *et al.,*

2008). Taken together, the above demonstrates the importance of efficient polyadenylation. Defects or mutations in this process can have significant effects on overall gene expression, as well as cell growth and viability.

• **Alternative Polyadenylation:** There are two types of regulation of polyadenylation, both of which involve PAS site choice and usage. Genes that possess a single poly(A) site are regulated in an "on/off" manner, in which the site is either used (i.e. "on") or inhibited (i.e. "off,"). Conversely, genes which contain multiple poly(A) sites employ alternative polyadenylation, in which one site is utilized, while the other sites are inhibited (Zhao *et al.,* 1999; Shatkin and Manley 2000). Bioinformatic analysis of human, mouse and rat cDNA and ESTs determined that 54% of human, 32% of mouse and 28% of rat genes undergo alternative polyadenylation (Tian *et al.,* 2005; Yan and Marr 2005). Many human genes exhibit tissue-specific differences in poly(A) site choice (Beaudoing and Gautheret 2001; Zhang *et al.,* 2005). For example, mRNAs from brain tissues generally have long 3'-UTRs due to the use of a downstream poly(A) site (Zhang *et al.,* 2005). Poly(A) site choice can also be affected by the cell's developmental stage, as 3'-UTRs can shorten in proliferating cells (Sandberg *et al.,* 2008) or lengthen during embryonic development (Ji *et al.,* 2009).

Genes that undergo alternative polyadenylation can be classified into three classes (Figure 1.4, Edwalds-Gilbert *et al.,* 1997; Tian *et al.,* 2005; Yan and Marr 2005; Neilson and Sandberg 2010). The first class involves 3'-UTRs that utilize one of two mutually exclusive terminal exons and are referred to as skipped-exon associated 3'-UTRs

(Figure1.4A). This type of regulation often results in the modification of the C-terminus of the resulting protein (Neilson and Sandberg 2010).

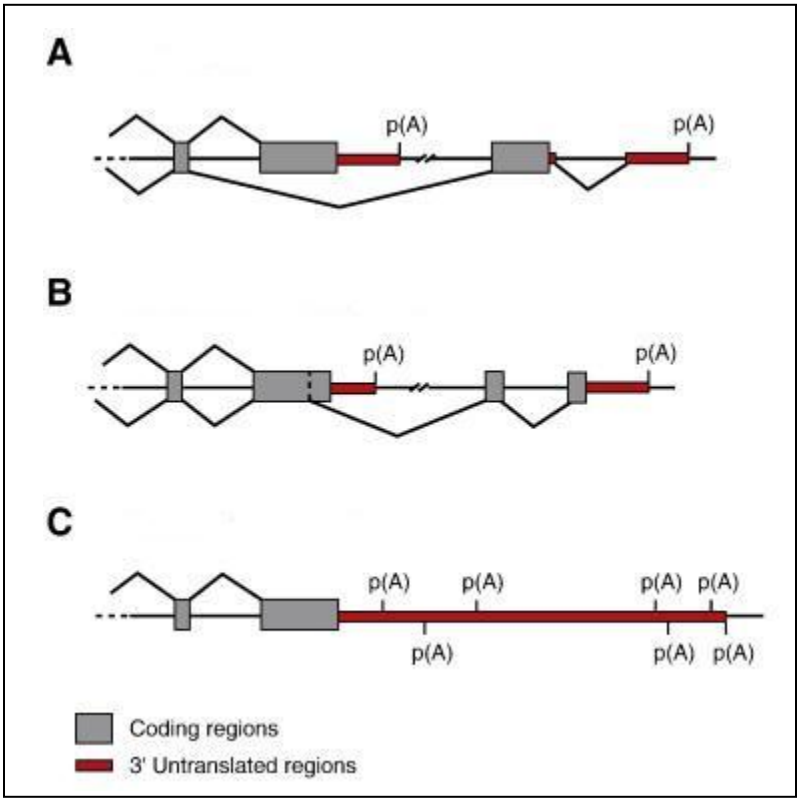

Figure 1.4 The Three Types Of Alternative Polyadenylation

The three types of alternative polyadenylation include (A) skipped-exon associated 3'-UTRs, which have two mutually exclusive poly(A) sites; (B) composite exons, where a single exon can contain competing splicing and polyadenylation signals, and (C) tandem UTRs that contain multiple poly(A) sites within the terminal exon. (Modified from Neilson and Sandberg 2010)

The second class, called composite exons, is defined by competing splicing and polyadenylation actions within a single exon. In these cases, the final mRNA transcript would contain at least part of the exon's original sequence (Figure 1.4B). The final class contains multiple poly(A) sites within the 3'-terminal exon, which are called tandem UTRs (Figure 1.4C). Unlike the first two classes, tandem UTRs generally have no direct

influence on the resulting protein's coding sequence. This class is also the most common type of alternative polyadenylation, accounting for more than half of known cases (Yan and Marr 2005; Neilson and Sandberg 2010).

• ***Cis/Trans*-Acting Factors:** Variation in the 3'-end of mRNAs due to alternative polyadenylation can result in the exclusion or inclusion of regulatory motifs- *cis*-acting sequence elements and/or *trans*-acting factor binding motifs (Wilusz and Spector 2010). *Cis*-acting elements include miRNA targets, which recruit small, non-coding microRNAs to the mRNAs of protein-coding genes, thereby marking them for posttranscriptional repression (Lewis *et al.,* 2005). AREs are AU-rich elements that are between 50-150 nucleotides and are commonly found in the 3'-UTRs of genes that encode proteins involved in cell growth or in the response to external stimuli (Barreau *et al.,* 2006). The presence of AREs promotes the destabilization of the mRNA via degradation of the poly(A) tail (Chen and Shyu 1995). Various ARE-binding proteins can bind to these sequences (Barreau *et al.,* 2006) and recruit/promote the association of the multi-protein exosome complex that is involved with 3'-5' degradation of mRNAs (Chen *et al.,* 2001; Mukherjee *et al.,* 2002; Gherzi *et al.,* 2004; Barreau *et al.,* 2006). GU-rich sequence elements (GREs) can also be found in the 3'-UTRs of short-lived transcripts (Vlasova *et al.,* 2008). These 11-mer sequences are bound by the RNA-binding CUG-binding protein (CUGBP1) and promote the decay of the mRNA, most likely through deadenylation (Paillard *et al.,* 2003; Moraes *et al.,* 2006; Vlasova *et al.,* 2008). It is worth noting that this protein has also been shown to inhibit poly(A) site usage when bound to a GU-rich

element in the late gene of human papillomavirus type 16 (Goraczniak and Gunderson 2008).

U1A and Polyadenylation Regulation: The complex interactions between 3'-end processing, alternative polyadenylation and *cis/trans*-acting factors allows for great flexibility in mRNA transcript formation. Consequently, these processes undergo extensive regulation. Several models of polyadenylation regulation have been proposed (Millevoi and Vagner 2010), but this work will focus on the model of regulation by PAP inactivation. The best-characterized example of this type of regulation is the autoregulation of the U1A pre-mRNA. Human U1A is a 32 kDa protein that is highly conserved among vertebrates (Boelens *et al.,* 1993). It is one of ten proteins bound to the small nuclear RNA (snRNA) of the U1 small nuclear ribonucleoprotein particle (U1 snRNP), Will and Luhrmann 1997; Klein Gunnewiek *et al.,* 2000). U1A has four domains: two RNA recognition motifs (RRMs)- an N-terminal RRM1 (amino acids [aa] 1-102) and a C-terminal RRM2 (aa 202-283) which are bridged by a linker region (aa 116-201) and a Polyadenylation Regulatory Domain or PAP Regulatory Domain (PRD, aa 103-115, Klein Gunnewiek *et al.,* 2000). As a component of U1 snRNP, the N-terminal RRM1 of U1A binds with high affinity to the stem-loop II (SLII) sequence AUUGCAC of the U1 snRNA (Scherly *et al.,* 1989; Lutz-Freyermuth *et al.,* 1990; Klein Gunnewiek *et al.,* 2000). This protein is predominantly located in the nucleus and is present in three populations. Approximately 85-95% is associated with U1 snRNP and not involved in polyadenylation regulation (Gunderson *et al.,* 1997 and 1998). The remaining two populations consist of pre-mRNA bound U1A, which is involved in

polyadenylation regulation (see below), and RNA-free U1A that has no known function (O'Connor *et al.,* 1997).

• **U1A Autoregulation:** U1A pre-mRNA contains a single poly(A) cleavage site (Figure 1.5A), making it the most characterized example of "on/off" poly(A) site regulation, as well as PAP inactivation. In vertebrates, the U1A protein is able to regulate its own expression level through a negative autoregulatory feedback loop (Boelens *et al.,* 1993; Gunderson *et al.,* 1994; Gunderson *et al.,* 1997). This is accomplished through the presence of a region known as the Polyadenylation Inhibitory Element (PIE) in the U1A pre-mRNA 3'-UTR (van Gelder *et al.,* 1993). This 50- nucleotide sequence is conserved in vertebrates, including human, mouse, *Xenopus,* fish and platypus (Boelens *et al.,* 1993), and consists of two seven-nucleotide loops (AUUGUAC and AUUGCAC) which closely or exactly match the AUUGCAC sequence of SLII of U1 snRNA (Boelens *et al.,* 1993). These sequences are 19-nucleotides apart and the second loop is 19-nucleotides upstream of the poly(A) signal (AUUAAA in the human gene, Figure 1.5A, Boelens *et al.,* 1993). While each of these loops can bind one molecule of U1A (via the N-terminal RRM1) independently with modest affinity, two molecules of U1A bind with high affinity, which is indicative of cooperative binding (van Gelder *et al.,* 1993; Klein Gunnewiek *et al.,* 2000). This homodimerization and cooperative binding is mediated through the PRD. Endogenous mouse U1A mRNA levels decreased when U1A was over-expressed and bound to PIE in mouse cells (Boelens *et al.,* 1993; van Gelder *et al.,* 1993) and human cells (Guan *et al.,* 2003).

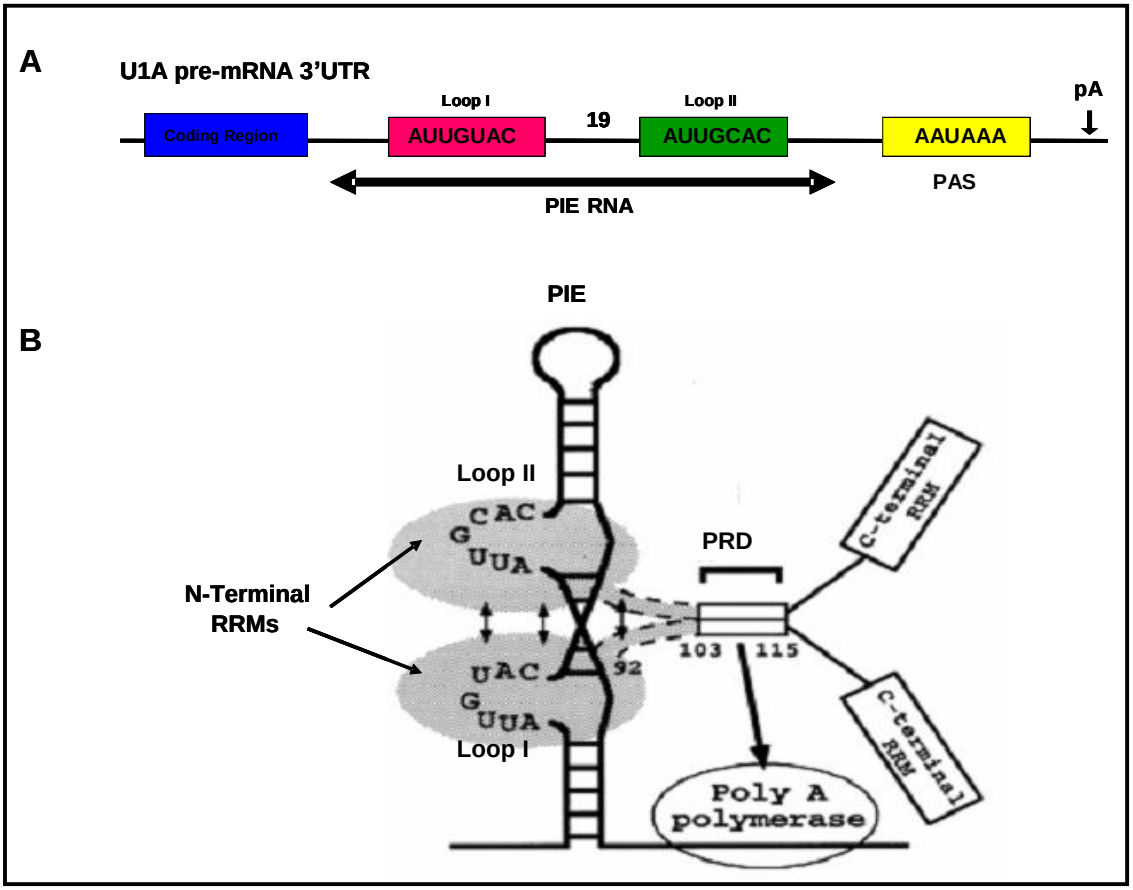

Figure 1.5 U1A Autoregulation

(A) The human U1A pre-mRNA 3'-UTR contains the Polyadenylation Inhibitory Element (PIE) which contains two loops (I and II), 19 nucleotides apart and loop II is 19 nucleotides upstream of the poly(A) signal, PAS. (B) PIE is shown bound by a homodimer of U1A: two molecules of U1A, one bound to each loop of PIE via their N-terminal RRM1s. The PRD, which aids in homodimerization and cooperative binding, interacts with residues within the C-terminal domain of PAP (not shown) and inhibits its activity. (Part B modified from Klein Gunnewiek *et al.*, 2000)

The resulting $(U1A)_2$-PIE complex is able to block polyadenylation by binding to and inhibiting PAP *in vitro and in vivo* (Figure 1.5B, Boelens *et al.*, 1993; Gunderson *et al.*, 1994). PAP inhibition results from direct interaction between the PRDs in U1A and the C-terminal 20 residues of PAP (Figure 1.5B, Gunderson *et al.*, 1997). In vertebrates, these 20 residues are the most highly conserved of the 240 total in the C-terminal domain

of PAP, and are sufficient for interaction with the $(U1A)_2$-PIE complex (Martin and Keller 1996; Gunderson *et al.,* 1997). The inactivation of PAP inhibits polyadenylation, thereby preventing transport of the U1A pre-mRNA to the cytoplasm. The transcript is instead rapidly degraded, which lowers the amount of mRNA available for translation, resulting in decreased U1A protein levels.

• **Other Examples of Autoregulation:** There have been several other examples of autoregulation of 3'-end processing reported. As mentioned above, CF I_m can regulate the pre-mRNA 3'-end processing of its own 68 kDa subunit *in vitro* (Brown and Gilmartin 2003). CF I_m can bind to a set of UGUAA sequence elements found in the 3'-UTR of the CF I_m-68 pre-mRNA. It is interesting to note that one of the elements overlaps with the AAUAAA PAS, suggesting inhibition is through steric hindrance (Brown and Gilmartin 2003). Autoregulation has also been found to control flowering in *Arabidopsis thaliana*. FCA is an RNA binding protein that promotes flowering (Macknight *et al.,* 1997) and it associates with the mRNA 3'-end processing factor, FY, and together they negatively autoregulate FCA expression through alternative polyadenylation (Quesada *et al.,* 2003 and 2005). Four different transcripts can be produced from the FCA pre-mRNA- α, β, γ, and δ via alternative splicing and alternative polyadenylation. FCA-γ codes for the functional full-length FCA protein, whereas the other three transcripts produce truncated, inactive isoforms (Macknight *et al.,* 1997). FCA binding promotes the usage of a proximal poly(A) site, which results in the production of the truncated isoforms. In addition, other proteins, including the SR proteins U2AF65 and SRp75, have protein domains similar to the PRD of U1A and

inactive PAP accordingly in cell culture when tethered to the 3' UTR of a reporter gene (Gunderson *et al.,* 1997; Ko and Gunderson 2002).

• **U1A Regulation of IgM:** U1A has recently been found to regulate the expression of IgM mRNA. The IgM heavy chain (μ) pre-mRNA, a skipped-exon associated 3'-UTR poly(A) configuration, contains two poly(A) sites- a downstream membrane specific site located in the 3' terminal exon and an upstream secretory specific site located within an upstream intron (Figure 1.6, Alt *et al.,* 1980; Phillips *et al.,* 2001). During B-cell differentiation, the IgM pre-mRNA is alternatively processed to encode for a membrane-bound receptor or a secreted antibody (Galli *et al.,* 1988; Edwalds-Gilbert *et al.,* 1997). The secretory poly(A) site is not expressed in undifferentiated B-cells, instead the alternative membrane-specific poly(A) site is used that allows for the inclusion of membrane encoding exons. Upon B-cell differentiation, there is an increase in cytoplasmic accumulation and stability of secretory mRNA since this is when secreted antibodies are needed. This regulation, unlike the "on/off" regulation used for U1A's single poly(A) site, involves the inhibition of one site so the other can be utilized. U1A inhibits usage of the secretory poly(A) site by binding to three sequence motifs located upstream of the μ secretory poly(A) site *in vivo* and *in vitro* (Figure 1.6, Phillips *et al.,* 2001). U1A binds to these sequence motifs via its RRM1 with 10-fold lower affinity than to U1 snRNA or PIE (Phillips and Gunderson 2003). This U1A-mediated inhibition of IgM polyadenylation is weaker than that of autoregulation of U1A, as this is more suitable for the regulation of heterologous RNAs (Phillips *et al.,* 2001; Phillips and Gunderson 2003).

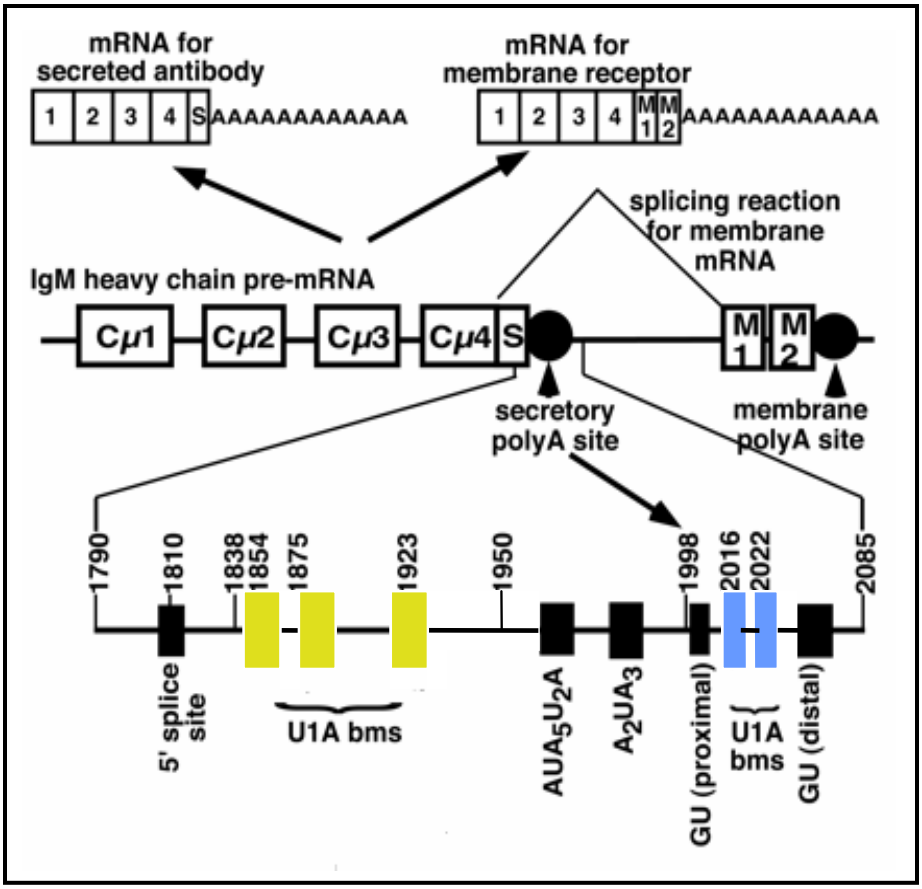

Figure 1.6 U1A Regulation of IgM Cleavage and Polyadenylation

Top shows a schematic of the IgM heavy chain pre-mRNA, which can be alternatively processed into a membrane or secretory form of mRNA. Below is an amplified look at the region surrounding the secretory poly(A) site, showing the location of the following: the 5' splice site, three upstream U1A binding motifs (bms, yellow), the hexanucleotide poly(A) signal and an adjacent upstream motif that aids in its activity, a proximal GU rich region, two additional U1A bms' (blue) and a distal GU-rich region. The two U1A bms' located within the GU-rich regions are involved in cleavage inhibition, while the three upstream U1A bms' are involved in polyadenylation inhibition. (Modified from Phillips *et al.*, 2004)

U1A also inhibits secretory poly(A) site expression *in vivo* by binding to two additional sequence motifs located in GU-rich regions downstream from the secretory poly(A) site (Figure 1.6, Phillips *et al.*, 2004). In contrast to U1A binding to the three upstream motifs, which inhibit polyadenylation, U1A binding to these two downstream motifs blocks cleavage by inhibiting the binding of CstF-64 to the adjacent GU-rich

regions (Phillips et al. 2004). Therefore, U1A regulates the expression of secretory poly(A) site usage at the level of both cleavage and polyadenylation (Phillips *et al*,. 2001 and 2004; Phillips and Gunderson 2003).

U1 snRNP and Polyadenylation Regulation: In vertebrates, the U1snRNP is the most abundant member of a group of five spliceosomal snRNPs (U1, U2, U4, U5 and U6) and consists of ten proteins stably bound to a 165-nucleotide U1 small nuclear RNA (U1 snRNA) (Will and Luhrmann 1997; Klein Gunnewiek *et al.,* 2000) which folds to form four stem-loops (SLI-IV, Figure 1.7A, Krol *et al., 1990*; Pomeranz Krummel *et al.,* 2009). Seven common Sm proteins bind as a donut-shaped complex to an Sm-binding site on the U1 snRNA located between SL3 and SL4 (Figure 1.7A, Pomeranz Krummel *et al.,* 2009; Kattah *et al.,* 2010). Interactions between the RNA backbone and the Sm proteins stabilize the core complex (Pomeranz Krummel *et al.,* 2009). The remaining three proteins are U1 specific- U1A, bound to SLII, U1-70K, bound to SL1, and U1C, which lacks an RRM and therefore can only bind after U1-70K and the Sm core have already bound (Gunnewiek *et al.,* 1995).

In general, U snRNPs function by nucleating and guiding assembly of the spliceosome onto the pre-mRNA. This occurs via short RNA:RNA base-pairing interactions between the U snRNA and the pre-mRNA substrate. U1 snRNP guides the formation of the E (early) spliceosomal complex by recognizing and binding to the 5' splice site of the pre-mRNA (Kramer 1996; Klein Gunnewiek *et al.,* 2000). This interaction involves base-pairing of the 5' end of U1 snRNA to conserved sequences in the 5' splice site (Figure 1.7B, Kramer 1996).

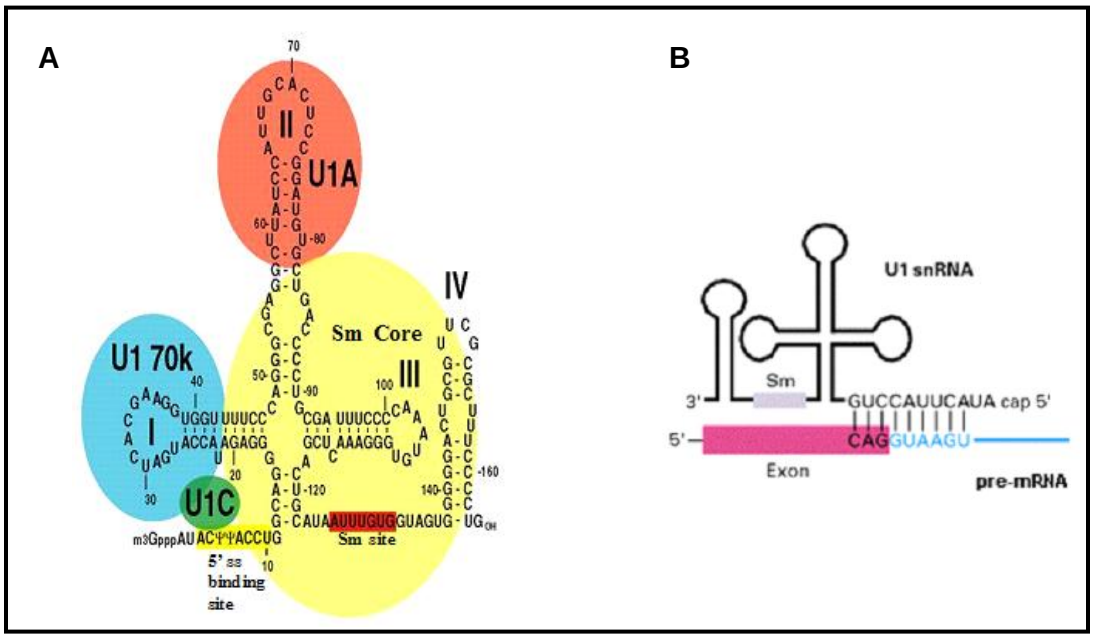

Figure 1.7 U1 snRNP Structure and Binding

A) U1 snRNA folds into a secondary structure with four stem-loops, I, II, III, IV. Sm proteins bind to the Sm site (red sequence) to form a globular core domain (yellow circle). U1-70K binds to stem-loop I through its RNA Recognition Motif (RRM), while U1A binds to stem-loop II through its N-terminal RRM1 domain. U1C binds through protein-protein interactions. (Modified from Nagai *et al.*, 2001)
B) U1 snRNP binds to pre-mRNA through RNA:RNA interactions between the U1 snRNA 5'-end sequence (yellow sequence in A) and the 5' splice site within the pre-mRNA. (Modified from Lodish *et al.*, 2000)

•Regulation of Multiple Genes by U1 snRNP: In addition to splicing, U1 snRNP has been implicated in poly(A) site regulation of many genes. The retrovirus HIV-1 has a duplicated poly(A) signal in both its 3' and 5' long terminal repeats (LTRs). The production of viral transcripts requires the use of the downstream poly(A) signal and suppression of the upstream 5'-LTR signal (Proudfoot 1991). U1snRNP blocks the poly(A) signal of the HIV-1 5' LTR by binding to the downstream major splice donor site, ie. the 5' splice site (Ashe *et al.*, 1997; Ashe *et al.*, 2000). Inhibition was dependent

on the presence of SLI of U1 snRNA, not SLII, indicating that U1-70K was necessary for inhibition (Ashe *et al.,* 2000).

Papillomaviruses are viruses of the family Papovaviridae, which have small double stranded circular DNA genomes. They are responsible for squamous epithelial and fibroepithelial tumors (both benign and malignant) in humans and animals. Papillomavirus gene expression can be divided into two phases, early and late (Spalholz and Howley 1989, Baker 1990). Early genes are responsible for viral replication and cellular transformation, while late genes encode for capsid proteins. Late genes are transcribed on the same transcript as early genes, but are only expressed in differentiated keratinocytes (Barksdale and Baker 1993; Zheng and Baker 2006). The Baker lab has shown through deletion and point mutagenesis that a sequence, AAGGUAAGU, homologous to a 5' splice site is required for inhibition of bovine papillomavirus type 1 (BPV-1) late gene expression (Furth and Baker 1991; Furth *et al.*, 1994). This sequence, located in the 3' UTR, binds to the 5' end of U1 snRNA in U1snRNP and inhibits usage of the late gene poly(A) site (Figure 1.8A, Furth and Baker 1991). Further studies have shown that it is a direct interaction between the U1 snRNP protein U1-70K with PAP that causes inhibition of the late poly(A) site in a manner similar to U1A autoregulation (Gunderson *et al,* 1998; Zheng and Baker 2006).

The late poly(A) site of Human papillomavirus type 16 (HPV-16) also has a sequence element that contains U1 snRNA binding sequences but they are much more complex than that of BPV-1.. (Figure 1.8B, Kennedy *et al.,* 1990), Upstream of the HPV16 late poly(A) site are two elements, a 5' and a 3' sub-element. The 5' sub-element has four weak non-consensus 5' splice sites that overlap with each other such that it is

likely only two U1 snRNAs can bind at one time (Furth *et al.*, 1994; Cumming *et al.*, 2003). The 3' sub-element contains a GU-rich region that binds the CUG-binding protein 1 (CUGBP1) and synergizes with the weak splice sites to inhibit expression *in vivo* (Goraczniak and Gunderson 2008).

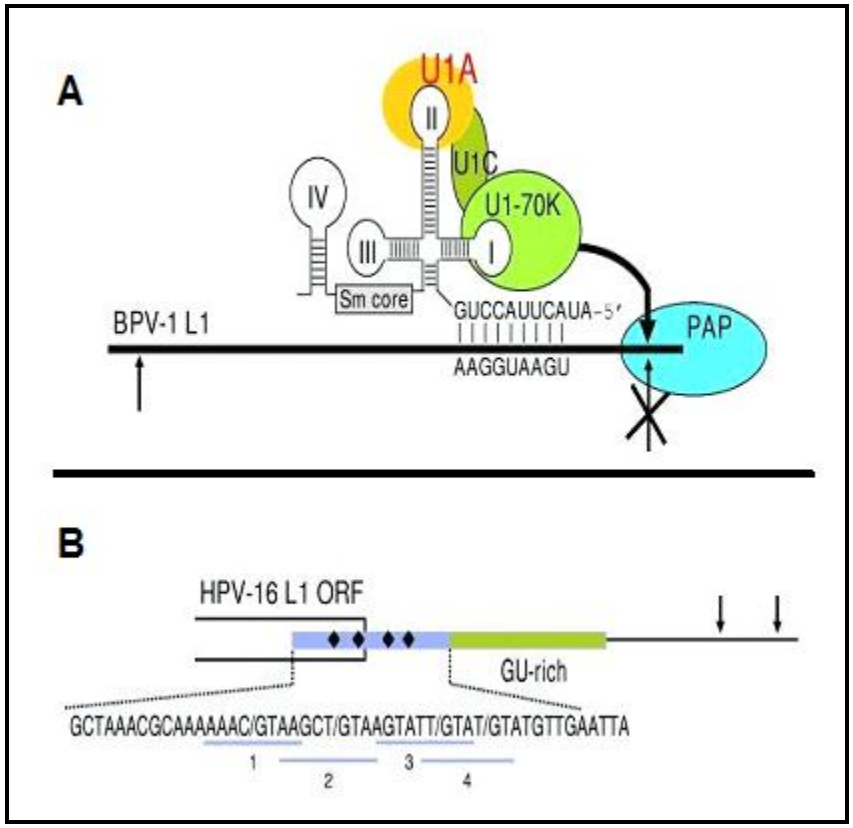

Figure 1.8 Papillomavirus Late Poly(A) Site Inhibition

A) 3' UTR of BPV-1 late pre-mRNA. A sequence similar to a 5' splice site base-pairs with the 5' end of U1 snRNA within the U1 snRNP. The U1 snRNP protein U1-70K interacts with PAP to inhibit usage of the late poly(A) site (second arrow).
B) 3' UTR of HPV-16 late pre-mRNA. This region has functionally independent 5' and 3' sub-elements. The 5' sub-element (blue region) has four weak non-consensus 5' splice sites (underlined), which are similar to the sequence element seen in BPV-1. The 3' sub-element (green region) contains a GU-rich region that by itself has no inhibitory activity but synergizes with the 5' sub-element to give enhanced inhibition of the late polyA site. (Modified from Zheng and Baker 2006)

U1-Sites: Sequences analogous to those found in papillomaviruses have more recently been found in a number of mammalian genes. These sites, hereafter referred to as U1-sites, are ten nucleotides long and closely resemble, or exactly match, the consensus 5' splice sites, but lack an associated downstream 3' splice site (Beckley *et al.,* 2001; Fortes *et al.,* 2003). These sites are conserved among mammalian U1A, MARK1, CCR5, Cdc7 and SRP40 genes (Gunderson lab, unpublished data). The effect of the U1-site on polyadenylation inhibition of the U1A pre-mRNA is the focus of the third and fourth chapters of this thesis and as such will be discussed in more detail therein.

In general, the Gunderson lab has, with collaborators, determined that U1snRNP binding to these U1-sites inhibits polyadenylation due to an interaction with PAP (Gunderson *et al,* 1998; Fortes *et al.,* 2003). This U1site:U1snRNP interaction then inhibits polyadenylation similar to that of U1A autoregulation (Gunderson *et al,* 1998; Beckley *et al.,* 2001; Fortes *et al.,* 2003). The length of the U1site:U1snRNP duplex is an important determinant of the strength of inhibition. A ten base-pair duplex is the most potent inhibitor of polyadenylation, while eight base-pairs are less potent and a duplex of seven base-pairs has no inhibitory activity (Abad *et al.,* 2008). Single point mutations (with the exception of G:U wobble base pairs) to nucleotides three-eight within the U1-site eliminated its inhibitory effectiveness, suggesting that the internal portion of the site is critical for activity (Abad *et al.,* 2008) Also, the inhibitory activity of these sites decreases if it is located within a secondary stem structure (see Chapter III, Abad *et al.,* 2008).

The exceptional potency of U1-site:U1 snRNP mediated inhibition has lead to the recent creation of a gene silencing technique called U1 Adaptors (Figure 1.9, Goraczniak

et al., 2009). Briefly, U1 adaptors are synthetic oligonucleotides divided into two domains: the first contains a sequence complimentary to the target gene's 3' terminal exon, while the second contains a U1-site sequence. The first domain hybridizes with the target pre-mRNA and the attached second domain (the U1-site) binds U1 snRNP. The presence of U1 snRNP, tethered by the U1 adaptor oligonucleotide at the target's 3' terminal exon, inhibits poly(A) tail addition in a gene specific manner, resulting in degradation of the unprocessed transcript.

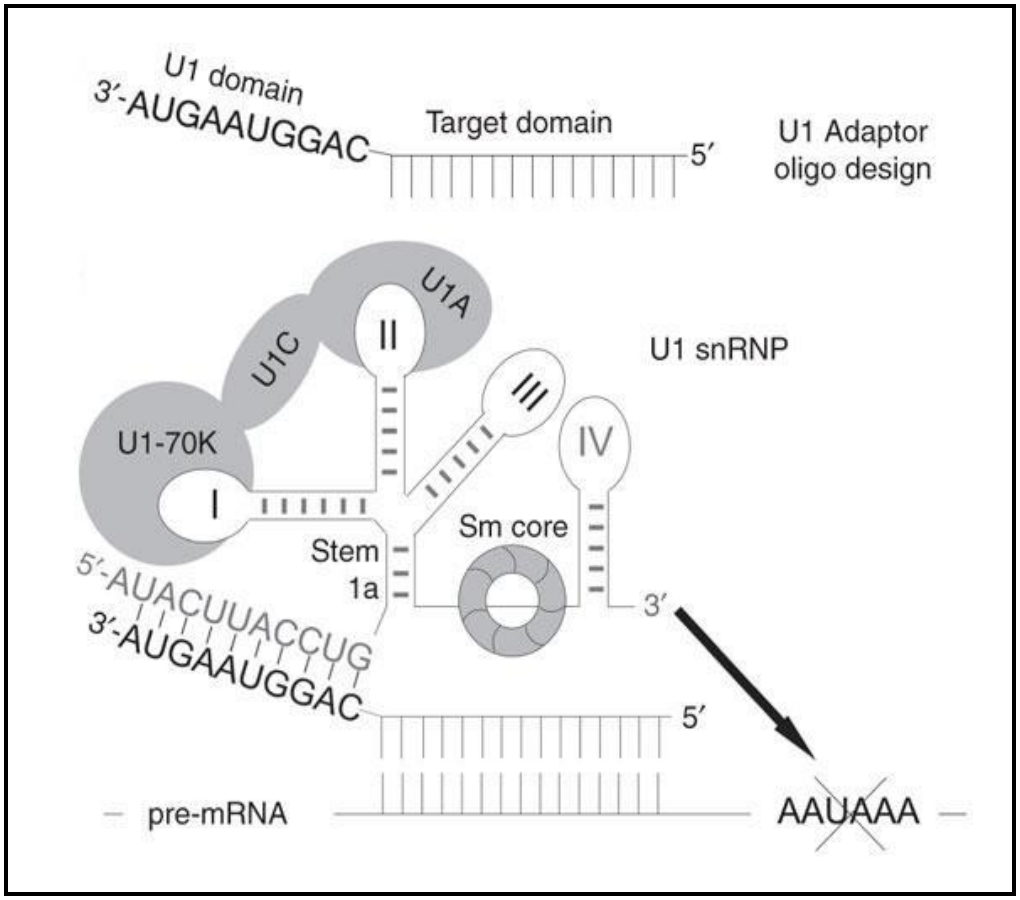

Figure 1.9 U1 Adaptor Design and Mechanism

Top shows a schematic of a U1 adaptor molecule, containing the target domain (complimentary to the target gene's 3' terminal exon) and a U1-site domain. Bottom shows a U1 adaptor molecule that has hybridized to the target pre-mRNA and attracted U1 snRNP via binding of the U1-site domain to the U1 snRNA of U1 snRNP. U1-70K then inhibits poly(A) tail addition by interacting with PAP, similar to U1A autoregulation. (Modified from Goraczniak *et al.*, 2009)

C. Summary

Variation in poly(A) site choice affects mRNA stability, export, translation and ultimately gene expression. The idea that poly(A) sites undergo regulation was once thought to be a relatively rare process. However, it is now known that between 30-50% of human genes possess more than one polyadenylation site (Beaudoing *et al*, 2000; Tian *et al.,* 2005; Yan and Marr, 2005). Consequently, polyadenylation regulation will likely become a very common phenomenon. U1A autoregulation is one of the best understood examples of "on/off" polyadenylation regulation. However, the recent discovery of U1-sites, and their role in regulation, is evidence that there is still much to be discovered concerning the mechanism and prevalence of U1A U1 snRNP and its associated subunits in regulating poly(A) sites.. The goal of this thesis is to investigate the contribution of the U1A U1-site to polyadenylation regulation.

Chapter II: Materials & Methods

Polymerase Chain Reaction (PCR)

PCR reactions were carried out in a final volume of 200 µl. Each reaction contained 20 µl of 10x PCR Buffer (500 mM KCl, 200 mM Tris, pH 8.4), 50 mM $MgCl_2$, 35 mM dNTPs, 100 pmol each of specific forward and reverse primers, 1 µg plasmid DNA template and 2 µl of homemade Taq Polymerase. The program for amplification consisted of once cycle at 95º C for five minutes, 30 cycles at 95º C for once minute; 50-55º C for one minute (temperature dependent on T_M of specific primers used); 72º C for one minute, once cycle at 72º C for five minutes, then hold at 4º C. Reactions were then purified via phenol/chloroform extraction, ethanol precipitated and resuspended in dH_2O.

Plasmid Constructs & Probe Templates

The cDNA constructs of U1A were made using Qiagen's pQE Trisystem. The plasmids contained a chicken actin promoter, Flag-Tagged U1A 5' UTR, U1A coding region, U1A 3' UTR and 150 nucleotides beyond the Poly(A) Site. The U1A 3' UTR contained various combinations of wild-type or mutant U1-site, U1-STEM region and PIE RNA. These mutants were made using specific PCR primers harboring the mutations, sometimes with the addition of an extra restriction enzyme site, then inserted into the wild-type vector via restriction digest. Probe templates were derived from these plasmids using PCR, followed by phenol/chloroform extraction, ethanol precipitation and restriction enzyme digestion. Reporter gene constructs contained an SV40 promoter, the coding region of the gene for Renilla Luciferase (RL), the U1A 3' UTR region and 260

nucleotides beyond the Poly(A) site. The U1A 3' UTR contained various mutations as in the cDNA constructs.

Recombinant Protein Preparation

Wild-type and/or mutant U1A and U1 snRNP were purified as previously described (Gunderson *et al.,* 1997 and 1998). Histidine-Tagged bovine poly(A) polymerase (bPAP) was transfected into BL21 LysS *E.* coli cells, plated and grown overnight. Four liters of LB media were inoculated with colonies and allowed to grow at 37º C, with shaking, for approximately four hours before induction with IPTG. Cells were allowed to grow for three additional hours before being spun down and harvested. The resulting pellet was resuspended in 100 mM Tris and 300 mM KCl with PMSF and glycerol, sonicated for 30 seconds and spun for 15 minutes at 15,000 rpm and 4º C. The supernatant was mixed with Ni-NTA Agarose beads for two and a half hours, washed twice with 1x GTK buffer (100 mM Tris, pH 8.0, 0.5 M KCl, 0.5 M EDTA, 50% v/v glycerol) plus 17 mM Imidazole. PAP was eluted in 400 mM Imidazole and protein concentrations of eluted fractions were determined by Bio-Rad assay.

***In vitro* Transcription with T7 or SP6 RNA Polymerase**

RNA probes were uniformly radiolabeled through *in vitro* transcription using either T7 or SP6 RNA polymerase and reagents from Ambion's Megascript T7 kit. Each 20 µl reaction contained the kit's 10x transcription buffer, 7.5 mM rATP, rCTP and rGTP, 3.75 mM rUTP, 2µl/reaction α-^{32}P-UTP (10mCi/ml, 3000Ci/mmol),

approximately 1-2 µg of digested DNA template and 2 µl of the kit's T7 enzyme mix or 1µl of SP6. Reactions were incubated at 37º C for 5-6 hours for T7 or 2-3 hours for SP6. Following incubation, 15 µl of 95% Formamide was added to each reaction, which were then heated for 4 minutes at 95º C and then loaded onto a 6% acrylamide RNA denaturing gel. After a 25 minute run, probes were cut from the gel in small slices and eluted overnight at 42º C in elution buffer (1% SDS, 20 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaOAc). Eluted RNA products were phenol/chloroform extracted and ethanol precipitated twice and finally resuspended in DEPC-treated water. The radioactivity of the probes was measured by a scintillation counter.

5'-End Labeling with T4 Polynucleotide Kinase (PNK)

Cold RNA probes for RNase digestion were generated using Ambion's Megascript T7 kit as described above. Purified products were treated with Calf Intestinal Phosphatase (CIP, Promega) in CIP Buffer (100 mM NaCl, 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM EDTA) for one hour at 37º C, phenol/chloroform extracted, ethanol precipitated and resuspended in DEPC-treated water. They were then incubated for 30 minutes at 37º C with 20 units of T4 Polynucleotide Kinase (NEB) in 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM EDTA with 2µl/reaction γ-^{32}P-ATP (10mCi/ml, 3000Ci/mmol). SDS (1%) was added to the reactions, which were then loaded onto a syringe containing G-50 matrix beads and centrifuged at 2,500 rpm for 5 minutes. Flow-thru was collected and RNA was phenol/chloroform extracted, ethanol precipitated, resuspended in 95% formamide and run on a 6% acrylamide RNA Denaturing gel. Probes were cut from the gel in small slices and eluted overnight at 42º C in elution buffer (1%

SDS, 20 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaOAc). Eluted RNA products were phenol/chloroform extracted and ethanol precipitated twice and finally resuspended in DEPC-treated water. The radioactivity of the probes was measured by a scintillation counter.

Gel Shift Assay

Electrophoretic Mobility Shift Assays (EMSAs) were performed as described (Klein Gunnewiek *et al.,* 2000; van Gelder *et al.,* 1993). 50-100,000 cpm of α-^{32}P-UTP-labeled RNA probes (wild-type or mutant) were added to 15 µl containing DRD buffer (120 mM KCl, 20% glycerol, 0.5 mM EDTA, 100 mM Tris, pH 8.0), 21 mM $MnCl_2$, 240 mM DTT, 4.5 µg BSA, 1 unit rRNasin (Promega) and 1.5 µg competitor tRNA. Appropriate amounts of either recombinant wild-type or "scrambled" U1A protein or wild-type U1 snRNP purified from HeLa cells, or both, were then added to the reactions. After gentle mixing, the reactions were immediately loaded onto a 6% (60:1) Native polyacrylamide gel using 1x Tris-Borate-EDTA (TBE) running buffer and run for two and a half hours at 55mA, 220V, 50W. Separated products were visualized by autoradiography and in some cases quantitated by phosphoimagery on an 860 Storm system.

Biotin Beads Pulldown Assay

RNA probes were generated via *in vitro* transcription using T7/SP6 RNA polymerase. Each 30 µl reaction contained 1x transcription buffer (40mM tris-HCl pH 7.9, 6mM $MgCl_2$, 2mM Spermidine, 10mM NaCl), .24M DTT, 25mM rNTPs, 0.5mM Biotin-rCTP, 0.3 units of rRNasin (Promega), 1 unit of T7/SP6 RNA polymerase, 2 µg of digested plasmid DNA and 2µl/reaction α-^{32}P-UTP (10mCi/ml, 3000Ci/mmol). In order for these probes to have a trace amount of α-^{32}P-UTP, the amount of radioactivity added is 1/10 the amount used when making uniformly labeled RNA probes. The probes were then incubated for one hour at 37ºC. Following incubation, 15 µl of 95% Formamide was added to each reaction, which were then heated for 4 minutes at 95º C and then loaded onto a 6% acrylamide RNA denaturing gel. After a 25 minute run, probes were cut from the gel in small slices and eluted overnight at 42º C in elution buffer (1% SDS, 20 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaOAc). Eluted RNA products were phenol/chloroform extracted and ethanol precipitated twice and finally resuspended in DEPC-treated water. The radioactivity of the probes was measured by a scintillation counter.

For each reaction, 20 µl of Streptavidin beads (Sigma), (20 µl bead slurry equals 10 ul of beads) were washed three times in one ml of binding buffer (150mM NaCl, 20 mM HEPES pH 7.9, 0.03% NP-40, 2 mM $MgCl_2$) by rotating for two minutes and spinning at room temperature for 30 seconds at 9000. The beads were then preblocked in 700 µl of binding buffer with 100 µg BSA, 50 µg tRNA and 100 µg glycogen, by rotating for 30 minutes, then washed three times with 1 ml binding buffer as above. The

appropriate amount of binding buffer was added to the beads in order to split them amongst the desired reaction tubes (40µl per reaction). Once split, the beads were spun for 30 seconds at 9000 rpm and the aqueous layer was removed. To each reaction tube, 100 µl binding buffer, 0.24M DTT, 10 µg tRNA and rRNasin (Promega), were added. After mixing, RNA probe was added and the reactions were rotated for 30 minutes at room temperature. The reactions were then spun down and the unbound RNAs were removed. The reactions were then washed three times with 600 µl of binding buffer. For each reaction, 72 µl binding buffer, 0.24M DTT, 12 µl 90% glycerol, 10 µg tRNA and rRNasin were added. After thorough mixing HeLa nuclear extract (NXT) and recombinant U1A (rU1A) were added. The reactions were then rotated for 30 minutes at room temperature. The reactions were again spun to remove unbound proteins and probes., then they were washed 3 times with 750 µl binding buffer with 237 mM NaCl. To elute, 40 µl of Laemmli buffer (0.1% SDS, 190 mM Glycine, 25 mM Tris) was added to each reaction and heated at 95°C for 4 minutes. Thirty µl was loaded onto a 12% SDS-PAGE gel to be analyzed via Western blot. The remaining 10 ul was added to 150 µl TESS (50 mM Tris pH 8.0, 5 mM EDTA, 0.5% SDS, 0.3 M NaOAc), extracted with phenol/chloroform and precipitated with ethanol and loaded on 8% RNA denaturing gel for analysis via Northern blot.

Western Blot

Protein samples were separated on a 12% SDS-polyacrylamide gel for two and a half hours at 55mA, 300V and 12 W, in 1x Laemli Buffer. The separated samples were then transferred to an Immobilon-P membrane (Millipore) in 1x Transfer Buffer (192 mM Glycine, 25 mM Tris, 20% v/v methanol) at 300 mA, 25V and 10W for 12 hours. After transfer, the membrane was blocked with 1x PBS, 0.1% Triton and 0.7% w/v milk powder for 30 minutes with rocking. This solution was then replaced with 60 mls of fresh blocking solution containing the primary antibody, and incubated for one hour with rocking. Anti-U1A antibody 856 was used at a 1:5000 dilution; anti-Flag antibody was used at a 1:10,000 dilution. The membrane was then washed twice with 1x PBS, 0.1 % Triton for eight minutes each, followed by incubation with the secondary anti-species horseradish peroxidase (HRP) antibody (Amersham) for one hour with rocking. Anti-rabbit was used in conjunction with anti-U1A, and anti-mouse was used with anti-Flag. The membrane was then washed two times each for eight minutes with 1x PBS, 0.1% Triton, 0.7% milk powder and 1x PBS, 0.1% Triton. The last wash of eight minutes was with 1x PBS alone. The membrane was then incubated for one minute with ECL Reagent (PerkinElmer) and products visualized by autoradiography.

Oligo Treatment of NXT

Anti-U1 oligo treated NXT and Northern and Western blotting were done as previously described (Gunderson et al. 1998). The anti-U1A antibody 856 (kindly provided by Iain Mattaj) and anti-Flag antibody (Sigma) were diluted 1:5,000, the anti-GAPDH antibody (Chemicon) was diluted 1:30,000.

Northern Blot

RNA probes for U1, U2, U4, U5 and U6 were made via T7 *in vitro* transcription in 20 µl reactions containing 1x transcription buffer (40mM tris-HCl pH 7.9, 6mM $MgCl_2$, 2mM Spermidine, 10mM NaCl), .24M DTT, 25mM rNTPs, .3 units of rRNasin (Promega), 1 unit of T7 RNA polymerase, 2 µg of digested plasmid DNA and 2µl/reaction α-^{32}P-UTP (10mCi/ml, 3000Ci/mmol). Probes were incubated for one hour at 37º C, then brought to 100 µl with 1% SDS, purified via phenol/chloroform extraction and transferred to a syringe containing a G50 matrix. After being spun for five minutes at 2,500 rpm the flow through was collected and SDS added to 1%. The radioactivity of the probes was measured by a scintillation counter.

RNA samples were run on an 8% (24:1) urea-acrylamide denaturing gel for one hour at 500mA, 1200V and 27W with 1/2x TBE buffer. The gel was then transferred to an Hybond N+ membrane (Amersham) in 1/2x TBE for three hours at 300mA, 20V and 5W. After transfer, membrane was washed for one minute in ddH_2O then placed in a Stratalinker and UV-crosslinked once on each side for 30 seconds. The membrane was

then incubated in 50 mls of prehybridization buffer (50% v/v Formamide, 5x SSC buffer [0.75 M NaCl, 75mM Na citrate pH 7.0], 5x Denhardts buffer [1mg/ml BSA, 1 mg/ml PVP polypropyl, 1mg/ml Ficoll-400], 1% SDS, 0.1 mg/ml carrier DNA) for 30 minutes at 42ºC with rocking. Ten million dpm of each RNA probe was then added to the prehybridization buffer and hybridization to the membrane occurred during an additional incubation for 12 hours, or overnight, at 42ºC with rocking. Hybridization was followed by two five minute washes with 2x SSC buffer (0.3 M NaCl, 30mM Na citrate pH 7.0) at room temperature, followed by two 30 minute washes with 2x SSC, 0.1% SDS at 55ºC, then a final wash with 2x SSC at room temperature for two minutes. The membrane was then wrapped in saran wrap to prevent desiccation and products were visualized by autoradiography.

Nonspecific Polyadenylation Assay

Nonspecific polyadenylation assays were performed as previously described (Gunderson *et al.*, 1994 and 1997). Each 22 µl reaction contained DRD buffer (120 mM KCl, 20% glycerol, 0.5 mM EDTA, 100 mM Tris, pH 8.0), 21 mM $MnCl_2$, 240 mM DTT, 2.5 µg BSA, 16 mM rATP, 1 µg tRNA, 4 M KCl, 10 % triton and 2 units of rRNasin (Promega). Reactions were prepared by adding 50-100,000 cpm of α-^{32}P-UTP-labeled RNA probes (wild-type or mutant) to a mixture containing the above reagents. Appropriate amounts of recombinant U1A, U1 snRNP and bPAP were added, the reactions gently mixed and then incubated at 37º C for 30 minutes. Resultant RNA products were extracted with TESS-tRNA and phenol/chloroform, then ethanol precipitated and resuspended in 95% formamide. They were separated via 8% denaturing

acrylamide gel, using 1x Tris-Borate-EDTA running buffer for two hours at 500mA, 1000V and 25W, then visualized by autoradiography and in some cases quantitated by phosphoimagery on an 860 Storm system.

RNase T1/V1 Digestion

For RNase T1 digestion without nuclear extract (NXT), 5'-end labeled RNA probes were resuspended in structure buffer (100 mM Tris, pH 7, 1 M KCl, 100 mM $MgCl_2$), incubated for 10 minutes at 50º C and cooled for 10 minutes at room temperature. Four µg of tRNA and appropriate amounts of recombinant U1A and/or U1 snRNP were then added to the mixture. Reactions were again incubated for 10 minutes at room temperature, followed by the addition of RNase T1 (Roche, 1U/µl, 0.0005U/reaction) and a final incubation at room temperature for 5 minutes. Reactions were stopped by the addition of TESS-tRNA and RNA was extracted by phenol/chloroform, precipitated with ethanol and resuspended in 95% formamide. Products were separated on an 8% acrylamide gel and visualized by autoradiography. For RNase T1 digestions with NXT, RNA probes were incubated in DRD buffer (120 mM KCl, 20% glycerol, 0.5 mM EDTA, 100 mM Tris, pH 8.0) with tRNA and NXT for 5 minutes at room temperature. Appropriate amounts of recombinant proteins were added, followed by a 5 minute incubation at room temperature. RNase T1 was then added and the protocol followed as for reactions without NXT. RNase V1 (Ambion) digestions, with and without NXT, were carried out as above, except for the final incubation in the presence of RNase V1 (0.1U/µl, 0.001U/reaction) being for 15 minutes at room temperature, or as per Ambion protocol.

Cell Culture, Transfection and Dual Luciferase Assay

HeLa cells were grown and passaged in D-MEM media containing 10% Fetal Bovine Calf Serum and antibiotics. 10 or 100 ng Renilla reporter constructs were transfected into 100,000 HeLa cells using Polyfect (Qiagen) on 24-well plates. Ten ng of Firefly Luciferase plasmid was co-transfected to control for transfection efficiency. Luciferase activity was measured 48 hours after transfection using Promega's Dual Luciferase Assay kit. Cells were washed with 1x PBS then lysed with the kit's Passive Lysis Buffer. After 15 minutes of rocking at room temperature, 10 µl of lysate was measured with Promega's "DLR-O-INJ" protocol in a $20/20^{n}$ Luminometer (Turner Biosystems). Relative luciferase activity was determined by normalizing the sample Renilla luciferase activity to the control Firefly luciferase activity.

Chapter III: A Bipartite U1-Site Represses U1A Expression By Synergizing With PIE To Inhibit Nuclear Polyadenylation

Guan F, Caratozzolo RM, Goraczniak R, Ho ES, Gunderson SI.
RNA. 2007 Dec;13(12):2129-40

This chapter was published in RNA Journal in 2007. Both F Guan and RM Caratozzolo contributed equally to this work. Figure 3.3, parts A-D were the result of experiments conducted during the completion of this thesis.

Introduction

Almost all eukaryotic mRNAs contain a post-transcriptionally added poly(A) tail that is important for many aspects of mRNA function. The poly(A) tail is added at the poly(A) site (pA) in the nucleus in a two-step reaction consisting of a large cleavage complex that cleaves the pre-mRNA into two fragments followed by poly(A) tail addition to the upstream fragment (Zhao et al. 1999). Typically, pA regulation involves the choice between two or more cleavage site's on a single pre-mRNA (Zhao *et al,*. 1999; Edmonds 2002) that can alter the coding region, miRNA binding sites, as well as stability and localization elements. Once considered rare, alternative poly(A) site choice is actually prevalent with current bioinformatically-based analyses of an ever increasing transcriptome database concluding that > 50% of human genes contain multiple pA's (Beadoing and Gautheret 2001; Tian *et al.,* 2005; Yan and Marr 2005). Genes with a single pA can also be upregulated or downregulated by modulating the activity of the single pA as an "on-off" mechanism. An example of this is the U1 snRNP-specific U1A

protein that negatively autoregulates itself by binding as two molecules to a Polyadenylation Inhibitory Element (PIE) found in its own 3'UTR (Boelens *et al.,* 1993; Gunderson *et al.,* 1994 and 1997). The regulatory mechanism involves the $(U1A)_2$-PIE RNA complex inhibiting the polyadenylation step by binding to and inhibiting the activity of poly(A) polymerase (PAP), the enzyme that adds the poly(A) tail to cleaved pre-mRNA. Without a poly(A) tail the U1A pre-mRNA fails to mature and is degraded by the nuclear exosome resulting in lower levels of U1A mRNA and protein.

More recently U1A was shown to regulate IgM heavy chain gene expression during B cell differentiation by inhibition of the secretory pA (Phillips *et al.,* 2001). Unlike the "on/off" autoregulation, the IgM system involves the choice between two poly(A) site's and U1A inhibits both the cleavage and the polyadenylation steps via binding to two downstream and three upstream non-consensus sites, respectively, of the secretory pA (Phillips *et al.,* 2001; Phillips *et al.,* 2004). A differentiation-specific increase in activity of the secretory pA is achieved by decreasing the levels of both total U1A and snRNP-free U1A, the latter a consequence of shifting more U1A into the U1 snRNP-bound population where it is unable to regulate polyadenylation (Phillips *et al.,* 2001; Milcarek *et al.,* 2003; Ma *et al.,* 2006).

Regulation of cleavage and polyadenylation in mammals by components of the core splicing and polyadenylation machineries is a recurrent theme that is not limited to U1A. The splicing factors SRP20 and the polypyrimidine tract binding protein activate an intron 4-located poly(A) site of the Calcitonin/CGRP gene resulting in activation of expression of the CGRP-specific mRNA (Lou *et al.,* 1998). hnRNPF and H also affect the IgM secretory pA (Arhin *et al.,* 2002) as does artificially raising CstF64 levels

(Takagaki *et al.,* 1996; Shell *et al.,* 2005). In vitro studies have shown the mammalian cleavage factor 1 (CFIm), via binding multiple sites upstream of the AAUAAA signal, can enhance pA activity of a number of genes including a CFIm subunit (Venkataraman *et al.,* 2005). As discussed next, U1 snRNP is also a potent regulator of poly(A) site's.

In mammals, U1 snRNP contains ten proteins bound to a U1 snRNA that functions early in splicing via a base-pairing interaction between U1 snRNA and the 5' splice site (ss) sequence (Will and Lührmann 1997). Separate from its role in splicing, U1 snRNP/U1 snRNA binding sites (herein called U1-sites) can inhibit gene expression by inhibiting poly(A) site activity. This was first shown in papillomaviruses where U1-sites potently inhibit expression of the viral late genes by inhibiting the late gene pA (Furth *et al.,* 1994) and later on it was demonstrated that U1 snRNP bound to the 5'ss in HIV-1 inhibits the poly(A) signal in the HIV-1 5'LTR (Ashe *et al.,* 2000). From these studies a potent gene silencing technology was developed where the artificial targeting of U1 snRNP to base-pair to a U1-site in the 3' terminal exon of either specific reporter or cellular genes resulted in strong silencing (typically 15-30 fold) of expression of that gene (Beckley *et al.,* 2001; Fortes *et al.,* 2003; Sajic *et al.,* 2007). The inhibitory mechanism involves the U1-70K subunit of U1-site-bound U1 snRNP inhibiting the polyadenylation activity of poly(A) polymerase (Gunderson *et al.,* 1998; Fortes *et al.,* 2003; Sajic *et al.,* 2007). Even if such inhibitory U1-sites match the consensus 5'ss, it was decided to call them U1-sites because they are fundamentally different in that: 1) they are not involved in splicing because there is a lack of an associated downstream 3'ss to splice to, and 2) the U1 snRNP:U1-site duplex must be $\geq$ 8bp with an abrupt loss of activity for 7bp duplexes (Fortes *et al.,* 2003). In contrast, splicing-active mammalian

5'ss sequences are highly degenerate, many of which make duplexes of < 6bp with U1 snRNP.

Although U1-sites cause a reduction in mRNA levels of their target gene, they are highly distinct from traditional mRNA stability elements in three ways. First, U1-sites block biosynthetic maturation of a gene-specific pre-mRNA in the nucleus, whereas mRNA stability elements influence stability of a pre-existing cytoplasmic pool of a gene-specific mature mRNA. Second, U1-sites only inhibit when placed in the 3' terminal exon (Beckley *et al.,* 2001; Fortes *et al.,* 2003), a restriction not applicable to mRNA stability elements (Wilusz *et al.,* 2001; Guhaniyogi and Brewer 2001). Third, U1-sites do not inhibit expression of mRNAs having a 3'-end histone stem-loop element in place of the poly(A) signal (Fortes *et al.,* 2003) whereas mRNA stability elements destabilize both types of mRNA (Pettitt *et al.,* 2002).

To date, naturally existing U1-sites have only been reported in papillomaviruses (Furth *et al.,* 1994; Cummings *et al.,* 2003) with no functional ones being reported in cellular genes. Here, bioinformatics, reporter gene and *in vitro* polyadenylation assays were used to identify and characterize a U1-site in the terminal exon of the human U1A gene. The U1-site is within a conserved bipartite element that represses U1-site activity by base-pairing to and trapping U1 snRNP in an inactive conformation. In its natural context the U1-site/bipartite element synergizes with the nearby PIE to negatively regulate U1A expression. This is the first functional U1 site to be identified in a cellular gene and of a synergistic action between two poly(A) site regulatory elements.

Results

Identification of a conserved U1 site in mammalian U1A gene 3'UTRs

To identify genes with U1-sites, a number of 3'UTR databases were searched, with the focus ultimately culminating on one located at http://bighost.area.ba.cnr.it/BIG/UTRHome (Pesole *et al.,* 2002) because: 1) it is well curated in that redundant sequences and cloning artifacts have been, for the most part, removed and, 2) it is searchable with very short, query sequences (< 15nts) with PatSearch software that, unlike standard BLAST search programs, will permit short (8-10nt) query sequences with mismatches. Positive hits were those 3'UTRs with a ≥ 8/10 uninterrupted match to the consensus U1 site sequence CAGGUAAGUA and where the match is conserved in the 3'UTRs from > two species. Surprisingly, the U1A gene, whose autoregulatory system have been extensively characterized in the past, was one of the top hits. The entire list of hits will be presented elsewhere (SIG, RG, ESH, unpublished results). Mammalian U1A 3' UTRs were aligned by ClustalW and divided into four regions (Figure 3.1), with region A having 15/56nts (~27%) identical in all seven species. Region B has 36/44nts (~82%) identical and contains an 8/10 match to the consensus U1-site that can make eight uninterrupted base-pairs with U1 snRNA, the minimal number needed for polyadenylation inhibition (Beckley *et al.,* 2001; Fortes *et al.,* 2003). Region C is only moderately conserved with 11/25nts (44%) being identical. Region D, containing the highly characterized PIE region, is the most conserved with 48/53nts (~91%) identical, with this conservation already being noted when U1A

autoregulation was first discovered (Boelens *et al.*, 1993). Strikingly, the distances between the U1 and PIE sites and the U1 and AUUAAA sites are also well conserved. This additional conservation of position made this putative U1-site unique among the top hits in our search and prompted us to determine whether it is functional. So why was the conserved pattern in regions B and C not reported in 1993? First, U1-sites had not yet been discovered and second, the only known U1A genes in 1993 were in human, mouse and *Xenopus*. Region B and C are missing in *Xenopus* whereas mouse and human 3'UTRs are too similar to make any sequence comparisons meaningful. Indeed additional bioinformatic analysis of lower vertebrate U1A genes (eg. *Xenopus laevis* and fish) shows that region B and C and the U1-site itself are not conserved whereas region D (the PIE site) is highly conserved (data not shown; request Supplementary material by emailing gunderson@biology.rutgers.edu). Thus the U1-site conservation in regions B and C is restricted to mammalian U1A genes.

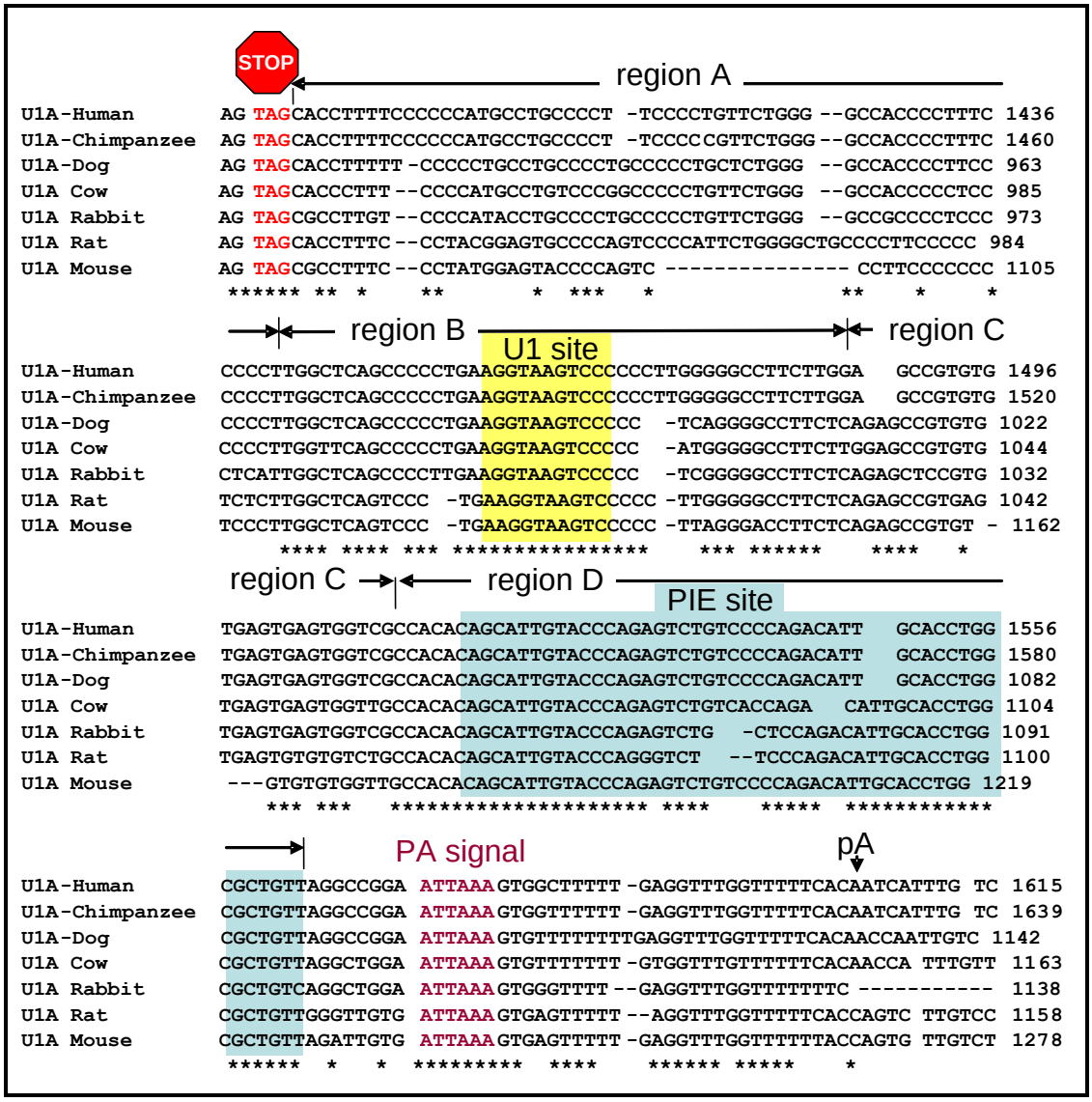

Figure 3.1 The U1-site Region is Conserved in Mammals.

Shown is a ClustalW alignment of seven mammalian U1A 3'UTRs: human (NM_004596), chimpanzee (XM_512674), dog (XM_533663), cow (BC112544), rabbit (AY387676), rat (NM_001008303) and mouse (BC003229). The numbers on the right are based on the accession numbers. The sequences span from the stop codon TAG (in red) to the poly(A) site (pA, and the human U1A poly(A) site is labeled with an arrowhead) plus several nucleotides of the genomic sequence past the cleavage site (except for rabbit U1A whose genome sequence is not available). The boundaries of regions A-D are labeled with arrows ending in a vertical bar, the putative U1 sites are highlighted with yellow, the PIE sites with green, and the AUUAAA poly(A) signals with dark pink. * means identical nucleotides. A ClustalW alignment of sequences downstream from the pA showed no significant conservation (data not shown).

FLAG-tagged U1A cDNA Expression Plasmids

To match as closely as possible the endogenous human U1A gene, U1-site function was tested in the context of the full-length natural human U1A mRNA by transfection of a Flag-tagged U1A cDNA expression plasmid under the control of a constitutive promoter (Figure 3.2A). Six matching plasmids were produced, each containing one of various combinations of wild type (wt), mutant (mt) or up-mutant (up) U1-sites with either a wt or mt PIE. The upU1-site is a perfect 10/10 match in place of the naturally occurring 8/10 match to the consensus binding site. The mtU1 and mtPIE sites were previously shown to be inactive for polyadenylation inhibition (Boelens *et al.,* 1993; Fortes *et al.,* 2003).

The U1 and PIE Sites Synergize But the U1-site Has No Activity When PIE is Mutated

Transfected cells were analyzed by Western blotting (an example is shown in Figure 3.2B), the signals quantitated with ImageQuant 5.2 software and the results graphed (Figure 3.2C) as inhibitory activities based on the following calculation. First, the FlagU1A signal was normalized by comparison with the co-transfected FlagPRP28 signal. Second, the double mutant mtU1/mtPIE plasmid had the highest expression level as it lacks inhibitory elements and so its inhibitory activity was set to 1.0 as the reference plasmid. Finally, dividing the normalized signal of the mtU1/mtPIE reference plasmid by the test plasmid gave the fold-inhibition. For example, the mtU1/wtPIE plasmid was 3.0-

fold less expressed than the mtU1/mtPIE plasmid thus, the mtU1/wtPIE plasmid and consequently its PIE site has a 3.0-fold inhibitory activity, a value consistent with the previously reported 3.1-fold inhibitory activity in vivo when PIE was inserted into a reporter plasmid (Boelens *et al.,* 1993). To confirm that the Western blot signal is dose dependent, a titration of HeLa nuclear extract (NXT) as compared to recombinant (r) U1A was done (Figure 3.2D). All of the Western blot analyses used throughout this work were done in this range. Supplementary material on the U1A protein preparations used in this work can be obtained by emailing gunderson@biology.rutgers.edu.

Mutation of the U1A U1 site in its natural context reduced inhibitory activity from 8-fold to 3-fold (a 2.7-fold decrease). It was therefore surprising that the U1-site showed no (1.2-fold) statistically-significant inhibitory activity when PIE was mutated. This lack of inhibitory activity sharply contrasted with previous work by members of this lab (Fortes et al., 2003) and others (Furth *et al.,* 1994; Beckley *et al.,* 2001) that single U1-sites inserted into reporter gene 3'UTRs strongly inhibited expression ranging from 15- to 32-fold. In those publications a wide variety of reporter genes were used, including BetaGalactosidase, Chloramphenicol Acetyl Transferase, Firefly and Renilla luciferase, GFP, YFP, RFP, as well as different reporter 3'UTR and poly(A) signal sequences (eg. Adenovirus, SV40, bovine growth hormone gene), indicating U1-sites maintain inhibitory activity in a variety of sequence contexts. One explanation, that such low activity is due to suboptimal base-pairing of U1 snRNA with the U1A U1-site, was ruled out for two reasons. First, the exact same 8/10 U1 site gave an 18-fold inhibition in a luciferase reporter context (Furth *et al.,* 1994) and second, replacing the U1-site in wtU1/mtPIE with a perfect 10/10 up-mutated U1 site (upU1/mtPIE) still gave only 1.4-

fold inhibition. Notably the upU1 site, like the wtU1, could also synergize with the wtPIE site. Results statistically the same as in Figure 3.2C were observed when the FlagPRP28 control plasmid was omitted or when the Flag tag on U1A was replaced with an HA tag indicating neither had an effect (data not shown). Furthermore, transfection of HeLa cells in place of HeLa-Tet and transfection of various amounts of each construct gave similar inhibitory activities, the latter demonstrating endogenous factors in vivo, such as U1A and U1 snRNP, were not limiting (data not shown). Thus the U1-site is only active in conjunction with a wtPIE site as alone it has little or no detectable activity.

The U1A Coding and 5'UTR Sequences are Dispensable

To determine whether the U1A coding and 5'UTR sequences affect expression, they were replaced with the Renilla coding region and reporter 5'UTR. As compared to the multi-step Western blot method, the Renilla system is simple to use, has a much more rigorous and accurate normalization method, and has a far-better dynamic range to measure protein expression levels and hence inhibitory activities. The activity of each Renilla plasmid was normalized to a co-transfected control Firefly plasmid and the inhibitory activity of the double mutant RL-mtU1/mtPIE was set to 1.0. As shown in Figure 3.2D, the pattern of inhibitory activity was nearly identical to that of the full length cDNA plasmids indicating the coding and 5'UTR regions of the U1A mRNA do not affect the inhibitory activity of the 3'UTR.

The U1 and PIE Sites Lead to Reduced mRNA Levels

As previously shown, U1 and PIE sites in other reporter gene contexts inhibit by reducing mRNA levels (Boelens *et al.*, 1993; Fortes *et al.*, 2003). To confirm this is the case here, total RNA from HeLa cells transfected with the various cDNA plasmids were analyzed by ribonuclease protection assay (RPA) where the RPA probe anneals to the 60nt Flag epitope tag in U1A but not to the Flag tag in the FlagPRP28 control or to endogenous U1A (Figure 3.2D). Parallel RPAs with an anti-GAPDH probe were done to control that the quality and amount of RNA used were the same. As can be seen in Figure 3.2E the affect on mRNA levels by the U1 and PIE sites closely tracks the protein levels seen in Figure 3.2D. Although reduced mRNA levels are likely a consequence of inhibiting mRNA biosynthesis by inhibiting poly(A) site activity, as this is consistent with prior work on each element, it was proposed that the juxtaposition of the U1 and PIE sites together in the natural U1A cDNA would affect mRNA stability. However, this was ruled out because actinomycin treatment of cells transiently transfected with the various U1A cDNA plasmids, followed by quantitative PCR analysis, indicated mRNA stability does not change (data not shown).

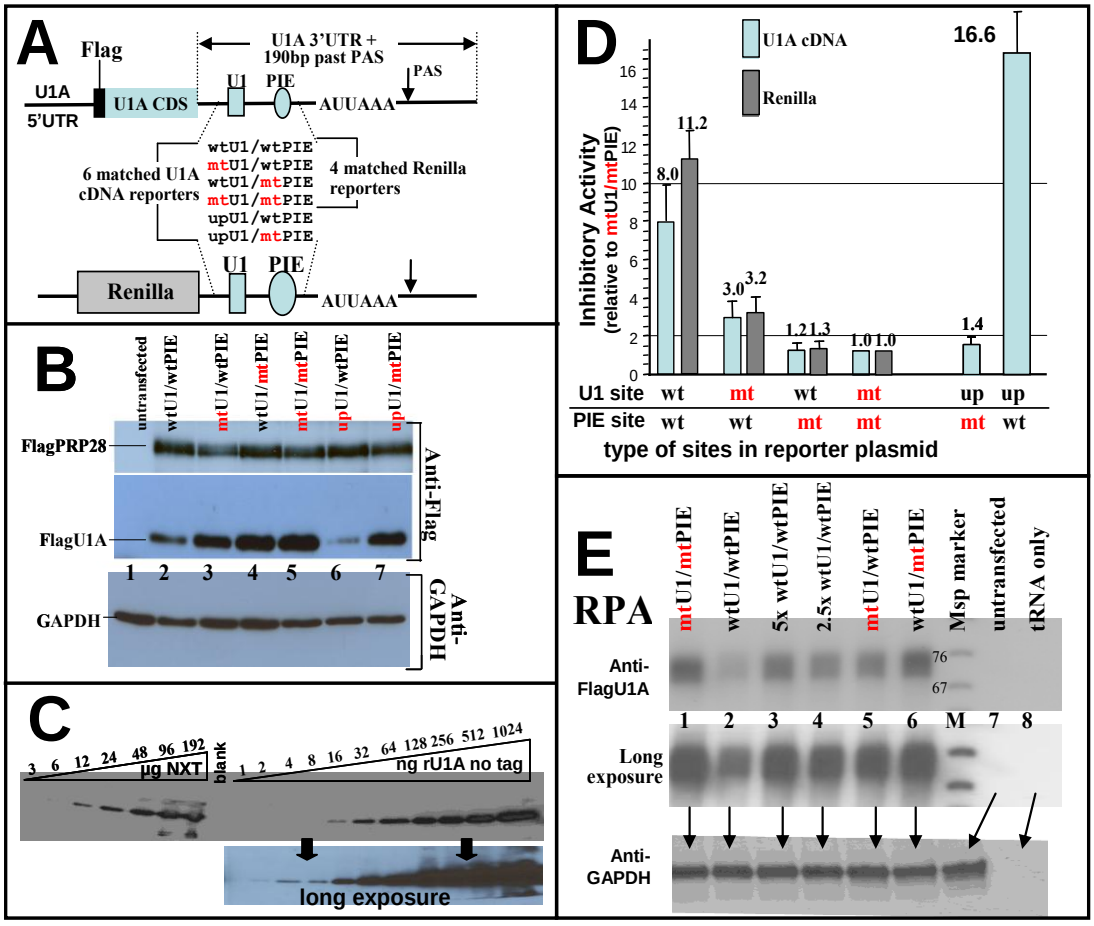

Figure 3.2 Inhibitory Activity of the U1 and PIE sites.

A) The six U1A cDNA plasmids contain a chicken actin promoter driving expression of the full-length human U1A cDNA where the 3' UTR has one of various combinations of the wildtype (wt), mutant (mt), or up mutant (up) U1-site combined with a wt or mt PIE site. The FLAG tag serves to distinguish transfected from endogenous U1A and was placed at the N terminus as this inactivates the N terminal RNA binding domain thus producing a U1A protein unable to autoregulate either the transfected U1A expression plasmid or the endogenous U1A gene (Gunderson *et al.*, 1997). To more closely match the endogenous gene, 190bp of human genomic U1A gene sequence starting from the natural poly(A) site (in this figure called PAS) to 190bp downstream were inserted into the corresponding Poly(A) site position of the cDNA plasmid. Four Renilla reporter plasmids (Promega) with a CMV promoter were modified by replacing the 3'UTR and polyadenylation signal sequences with the corresponding sequences from the U1A cDNA including 190bp past the poly(A) site. B) Western blot. Each U1A cDNA plasmid (0.5µg) in panel A was transfected into = four million Hela-Tet cells along with 0.5µg Flag-PRP28 expression construct that controls transfection efficiency. Hela-Tet cells from the "Tet-Off" system were used because the FlagPRP28 reporter is under the control of a Tet-responsive promoter. After 48 hours, cells were lysed in laemmli buffer and

10µg of protein loaded on a protein gel for Western blotting followed by sequential probing with anti-Flag (upper two panels are taken from the same exposure) and anti-GAPDH (lower panel) antibodies as shown. The blot was then probed with anti-U1A antibody (data not shown). C) The dynamic range of the Western blotting protocol was analyzed by performing a serial dilution of HeLa NXT and of purified recombinant (r) untagged U1A. The blot was probed with anti-U1A856 antibody. D) Western blot results from eight independent transfection experiments were quantitated and summarized as a graph with the inhibitory activity of the double mutant being set to 1.0. Also shown are results of eight independent transient transfections of each Renilla reporter where Renilla activity was normalized to a co-transfected Firefly reporter with the inhibitory activity of the double mutant being set to 1.0. The relative inhibitory activities were determined as described in the text. E) RPA analysis of total RNA from transfected cells was used to measure mRNA levels produced from the U1A cDNA plasmids. The autoradiograph is representative of the data we obtained when analyzing total RNA from other independent transfections. The 32P-labelled anti-Flag probe specifically recognizes the Flag tag in the U1A cDNA but not that of the FlagPRP28 plasmid. The protected RPA product for the anti-Flag probe runs as a broad series of bands from 70-75nts because the 3' end is AU rich and so is "nibbled" back by the Rnase digestion step. Also shown is a longer exposure of the anti-Flag probed RPA gel. Lanes 1, 2, 5, 6 and 7 are as in Panel B. Lanes 3 and 4 are total RNA from cells transfected with 5x and 2.5x as much wtU1/wtPIE plasmid so as to assess the response of the assay to increasing amounts of mRNA. The same total RNA samples were also probed on a different gel with 32P-labelled anti-GAPDH as a control. Lane M is a size marker of ^{32}P-labeled Msp1-digested pBR322 with the 67nt and 76nt length bands indicated.

The Isolated U1A U1-Site Strongly Inhibits Renilla Expression

As mentioned above, it was unexpectedly found that the U1A site has no activity on its own and only modest activity when PIE is wild type. In order to understand these unexpected results, the 3'UTR of RL/wtU1/wtPIE was reduced to a minimum to make the RLm plasmid (m=minimal) where sequences necessary for poly(A) site activity were retained while the upstream regions A, B, C and D were deleted (Figure 3.3A). As RLm lacks inhibitory elements its inhibitory activity was set to 1.0 as the reference plasmid.

Regions A, B and C had no intrinsic inhibitory activity when each was inserted into RLm (data not shown). Insertion of wtPIE (region D) to make RLm/wtPIE decreased expression 2.8-fold relative to insertion of a mutated PIE (RLm/mtPIE) indicating PIE has a 2.8-fold inhibitory activity, nearly the same value as it had in Figure 3.2. In contrast, insertion of just part of region B, namely the U1A U1-site (10nts), to make RLm/wtU1/wtPIE and RLm/wtU1/mtPIE, caused a 12-fold repression both when PIE was wild type (12.2 divided by 1.00) and when PIE was mutated (33.4 divided by 2.8). This 12-fold effect was specific as two matching control plasmids with a mutated U1-site gave no affect. The 12-fold magnitude of inhibition is consistent with what has previously been observed for inhibition mediated by the 5' end U1 snRNA base-pairing to a U1-site. To rigorously demonstrate that such base-pairing is occurring *in vivo* a compensatory or "suppressor" U1 snRNA was expressed that has its 5' end complementary to the mutated U1-site. If the suppressor U1 snRNA restores inhibition to a reporter with such a mutated U1-site then it can be concluded that inhibition is mediated by base-pairing to U1 snRNA. Such suppressor U1 snRNA experiments have been done by for reporter genes as well as for the papillomavirus late genes (Furth *et al.,* 1994; Beckley *et al.,* 2001; Fortes *et al.,* 2003). Expression of a suppressor U1 snRNA restored inhibitory activity to the mutant U1-site demonstrating base-pairing of the 5' end of U1 snRNA to the U1-site is required for inhibition (data not shown).

A Bipartite Element Consisting of a U1-site Repressed by a U1-STEM Structure

Next, experiments were conducted to determine why the U1A U1-site in Figure 3.3A was now highly inhibitory and its activity independent of PIE. By use of reporter assays, it was recently demonstrated that a highly active U1-site could be fully repressed when completely base-paired within a stem structure (Fortes *et al.,* 2003). Inspection of the natural U1A 3'UTR identified a putative secondary structure in region B, here designated the U1-STEM (Figure 3.3B), where the upstream half of the U1-site is base-paired to the U1-STEM while the downstream half is in an exposed loop. Strikingly, the putative base-pairing pattern is perfectly conserved in all seven mammals (data not shown; supplementary material is available by emailing gunderson@biology.rutgers.edu) and these sequences are the most conserved part of the sequences flanking the U1A U1-site. To test the role of the U1-STEM, 18nt of the wild type U1-STEM (wtSTEM) were inserted into the four U1-site-containing plasmids shown in Figure 3.3A and the four resulting plasmids tested for activity in transfected cells. Pairwise comparison of Panel A with Panel C plasmids indicated the 18nt wtSTEM caused a 10-fold repression when PIE was wild type (33.4-fold to 3.4-fold) and when PIE was mutant (12.2-fold to 1.2 fold). This repression was specific to the U1-site because the wtSTEM had no affect on PIE activity (still ~3-fold) and had no activity on its own (compare RLm/mtU1/mtPIE to RLm/mtU1/wtSTEM/mtPIE). To demonstrate base-pairing was the basis of repression, as opposed to changing the spacing between elements, four additional plasmids were made where an 18nt mutated U1-STEM (mtSTEM) sequence was inserted where the mutation should significantly reduce base-pairing potential. The mtSTEM had no

activity on its own and did not affect PIE activity which was still ~3-fold. As predicted, the mtSTEM restored U1-site activity both when PIE was wild type (11.6 to 3.4-fold) and when PIE was mutated (4.7- to 1.2-fold), thus strongly supporting the U1-STEM:U1-site base-pairing model. Although this restoration was specific, its magnitude was not quite at the level of the Figure 3.3B plasmids which have a deleted U1-STEM. This suggests that the mtSTEM could still have residual base-pairing to the U1 site. Interestingly, the pattern of the wtSTEM-containing plasmids in Figure 3.3C as compared to the Renilla plasmids in Figure 3.2C indicate synergy between the U1 and PIE sites is lost. Insertion of region C (but not region A) restored this synergy to the levels seen in Figure 3.2C (data not shown; supplementary material can be obtained by emailing gunderson@biology.rutgers.edu). Thus, these assays have identified 1) a bipartite U1 element comprised of a highly inhibitory U1 site being repressed via base-pairing to a U1-STEM and 2) that region C is needed for synergy between the bipartite U1 element and PIE.

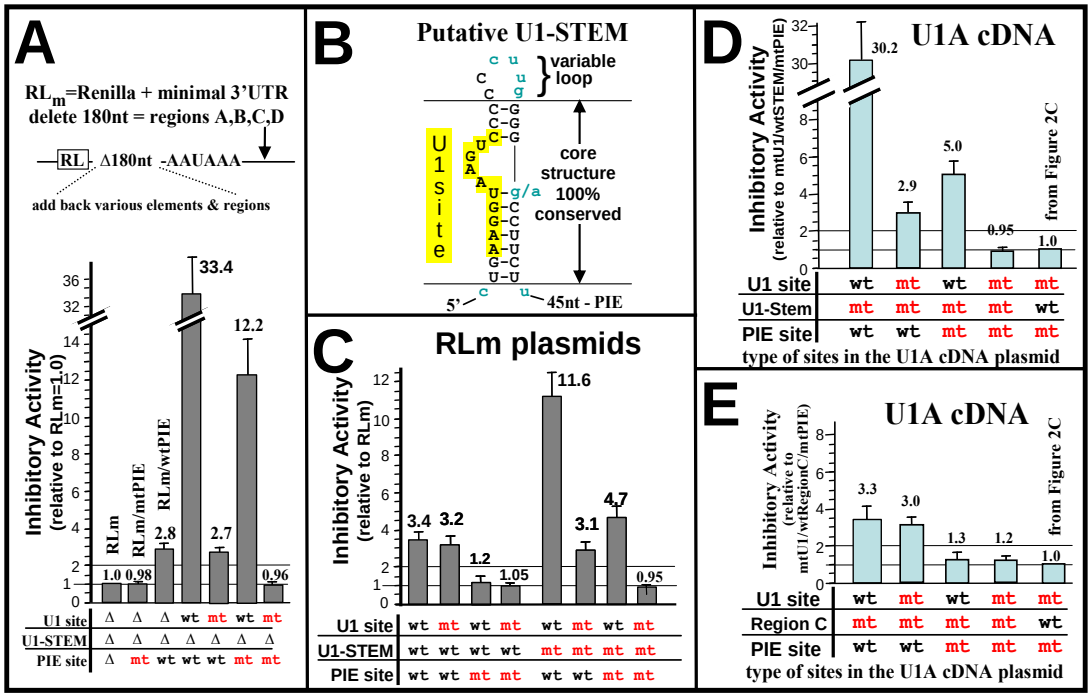

Figure 3.3 The U1 site is Suppressed by Base-pairing to a U1-STEM and Region C Contributes to Synergy.

A) RLm is a deletion mutant of the Figure 3.2A Renilla reporter that retains the minimal core poly(A) signal sequences while deleting upstream regions A, B, C and D. Each upstream region was then reinserted into RLm either alone or in combination with other regions and the inhibitory activity measured by transfection as in Figure 3.2. The plasmids were analyzed as described in Figure 3.2. As shown the U1A U1-site has high inhibitory activity suggesting that flanking sequences present in Figure 3.2, but absent here, are repressing U1 site activity. B) Conservation in mammalian U1A genes of base-pairing between the U1-STEM and the U1-site. The specific sequence shown is from human U1A. The C:G of the variable loop is shown to not be base-paired because it is not conserved in two of the seven mammalian U1A genes as shown in Figure S2. Perfectly conserved nucleotides are in uppercase black font, non-conserved in lowercase green font. C) Analysis of the U1-STEM or a mutated U1-STEM when inserted into the four U1-site-containing plasmids in Figure 3.3A. D) Analysis of the U1-STEM mutation (Panel D) and the region C mutation (panel E) in the context of the U1A cDNA plasmid. The assays and analysis were done as in Figure 3.2.

The Bipartite Element and Region C Have the Same Activity in the Context of the U1A mRNA

To test whether the U1-STEM represses in the context of the natural U1A mRNA, four plasmids were made that identically match the four Figure 3.2A U1A cDNA plasmids (the two upU1 site plasmids were not included) except the U1-STEM was mutated with the same mutation as in Figure 3.3C. An additional set of four U1A cDNA plasmids were made where region C was mutated rather than the U1-STEM. All eight plasmids were analyzed by transfection as in Figure 3.2 and the results are graphed in Figures 3.3D (U1-STEM mutants) and 3.3E (region C mutants). Pairwise comparison of the plasmids in Figure 3.2C with those in Figure 3.3D-E indicate that by themselves the U1-STEM and Region C have no activity and do not affect the activity of PIE which remained at 3-fold. In contrast, mutation of the U1-STEM specifically derepressed the activity of the U1-site both when PIE was mutated (1.2-fold in Figure 3.2C to 5.0-fold) or wild type (8-fold in Figure 3.2C to 30.2-fold). Mutation of Region C specifically caused a loss of synergy between the U1 and PIE sites. Therefore, both the U1-STEM and region C function in the full-length U1A mRNA, have no inhibitory activity on their own and have no affect on PIE. Furthermore, the U1-STEM specifically represses the U1 site while region C specifically affects synergy.

Synergy can be reconstituted in HeLa NXT

To elucidate the mechanism of inhibition seen *in vivo*, *in vitro* cleavage and polyadenylation assays were next performed in HeLa nuclear extract (NXT) with RNA substrates derived from the U1A 3'UTR and poly(A) site region extending from region B to 80nts past the poly(A) site (see Figure 3.5A for a schematic of the RNA). This assay system was chosen as it contains all the factors necessary to reconstitute U1A:PIE autoregulation by itself and U1 snRNP:U1-site-mediated polyadenylation inhibition by itself. Previous work by members of this lab showed U1A-mediated inhibition requires addition of recombinant (r) U1A (Boelens *et al.,* 1993) because nearly all (~95%) of endogenous U1A is in the U1 snRNP and therefore is unavailable to bind PIE (Boelens *et al.,* 1993, *Ma et al.,* 2006). The various rU1As used in Figures 3.4 and 3.5 were judged to be pure based on coomassie stained gels and Western blotting (data not shown; supplementary material can be obtained by emailing gunderson@biology.rutgers.edu).

Inhibition of cleavage/polyadenylation of the wtU1/wtPIE RNA was seen upon addition of 5ng of rU1A (Figure 3.4A, lane 3) and increased amounts of rU1A gave increased inhibition. If the inhibitory activity of the $(U1A)_2$-PIE complex on the wtU1/wtPIE RNA is independent of the U1-site then it would be expected that the mtU1/wtPIE RNA would have the same inhibitory response to added rU1A. Instead, 12.5ng of U1A were needed to observe inhibition of the mtU1/wtPIE RNA (lane 11) indicating the wtU1 site strengthens the rU1A-mediated inhibition. Most importantly, this increase in inhibition correlates well with the 2.7 fold difference in synergistic inhibitory activity that the U1 and PIE sites exhibit *in vivo* (Figure 3.2) indicating

synergy has been reconstituted with this *in vitro* assay. Additional controls underscored the specificity of these results as addition of rU1A did not inhibit the wtU1/mtPIE RNA (lanes 16-18) or the mtU1/mtPIE RNA (data not shown).

To understand the basis of the synergy, up to a 50-fold stoichiometric excess of purified U1 snRNP over RNA substrate was added to these assays and no affect was seen, indicating U1 snRNP is not limiting (data not shown). To eliminate base-pairing by U1 snRNP, a well-known method where NXT is pre-treated with a U1 oligo complementary to nts 1-12 of U1 snRNA was used (Krämer and Keller 1990). Rnase H intrinsic to NXT degrades the RNA strand of the RNA:DNA duplex, that is nts 1-12 of U1 snRNA are removed, thereby preventing U1 snRNP:U1-site base-pairing. As a control, NXT was also pre-treated with an unrelated oligo (non-specific oligo). As shown in the Northern blot in Figure 3.4B, the U1-oligo quantitatively and specifically removed ~10nt off of U1 snRNA. Figure 3.4B lanes 4-9 reproduces published data where U1-oligo treatment strongly derepresses polyadenylation of an Adenovirus L3 polyadenylation substrate RNA that artificially contains a U1 site (AdL3+wtU1) (Gunderson *et al.*. 1998). Two controls demonstrated that the derepression is specific. First, the U1 oligo had no affect on a matching substrate having a mutated U1 site (AdL3+mtU1). Second, mock treatment with a non-specific oligo had no affect. Next, these oligo-treated NEs were used to analyze the U1A RNA substrates. In contrast to the AdL3+wtU1 RNA, polyadenylation of the wtU1/wtPIE U1A RNA was not derepressed, as illustrated by the inhibition profile of U1 oligo- and nonspecific oligo- treated reactions, which were the same (Figure 3.4C compare lanes 2, 9, 16 and 22). However, the U1 oligo treatment did cause loss of synergy (Figure 3.4C compare lanes 16-21 with

22-27) as compared to the nonspecific-oligo-treated reactions (compare lanes 2-7 with 9-12). Thus synergy can be reconstituted in vitro by addition of rU1A to HeLa NE and synergy requires the 5' end of U1 snRNP and the U1 and PIE sites. Next, the binding of these factors to the RNA was examined.

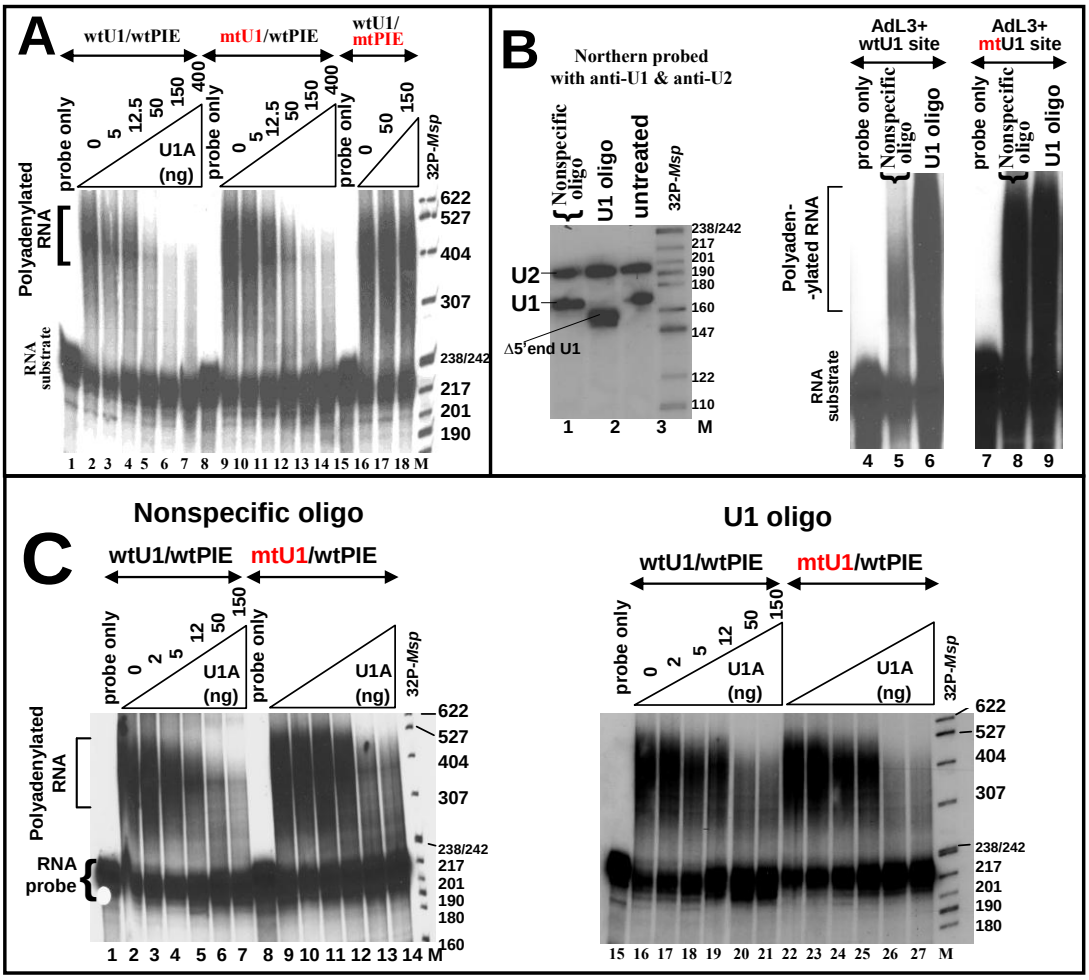

Figure 3.4 RNA-bound U1A and U1 snRNP synergistically inhibit polyadenylation in vitro.

A) Polyadenylation assays were done to test the inhibition of various U1A-3'UTR-derived RNA substrates by U1A protein. The sequence of the RNA probes are the same as in Figure 3.5. Hela NXT with zero (lanes 2, 9 and 16) or increasing amounts of rU1A were added to the probes at the start of the polyadenylation reaction. After one hour, the polyadenylated RNAs were extracted and separated by denaturing PAGE. Lanes 1, 8 and 15 are probe + buffer but no NXT and so do not get a poly(A) tail added. The positions of the RNA substrates and polyadenylated RNA

products are shown. The size marker lane is ^{32}P-labeled Msp1-digested pBR322 with nucleotide lengths indicated. Panels B-C: Removal of the 5' end of U1 snRNA results in loss of synergistic inhibition. B) Lanes 1-3 are a Northern blot of total RNA from HeLa NXT treated with a nonspecific oligo (lane 1) or a U1 oligo (lane 2) or untreated (lane 3). The Northern blot was probed with anti-U1 snRNA (anti-U1) and anti-U2 snRNA (anti-U2) probes. The size marker lane is as in panel A. As described in the text and Panel A, lanes 4-9 are polyadenylation reactions but with oligo-treated NXT and different RNA substrates. Lanes 4 and 7 are probe + buffer but no NXT and so do not get a poly(A) tail added. C) As described in the text and Panel B, polyadenylation reactions with U1A-derived RNA substrates were done with HeLa NXT treated with a nonspecific oligo (lanes 1-14) or the U1 oligo (lanes 15-27). Note the rU1A titration is slightly different than in Panel A. Lanes 1, 8 and 15 are probe + buffer but no NXT and so do not get a poly(A) tail added.

Synergy is Not Based on Enhanced Binding of U1 snRNP or U1A to the 3'UTR

One straightforward model to explain synergy is that the U1-STEM occludes binding of U1 snRNP to the U1 site and that the U1A:PIE complex overcomes this occlusion. To test this, biotin-selection ("pull down") assays were done using the same RNA probes as in Figure 3.4 except that they were double-labeled with biotin-CTP and ^{32}P-UTP (Figure 3.5). Biotin-U1A-RNA probes bound to streptavidin beads were mixed with HeLa NXT and zero or increasing amounts of recombinant U1A (rU1A) protein. After wash steps, specifically bound protein(s) and/or RNAs were eluted with SDS buffer and analyzed by Western and Northern blotting (Figure 3.5B). The U1A Western blot signal corresponds to exogenously added rU1A + U1 snRNP-bound U1A + snRNP-free-U1A in the NXT. The Northern membrane was probed with ^{32}P-anti-U1 snRNA to measure the amount of U1 snRNP pulled down. The Northern blot also contains signals from the ^{32}P-labeled, biotin-RNA probe that co-eluted from the beads which controlled for: 1) uniform transfer of RNAs to the Northern membrane, and 2) that similar amounts of U1A RNA probe were used for each pull down and that the RNA had not been

degraded during all of the incubation, wash and elution steps. Increasing both NXT (lanes 15-17) and rU1A (lanes 5-12) gave increased signals demonstrating that the U1A RNA probe was in stoichiometric excess over both U1A and U1 snRNP, a requirement for detecting changes in binding. Specificity of binding was confirmed as mutation of the U1-site (lanes 9-12) and both sites (lane 18) resulted in a significant reduction in binding. Probing the Northern to detect other spliceosomal U snRNAs (U2, U4, U5 and U6) demonstrated their binding was very low to not detectable (for an example see U2 snRNA in Figure 3.5D). The mtU1/wtPIE probe (lane 9) but with no added rU1A gave a visible Western blot signal only upon longer exposures consistent with previous data that the amount of snRNP-free U1A in HeLa NE is relatively low (Boelens *et al.,* 1993).

Having established that the assay conditions were both specific and in probe excess, the key question could be addressed, namely does increasing the U1A:PIE complex lead to changes in the levels of U1-site-bound U1 snRNP. As expected, increasing amounts of rU1A added to the wtU1/wtPIE probe gave an increasing Western blot signal of U1A, however, the Northern blot shows no change of U1 snRNA bound to the probe (lanes 5-8). This indicates that 1) the amount of probe is in excess over U1A protein, and 2), the binding of rU1A to the PIE site has no affect on the binding of U1 snRNP to the U1-site. Likewise, if U1 snRNP is affecting binding of U1A to PIE then the Western blot signals in lanes 6-8 would be either increased or decreased relative to lanes 10-12. Instead quantitation of these signals, which included normalization with the eluted RNA probes shown on the Northern blot panel, indicates there is no difference. Finally, the signals in lanes 5 and 13 are of the same intensity indicating that the presence or absence of the PIE site has no detectable affect on U1 snRNP binding. Thus these data

support the conclusion that synergy is not due to enhanced binding of U1 snRNP or U1A to the 3'UTR.

To clearly distinguish rU1A from endogenous, these experiments were repeated with a C-terminal Flag tagged rU1A protein (rU1Aflag) as shown in Figure 3.5C. The various rU1A preparations were judged to be nearly 100% pure and Western blots titrating recombinant U1A versus endogenous U1A in NXT were done to calibrate the amounts of protein and to avoid interfering of their Western blot signals (Figure 3.2B lower panel). As can be seen most of the endogenous U1A signal is coming from U1 snRNP-bound U1A (compare lanes 9-12 with 13-15). Northern blotting showed addition of rU1Aflag had no affect on the U1 snRNA signal (data not shown), which is in agreement with the Northern in Figure 3.5B. To demonstrate that the specificity of U1 snRNP binding, the 5'-end of U1 snRNA was removed and it was observed that U1 snRNP binding is completely absent (Figure 3.5D). In closing, it is worth noting the striking lack of correlation between binding of U1 snRNP to the U1 site and the inhibitory activity measured *in vivo*. In particular, the upU1/wtPIE plasmid gave a 16-fold inhibitory activity *in vivo* as compared to the wtU1/mtPIE (1.2-fold) or the wtU1/wtPIE (8-fold) plasmids (Figure 3.2). Nevertheless, all three RNA substrates bound U1 snRNP with approximately the same efficiency (Figure 3.5B). Furthermore, it was expected that the U1-STEM would inhibit U1 snRNP binding, however biotin pull-down assays comparing RNA substrates with a wtSTEM to those with a mtSTEM showed no difference in U1 snRNP binding (data not shown). As described below, a m model is proposed where U1 snRNP is bound to the wild type bipartite element but is unable to inhibit unless the PIE site is wild type and occupied by U1A.

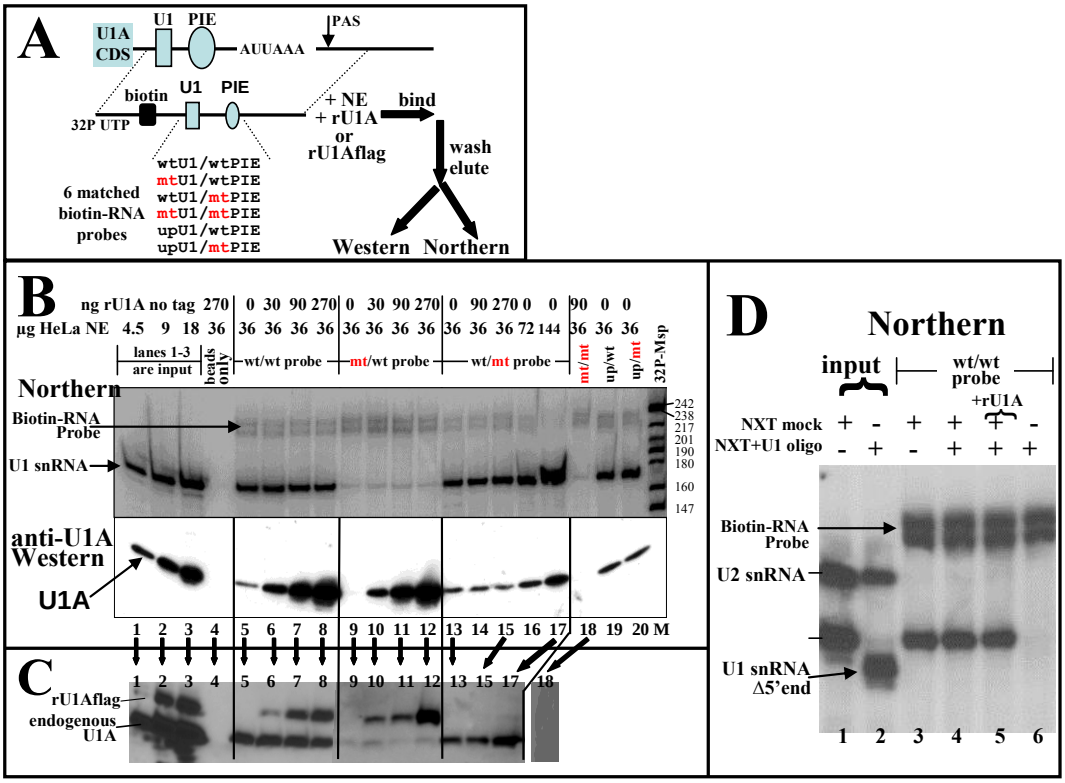

Figure 3.5. U1A and U1 snRNP Don't Significantly Affect Each Other's Binding in HeLa NXT

A) Schematic of a biotin-selection protocol that is described in the text. The RNAs are double labeled with ^{32}P-UTP and biotin (black square) and are derived from the U1A 3'UTR. (In this figure, the poly(A) site is called PAS) B) Shown are the Northern and Western blot results of the biotin-selection assay. The type of double labeled biotin-RNA-probe and the amount of rU1A and NE added to each reaction are indicated. As described in the text, these double labeled probes also appear on the Northern blot and serve as an internal control. To detect the amount of unlabeled U1 snRNP in the pull down, the Northern blot was probed with anti-U1 snRNA probe. The Western blot was probed with anti-U1A antibody. Lanes 1-3 are NXT inputs showing the linear response of both the Northern and Western blots in this range. Lane 4 shows the background binding to beads without any RNA probe. The amounts of NXT and exogenously added rU1A are indicated. Also indicated are the positions of the double labeled RNA probes, U1 snRNA and U1A. The size marker lane is as in Figure 3.4. The entire experiment was repeated 5x with results consistent to the example shown here. C) Panel B was repeated but rU1Aflag was used in place of rU1A. Three times less rU1Aflag was used so that its Western blot signal would not mask that of endogenous U1A and film exposures were longer to help visualize the weak endogenous U1A signal, especially in lanes 9-12. Lane 18 was spliced in from a different part of the same autoradiograph.

D) The biotin assay was done with HeLa NXT either mock treated with a non-specific oligo (lane 1) or U1 oligo treated to remove the 5' end of U1 snRNA. Lane 1 contains twice as much RNA as lane 2 so as to confirm the assay is dose dependent. The Northern blot was probed with anti-U1 and anti-U2 snRNA probes. Lanes 4 and 5 are controls where the mock treated and U1 oligo treated reactions were mixed just prior to selection in order to show the U1 oligo does not affect the biotin selection step. Lane 5 also included 270ng rU1A during the incubation step with the biotinylated RNA. The Northern blot shows U1 oligo treatment caused complete loss of binding of U1 snRNA to the wt/wt biotin probe as compared to mock treatment. Results were the same for the other 3 types of probes (data not shown).

Discussion

An evolutionarily conserved 10nt binding site for the U1 snRNP (U1-site) was identified in the U1A gene's 3'UTR and shown to function in repressing U1A expression. The U1-site is within a bipartite element that also contains a U1-STEM that base-pairs to and represses the activity of the U1-site. This is the first example of an endogenous cellular gene having a functional U1-site. Up to now only two types of U1-sites had been reported: U1-sites in papillomaviruses (Furth *et al.,* 1994; Cummings *et al.,* 2003) and artificial U1-sites used to silence expression of reporter and specific endogenous genes (Beckley *et al.,* 2001, Fortes et al., 2003). Prior work from multiple labs showed that these U1-sites use the same mechanism, namely inhibition of pre-mRNA maturation by inhibition of nuclear polyadenylation resulting in lower mRNA levels. This prior work also showed that inhibition requires the U1-site to be in the 3' terminal exon and able to form a duplex of 8-10bp with U1 snRNP (Gunderson *et al.,* 1998; Beckley *et al.,* 2001; Fortes *et al.,* 2003). The U1A gene's U1-site described here has these features.

Synergy and the bipartite element

The U1A U1-site is found within an unusual bipartite element where the U1-site is repressed by a U1-STEM sequence. Instead of simply blocking U1 snRNP binding, the U1-STEM allows binding but does not permit U1 snRNP to be active unless the nearby PIE is bound by U1A protein. Although the net affect is synergistic, this analysis indicates that it is a property of the ternary complex being more inhibitory to the polyadenylation machinery than each complex individually. This is supported by the *in vitro* polyadenylation assays in HeLa NXT. So why is the ternary complex more inhibitory? The most straightforward explanation is that the U1A:PIE complex releases U1 snRNP from inhibition by the U1-STEM. In this case what has been called synergy is in fact only an apparent synergy of rescuing U1 snRNP from the U1-STEM repressor, rather than a synergy between the inhibitory activities of the U1 and PIE sites. However an alternative explanation, that a unique, but as yet uncharacterized, feature of the ternary complex has enhanced polyadenylation inhibitory activity, must also be considered. For example, U1A in the U1 snRNP may "jump" over and bind to the PIE site. Although this cannot be ruled out it seems unlikely for several reasons. First, PIE requires two molecules of U1A to inhibit and U1 snRNP has only one U1A molecule. Thus the “jumping" would only activate a PIE that already had one molecule of bound U1A. Second, it does not explain the high inhibitory activity seen when the U1-STEM and PIE are mutated as mutation of PIE would prevent such jumping. Furthermore, when U1 snRNP was studied separate from PIE (as in papillomaviruses and HIV-1) this lab and others showed that the U1-70K subunit of U1 snRNP is necessary and sufficient for

polyadenylation inhibition and U1A makes no contribution (Gundeson *et al.,* 1998; Ashe *et al.,* 2000; Beckley *et al.,* 2001; Sajic *et al.,* 2007). Thus the simpler model is that U1A releases the U1-STEM mediated inhibition of the bound U1 snRNP so that U1-70K can now inhibit polyadenylation. However, a thorough understanding of the inhibitory mechanism will require reconstitution of this system with purified factors and a better understanding of what is the repressor itself, a topic discussed below.

Trapping U1 snRNP in a non-productive complex and parallels with other genes

An important question is whether U1-STEM-mediated repression requires a trans-acting factor or is repression completely RNA mediated. Such a hypothetical trans-acting factor could be either pre-assembled in the repressor complex or recruited during the binding of U1 snRNP. A related question is how the U1-STEM traps and holds U1 snRNP if only 4 of 10 nts of the U1 site are available to base-pair to U1 snRNP. Perhaps U1 snRNP base-pairs to these four nts and then makes additional protein-RNA or RNA-RNA contacts with the U1-STEM or to sequences nearby. Alternatively, the U1-STEM could open to allow U1 snRNP base-pairing to the full U1-site and then additional interactions hold U1 snRNP in an inactive conformation. Additionally, the role of region C in promoting synergy and why it depends on the U1A-PIE complex will need to be elucidated.

In the context of splicing, the U1 snRNP:5'ss complex is frequently a target of regulation by nearby splicing enhancer or suppressor elements that exert their affect through binding of splicing regulatory factors, most typically members of the SR and

hnRNP family (Black 2003). In contrast, examples of regulated splicing based on RNA secondary structure are far less frequent with most examples involving the branchpoint and 3'ss elements rather than the 5'ss. One 5'ss-related example that does have similarities with the U1-STEM repressor involves the S. cerevisiae L30 protein (formerly called L32). L30 negatively autoregulates its own expression by binding to and inhibiting splicing of a single intron found in its own pre-mRNA (Vilardell and Warner 1994). U1 snRNP is bound to and trapped by an L30 protein:L30 pre-mRNA complex where U1 snRNP base-pairs to just half of the 5'ss (nts 2-6) while the remaining part of the 5'ss is in a stem structure. The trapping mechanism blocks U1 snRNP from making interactions to carry out the ATP-dependent spliceosome assembly steps, consequently intron removal is blocked leading to a negative autoregulatory feedback of L30 on its own expression. However, whether L30 or the RNA:5'ss stem or both directly repress U1 snRNP and what part of U1 snRNP is repressed has remained unresolved. Another example of a trapped U1 snRNP is the Drosophila P-element Somatic Inhibitor (PSI) protein which, in somatic cells, inhibits splicing of the P element intron 3 by tethering U1 snRNP to a decoy 5'ss sequence near the authentic 5'ss (Labourier et al. 2001). Interestingly, PSI tethers and inactivates U1 snRNP by binding to U1-70K, a U1 snRNP subunit that makes numerous key interactions with splicing regulatory proteins (Black 2003) and is also responsible for inhibition of polyadenylation as it contains four polyadenylation regulatory domains (PRDs) that directly bind to and inhibit poly(A) polymerase (Gunderson *et al.*, 1998). If one uses the PSI analogy, the U1 snRNP trapping mechanism could plausibly involve the PRDs of U1-70K being masked by the U1-STEM or a hypothetical polypeptide that binds at or near the U1-STEM. Current studies are underway to test these possibilities.

Prevalence and regulation of U1 sites

As this is the first example of a 3' terminal U1-site in a cellular gene it is only natural to wonder how many more there are. The 3'UTR database search yielded other genes with conserved U1-sites that are now being investigated. Whether these U1 sites are autonomous or depend on nearby elements will require a combination of bioinformatics, sequence conservation and testing candidate U1-sites in reporter genes. Candidate U1-sites deemed to be inactive by reporter gene assays will have to be scrutinized for potential base-pairing interactions with flanking elements. In the case of papillomavirus U1-sites, it is evident that a differentiation dependent repression of the U1-site must occur in keratinocytes in order to release the viral poly(A) site for high expression of the viral late genes necessary to make virions. The difficulty in recapitulating this release in cultured cells has hindered elucidation of the mechanism.

Internal poly(A) sites that occur in introns are a second class of U1-site-containing genes where the upstream unused 5'ss can in fact be a potential U1 site able to inhibit the internal pA. Recent bioinformatic analysis estimates that 20% of human genes have such internal poly(A) sites (Tian *et al.*, 2007). The IgM gene represents a well-characterized example of a regulated internal poly(A) site, the secretory pA, and indeed mutation of its unused 5'ss leads to increased expression in reporter genes containing the secretory pA region. As this same pA is also regulated by U1A binding sites (Phillips et al. 2001; 2004) it will be of interest to determine how they interact with the upstream U1 snRNP:5'ss complex.

Significance for Autoregulation of U1A

The question remains, why does the U1A gene have such a complex collection of inhibitory elements? One possibility, that the bipartite U1 site is acting to amplify the negative autoregulatory feedback of the U1A:PIE complex, is not consistent with our previous finding that the U1A autoregulatory system could be made stronger by simple amino acid substitutions in the U1A protein (Guan *et al.*, 2003). Another possibility is that it may be important to integrate regulation of U1 snRNP and U1A levels. A third possibility hinges on whether a trans-acting factor is found to bind and affect the bipartite element. This factor could in principle allow other proteins or pathways to regulate U1A levels leading to downstream affects on expression levels of genes whose poly(A) sites are regulated by U1A.

Chapter IV: Analysis to Determine the Requirements for U1 snRNP Binding to the U1-Site and Establish RNA Secondary Structure of the U1 site:U1-STEM

(A manuscript of this chapter will be submitted in 2011 to JBC with two authors: Caratozzolo RM and Gunderson SI.)

Introduction

It has previously been established that the polyadenylation of U1A mRNA is regulated via a negative feedback loop, which requires two molecules of U1A to bind to a regulatory element, (PIE) in its 3'UTR (Boelens *et al.,* 1993; van Gelder *et al.,* 1993; Gunderson *et al.,* 1994 & 1997). In addition to PIE (Polyadenylation Inhibitory Element), a highly conserved regulatory element (Chapter I, Boelens *et al.,* 1993), it was recently determined that polyadenylation inhibition requires a second conserved regulatory element – the U1-Site (Chapter III, Guan *et al.,* 2007). As discussed in the Introduction and in Chapter III, U1-Sites, as found in papillomaviruses, are effective inhibitors of polyadenylation, and work via base-pairing with the 5'-end of U1 snRNP's U1 snRNA (Gunderson *et al.,* 1998; Guan *et al.,* 2007). The U1-70k subunit of the bound U1 snRNP then interacts with PAP, thereby inhibiting its polyadenylation activity (Gunderson *et al.,* 1998).

The U1-Site in the U1A 3'UTR was found to work in conjunction with U1 snRNP and PIE to increase polyadenylation inhibition, forming a repression complex (Guan *et al.,* 2007). This synergistic effect was seen despite the U1-Site's repression by the U1-STEM. As demonstrated in Chapter III, the cooperativity between the U1-Site and PIE was not due to enhanced binding by U1 snRNP or PIE. In fact, U1 snRNP bound with equal affinity to various RNA substrates, including one with a mutated PIE site, in the

absence of a U1-STEM. Additionally, there was no difference in binding of U1 snRNP when the U1-STEM was wild-type versus mutated.

As a whole, this data lead to the following model: U1 snRNP present in the repression complex bound to the U1-site, is unable to inhibit polyadenylation without the presence of U1A bound to PIE. In essence, U1 snRNP would be trapped in an inactive conformation and incapable of contributing to polyadenylation inhibition until it is released. Examples of this trapping have been established in several genes, including the *S. cerevisiae* L30 protein (Vilardell and Warner 1994, Chapter III) and the Drosophila P-element Somatic Inhibitor (PSI) protein (Labourier *et al.,*. 2001, Chapter III) although in each case the trapping mechanism has remained unsolved. To determine the mechanism of U1 snRNP release, three possible models were postulated. In the first model, a *trans*-acting factor could bind to the complex, thereby releasing U1 snRNP. This factor could bind simultaneously with the formation of the repression complex or be recruited later. In the second model, U1A alone is the *trans*-acting factor that directly leads to U1 snRNP release. For example, U1 snRNP could transiently bind to the four free nucleotides of the U1-Site and upon U1A binding to PIE, the U1A facilitates opening of the U1-STEM, thereby allowing U1 snRNP to bind to the full nine nucleotides of the U1-Site. Once fully bound, U1 snRNP would then be able to actively inhibit polyadenylation. In the third model, the 3'UTR of the repression complex is sufficient to trap U1 snRNP. When U1A binds to PIE, a remodeling of the 3'UTR RNA (most likely Region C as this region is needed for synergy, see Chapter III) would allow U1 snRNP to bind to the U1-site leading to active inhibition of polyadenylation.

To assess the first model, multiple biotin pull down and UV-crosslinking assays were done, in both the presence and absence of nuclear extract, to search for a *trans*-acting factor. No such factor was found, as there were no explicit differences in proteins bound (biotin assay) or the UV crosslinking pattern with or without extract (data not shown). As a result, the latter two models regarding the trapping of U1 snRNP became the focus of further experiments. To examine these possibilities, it was necessary to investigate further the relationship between U1A and U1 snRNP binding and its effects on RNA structure. Specifically, multiple U1-STEM mutants were used to resolve the effect of U1A on U1 snRNP binding and to determine other inherent properties of the U1-STEM itself. In addition, RNase digestion assays were used to confirm the physical presence of the U1-STEM and demonstrate that Region C does undergo remodeling. These data support the model that all inhibitory characteristics of the repression complex are due solely to the complex itself and not to any outside factors.

Results

U1 snRNP Binding Increases in the Presence of rU1A *in vitro*

In order to examine the effect of a mutated U1-STEM on polyadenylation inhibition of U1A pre-mRNA, the binding capability of rU1A and U1 snRNP were first determined, using gel shift assays (Electrophoretic mobility shift assays, EMSA), in the context of an intact U1-STEM. The ability of rU1A-PIE complexes to give discernable shifts has been well-established previously (van Gelder *et al.,* 1993, Gunderson *et al.,* 1994, Gunderson *et al.,* 1997). RNA probes were derived from the U1A 3’ UTR (Figure

4.1) and radiolabeled with α-^{32}P-UTP. These probes contained a wild-type (wt) U1-site, located within an intact U1-STEM, in combination with either a wt or mutant (mt) PIE site (wt/wt/wt vs. wt/wt/mt). These shifts were conducted in a "pure" system, using only purified factors, without nuclear extract. As protein was added and bound to the probe, different complexes form and as a result the probe "shifts" upward, as its mobility is slowed. The binding of rU1A results in two shift bands: the first complex, which is faster and therefore the lower band, is a monomer, consisting of the probe plus one molecule of rU1A bound to PIE, $(rU1A)_1$-PIE. The second, higher band, is slower and represents the dimer, which has two molecules of rU1A bound to PIE, $(rU1A)_2$-PIE. The binding of U1 snRNP to the RNA probe results in a higher band since it will be much slower due to the higher molecular weight of U1 snRNP (approximately 300,000 kDa) compared to two molecules of rU1A (66 kDa). The highest band represents the trimer, which has U1 snRNP bound at the U1-site in addition to the two molecules of rU1A bound to PIE, $(rU1A)_2$-PIE;U1 snRNP-U1-Site.

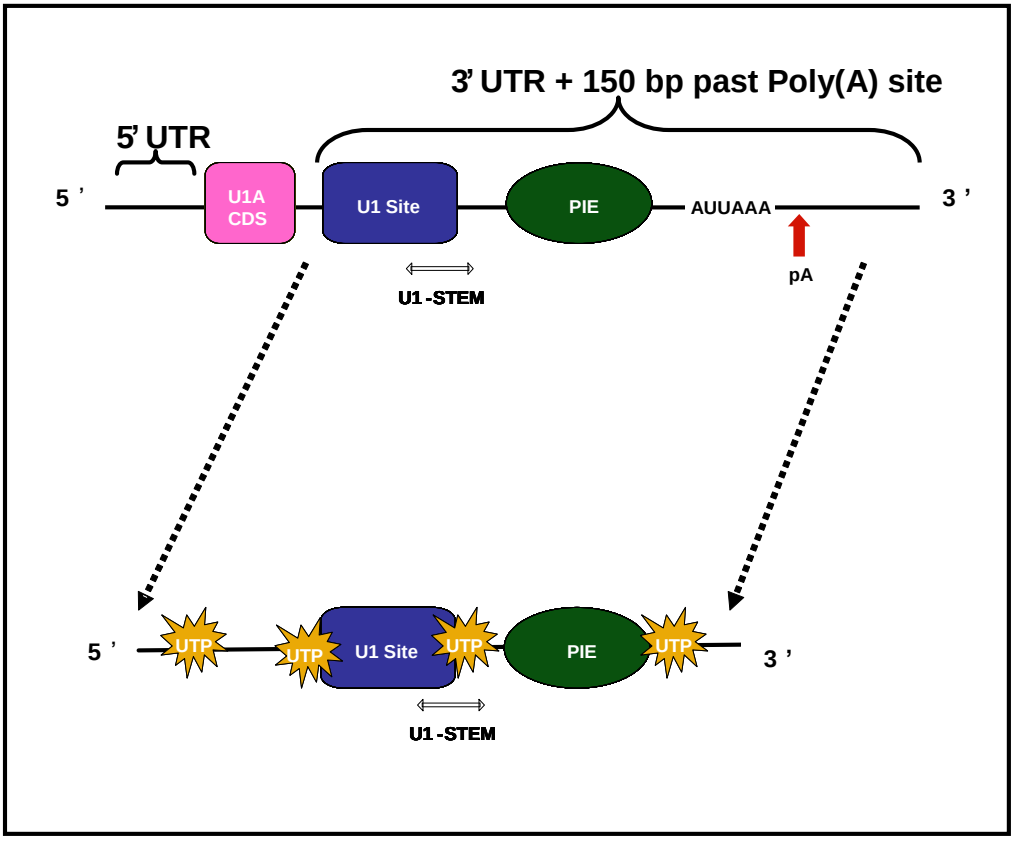

Figure 4.1: Structure of RNA Probes

U1A cDNA constructs contained the 5' UTR, coding sequence and 3' UTR including 150 base pairs (bp) beyond the poly(A) site (pA). Radiolabeled (α-^{32}P-UTP) RNA probes were derived from the U1A 3' UTR, which contained a wild-type (WT) or mutant (MT) version of the following RNA elements in 5' to 3' order: U1-site, U1-STEM, and PIE. For example: the wt/wt/mt RNA probe means a 3' UTR probe containing a wt U1-site, wt U1-STEM and mutant PIE.

Gel shift titrations were performed with each probe to look at rU1A and U1 snRNP binding abilities alone (Figures 4.2A and 4.2B). Protein concentrations ranged from 50-1000 ng. For the wt/wt/wt probe, as rU1A concentration increased there was an initial trend toward monomer over dimer formation from 50-200 ng (Figure 4.2A, lanes 2-4). The switch to dimer-only formation occurred at 400 ng (lane 5) and correspondingly increased with increased rU1A concentration (lanes 6-8) until all probe was bound (as seen by the disappearance of the input probe band). In contrast, wt/wt/mt probe (Figure

4.2B, lanes 2-8) shows no rU1A binding, even at the highest concentration (intensity of input band remains constant).

U1 snRNP binding, in both wt/wt/wt and wt/wt/mt probes, was not seen until 400 ng (Figure 4.2A & 4.2B, lane 13), and thereafter increased as U1 snRNP concentration increased (lanes 14-16). The mutated PIE site resulted in only a slight increase in binding of U1 snRNP (compare U1 snRNP band intensities, Figures 4.2A & 4.2B, lanes 13-16). However, unlike rU1A, the probe does not become saturated with protein, as seen by the presence of the input probe band even at high concentration (compare probe band levels in Figures 4.2A & 4.2B, lanes 6-8 to lanes 14-16). This indicates that U1 snRNP binds with lower affinity to the U1-site than rU1A does to PIE. The faint lower bands seen in lanes 11-16 of Figure 4.2A are most likely due to U1A that has fallen off of U1 snRNP, and is therefore able to freely bind PIE (as demonstrated by Bach *et al.,* 1990).

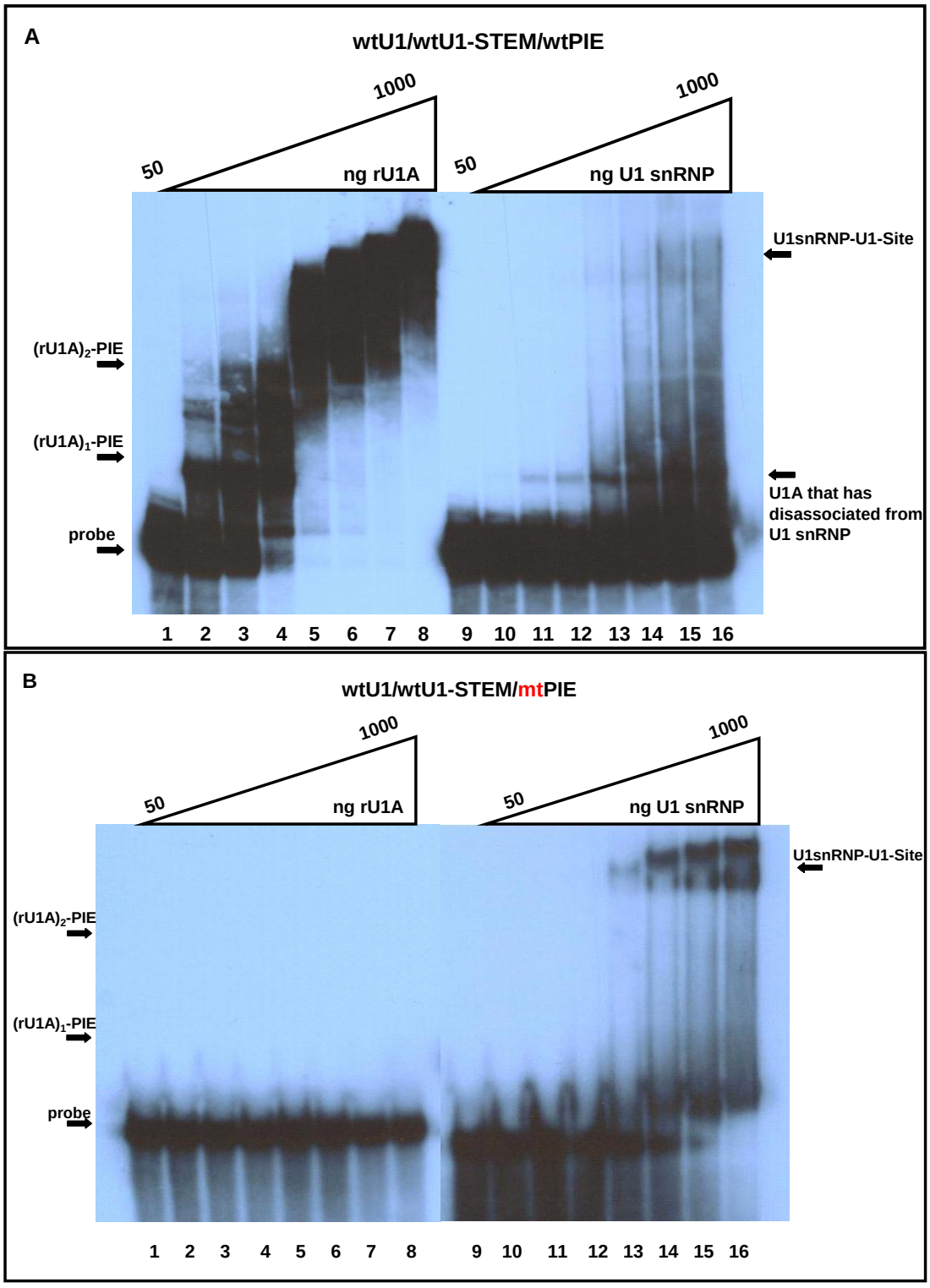

Figure 4.2: rU1A & U1 snRNP Binding Capabilities with an Intact U1-STEM

Gel shift titrations were performed using purified factors and probes. wt/wt/wt (panel A) or wt/wt/mt (panel B) RNA probes were incubated with increasing

amounts (50, 100, 200, 400, 600, 800 & 1000ng) of rU1A (lanes 2-8) or U1 snRNP (lanes 10-16). Lanes 1 & 9 are input lanes in the absence of added purified factor. As indicated by arrows on the left, the lowest band is the unbound probe, above that is the monomer, ($rU1A_1$)-PIE, consisting of the probe plus one molecule of rU1A bound to PIE. The dimer, ($rU1A_2$)-PIE, consists of the probe plus two molecules of rU1A bound to PIE. (A) wt/wt/wt probe bound rU1A at the lowest concentration (lane 2) and, as rU1A concentration increased so did monomer and dimer formation (lanes 3-8) until all probe was in the dimer conformation (as seen by the disappearance of the free probe band in lanes 7 & 8). U1 snRNP (lanes 9-16) bound with lower affinity as even high U1 snRNP concentration showed far less binding when compared to low concentrations of rU1A. (B) rU1A does not bind to the wt/wt/mt RNA that contains a mutant PIE site (lanes 1-8). U1 snRNP binds to RNA with a mutant PIE site (lanes 9-16) with affinity similar to RNA with a wild-type PIE site (see panel A).

After the above titrations, gel shifts were done in the presence of both rU1A and U1 snRNP using only wt/wt/wt probe (Figure 4.3). A low enough concentration of U1 snRNP was used such that it would have poor binding on its own, therefore, any increase in binding could be attributed to the presence of rU1A. As shown previously in Figure 4.2A, increased rU1A concentration resulted in greater dimer formation (lanes 2-4), and U1 snRNP bound poorly at 100 ng (lane 5). However, Figure 4.3 shows that U1 snRNP bound with much higher affinity, at 100 ng, when in the presence of rU1A, and binding increased as rU1A concentration increased (lanes 6-8). This suggests that rU1A increases the binding affinity of U1 snRNP for the U1-Site.

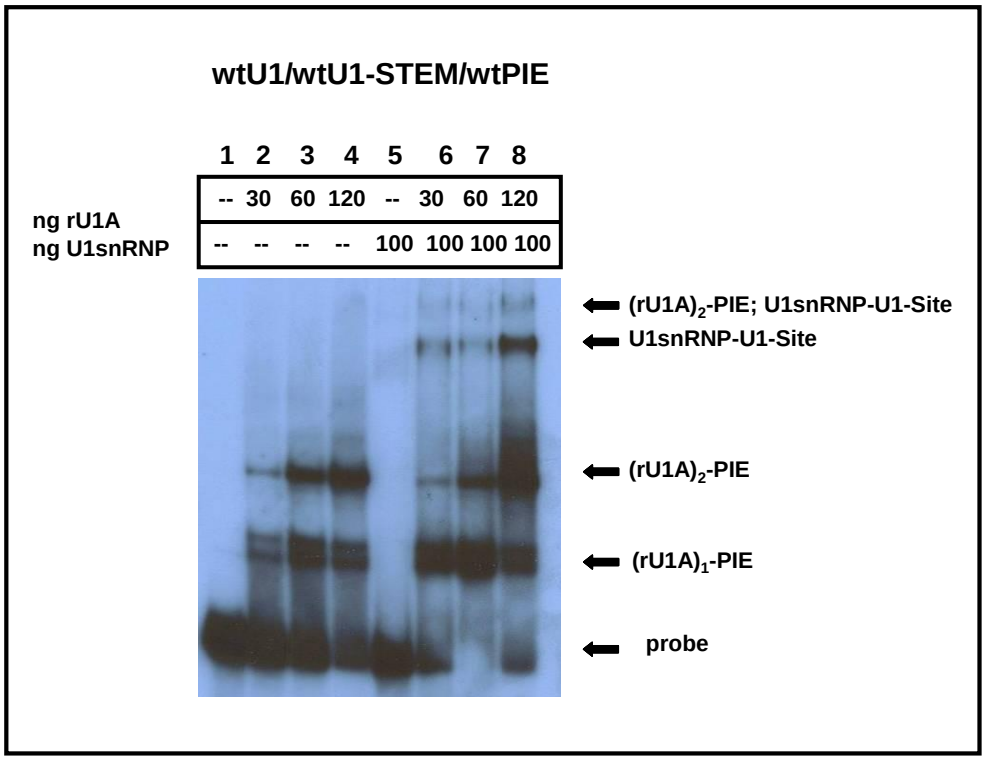

Figure 4.3: U1 snRNP Binding Increases In The Presence of rU1A

Gel shift assay using purified factors and wt/wt/wt probe from Figure 4.1: probe was incubated with increasing amounts (30, 60 & 120 ng) of rU1A alone (lanes 2-4) or in the presence of 100ng of U1 snRNP (lanes 6-8). Lane 1 is the input lane, while lane 5 has 100ng of U1 snRNP alone. Positions of unbound probe, monomers & dimers are indicated as in Figure 4.2 with the addition of the trimer ($rU1A_2$)-PIE:U1 snRNP-U1-Site, the probe with two molecules of rU1A bound to PIE plus U1 snRNP bound at the U1-site. Lanes 6, 7 & 8 show increased U1 snRNP binding in the presence of rU1A (compare lanes 6- 8 to lane 5).

To determine if the effect seen in Figure 4.3 was specific to rU1A, the gel shift was repeated using "scrambled" rU1A, S-rU1A. Scrambled rU1A has been extensively used by the Gunderson lab, including in several publications (Gunderson *et al.*, 1997 and Klein Gunnewiek *et al.*, 2000). This protein has a mutated PAP regulatory domain (PRD) (the order of the 13 amino acids in the PRD was "scrambled") which results in a loss of cooperative binding ability between two molecules of rU1A (triple the amount of protein was used in order to be comparable to the wt protein). This results in lower affinity for

dimer formation, as seen in Figure 4.4 (lanes 2-4), when compared to Figure 4.3 (lanes 2-4). U1 snRNP, in the presence of S-rU1A showed only minimal binding (compare lanes 6-8 to lane 5 and to Figure 4.3). This substantiates the idea that rU1A plays a direct role in U1 snRNP binding to the U1-site. It also implies rU1A specificity, in that the PRD, not just any RNA Binding Protein (RBP), is required.

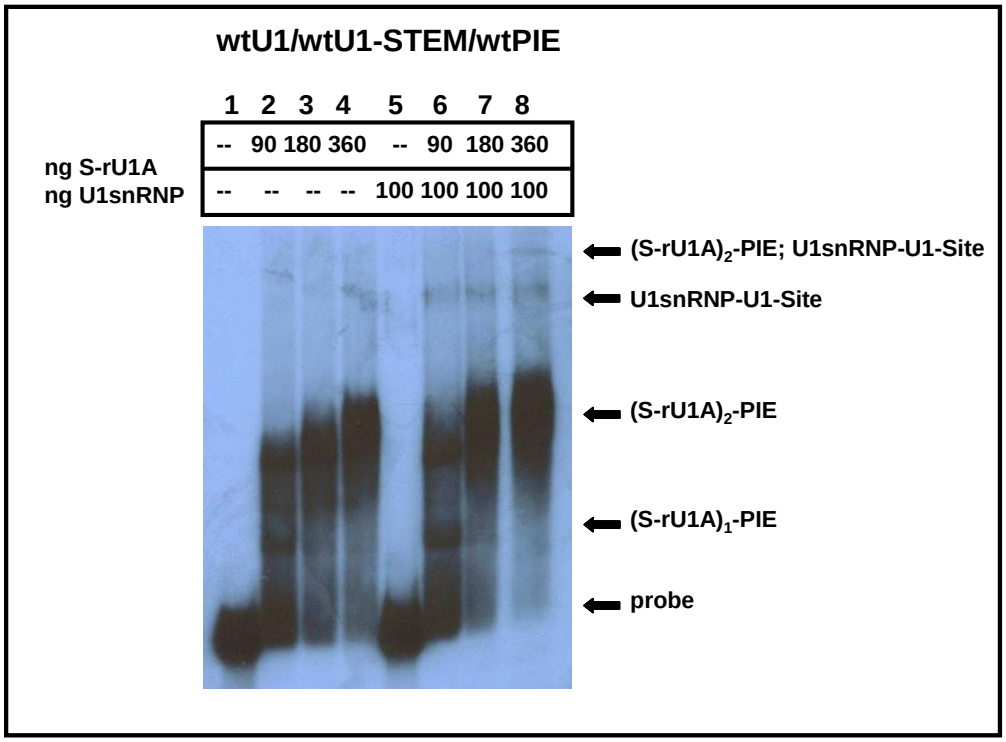

Figure 4.4: Effects of "Scrambled" rU1A On U1 snRNP Binding

Gel shift assay using purified factors and wt/wt/wt probe from Figure 4.1: probe was incubated with increasing amounts (90, 180 & 360 ng) of "scrambled" rU1A, S-rU1A, alone (lanes 2-4) or in addition to 100ng of U1 snRNP (lanes 6-8). Lane 1 is the input lane, while lane 5 has 100ng of U1 snRNP alone. S-rU1A binding is not as tight as wt rU1A, as seen by the smearing in lanes 2-4 and 6-8, compared to Figure 4.3. Lanes 6, 7 & 8 show no increase of U1 snRNP binding in the presence of "scrambled" rU1A (compare lanes 6-8 to lane 5), and compare to Figure 4.3.

A Mutated U1-STEM Results In Greater U1 snRNP Binding and Increased Polyadenylation Inhibition

Previously it was determined *in vivo* that the U1-site is repressed by a U1-STEM structure (Guan *et al.,* 2007). To further examine the mechanism of repression, the next set of assays were conducted *in vitro*, using only purified factors, in the context of a mutated U1-STEM. A U1-STEM mutant was constructed (as described above) by mutating seven nucleotides downstream of the U1-Site, five of which naturally base-pair with the U1-Site (Figure 4.5). The U1-Site itself was left unchanged, leaving it fully accessible to be bound by U1 snRNP. Gel shift titrations with rU1A were run to establish that the loss of the U1-STEM has no effect on rU1A binding to PIE (Figure 4.6A). Upon comparison, rU1A binds with the same affinity to both the wild-type and mutant probes (compare lanes 2-8 to lanes 10-16). Therefore, rU1A does not require an intact U1-STEM to interact and bind with PIE. This agrees with previously published results (Boelens *et al.,* 1993 and van Gelder *et al.,* 1993) which examined rU1A binding in the presence of a minimal 3' UTR.

Conversely, U1 snRNP binds with much greater affinity to the mutant than to wild-type (Figure 4.6B). Binding is seen as early as 100 ng (lane 11) in the mutant, whereas wild-type shows binding at 200 ng (lane 5). While both probes showed increased binding as protein concentration increased, the mutant had a much stronger attraction. This is evidenced by comparing the signal intensities at each concentration (i.e. compare lane 6 to lane 14, etc.). Signal intensity at higher concentrations is much stronger for the mutant, indicating greater binding of U1 snRNP to the mutant probe. This is

corroborated by the decrease in intensity of the input probe band for the mutant as U1 snRNP concentration increases, while wild-type input probe levels remain more constant. Therefore, less wild-type probe is being bound. Overall, the mutant demonstrated a higher affinity for U1 snRNP, which is consistent with the model that the U1-STEM blocks access to the U1-Site.

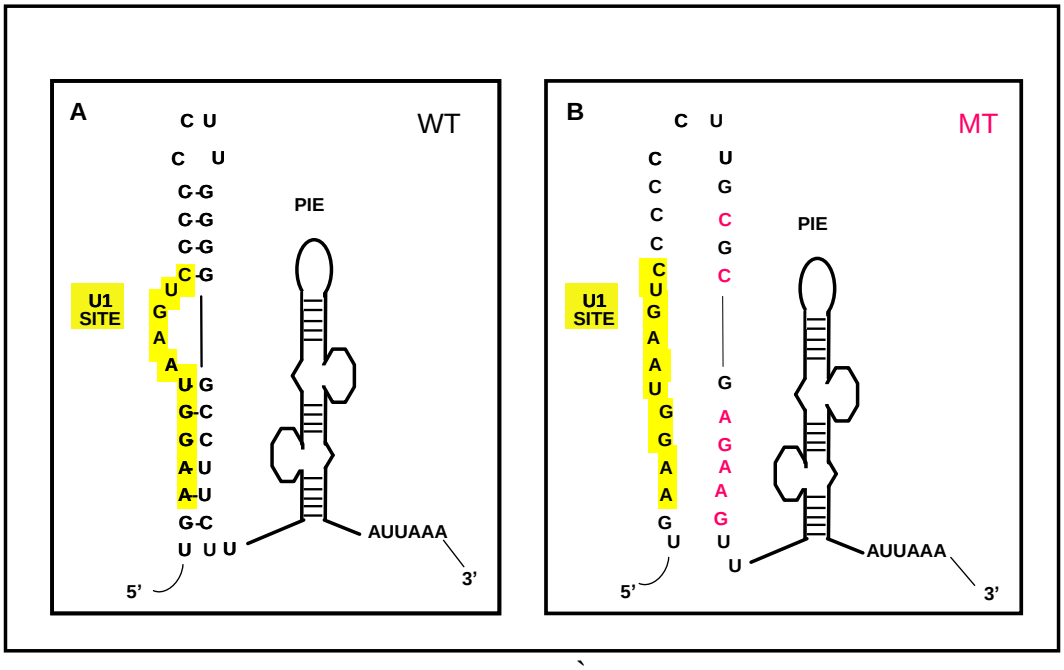

Figure 4.5: Sequence of U1-STEM Mutant

Shown is part of the sequence of the human U1A 3' UTR wt (panel A) and mutated U1-STEM (panel B) probes. The U1-site (highlighted in yellow) is located within the U1-STEM, upstream of PIE. The mutated bases of the U1-STEM probe contained an altered sequence (in red) which abated the base-pairing of the U1-STEM. This leaves the U1-site completely accessible while leaving PIE intact.

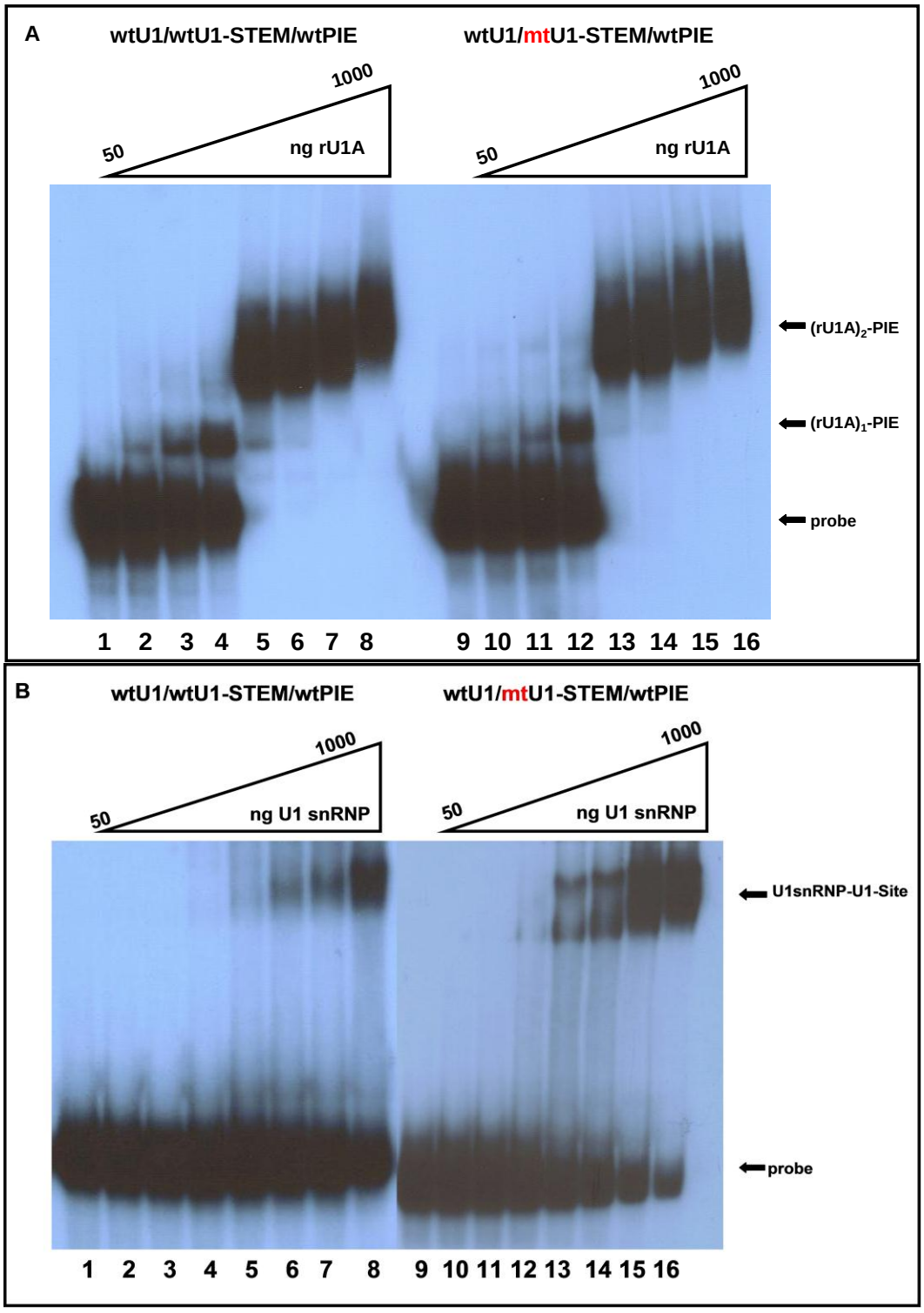

Figure 4.6: Establishing the Affect of a Mutated U1-STEM *in vitro*

Gel shift titrations were done with both mutants as in Figures 4.2, 4.3 and 4.4 (A) The mutated U1-STEM probe (see Figure 4.6 for mutant U1-STEM) had a similar binding affinity as the wt, to rU1A. Monomer formation began for each probe at 50

ng (Lanes 2 & 10), followed by full dimer formation at 400 ng (Lanes 5 & 13). This implies that the U1-STEM has no role in the binding of rU1A to PIE. (B) U1 snRNP binding is greatly increased in the mutant. Binding is first seen faintly at 100 ng (lane 11) in the mutant, and amplified steadily as concentration increased. Also note the decrease in the input probe band, starting at 400 ng (lane 13). The wt has no binding until 200 ng (lane 4), and while binding increases as more protein is added, it is with less intensity (compare signal strength of lanes 13-16 to lanes 5-8).

It was already determined that U1 snRNP binds better to wild-type RNA in the presence of rU1A, as shown above, therefore the next step was to see if the same is true for the mutant RNA. The gel shift in Figure 4.3 was repeated using both wild-type and mutant probes, the results of which are seen in Figure 4.7. As in Figure 4.6B, the mutant probe bound U1 snRNP with higher affinity than wild-type (Figure 4.7, compare lane 13 to lane 5). Additionally, binding of U1 snRNP to the wild-type probe increased as rU1A was added (compare lanes 6-8 to lane 5). This coincided with the results seen in Figure 4.3. Furthermore, U1 snRNP binding to the mutant also increased when rU1A was added (compare lanes 14-16 to lane 13). In both cases, all three complexes (monomer, dimer and trimer) formed when rU1A was first added (30 ng, lanes 6 and 14), then as rU1A concentration increased, all monomer complexes became dimers and/or trimers (lanes 7, 8, 15 and 16). These trends were verified by quantification of signal intensities for each complex using ImageQuant software (data not shown). The transformation of monomer to dimer/trimer is seen at rU1A concentration of 120 ng for the wild-type (lane 8) and at 60 ng (lane 15) for the mutant. In addition, all mutant probe was bound at 60 ng, whereas wild-type probe was not fully bound until 120 ng (compare input probe in lanes 15 and 16 to lanes 7 and 8). Therefore, the mutant transitions into the higher complexes more easily than wild-type. Most importantly, there are no additive effects, since trimer formation was 3-4x greater in the mutant at all rU1A concentrations, when compared to

wild-type (data not shown). Taken as a whole, these gel shift assays demonstrate that the U1-STEM mutant binds U1 snRNP with greater affinity than wild-type whether alone or in the presence of rU1A. This argues against the model that rU1A binding to PIE directly opens the U1-STEM. Instead, U1 snRNP binding to the U1-Site is able to increase if the RNA is simply remodeled.

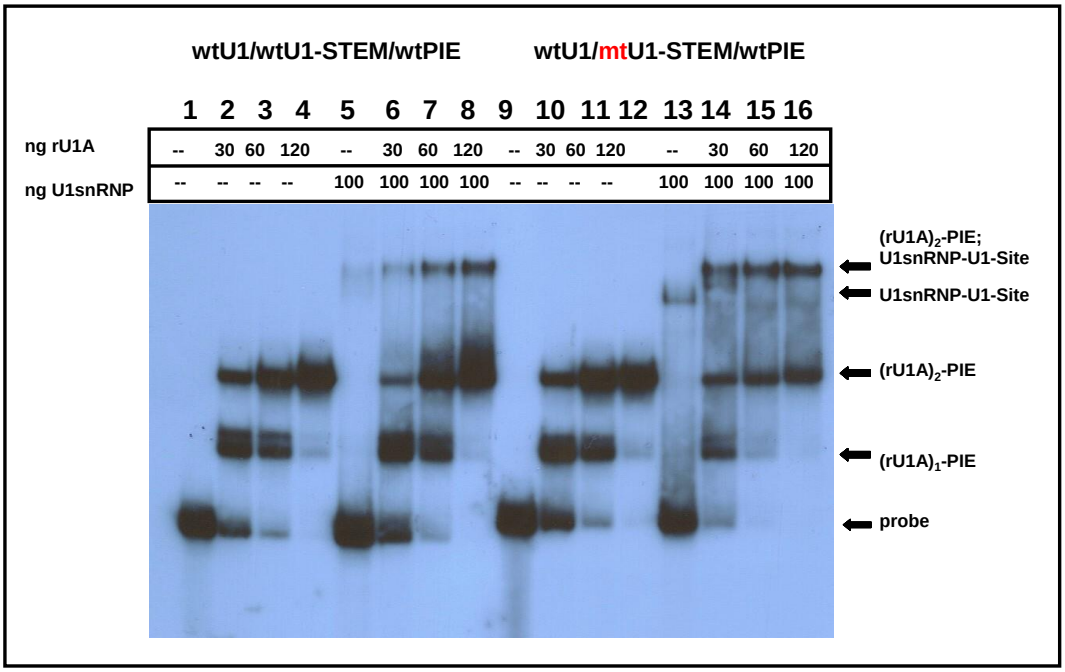

Figure 4.7: Comparison of WT and MT U1-STEM Binding Capabilities

Gel shift assays were done as in Figures 4.3 and 4.4. Lanes 1 & 9 are input lanes; lanes 2, 3, 4 10, 11 & 12 contain 30, 60 and 120 ng of rU1A; Lanes 5 & 13 have 100ng of U1 snRNP; lanes 6, 7, 8, 14, 15 & 16 contain 30, 60 and 120 ng of rU1A plus 100ng of U1 snRNP. U1 snRNP bound the wt probe poorly unless in the presence of rU1A (compare lane 5 to lanes 6-8 and see Figure 4.3). A mutated stem resulted in greater U1 snRNP binding when alone (compare lane 13 to lane 5) and when paired with rU1A (compare lanes 14-16 to lanes 6-8).

Once it was established that U1 snRNP bound more readily to the U1-STEM mutant, polyadenylation assays (Poly(A) assays), were conducted to determine the

inhibitory capability of the mutant RNA. Poly(A) assays are a measure of poly(A) tail length, where a short tail indicates strong inhibition, and a long tail indicates weak inhibition. Wild-type and mutant probes were incubated with recombinant Poly(A) polymerase (rPAP) at a concentration of 50 ng (Figure 4.8, concentration determined via titration analysis, data not shown). The samples in lane 2 (wild-type) and lane 7 (mutant) contained only probe and rPAP, and therefore have the longest poly(A) tails, since no inhibitory proteins were present. Inhibition occurred when rU1A and U1 snRNP were added, both individually (lanes 3 and 4 for wild-type, lanes 8 and 9 for mutant), and together (lanes 5 and 10), as indicated by the shorter tail lengths. The combination of rU1A and U1 snRNP resulted in the shortest tail (highest inhibition) for the wild-type probe (lane 5). This coincides with this lab's previously published work, which established, in cell culture assays, that PIE and the U1-site work synergistically to increase inhibition (Guan *et al.*, 2007). Overall, tail lengths for the mutant were shorter than the wild-type, when rU1A and U1 snRNP were added individually (compare lane 8 to lane 3, and lane 9 to lane 4), indicating inhibition was stronger in the mutant. There was no discernable difference between tail lengths for both probes when U1 snRNP and rU1A were added together (lanes 5 and 10) and when U1 snRNP was added alone to the mutant (lane 9). The lack of additive effects may be the result of saturation, which corresponds with the gel shift assay results seen above. This suggests that a mutated U1-STEM is sufficient to cause increased inhibition, likely a result of the increased binding affinity of U1 snRNP to the now accessible U1-Site.

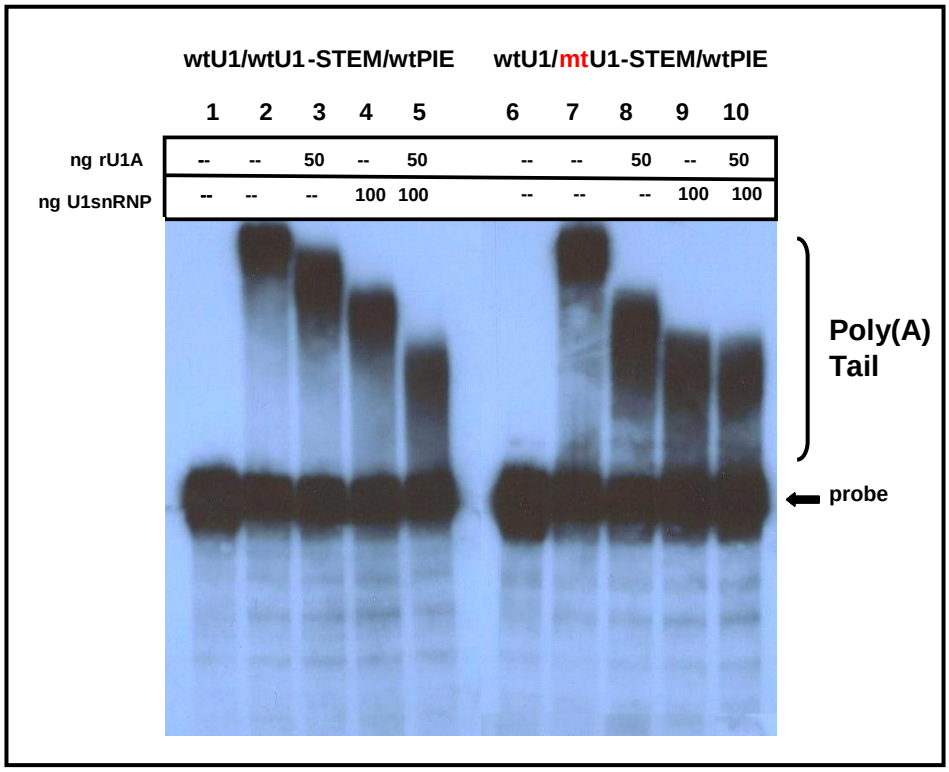

Figure 4.8: Effect of a Mutated U1-STEM on Polyadenylation

Polyadenylation assays were done using the probes from Figure 4.5. All samples were incubated with 50 ng recombinant Poly(A) Polymerase (rPAP), except for input lanes 1 & 6. Lanes 3 & 8 have 50 ng rU1A added; lanes 4 & 9 have 100 ng U1 snRNP added, and lanes 5 & 10 contain both rU1A and U1 snRNP. Inhibition is measured by the length of the poly(A) tail, as indicated above on the right. An increase in inhibition was illustrated by shorter tails (compare lanes 3-5 to lane 2 and lanes 8-10 to lane 7) in the presence of rU1A and/or U1 snRNP for both probes. Tail lengths for the mutant probe are shorter than those of the wild-type probe for each combination, with the exception of lanes 5 and 10, which are approximately equal.

The RNA Structure of the U1-STEM

Once it was established that mutating the U1-STEM region increased U1 snRNP binding and inhibition, the next step was to determine the physical presence of the U1-STEM. RNase digestion is a commonly used method to determine secondary structure. Given that the U1A 3' UTR RNA sequence contained multiple Guanine (G) residues,

RNase T1, which cleaves at the 3' end of single-stranded G's, was chosen along with RNase V1, which cleaves double-stranded RNA. The probes used for these assays (Figure 4.9) were similar to those in Figure 4.5, except they were 5'-end-labeled with radioactive γ-^{32}P- ATP and had 50 nucleotides of 3' UTR sequence added upstream of the U1-site, which helps the probe to more closely mimic the natural context of the gene as well as to increase the distance between the end-label and the U1-STEM. Single-stranded (ss) G residues that should be susceptible to RNase T1 cleavage are indicated by arrows. Those residues that should be affected by the binding of rU1A or U1 snRNP are indicated by arrows with either circle (rU1A) or diamond (U1 snRNP) heads. Double-stranded (ds) G residues that may be partially susceptible, due to proximity to bulges/loops, are indicated by asterisks. Double-stranded regions susceptible to cleavage by RNase V1 are shaded in green. Mutations made to open up the U1-STEM are indicated by red stars.

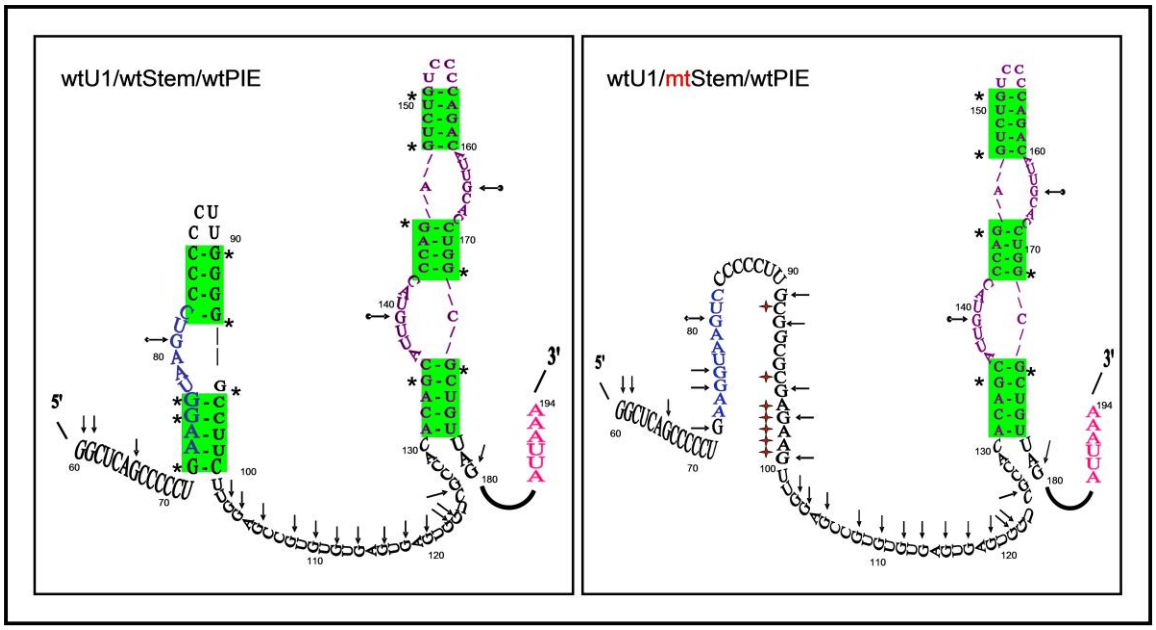

Figure 4.9: Sequence of RNase Digestion Probes

Sequences of wild-type and mutated U1-STEM probes used in RNAse T1/V1 analysis. These probes were the same as in Figure 4.5, except they were radioactively end-labeled with γ-ATP and approximately 50 nucleotides of 3'UTR sequence was added to increase the distance between the end-label and the U1 site. The U1 site is in blue, PIE is in purple and the poly(A) signal is in pink. Mutated nucleotides, which open the U1-STEM are marked with red stars. Black arrows indicate nucleotides susceptible to RNase T1 cleavage (3' of ss G residues); arrows with diamond ends indicate nucleotides that should be effected when U1 snRNP binds to the U1-site, while arrows with circle ends indicate residues that should be effected when rU1A binds to PIE. Nucleotides which may have partial digestion, due to proximity to bulges/loops are indicated by asterisks. Shaded green regions indicate regions susceptible to RNase V1 cleavage (ds RNA).

Four gels were run using wt and mt probes with both RNase T1 and RNase V1. The gels in Figures 4.10 and 4.11 are representative of all the data obtained from these assays. The gels for wt and mt are the same in both Figures; however, the section concerning the U1-STEM region has been excised for clarity in Figure 4.10. The top panel of Figure 4.10 has two simplified drawings of the U1-STEM region, in both wt and mt probe, to show the exact G residues susceptible to RNase T1 cleavage. The wt probe has free, ss G residues at positions 67, 81 and 95, while ds residues at 73, 76, 77 and 91-

94 should be unaffected. As seen in the Figure, the residue at position 81 (band above the blue box) is cleaved the strongest, followed by positions 73 (lower band in the bottom blue box) and 67 (band below bottom blue box). The cleavage at position 73 is most likely due to nucleotides present at the boundary the U1-STEM being in a constant state of flux between ss and ds character. Equally, the lack of cleavage at position 95 (top blue box) is probably a result of its presence between two ds residues, making it unavailable for cleavage. The remaining residues at positions 76 and 77 (top bands in bottom blue box), and 91-94 (red box) show weak cleavage, indicating that these nucleotides are double-stranded.

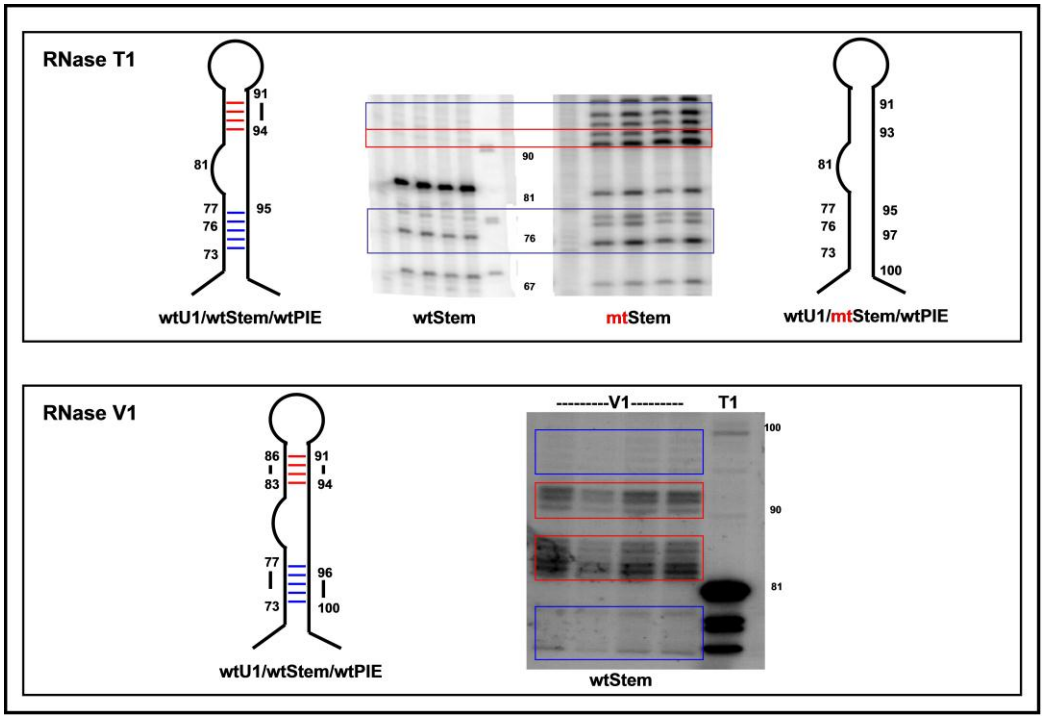

Figure 4.10: Determining the RNA Secondary Structure of the U1-STEM Via RNase Digestion

RNase T1/V1 digestions with diagrams of wt and mt U1-STEM. The numbers on the diagrams specify the individual nucleotides of the U1-STEM that could be susceptible to RNase cleavage. Red boxes correspond to the upper portion of the U1-STEM, and blue boxes to the lower portion. *Top:* Nucleotide positions on the diagrams correspond to G residues. RNase T1 cut the wt strongly only at residue 81, located within the open portion of the U1-site. The mutant was cut at several residues- 73, 76, 77, 81, 91, 93, 95, 97 and 100, of which only 97 is a mutation. The lack of these residues in the wt film indicate that they are not available for RNase T1 cleavage in a normal context. (Digestion films are portions taken from Figure 3C and simplified to focus solely on the U1-STEM region) *Bottom:* Nucleotide positions correspond to double-stranded sections of the U1-STEM. RNase V1 was used on the wt probe only (last lane was a control using RNase T1). RNase V1 digestion was strongest at residues 83-86 and 91-94, which correspond to the upper portion of the U1-STEM. Weak digestion is seen for the lower section of the U1-STEM at residues 73-77 and 96-100.

The mutant U1-STEM probe has free, ss G residues at positions 67, 73, 77, 81, 91, 93, 95 and 97 (mutation). While strong cleavage is seen at all target positions, the most significant are the bands at positions 91 and 93 (red box) and 95 (upper blue box).

Cleavage at these positions in the mutant, but not the wild-type, indicates that in the normal context they are not accessible for RNase cleavage, as they are in the U1-STEM structure.

To verify the RNase T1 results, RNase V1 assays were done using wild-type probe only, shown in the bottom panel of Figure 4.10. The diagram on the left indicates the ds regions of the U1-STEM that would be susceptible to RNase V1 cleavage. The results of RNase V1 digestion are shown on the right. Probe was also digested with RNase T1 as a control (last lane). Strong digestion was seen corresponding to the upper region of the U1-STEM, residues 83-86 and 91-94 (red boxes). Conversely, bands corresponding to the lower region of the U1-STEM, 73-77 and 96-100 (blue boxes) show only weak digestion. A possible explanation for this disparity is that the upper region of the U1-STEM is more stable (it has four G-C bonds in a row) than the lower region (it has three G-C bonds out of five total, not all together). The presence of bands at positions 73, 76 and 77 in the wild-type after RNase T1 digestion also supports this model. This topic will be further addressed with Figures 4.12 and 4.13 below.

Effects of rU1A and U1 snRNP Binding on RNA Structure

After determining the physical presence of the U1-STEM, the next step was to determine if the addition of rU1A and/or U1 snRNP altered the digestion patterns (Figure 4.11). Multiple assays were conducted both with and without nuclear extract. There was an overall trend of increased digestion in the presence of nuclear extract (data not shown) and all films shown in Figures 4.10-4.11 are representative of assays conducted with nuclear extract. Both wild-type (left side, lanes 1-5) and mutant probes (right side, lanes

6-10) were used. In both films, the first lane is the input lane (lanes 1 and 6), while all remaining lanes were treated with 1:2000 RNase T1 (lanes 2-5 and lanes 7-10). The last three lanes for each film had either rU1A, U1 snRNP or both added as follows: rU1A alone (lanes 3 and 8), U1 snRNP alone (lanes 4 and 9) and both rU1A and U1 snRNP (lanes 5 and 10). The last lane in the wild-type film is the MSP marker used to determine nucleotide position.

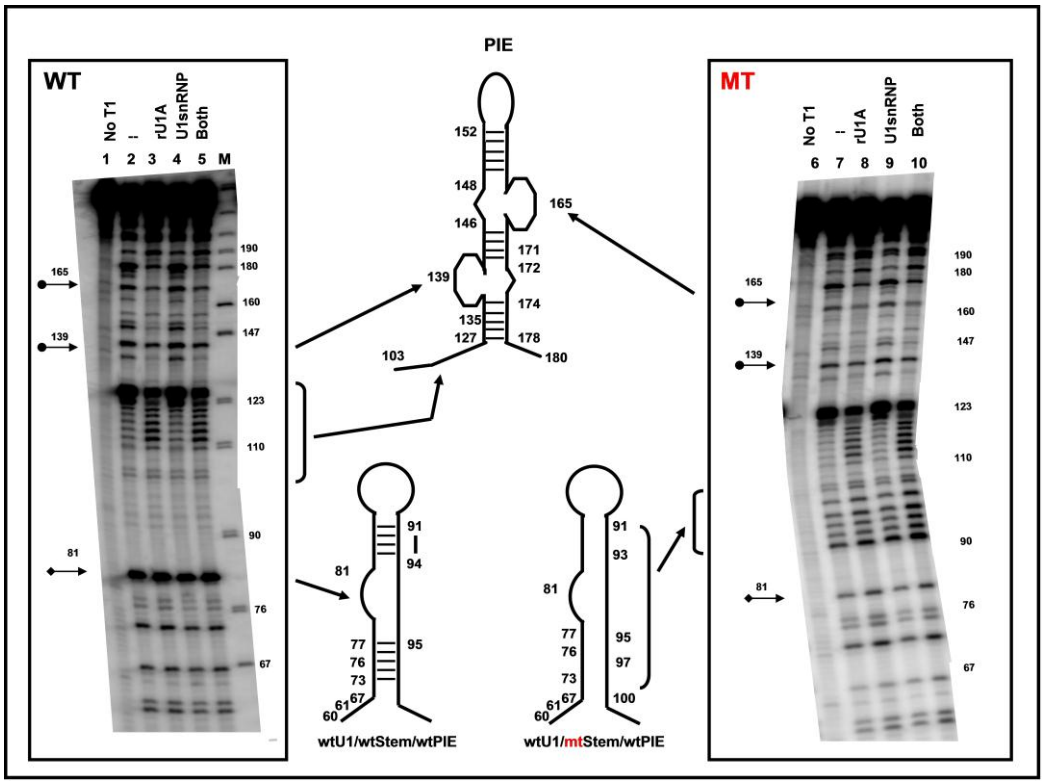

Figure 4.11: Mapping the U1-STEM Region Via RNase Digestion

RNase T1 mapping of wt and mt U1-STEM probe. Probes were incubated with 1:2000 T1 for 5 min. Lanes 1 and 6 are input lanes; lanes 2 and 7 have T1 only added; lanes 3 and 8 have 1200ng of rU1A added; lanes 4 and 9 have 780ng of U1 snRNP added, lanes 5 and 10 have both rU1A and U1 snRNP added. Lanes 1 – 5 correspond to the wild-type probe, while lanes 6-10 correspond to the mutated U1-STEM probe. The last lane in the wt film is the MSP marker, labeled M. Nucleotide 81 (arrow with diamond end) is located in the middle of the U1 site and the intensity of the band slightly decreases when U1 snRNP is added (compare lane 4 to lane 2,

and lane 9 to lane 7). Nucleotides 139 & 165 (arrows with circle ends) are located within the two rU1A binding loops of PIE, and the intensity of these bands greatly decreases when rU1A is added (compare lanes 3 & 5 to lane 2, and compare lanes 8 & 10 to lane 7). The U1-STEM runs from nucleotides 73 to 99, with possible RNase T1 targets at nucleotides 73, 76, 77, 81, 91, 93, 95 and 97. Bands at nucleotide position 81 are the most intense bands visible from this region in the wt probe. Bands are visible for positions 73, 76 and 77 but with decreased intensity. All target bands are visible in the mt U1-STEM probe.

The diagrams in between the two films indicate regions that should have altered digestion patterns after the addition of protein. The G residue at position 81 (indicated on the films by diamond-headed arrow) is located within the U1-STEM loop that bulges out and, more importantly, is located in the middle of the U1-site that U1 snRNP binds to. Two G residues, positions 139 and 165 (indicated on the films by circle-headed arrows), are located within the two U1A binding loops of PIE. Consequently, the digestion pattern for all three nucleotides were expected to change once the corresponding protein was bound, changing from strong digestion (dark band) to weak/no digestion (faint/no band), since the protein would block access to the site. The region from 100-126 corresponds to the region between the U1-STEM and PIE (hereafter referred to as Region C). Nucleotides at positions 90-100 indicate the mutated region created to open the U1-STEM.

The digestion pattern for the wild-type probe, absent any added proteins (lane 2), showed strong digestion at all expected, free ss G residues: 60, 61 and 67 (upstream of the U1-STEM; 73 and 81 (within U1-STEM); 103, 104, 106, 109, 111, 113, 115, 117, 119, 121, 123, 124 and 127 (Region C); 139, 146, 148, 152, 165 and 174 (within PIE); and 180 (downstream of PIE). The three residues that should be affected by protein binding (81, 139 and 165) are indicated on the films by arrows, as mentioned above. The

digestion pattern for the mutant probe (lane 7) matches that of the wild-type, with the addition of bands at positions 91, 93, 95, 97 and 100, which correspond to the mutated U1-STEM region.

The binding of rU1A to both the wt (lanes 3 and 5) and the mt (lanes 8 and 10) resulted in drastic reductions to band intensity at positions 139 and 165 (compare to lane 2 and lane 7, respectfully). This is indicative of rU1A binding to the two loops of PIE and occluding those nucleotides from RNase T1 digestion. The binding of rU1A also results in varying degrees of decreased digestion at positions 146, 148, 152 and 174. As all of these residues are located at the boundary of one of the two rU1A binding loops in PIE (with the exception of 152, which is at the boundary of the top hairpin loop), these changes are likely the result of changes in the structure of PIE after rU1A has bound.

The addition of U1 snRNP to the both the wt (lanes 4 and 5) and mt probes (lanes 9 and 10) resulted in no significant change in cleavage at position 81 (compare to lane 2 and lane 7, respectfully). Given that it was shown in Figures 4.2 and 4.6 that U1 snRNP does indeed bind to both the wt and mt probes, this was an extremely surprising result. This was not thought to be due to error within the assay or with the probes, seeing as rU1A was able to bind as expected. This data supports the previously stated conclusion that no trans-acting factor is required since these assays were conducted in the presence of nuclear extract, while the earlier results from Figure 4.6-4.8 were conducted in pure systems. Another possibility is that the binding of U1 snRNP to the U1-Site is too transitory to be measured by this assay, or that U1 snRNP is actually binding elsewhere.

rU1A Binding May Cause Remodeling of Region C

Another surprising result was the effect of rU1A binding on residues in Region C. In the absence of any external protein, the residues at positions 103, 104, 106, 109, 111, 113, 115, 117, 119 and 121 demonstrated weak/moderate digestion, in both wt and mt probes (Figure 4.11, lanes 2 and 7). Conversely, cleavage was especially prominent at nucleotides 123, 124 and 127, positions directly upstream of PIE. Upon rU1A binding, there is a dramatic shift within this region: the intensity of bands at positions 123, 124 and 127 decrease, while cleavage increases for the remaining residues in Region C (compare lane 3 to lane 2 for wt, and lane 8 to lane 7 for the mt). These results imply that this portion of the U1A RNA may fold in on itself via Region C, bringing the U1-STEM and PIE close together. The binding of rU1A to PIE may then disrupt this structure and release Region C, thereby exposing the U1-site and U1-STEM.

The Upper Part of the U1-STEM is More Essential in Inhibiting the U1-Site

To further address the idea that the lower portion of the U1-STEM is unstable, a set of six mutants was created by opening the U1-STEM via mutations of one or more nucleotides (Figure 4.12). These mutations incrementally abate base-pairing, thereby increasing access to the U1-site. The mutants, -2b, -1, -2t, -4, -7 and -10, were named based on the number of nucleotides mutated and in the case of -2b and -2t their position (b – bottom portion of the U1-STEM, t – top portion of the U1-STEM). The -10 mutant, which completely opens the U1-STEM is similar to the original U1-STEM mutant used previously (see Figure 4.5) but differs at eight nucleotides. Two mutants (+2 and +4),

added nucleotides to sequentially close the loop within the U1-STEM, which then occluded the entire U1-site. All mutants bound rU1A with affinity equal to that of the wild-type (data not shown). Alternatively, U1 snRNP binding varied among the mutants. Of the six mutants that opened the U1-STEM, -7 and -10 had the strongest binding, starting at 400 and 200 ng, respectively, while -4, -2t and -1 had minimal binding starting at 600 ng (data not shown). These results were as expected since greater accessibility to the U1-site should increase binding affinity. The -2b mutant, which opened the bottom portion of the U1-STEM and the two mutants that closed the U1-STEM, +2 and +4, had only nominal binding, even at high concentrations (data not shown). While the mutants that closed the U1-STEM were not expected to have high binding affinity, the lack of binding by the -2b mutant was surprising. This data was confirmed by later experiments, as explained below.

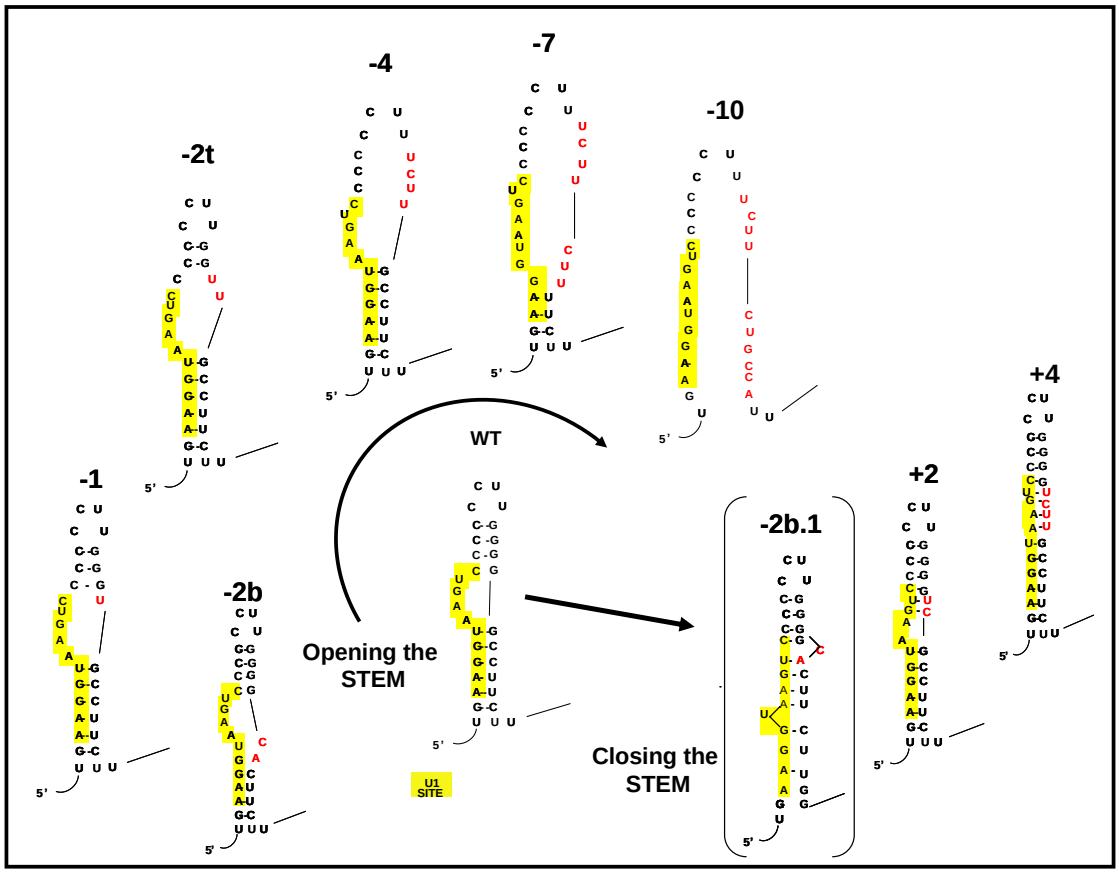

Figure 4.12: Sequences of Mutants to Open and Close the U1-STEM

Structure of WT U1A 3' UTR and eight U1-STEM mutants created to examine the effect opening or closing the U1-STEM has on U1 snRNP binding and polyadenylation inhibition. Six mutants opened the U1-STEM, starting from the lower left and moving clockwise: -2b has two NT mutated to partially open lower part of the U1-STEM; -1 has one NT mutated to partially open the upper part of the U1-STEM; -2t has two NT mutated to partially open the upper part of U1-STEM; -4 has four NT mutated to completely open the upper part of the U1-STEM; -7 has seven NT mutated to partially open the bottom portion of the U1-STEM in addition to the fully open upper part and -10 has ten NT mutated to completely open the STEM (this differs from the original U1-STEM mutant, seen in Figure 4.5, by eight NTs). Two mutants were created to sequentially close the U1-site loop of the U1-STEM: +2 has two NT added to partially close the U1-STEM at the top and +4 has four NT added to completely close the U1-STEM.

To determine the effects of these mutants *in vivo*, Renilla reporter plasmids were constructed for each mutant. These plasmids are similar to those in Figure 4.1, but contained the coding region for Renilla Luciferase (RL), instead of U1A. A series of

luciferase assays using the eight mutants along with the wild-type, a double mutant (mtU1-Site/wtU1-STEM/mtPIE) and the original U1-STEM mutant were undertaken and the mean values were calculated (+/- SEM) using GraphPad software (San Diego, CA), and expressed as Fold Inhibition (Figure 4.13). As the double mutant lacks inhibitory elements, resulting in the highest levels of polyadenylation, its value was set to 1 and all other samples were normalized to that value. The fold inhibition for all constructs were the following: wt 14.4 (+/- 4.5); double mutant 1.00 (+/- 0.00); mt U1-STEM 35.6 (+/- 7.1); -2b mt 3.7 (+/- 0.9); -1 mt 11.1 (+/- 3.8); -2t mt 56.8 (+/- 13.7); -4 mt 42.8 (+/- 10.9); -7 mt 46.9 (+/- 8.9); -10 mt 37.3 (+/- 5.6); +2 mt 4.6 (+/- 0.6) and +4 mt 3.7 (+/- 0.9). The original U1-STEM mutant and the -10 mutant had almost identical fold inhibition values, which was to be expected since they are structurally identical. This also indicates that the sequence of the mutation is not important for inhibition. The low levels of inhibition produced by the +2 and +4 mutants were consistent with a closed U1-STEM. The -2b mutant had a fold inhibition similar to that of the closed U1-STEM mutants. This also correlates with the earlier experiments noted above, that showed this mutant had minimal binding to U1 snRNP. Overall, this implies that the lower portion of the U1-site is not sufficient to bind U1 snRNP and that the lower portion of the U1-STEM is not inhibitory. Alternately, the mutant -2b could remodel and form the structure labeled -2b.1 in Figure 4.12. This would result in closing of the U1-STEM, similar to the +2 and +4 mutants. In this case, similar inhibitory patterns between the three mutants is to be expected. In contrast, the upper U1-STEM mutants all had increased levels of inhibition, with a drastic increase seen from -1 to -2t. Taken together, this data suggests

that the upper portion of the U1-STEM is essential to partially block binding of U1 snRNP to the U1-Site.

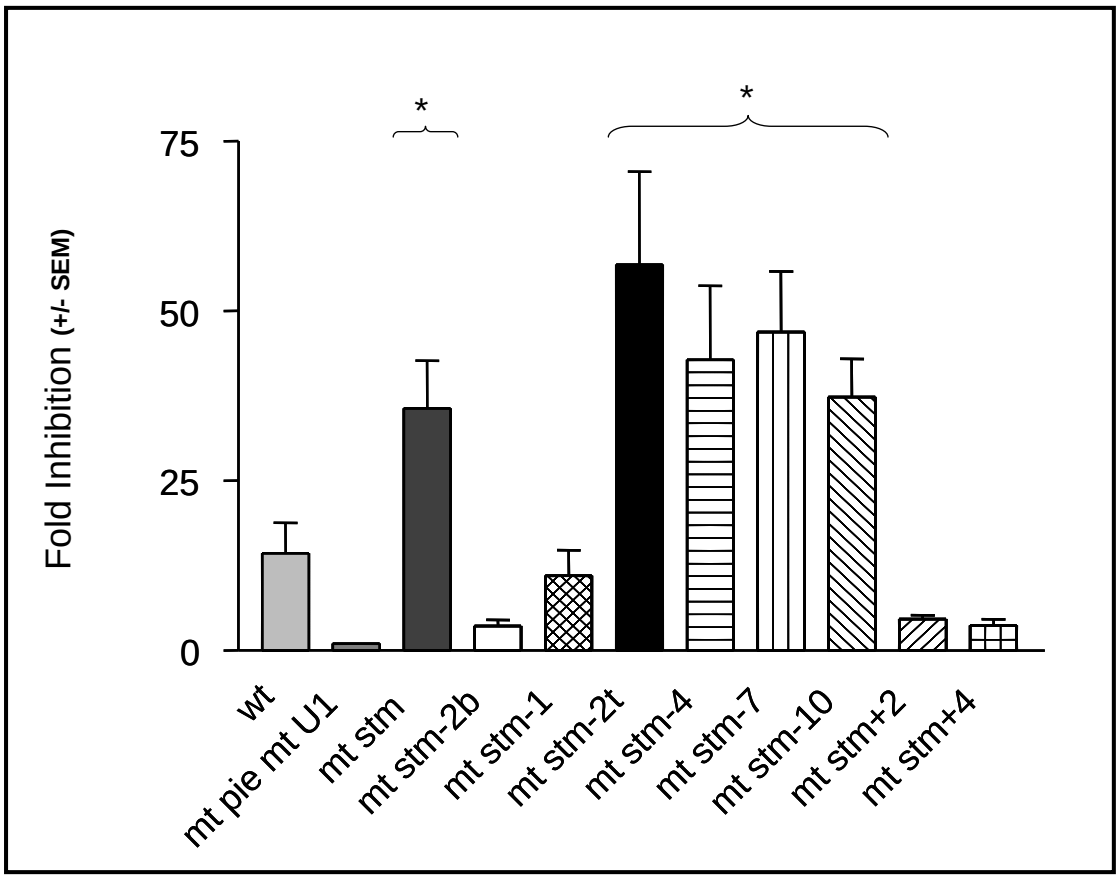

Figure 4.13: Effects of Opening/Closing the U1-STEM

Graphical representation of Luciferase transfections using the mutants in Figure 4.12 in addition to a double mutant (mtU1site/wtU1-STEM/mtPIE) and the original U1-STEM mutant shown in Figure 4.5. All values were normalized to the double mutant, which was set to 1. The four mutants that open the U1-STEM by two nucleotides or more, -2t, -4, -7 and -10 and the original U1-STEM mutant all showed statistically similar inhibitory levels (indicated by the asterisks). As the -10 mutant and the original U1-STEM mutant both completely open the U1-STEM, they have the same level of inhibition. The two mutants that close the U1-STEM show low levels of inhibition, as does the -2b mutant. Since the -2b mutant gives a lower level of inhibition compared to the -1 mutant, it may be that opening of the upper part of the U1-STEM is more important than opening the bottom portion. This opening my cause some sort of remodeling that allows U1 snRNP to bind with more affinity, thereby increasing inhibition. Values are a mean +/- SEM of 4 to 5 transfections. Graphs and SEM were calculated using GraphPad software (San Diego, CA).

Discussion

The data presented in this chapter is a continuation of what was previously discussed in Chapter III. In Guan *et al.,* 2007, a conserved U1-site present in the U1A 3'UTR was determined to act synergistically with PIE to increase polyadenylation inhibition. While it was observed that a wild-type PIE, with two molecules of U1A was required for U1 snRNP aided inhibition, the exact mechanism behind this necessity remained uncertain. The simplest explanation, i.e. the presence of a trans-acting factor, was discounted since synergy was reconstituted *in vitro* (Guan *et al.,* 2007) and subsequent biotin pulldowns and UV-crosslinking assays failed to uncover any novel proteins. As a result, the focus of these experiments shifted to examine in more detail how U1A binding effects the binding of U1 snRNP and also what, if any, effect it has on the RNA secondary structure of the 3'UTR.

The first step towards this goal was to determine the binding capabilities of each protein alone in the context of a fully wild-type U1-STEM. rU1A and U1 snRNP both bound in accordance with previously published results (van Gelder *et al.,* 1993; Gunderson *et al.,* 1994; Gunderson *et al.,* 1997; Gunderson *et al.,* 1998) when the RNA had either a wild-type PIE (Figure 4.2A) or a mutant PIE (Figure 4.2B). When binding was next examined in the presence of both proteins, a significant increase in U1 snRNP binding was seen (Figure 4.3, lanes 6-8). When 100ng of U1 snRNP was added alone, there was only minimal binding (lane 5 and see Figure 4.2B, lane 11), but once rU1A was added, U1 snRNP complexes with the U1-Site and PIE were amplified (Figure 4.3 compare lanes 6-8 to lanes 2-4).

This data is a direct contradiction to what was shown in Chapter III. As seen in Figure 3.5B, increasing amounts of rU1A had no effect on U1 snRNP binding (lanes 5-8). This disparity may be a product of the assays themselves. The western and northern blots measure the binding of each protein separately, rU1A and U1 snRNP's U1 snRNA respectively, while the gel shift assays measure the binding of the protein complex as a whole. Further experiments are needed to fully understand the relationship between rU1A binding and U1 snRNP, which will be addressed later on.

The apparent stimulatory effect of rU1A was further tested using "scrambled" rU1A. Only a slight increase in U1 snRNP binding is seen in the presence of this protein, which has lost its cooperative binding ability (Figure 4.4, lanes 5-8). As it has already been established that U1 snRNP has no inhibitory activity without the existence of two molecules of U1A in PIE, this data serves as further confirmation that U1A specifically stimulates U1 snRNP, as compared to any other RNA binding protein.

A possible explanation for the stimulatory effect of rU1A is that by binding to PIE, the U1-STEM is opened which allows U1 snRNP better access to the U1-Site. If the increase in U1 snRNP binding is solely due to the U1-STEM being open, then a fully mutated U1-STEM should result in greater binding independent of rU1A. Gel shifts using a mutated U1-STEM probe did show an increase in U1 snRNP binding in the absence of rU1A (Figure 4.6B and Figure 4.7, compare lanes 5 and 13). However, there was still an obvious stimulatory effect when rU1A was added (Figure 4.7, lanes 13-16). This effect is not seen when measuring inhibitory activity of the mutated U1-STEM via Poly(A) assay (Figure 4.8). The level of inhibition of the mutant when both proteins were present

was equal to that of U1 snRNP alone and to the wild-type when both proteins were present (compare lanes 9 and 10 to lane 5).

This apparent saturation implies that there is no additive effect when rU1A is added to a mutated U1-STEM. This corresponds with previous results seen in Figure 3.3C, in which a mutated U1-STEM, in the presence of a wild-type PIE and U1-site, was sufficient to increase inhibition almost four-fold over a fully wild-type probe (Guan *et al.,* 2007). In essence, while the presence of rU1A may increase U1 snRNP binding affinity, it does not also increase the inhibitory effect of U1 snRNP bound to the U1-site. A likely explanation for this effect is that binding of rU1A to PIE makes the U1-site more accessible to binding by U1 snRNP. This could be due to the U1-STEM being opened or thru remodeling of the RNA structure. This remodeling could help to stabilize U1 snRNP while it binds to the partially occluded U1-site. Since increased levels of rU1A were able to stimulate U1 snRNP binding to a mutated U1-STEM, their effect is not due to only opening up the U1-STEM. Instead, remodeling of the 3'UTR regions ABCD (see Figure 3.1) is most likely involved and allows for stabilizing RNA-Protein, Protein-Protein or RNA-RNA interactions.

The possibility of remodeling was further investigated via RNase digestion. This technique also allowed for the confirmation of the secondary structure of the U1-STEM region, which up to this point had been strictly theoretical. Comparison of the digestion patterns after RNase T1 and V1 incubation indicated that nucleotides located within the U1-STEM region were double-stranded (Figure 4.10). The upper portion of the U1-STEM (detectable residues 91-94), which is comprised of four C-G pairs, gave the strongest RNase V1 cleavage signal (which cuts double-stranded RNA) and the weakest

RNase T1 signal (which cuts single stranded G residues). The lack of a signal for the lower region (detectable residues 73, 76, 77 and 95) indicates that it's base-pairing is not as stabile as the upper portion. While three of the five base-pairs in the lower region are G-C bonds, they are separated by A-U base-pairs, and therefore more likely susceptible to breakage. This instability of the lower region could be what contributes to remodeling, or opening, of the U1-STEM after rU1A binds to PIE. If the base-pairs in this lower region were disrupted, it would provide 9/10 nucleotides for the U1 snRNA of U1 snRNP to bind to, thereby increasing U1 snRNP's binding affinity for the U1-site.

A comparison of the digestive pattern of wild-type and mutant probes with and with out the addition of rU1A does show a moderate increase in cleavage of lower region nucleotides 73, 76 and 77 (Figure 4.11, compare lanes 2 and 3, and lanes 7 and 8). While no significant change in the upper region residues was seen. This is further confirmation that rU1A helps to open the bottom portion of the U1-STEM. The addition of U1 snRNP (lanes 4 and 9) showed a decrease in digestion at positions 73, 76 and 77, which is indicative of U1 snRNP binding to the U1-site. Surprisingly, the presence of both proteins did not result in any considerable pattern changes, instead resembling those of rU1A added alone (compare lane 5 to lanes 3 and 4, and lane 10 to lanes 8 and 9). Even more surprising is the lack of change at position 81. While digestion is significantly less in the mutant, the relative levels do not drastically decrease in the presence of U1 snRNP, as would be excepted if U1 snRNP had full access to the U1-site. The presence of U1 snRNP and rU1A together also resembled that of rU1A alone (compare lanes 8, 9 and 10), which does not agree with the shift data which shows rU1A increasing U1 snRNP

binding affinity. This discrepancy is further seen in the wild-type probe, since there is no change to position 81 when one or both proteins were added (lanes 3-5).

One possible model is that the U1 snRNA of U1 snRNP only makes contacts with the lower portion of U1-STEM, up to around residue 78. Alternatively, the U1-site could attract U1 snRNP to the U1-STEM region, but since the site is partially occluded, U1 snRNP insinuates itself somewhere nearby. This type of interaction could then be strengthened due to remodeling of Region C (nucleotides 100-130). Residues in this region became available for cleavage when rU1A was present (lanes 3 and 5 and lanes 8 and 10). No effect is seen in the presence of U1 snRNP (lanes 4 and 9). Overall, it seems likely that rU1A binding induces a conformational change within Region C which allows for U1 snRNP to make transitory contacts with the partially available U1-site and helps to stabilize the entire repression complex.

If the lower portion of the U1-STEM is unstable, alternating between both open (single stranded) and closed conformations, it's role in suppressing the U1-site would be limited. The upper portion, which is more stable and not susceptible to remodeling, should then be the main cause of repression. Luciferase studies with the multiple U1-STEM mutants (Figure 4.12) demonstrated that any mutation to the upper portion was equal to that of mutating the entire U1-STEM (Figure 4.13). Therefore, the U1-STEM inhibitory activity resides in the upper portion.

Studies of the PIE region of U1A RNA have shown that the RNA samples multiple conformations until U1A binds, capturing the RNA in its desired formation (Gubser & Varani *et al.,* 1996, Shajani *et al.,* 2007, Qin *et al.,* 2010). Moreover, while the upper stem-loop of PIE is able to decouple its motion from the lower region, binding

of U1A to the RNA eventually results in the RNA becoming more rigid and stable. Applying this premise to the U1-STEM, it then seems likely that the top, inhibitory, portion, which is more stable, remains intact while the bottom portion and Region C are free to change conformation until U1A binds and stabilizes the RNA, providing a scaffold for U1 snRNP to bind to. Additional binding assays will be necessary to determine the exact location of U1 snRNP binding.

Chapter V: Summary and Conclusion

The majority of Eukaryotic mRNA's are post-transcriptionally modified at their 3'-ends to possess a poly(A) tail. This tail is essential for many mRNA functions and as such, is subject to precise regulation. This control of mRNA levels ultimately allows the cell to control gene expression. Polyadenylation regulation can involve the use of one poly(A) site out of many on a single mRNA (alternative polyadenylation, Chapter I), or a lone site may be controlled in an "on/off" manner. The U1A gene's expression is one of the best studied examples of "on/off" polyadenylation regulation. U1A autoregulates itself by binding to a regulatory element, PIE, located within its 3'-UTR. This mechanism has been greatly studied by members of this lab and others. This work expands on that research, demonstrating that the inhibition of U1A is not solely due to PIE, but instead requires an additional regulatory element, a U1-site.

Validation of the first U1-site in a eukaryotic gene

U1-sites were originally found only in papillomaviruses, but a search of 3'UTR databases found potential U1-sites in U1A, as well as other cellular genes (Figure 3.1). The results presented here represent the first characterization of a functional U1-site in an endogenous eukaryotic gene. It was revealed, through cDNA assays, that the U1A gene's U1-site was only active in the presence of wild-type PIE, indicating that U1A protein bound to PIE had to be present for the U1 site to effect inhibition (Figure 3.2). When both the U1-site and PIE were wild-type, they worked together synergistically to increase inhibition. Renilla assays using a minimal 3'UTR containing an isolated U1-site gave

greater levels of inhibition, compared to the same U1 site in the U1A cDNA context (Figure 3.3) indicating the presence of an undiscovered U1-site inhibitor. Further analysis led to the discovery of the U1-STEM that partially masked the inhibitory activity of the U1-site, leading to the reduced levels seen in the cDNA assays. The region between the U1-STEM and PIE (called Region C, Figure 3.1) was also determined to be required for synergistic inhibition of the U1-site and PIE (Figure 3.3).

Further testing by biotin pulldown assays showed that the synergy was not due to an increase in U1 snRNP, or U1A, binding (Figure 3.5). In fact, U1 snRNP was able to bind even in the presence of a mutated PIE, indicating that U1A is not required for U1 snRNP to bind to the U1-site. This data led to the model, proposed at the end of Chapter 3, that U1 snRNP binds to the partially exposed U1-site and remains inactive until U1A binds to PIE. Upon U1A binding, the U1-STEM would open allowing U1 snRNP to bind more fully, thereby becoming active in increasing inhibition. Alternatively, a *trans*-acting factor could bind to the 3'UTR and trap U1 snRNP in an inactive conformation. U1A binding would then force out this factor and allow U1 snRNP to become active.

Evidence for an intrinsic inhibitory mechanism

The possible role of a novel *trans*-acting factor was ruled out quickly once various assays produced no viable suspects. The trapping of U1 snRNP then had to be due to the structure of the mRNP complex itself. To further examine this possibility, multiple binding assays were conducted to determine the requirements for U1 snRNP binding to the U1-site (Figures 4.2 thru 4.7). Surprisingly, these experiments showed an

increase in U1 snRNP binding when U1A was added, regardless of whether the U1-STEM was wild-type or mutated (Figures 4.3 and 4.7), thus strongly arguing against a simple occlusion model where the U1 stem:U1 site blocks binding of U1 snRNP. RNA structure probing firmly established the RNA secondary structure of the U1-STEM and its base pairing with the U1-site (Figure 4.10). While these digestions corroborated what was known about U1A binding to PIE, they failed to identify which bases are directly bound by U1 snRNP. Residues that were expected to be protected from digestion by U1 snRNP binding remained fully sensitive to digestion, unlike PIE residues, which visibly showed a reduction in RNase sensitivity when rU1A was present (Figure 4.11). The earlier binding assays clearly demonstrated that U1 snRNP is indeed binding to the mRNA, so why did the RNase digestions not reflect this? One simple answer is that the assay used (RNase protection) is technically limited, for example, if U1 snRNP is binding near or at the ends of the RNA probe. Another possibility is U1 snRNP is conformationally switching its binding between two or more sites which would dilute out any RNase protection signal. An additional possibility may be found in Region C. This region was dramatically affected by the presence of rU1A, which suggests that this region may remodel itself upon U1A binding to PIE, which in turn could affect the availability of the U1-site.

Examination of different domains of the U1-STEM

A panel of eight U1-STEM mutants was tested to assess whether different parts of the U1-STEM contribute differently to U1-site repression (Figure 4.12). These *in vivo* assays showed that any mutation to the upper portion of the U1-STEM resulted in inhibition equal to that of mutating the entire U1-STEM (Figure 4.13). Thus, it can be concluded that the upper portion of the U1-STEM is more important in U1-site repression as compared to the lower portion of the U1-STEM. Furthermore, the RNase digestion data provide evidence for the instability of the lower regions of the U1-STEM relative to the upper regions. Residues present in the lower region existed both in single stranded and double stranded form, a stark contrast to the upper region that gave digestion patterns consistent with a stable stem-loop structure (Figure 4.11).

Model for the U1A gene's poly(A) site

Taken together, this data suggests an altered model of U1-site mediated polyadenylation inhibition (Figure 5.1). When U1A is not present, the upper portion of the U1-STEM remains fixed and closed, while the lower portion is free to change conformation. U1 snRNP could bind to the partially exposed U1-site, but would remain in an inactive state. Region C can fold into a closely packed conformation while remaining single-stranded (RNase V1 digestions of Region C did not show any cleavage [data not shown] indicating the region is not double-stranded when U1A is absent). This conformation would bring PIE into close proximity of the U1-STEM, which would help

to keep U1 snRNP bound in an inactive state. Once U1A binds to PIE, Region C would remodel by opening up, creating space between the U1-STEM and PIE. U1 snRNP can now move around within the U1-STEM, due to the increased distance between it and PIE. U1 snRNP would also be free to interact with the U1A proteins bound to PIE. These protein-protein interactions would then lead to an increase in polyadenylation inhibition.

Broader function of U1A

An obvious question remains – why does U1A mRNA need such a complex regulatory mechanism? The high level of conservation of PIE and the U1-site (Figure 3.1), as well as the U1-STEM (Figure 3.3) indicates that these regions are necessary for function. If U1A's primary function was simply as a component of U1 snRNP, then this extreme control of its mRNAs regulation seems excessive. However, if U1A instead plays a much broader role in cellular function, this type of regulation would be expected. The discovery that U1A regulates IgM poly(A) site usage (Introduction, Figure 1.6) is a clue to what that role may be. IgM may just be the first of multiple genes that U1A regulates, marking U1A as a master regulator of not only poly(A) sites, but of gene expression overall.

In regards to its own regulation, further studies will be needed to examine more closely the physical binding of U1 snRNP to the U1-STEM region. Nonetheless, this U1-site in the U1A gene is the first validated example of a functional U1-site in an endogenous gene in any known eukaryote. Studies to characterize the U1-sites of other genes found in the original 3'UTR search may also provide insight into how these sites

work in a broader context. The potency of this U1-site, plus those found in papillomaviruses, as well as the preliminary success seen with the U1 Adaptor technology indicate that U1-sites may play a broader role in regulation than previously imagined.

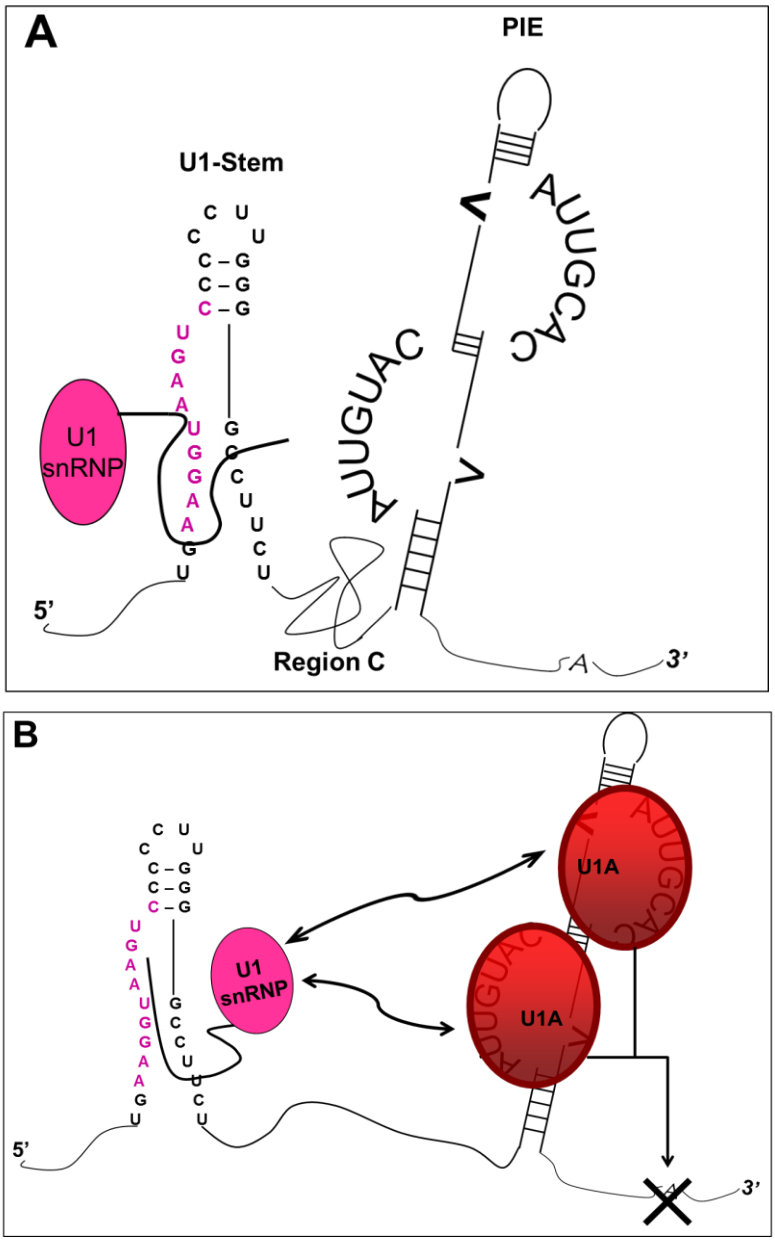

Figure 5.1 Model For Polyadenylation Inhibition

A. Illustration of the proposed model of polyadenylation inhibition. (A) The U1-STEM (with the U1-site in purple) is shown with only the upper portion base-paired, as indicated by the RNase V1 digestion data. U1 snRNP is able to bind to the lower portion of the exposed U1-site, but remains inactive. The two loops of PIE are unoccupied by U1A, and Region C is folded in on itself, bringing PIE into close proximity of the U1-STEM. (B) Once PIE is bound by two molecules of U1A, Region C is remodeled and opens up, allowing U1 snRNP to make better contacts with the U1-site. Protein-protein interactions (indicated by arrows) between U1 snRNP and U1A could then stabilize the mRNA and result in increased inhibition of polyadenylation.

References

Abad X, Vera M, Jung SP, Oswald E, Romero I, Amin V, Fortes P, Gunderson SI. 2008 Requirements for gene silencing mediated by U1 snRNA binding to a target sequence. Nucleic Acids Res. Apr;36(7):2338-52.

Alt FW, Bothwell AL, Knapp M, Siden E, Mather E, Koshland M, Baltimore D. 1980. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. Cell. Jun;20(2):293-301.

Amrani N, Minet M, Le Gouar M, Lacroute F, Wyers F. 1997 Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. Mol Cell Biol. Jul;17(7):3694-701.

Arhin GK, Boots M, Bagga PS, Milcarek C, Wilusz J. 2002 Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. Nucleic Acids Res. Apr 15;30(8):1842-50.

Ashe MP, Pearson LH, Proudfoot NJ. 1997. The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. EMBO J. Sep 15;16(18):5752-63.

Ashe MP, Furger A, Proudfoot NJ. 2000, Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. RNA. Feb;6(2):170-7.

Aström A, Aström J, Virtanen A. 1991 A simple procedure for isolation of eukaryotic mRNA polyadenylation factors. Eur J Biochem. Dec 18;202(3):765-73.

Awasthi S, Alwine JC. 2003 Association of polyadenylation cleavage factor I with U1 snRNP. RNA. Nov;9(11):1400-9.

Bagga PS, Ford LP, Chen F, Wilusz J. 1995. The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. Nucleic Acids Res. May 11;23(9):1625-31.

Balbo PB, Bohm A. 2007 Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. Structure. Sep;15(9):1117-31.

Ballantyne S, Bilger A, Astrom J, Virtanen A, Wickens M. 1995 Poly (A) polymerases in the nucleus and cytoplasm of frog oocytes: dynamic changes during oocyte maturation and early development. RNA. Mar;1(1):64-78.

Baker CC. 1990. Bovine papillomavirus type I transcription, p. 91-112. *In* H. Pfister (ed.), Papillomavirus and human cancer. CRC Press, Inc., Boca Raton, Fla.

Barabino SM, Hübner W, Jenny A, Minvielle-Sebastia L, Keller W. 1997 The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. Genes Dev. Jul 1;11(13):1703-16.

Barabino SM, Ohnacker M, Keller W. 2000 Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. EMBO J. Jul 17;19(14):3778-87.

Bard J, Zhelkovsky AM, Helmling S, Earnest TN, Moore CL, Bohm A. 2000 Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. Science. Aug 25;289(5483):1346-9.

Barksdale SK, Baker CC. 1993. Differentiation-specific expression from the bovine papillomavirus type 1 P2443 and late promoters. J Virol. Sep;67(9):5605-16.

Barnard DC, Ryan K, Manley JL, Richter JD. 2004 Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. Cell. Nov 24;119(5):641-51.

Barreau C, Paillard L, Osborne HB. 2006 AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. Jan 3;33(22):7138-50. Print 2005.

Beaudoing E, Gautheret D. 2001 Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data. Genome Res. Sep;11(9):1520-6.

Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. 2000. Patterns of variant polyadenylation signal usage in human genes. Genome Res. Jul;10(7):1001-10.

Beckley SA, Liu P, Stover ML, Gunderson SI, Lichtler AC, Rowe DW. 2001 Reduction of target gene expression by a modified U1 snRNA. Mol Cell Biol. Apr;21(8):2815-25.

Benoit B, Juge F, Iral F, Audibert A, Simonelig M. 2002 Chimeric human CstF-77/Drosophila Suppressor of forked proteins rescue suppressor of forked mutant lethality and mRNA 3' end processing in Drosophila. Proc Natl Acad Sci U S A. Aug 6;99(16):10593-8.

Bentley DL. 2005 Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol. Jun;17(3):251-6

Bienroth S, Wahle E, Suter-Crazzolara C, Keller W. 1991 Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. J Biol Chem. Oct 15;266(29):19768-76.

Bienroth S, Keller W, Wahle E. 1993 Assembly of a processive messenger RNA polyadenylation complex. EMBO J. Feb;12(2):585-94.

Beaudoing E, Gautheret D. 2001 Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data. Genome Res. Sep;11(9):1520-6.

Black DL. 2003 Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem. ;72:291-336.

Boelens WC, Jansen EJ, van Venrooij WJ, Stripecke R, Mattaj IW, Gunderson SI. 1993. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. Cell. Mar 26;72(6):881-92.

Borman AM, Michel YM, Kean KM. 2000 Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. Nucleic Acids Res. Nov 1;28(21):4068-75.

Bousquet-Antonelli C, Presutti C, Tollervey D. 2000. Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell. Sep 15;102(6):765-75.

Brown CE, Sachs AB. 1998. Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. Mol Cell Biol. Nov;18(11):6548-59.

Brown KM, Gilmartin GM. 2003 A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol Cell. Dec;12(6):1467-76.

Chen CY, Shyu AB. 1995 AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci. Nov;20(11):465-70

Chen CY, Gherzi R, Ong SE, Chan EL, Raijmakers R, Pruijn GJ, Stoecklin G, Moroni C, Mann M, Karin M. 2001 AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell. Nov 16;107(4):451-64.

Chen F, MacDonald CC, Wilusz J. 1995. Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acids Res. Jul 25;23(14):2614-20.

Connelly S, Manley JL. 1988 A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. Apr;2(4):440-52.

Cumming SA, McPhillips MG, Veerapraditsin T, Milligan SG, Graham SV. 2003. Activity of the human papillomavirus type 16 late negative regulatory element is partly due to four weak consensus 5' splice sites that bind a U1 snRNP-like complex. J Virol. May;77(9):5167-77.

Danckwardt S, Hentze MW, Kulozik AE. 2008. 3' end mRNA processing: molecular mechanisms and implications for health and disease. EMBO J. Feb 6;27(3):482-98.

Danckwardt S, Kaufmann I, Gentzel M, Foerstner KU, Gantzert AS, Gehring NH, Neu-Yilik G, Bork P, Keller W, Wilm M, Hentze MW, Kulozik AE. 2007 Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. EMBO J. Jun 6;26(11):2658-69.

Deo RC, Bonanno JB, Sonenberg N, Burley SK. 1999 Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell. Sep 17;98(6):835-45.

de Vries H, Rüegsegger U, Hübner W, Friedlein A, Langen H, Keller W. 2000 Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. Nov 1;19(21):5895-904.

Edmonds M. 2002 A history of poly A sequences: from formation to factors to function. Prog Nucleic Acid Res Mol Biol.;71:285-389

Edwalds-Gilbert G, Veraldi KL, Milcarek C. 1997 Alternative poly(A) site selection in complex transcription units: means to an end? Nucleic Acids Res. Jul 1;25(13):2547-61.

Eldad N, Yosefzon Y, Arava Y. 2008 Identification and characterization of extensive intra-molecular associations between 3'-UTRs and their ORFs. Nucleic Acids Res. Dec;36(21):6728-38.

Fitzgerald M, Shenk T. 1981. The sequence 5'-AAUAAA-3'forms parts of the recognition site for polyadenylation of late SV40 mRNAs. Cell. Apr;24(1):251-60.

Fong N, Bentley DL. 2001 Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev. Jul 15;15(14):1783-95.

Ford LP, Bagga PS, Wilusz J. 1997 The poly(A) tail inhibits the assembly of a 3'-to-5' exonuclease in an in vitro RNA stability system. Mol Cell Biol. Jan;17(1):398-406.

Fortes P, Cuevas Y, Guan F, Liu P, Pentlicky S, Jung SP, Martínez-Chantar ML, Prieto J, Rowe D, Gunderson SI. 2003 Inhibiting expression of specific genes in mammalian cells with 5' end-mutated U1 small nuclear RNAs targeted to terminal exons of pre-mRNA. Proc Natl Acad Sci U S A. Jul 8;100(14):8264-9

Furth PA, Baker CC. 1991. An element in the bovine papillomavirus late 3' untranslated region reduces polyadenylated cytoplasmic RNA levels. J Virol. Nov;65(11):5806-12.

Furth PA, Choe WT, Rex JH, Byrne JC, Baker CC. 1994. Sequences homologous to 5' splice sites are required for the inhibitory activity of papillomavirus late 3' untranslated regions. Mol Cell Biol. Aug;14(8):5278-89.

Galli G, Guise J, Tucker PW, Nevins JR. 1988. Poly(A) site choice rather than splice site choice governs the regulated production of IgM heavy-chain RNAs. Proc Natl Acad Sci U S A. Apr;85(8):2439-43

Gallie DR. 1991 The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev. Nov;5(11):2108-16.

Gherzi R, Lee KY, Briata P, Wegmüller D, Moroni C, Karin M, Chen CY. 2004 A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery. Mol Cell. Jun 4;14(5):571-83.

Gil A, Proudfoot NJ. 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. Cell. May 8;49(3):399-406.

Gilmartin GM. 2005. Eukaryotic mRNA 3' processing: a common means to different ends. Genes Dev. Nov 1;19(21):2517-21

Goraczniak R, Behlke MA, Gunderson SI. 2009 Gene silencing by synthetic U1 adaptors. Nat Biotechnol. Mar;27(3):257-63.

Goraczniak R, Gunderson SI. 2008 The regulatory element in the 3'-untranslated region of human papillomavirus 16 inhibits expression by binding CUG-binding protein 1. J Biol Chem. Jan 25;283(4):2286-96.

Guan F, Palacios D, Hussein RI, Gunderson SI. 2003 Determinants within an 18-amino-acid U1A autoregulatory domain that uncouple cooperative RNA binding, inhibition of polyadenylation, and homodimerization. Mol Cell Biol. May;23(9):3163-72.

Guhaniyogi J, Brewer G. 2001 Regulation of mRNA stability in mammalian cells. Gene. Mar 7;265(1-2):11-23.

Gunderson SI, Beyer K, Martin G, Keller W, Boelens WC, Mattaj LW. 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. Cell. Feb 11;76(3):531-41.

Gunderson SI, Vagner S, Polycarpou-Schwarz M, Mattaj IW. 1997. Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. Genes Dev. Mar 15;11(6):761-73.

Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. 1998. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. Mol Cell. Jan;1(2):255-64.

Gunnewiek JM, van Aarssen Y, Wassenaar R, Legrain P, van Venrooij WJ, Nelissen RL. 1995 Homodimerization of the human U1 snRNP-specific protein C. Nucleic Acids Res. Dec 11;23(23):4864-71.

He X, Moore C. 2005 Regulation of yeast mRNA 3' end processing by phosphorylation. Mol Cell. Sep 2;19(5):619-29.

Higgs DR, Goodbourn SE, Lamb J, Clegg JB, Weatherall DJ, Proudfoot NJ. 1983. Alpha-thalassaemia caused by a polyadenylation signal mutation. Nature. Nov 24-30;306(5941):398-400.

Hirose Y, Manley JL. 1998 RNA polymerase II is an essential mRNA polyadenylation factor. Nature. Sep 3;395(6697):93-6.

Hirose Y, Manley JL. 2000. RNA polymerase II and the integration of nuclear events. Genes Dev. Jun 15;14(12):1415-29

Hofmann I, Schnölzer M, Kaufmann I, Franke WW. 2002 Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in Xenopus laevis oocytes. Mol Biol Cell. May;13(5):1665-76.

Hu J, Lutz CS, Wilusz J, Tian B. 2005. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA. Oct;11(10):1485-93

Ji Z, Tian B. 2009a Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. PLoS One. Dec 23;4(12):e8419

Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009b Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci U S A. Apr 28;106(17):7028-33.

Kahvejian A, Svitkin YV, Sukarieh R, M'Boutchou MN, Sonenberg N. 2005 Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev. Jan 1;19(1):104-13.

Kambach C, Mattaj IW. 1992 Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA. J Cell Biol. Jul;118(1):11-21.

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. Jul 15;21(14):1779-89.

Kattah NH, Kattah MG, Utz PJ. 2010 The U1-snRNP complex: structural properties relating to autoimmune pathogenesis in rheumatic diseases. Immunol Rev. Jan;233(1):126-45.

Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. 2004 Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO J. Feb 11;23(3):616-26.

Keller RW, Kühn U, Aragón M, Bornikova L, Wahle E, Bear DG 2000 The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. J Mol Biol. Mar 31;297(3):569-83

Keller W, Bienroth S, Lang KM, Christofori G. 1991 Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. EMBO J. Dec;10(13):4241-9.

Kennedy IM, Haddow JK, Clements JB. 1990. Analysis of human papillomavirus type 16 late mRNA 3' processing signals in vitro and in vivo. J Virol. Apr;64(4):1825-9.

Keon BH, Schäfer S, Kuhn C, Grund C, Franke WW. 1996 Symplekin, a novel type of tight junction plaque protein. J Cell Biol. Aug;134(4):1003-18.

Kerwitz Y, Kühn U, Lilie H, Knoth A, Scheuermann T, Friedrich H, Schwarz E, Wahle E. 2003 Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. EMBO J. Jul 15;22(14):3705-14.

Kim H, Lee Y. 2001 Interaction of poly(A) polymerase with the 25-kDa subunit of cleavage factor I. Biochem Biophys Res Commun. Nov 30;289(2):513-8.

Klein Gunnewiek JM, Hussein RI, van Aarssen Y, Palacios D, de Jong R, van Venrooij WJ, Gunderson SI. 2000. Fourteen residues of the U1 snRNP-specific U1A protein are required for homodimerization, cooperative RNA binding, and inhibition of polyadenylation. Mol Cell Biol. Mar;20(6):2209-17.

Ko B, Gunderson SI. 2002. Identification of new poly(A) polymerase-inhibitory proteins capable of regulating pre-mRNA polyadenylation. J Mol Biol. May 17;318(5):1189-206.

Krämer A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. Annu Rev Biochem. 65:367–409.

Krämer A, Keller W. 1990 Preparation and fractionation of mammalian extracts active in pre-mRNA splicing. Methods Enzymol. ;181:3-19.

Krol A, Westhof E, Bach M, Lührmann R, Ebel JP, Carbon P. 1990 Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. Nucleic Acids Res. Jul 11;18(13):3803-11.

Kyburz A, Friedlein A, Langen H, Keller W. 2006 Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. Mol Cell. Jul 21;23(2):195-205.

Kyriakopoulou CB, Nordvarg H, Virtanen A. 2001 A novel nuclear human poly(A) polymerase (PAP), PAP gamma. J Biol Chem. Sep 7;276(36):33504-11.

Labourier E, Adams MD, Rio DC. 2001 Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. Mol Cell. Aug;8(2):363-73.

Lewis BP, Burge CB, Bartel DP. 2005 Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. Jan 14;120(1):15-20.

Li Y, Chen ZY, Wang W, Baker CC, Krug RM. 2001 The 3'-end-processing factor CPSF is required for the splicing of single-intron pre-mRNAs in vivo. RNA. Jun;7(6):920-31.

Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E. 2000. Molecular Cell Biology. 4th ed. New York: W. H. Freeman & Co.

Logan J, Falck-Pedersen E, Darnell JE Jr, Shenk T. 1987 A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc Natl Acad Sci U S A. Dec;84(23):8306-10.

Lou H, Neugebauer KM, Gagel RF, Berget SM. 1998 Regulation of alternative polyadenylation by U1 snRNPs and SRp20. Mol Cell Biol. Sep;18(9):4977-85.

Lutz CS, Murthy KG, Schek N, O'Connor JP, Manley JL, Alwine JC. 1996 Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. Genes Dev. Feb 1;10(3):325-37.

Lutz-Freyermuth C, Query C C, Keene J D. 1990. Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. Proc Natl Acad Sci USA.;87:6393–6397.

Ma J, Gunderson SI, Phillips C. 2006 Non-snRNP U1A levels decrease during mammalian B-cell differentiation and release the IgM secretory poly(A) site from repression. RNA. Jan;12(1):122-32.

MacDonald CC, Wilusz J, Shenk T. 1994 The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Mol Cell Biol. Oct;14(10):6647-54.

Macknight R, Bancroft I, Page T, Lister C, Schmidt R, Love K, Westphal L, Murphy G, Sherson S, Cobbett C, Dean C. 1997 FCA, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. Cell. May 30;89(5):737-45.

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. 2006 Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature. Dec 14;444(7121):953-6.

Mandel CR, Bai Y, Tong L. 2008. Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci. Apr;65(7-8):1099-122.

Martin G, Keller W. 1996 Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. EMBO J. May 15;15(10):2593-603.

Martin G, Keller W, Doublié S. 2000 Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. EMBO J. Aug 15;19(16):4193-203.

Martin G, Möglich A, Keller W, Doublié S. 2004 Biochemical and structural insights into substrate binding and catalytic mechanism of mammalian poly(A) polymerase. J Mol Biol. Aug 20;341(4):911-25.

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL. 1997 The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature. Jan 23;385(6614):357-61.

McCracken S, Longman D, Johnstone IL, Cáceres JF, Blencowe BJ. 2003 An evolutionarily conserved role for SRm160 in 3'-end processing that functions independently of exon junction complex formation. J Biol Chem. Nov 7;278(45):44153-60.

McDevitt MA, Hart RP, Wong WW, Nevins JR. 1986. Sequences capable of restoring poly(A) site function define two distinct downstream elements. EMBO J. Nov;5(11):2907-13.

McLauchlan J, Gaffney D, Whitton JL, Clements JB. 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucleic Acids Res. Feb 25;13(4):1347-68

Meyer S, Urbanke C, Wahle E. 2002 Equilibrium studies on the association of the nuclear poly(A) binding protein with poly(A) of different lengths. Biochemistry. May 14;41(19):6082-9.

Milcarek C, Martincic K, Chung-Ganster LH, Lutz CS. 2003 The snRNP-associated U1A levels change following IL-6 stimulation of human B-cells. Mol Immunol. Mar;39(13):809-14.

Millevoi S, Vagner S. 2010 Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res. May;38(9):2757-74

Millevoi S, Geraghty F, Idowu B, Tam JL, Antoniou M, Vagner S 2002 A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. EMBO Rep. Sep;3(9):869-74.

Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S. 2006 An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. EMBO J. Oct 18;25(20):4854-64.

Moraes KC, Wilusz CJ, Wilusz J. 2006 CUG-BP binds to RNA substrates and recruits PARN deadenylase. RNA. Jun;12(6):1084-91

Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ. 1998 The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. Genes Dev. Aug 15;12(16):2522-34.

Moore MJ. 2005 From birth to death: the complex lives of eukaryotic mRNAs. Science. Sep 2;309(5740):1514-8.

Mukherjee D, Gao M, O'Connor JP, Raijmakers R, Pruijn G, Lutz CS, Wilusz J. 2002 The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. EMBO J. Jan 15;21(1-2):165-74.

Murthy KG, Manley JL. 1995 The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes Dev. Nov 1;9(21):2672-83.

Nagai K, Oubridge C, Jessen TH, Li J, Evans PR. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature. Dec 6;348(6301):515-20.

Natalizio BJ, Muniz LC, Arhin GK, Wilusz J, Lutz CS. 2002. Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals. J Biol Chem. Nov 8;277(45):42733-40.

Neilson JR, Sandberg R. 2010 Heterogeneity in mammalian RNA 3' end formation. Exp Cell Res. May 1;316(8):1357-64

Niwa M, Rose SD, Berget SM. 1990 In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev. Sep;4(9):1552-9.

Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A. 2005 Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat Struct Mol Biol. Feb;12(2):144-51

Oberg D, Fay J, Lambkin H, Schwartz S. 2005. A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. J Virol. Jul;79(14):9254-69.

O'Connor JP, Alwine JC, Lutz CS. 1997. Identification of a novel, non-snRNP protein complex containing U1A protein. RNA. Dec;3(12):1444-55.

Paillard L, Omilli F, Legagneux V, Bassez T, Maniey D, Osborne HB. 1998 EDEN and EDEN-BP, a cis element and an associated factor that mediate sequence-specific mRNA deadenylation in Xenopus embryos. EMBO J. Jan 2;17(1):278-87.

Pesole G, Liuni S, Grillo G, Licciulli F, Mignone F, Gissi C, Saccone C. 2002 UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Nucleic Acids Res. Jan 1;30(1):335-40.

Pettitt J, Crombie C, Schümperli D, Müller B. 2002 The Caenorhabditis elegans histone hairpin-binding protein is required for core histone gene expression and is essential for embryonic and postembryonic cell division. J Cell Sci. Feb 15;115(Pt 4):857-66.

Phillips C, Jung S, Gunderson SI. 2001. Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA. EMBO J. Nov 15;20(22):6443-52.

Phillips C and Gunderson SI. 2003. Sequences adjacent to the 5’ splice site control U1A binding upstream of the IgM heavy chain secretory poly(A) site. J Biol Chem. 278:22102-11.

Phillips C, Pachikara N, Gunderson SI. 2004. U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions. Mol Cell Biol. Jul;24(14):6162-71.

Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K. 2009 Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature. Mar 26;458(7237):475-80.

Preiss T, Hentze MW. 1998 Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature. Apr 2;392(6675):516-20.

Preker PJ, Ohnacker M, Minvielle-Sebastia L, Keller W. 1997 A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor. EMBO J. Aug 1;16(15):4727-37.

Proudfoot N. 1991. Poly(A) signals. Cell. Feb 22;64(4):671-4.

Proudfoot NJ, Brownlee GG. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. Nature. Sep 16;263(5574):211-4.

Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. Cell Feb 22;108(4):501-12.

Quesada V, Macknight R, Dean C, Simpson GG. 2003 Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time. EMBO J. Jun 16;22(12):3142-52.

Quesada V, Dean C, Simpson GG. 2005 Regulated RNA processing in the control of Arabidopsis flowering. Int J Dev Biol.;49(5-6):773-80.

Raabe T, Murthy KG, Manley JL. 1994 Poly(A) polymerase contains multiple functional domains. Mol Cell Biol. May;14(5):2946-57.

Rappsilber J, Ryder U, Lamond AI, Mann M. 2002 Large-scale proteomic analysis of the human spliceosome. Genome Res. Aug;12(8):1231-45.

Rosonina E, Kaneko S, Manley JL. 2006 Terminating the transcript: breaking up is hard to do. Genes Dev. May 1;20(9):1050-6

Rüegsegger U, Beyer K, Keller W. 1996 Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. J Biol Chem. Mar 15;271(11):6107-13.

Rüegsegger U, Blank D, Keller W. 1998 Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Mol Cell. Jan;1(2):243-53.

Ryan K, Murthy KG, Kaneko S, Manley JL. 2002 Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. Mol Cell Biol. Mar;22(6):1684-92.

Ryan K, Calvo O, Manley JL. 2004 Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA. Apr;10(4):565-73.

Ryner LC, Takagaki Y, Manley JL. 1989 Multiple forms of poly(A) polymerases purified from HeLa cells function in specific mRNA 3'-end formation. Mol Cell Biol. Oct;9(10):4229-38.

Sachs A and Wahle E. 1993. Poly(A) tail metabolism and function in eukaryotes. J Biol Chem. Nov 5: 268(31):22955-8.

Sachs AB. 2000 Physical and functional interactions between the mRNA cap structure and the polyA tail. In: Sonenberg N, Hershey JWB, Mathews MB, editors. Translational Control of Gene Expression. Cold Spring Harbor: CSHL press.

Sachs AB, Sarnow P, Hentze MW. 1997. Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell. Jun 13;89(6):831-8

Sadowski M, Dichtl B, Hübner W, Keller W. 2003 Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. EMBO J. May 1;22(9):2167-77.

Sajic R, Lee K, Asai K, Sakac D, Branch DR, Upton C, Cochrane A. 2007 Use of modified U1 snRNAs to inhibit HIV-1 replication. Nucleic Acids Res. ;35(1):247-55.

Salisbury J, Hutchison KW, Graber JH. 2006 A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif. BMC Genomics. Mar 16;7:55.

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008 Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science. Jun 20;320(5883):1643-7.

Scherly D, Boelens W, van Venrooij W J, Dathan N A, Hamm J, Mattaj I W. 1989. Identification of the RNA binding segment of human U1A protein and definition of its binding site on U1 snRNA. EMBO J. 8:4163–4170.

Shatkin AJ, Manley JL. 2000. The ends of the affair: capping and polyadenylation. Nat Struct Biol. Oct;7(10):838-42.

Sheets MD, Ogg SC, Wickens MP. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res. Oct 11;18(19):5799-805.

Shell SA, Hesse C, Morris SM Jr, Milcarek C. 2005 Elevated levels of the 64-kDa cleavage stimulatory factor (CstF-64) in lipopolysaccharide-stimulated macrophages influence gene expression and induce alternative poly(A) site selection. J Biol Chem. Dec 2;280(48):39950-61.

Simonelig M, Elliott K, Mitchelson A, O'Hare K. 1996 Interallelic complementation at the suppressor of forked locus of Drosophila reveals complementation between suppressor of forked proteins mutated in different regions. Genetics. Apr;142(4):1225-35.

Sittler A, Gallinaro H, Jacob M. 1994. Upstream and downstream cis-acting elements for cleavage at the L4 polyadenylation site of adenovirus-2. Nucleic Acids Res. Jan 25;22(2):222-31.

Spalholz BA and Howley PM. 1989. Papillomavirus-host cell interactions, p. 27-53. *In* G. Klein (ed.), Advances in viral oncology, vol. 8. Raven Press, New York.

Takagaki Y, Seipelt RL, Peterson ML, Manley JL. 1996 The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell. Nov 29;87(5):941-52.

Takagaki Y, Manley JL. 2000 Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol Cell Biol. Mar;20(5):1515-25.

Thompson JD, Higgins DG, Gibson TJ. 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. Nov 11;22(22):4673-80.

Thuresson AC, Aström J, Aström A, Grönvik KO, Virtanen A. 1994 Multiple forms of poly(A) polymerases in human cells. Proc Natl Acad Sci U S A. Feb 1;91(3):979-83.

Tian B, Hu J, Zhang H, Lutz CS. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. Jan 12;33(1):201-12.

Tian B, Pan Z, Lee JY. 2007 Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. Feb;17(2):156-65.

Topalian SL, Kaneko S, Gonzales MI, Bond GL, Ward Y, Manley JL. 2001 Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. Mol Cell Biol. Aug;21(16):5614-23.

Vagner S, Rüegsegger U, Gunderson SI, Keller W, Mattaj IW. 2000a Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP. RNA. Feb;6(2):178-88.

Vagner S, Vagner C, Mattaj IW. 2000b The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes Dev. Feb 15;14(4):403-13.

van Gelder C W G, Gunderson S I, Jansen E J R, Boelens W C, Polycarpou-Schwarz M, Mattaj I W, van Venrooij W J. 1993. A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. EMBO J. 12:5191–5200.

van Gelder CW, Thijssen JP, Klaassen EC, Sturchler C, Krol A, van Venrooij WJ, Pruijn GJ. 1994. Common structural features of the Ro RNP associated hY1 and hY5 RNAs. Nucleic Acids Res. Jul 11;22(13):2498-506.

Venkataraman K, Brown KM, Gilmartin GM. 2005 Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. Jun 1;19(11):1315-27.

Vilardell J, Warner JR. 1994 Regulation of splicing at an intermediate step in the formation of the spliceosome. Genres Dev. Jan;8(2):211-20.

Vilardell J, Warner JR. 1997 Ribosomal protein L32 of Saccharomyces cerevisiae influences both the splicing of its own transcript and the processing of rRNA. Mol Cell Biol. Apr;17(4):1959-65.

Vinciguerra P, Stutz F. 2004. mRNA export: an assembly line from genes to nuclear pores. Curr Opin Cell Biol. Jun;16(3):285-92.

Vlasova IA, Tahoe NM, Fan D, Larsson O, Rattenbacher B, Sternjohn JR, Vasdewani J, Karypis G, Reilly CS, Bitterman PB, Bohjanen PR. 2008 Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1. Mol Cell. Feb 1;29(2):263-70.

Wahle E, Rüegsegger U. 1999 3'-End processing of pre-mRNA in eukaryotes. FEMS Microbiol Rev. Jun;23(3):277-95.

Wahle E. 1995. Poly(A) tail length control is caused by termination of processive synthesis. J Biol Chem. Feb 10;270(6):2800-8.

Wahle E, Martin G, Schiltz E, Keller W. 1991 Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase. EMBO J. Dec;10(13):4251-7.

Wells SE, Hillner PE, Vale RD, Sachs AB. 1998 Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell. Jul;2(1):135-40.

Whitelaw E, Proudfoot N. 1986 Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. EMBO J. Nov;5(11):2915-22.

Wickens M, Stephenson P. 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. Science. Nov 30;226(4678):1045-51.

Wickens M, Anderson P, Jackson RJ. 1997. Life and death in the cytoplasm: messages from the 3' end. Curr Opin Genet Dev.Apr;7(2):220-32.

Will C L, Luhrmann R. 1997. Protein functions in pre-mRNA splicing. Curr Opin Cell Biol. 9:320–328.

Wilusz CJ, Wormington M, Peltz SW. 2001. The cap-to-tail guide to mRNA turnover. Nat Rev Mol Cell Biol. Apr;2(4):237-46.

Wilusz JE, Spector DL. 2010 An unexpected ending: noncanonical 3' end processing mechanisms. RNA. Feb;16(2):259-66.

Yan J, Marr TG. 2005 Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. Genome Res. Mar;15(3):369-75.

Zarkower D, Wickens M. 1988. A functionally redundant downstream sequence in SV40 late pre-mRNA is required for mRNA 3'-end formation and for assembly of a precleavage complex in vitro. J Biol Chem. Apr 25;263(12):5780-8.

Zhang H, Lee JY, Tian B. 2005 Biased alternative polyadenylation in human tissues. Genome Biol 6(12):R100.

Zhang Z, Fu J, Gilmour DS. 2005 CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. Genes Dev. Jul 1;19(13):1572-80.

Zhang Z, Gilmour DS. 2006 Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Mol Cell. Jan 6;21(1):65-74.

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. Jun;63(2):405-45.

Zhao W, Manley JL 1996 Complex alternative RNA processing generates an unexpected diversity of poly(A) polymerase isoforms. Mol Cell Biol. May;16(5):2378-86.

Zheng ZM, Baker CC. 2006. Papillomavirus genome structure, expression, and post-transcriptional regulation. Front Biosci. Sep 1;11:2286-302.

CURRICULUM VITAE

Rose Marie Caratozzolo

Education

2002	B.A	Rutgers College, Rutgers University
2011	Ph.D	Rutgers, The State University of New Jersey - New Brunswick

Publications

*Guan F, ***Caratozzolo RM**, Goraczniak R, Ho ES, Gunderson SI. 2007 A bipartite U1 site represses U1A expression by synergizing with PIE to inhibit nuclear polyadenylation. RNA. Dec;13(12):2129-40.
*Acknowledged that first two authors made equal contributions.

Major Conference Posters

Caratozzolo RM, Lee R, Gunderson SI, Tallarico A, Murakami A, Noonan K, Marasco W. HIV-1 Nef Protein Induces Exon Skipping Of Grb2, A Human Signal Transduction Gene. 9th Annual Meeting of the RNA Society, 2004

Shi J, **Caratozzolo RM**, Gunderson SI, Scotto KW. Caffeine Regulates Alternative Splicing Of A Subset Of Genes. Eukaryotic mRNA Processing, Cold Spring Harbor NY, 2005

Caratozzolo RM, Shi J, Hu Z, Scotto KW, Gunderson SI. In Vitro Studies To Reconstitute Caffeine Induced Alternative Splicing. Eukaryotic mRNA Processing, Cold Spring Harbor NY, 2005

Caratozzolo RM, Gunderson SI. Determining the Molecular Basis of Synergy Between The U1 Site In The U1A 3' UTR And The Nearby PIE Site. 13th Annual Meeting of the RNA Society, 2008

Goraczniak R, **Caratozzolo RM**, Gunderson SI. Combinatorial Control of 3' End Processing. 14th Annual Meeting of the RNA Society, 2009