THE *DROSOPHILA MELANOGASTER* PROTEIN SUPPRESSOR OF SABLE [SU(S)] NEGATIVELY REGULATES TRANSPOSON-CONTAINING TRANSCRIPTS BY A MECHANISM THAT INVOLVES DIRECTLY BINDING TO ITS RNA TARGET

Lonna Finnic Mollison

A dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Genetics and Molecular Biology

Chapel Hill
2016

Approved by:

Lillie Searles

Robert Duronio

William Marzluff

Mark Peifer

Corbin Jones

ABSTRACT

Lonna Finnic Mollison: The *Drosophila melanogaster* Protein Suppressor of sable [Su(s)] Negatively Regulates Transposon-Containing Transcripts by a Mechanism that Involves Directly Binding to its RNA Target
(Under the direction of Lillie Searles)

RNA quality control systems operate at various stages of gene expression to prevent aberrant RNAs from accumulating. The nuclear pathways that lead to the identification and elimination of defective pre-mRNAs are incompletely understood, especially in multicellular organisms. The Suppressor of sable (Su(s)) protein of *D. melanogaster* plays a role in this process. Su(s) is a nuclear RNA-binding protein that negatively regulates the accumulation of RNA from genes that contain transposon insertions in the 5' transcribed region. Previous studies have shown that the Su(s)-regulatory pathway induces premature transcription termination and degradation of the resulting RNAs. Here, I present *in vitro* and *in vivo* evidence that Su(s) recognizes specific sequences in one of its biological targets. I found that a U-rich element is efficiently bound by Su(s) and a G-rich element appears to be a weaker binding site. The results of reporter gene analysis confirmed that the U-rich and G-rich elements are relevant regulatory sequences. However, a GUA-rich element that contributes significantly to this regulation is not a Su(s) binding site. These results indicate that this regulation depends on the direct binding of Su(s) and, possibly, one or more other proteins to the RNA.

To my sister Tina and my Father

ACKNOWLEDGEMENTS

I would like to thank Dr. Lillie Searles for allowing me to join her lab and for her continued commitment to my success. Lillie has been a patient, kind, and understanding mentor with a unique ability to bring out the best in me while challenging me to become more cautious and well reasoned in my scientific viewpoint. I have grown immensely from her mentorship and guidance.

I owe so much gratitude and appreciation to my family. First and foremost I want to thank, my parents, who lovingly raised eight children. They were the first example of the type of person I should strive to be. They instilled in me the attributes that are required to persevere through the rigors of the PhD journey and for this I am forever grateful. Although my father has passed, my mother continues to find the strength to encourage her adult children. She has been my biggest champion, supporter and friend, I love you Mommy. To my seven siblings, who have always protected their youngest sister, I thank you for your love and support. My sister Shona, who encouraged me to go back to school after I had gone through a layoff, thank you for being my best friend. To my older sisters Eva, Reva and Kleva, I admire your strength to raise children and pursue your own career and I thank you for your love and support over the years. To my only brother Joel, I appreciate your sense of humor and your kind nature. To my sister Tina, I miss you every day but you inspired me to reach for my goals. Finally I want to thank my nephew Trenton and my friends, for providing balance, fun, and a voice of reason.

PREFACE

The work presented in Chapter II determines that the RNA binding activity of Su(s) contributes to its regulatory activity. While I performed the majority of the experimental procedures used to investigate the Su(s)/*αβ* RNA interaction *in vitro* and *in vivo*, a research specialist in the Searles lab, Paul Brewer-Jensen, made significant contributions to the reporter gene analysis. I generated and tested the reporter constructs for Region 2 of *αβ* RNA, which contains *invader1* sequences. These constructs determined that Su(s) binding sites are important for regulation of Region 2. However, Paul performed repeat experiments to produce publication-quality Northern blot images. Additionally, all reporter constructs for Region 1, which contain *Dm88* sequences, were generated and tested by Paul. Therefore, Paul's efforts truly helped to create a more comprehensive analysis of the Su(s) mediated regulation of *αβ* RNA, and I thank him for his contributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

$\alpha\beta$	alpha/beta
APT	Associated with Pta1
ARMs	Arginine Rich Motifs
CBC	Cap-Binding Complex
CCCH	Cysteine-Cysteine-Cysteine-Histidine
cDNA	complimentary DNA
CE	Capping Enzyme
CFI	Cleavage Factor I
CLIP	Crosslinking-Immunoprecipitation
CPF	Cleavage and Polyadenylation Factor
CPSF	Cleavage and Polyadenylation Specificity Factor
CPM	Counts Per Minute
CstF	Cleavage Stimulation Factor
CTD	C-Terminal Domain
CUTS	Cryptic Unstable Transcripts
DNA	Deoxyribonucleic acid
EJC	Exon-Junction Complex
EMSA	Electrophoretic Mobility Shift Assay
eRNAs	enhancer RNAs
HIV	Human Immunodeficiency Virus
kDA	kilodaltons
LTR	Long Terminal Repeats

lRNAs	long RNAs
m^7G	7-methylguanosine cap
mRNA	messenger RNA
miRNA	microRNAs
ncRNAs	non-coding RNAs
NEXT	Nuclear-Exosome Targeting Complex
NNS	Nrd1-Nab3-Sen1
NRQC	Nuclear RNA Quality Control
NXF1	Nuclear Export Factor
PAP	Poly (A) polymerase
PAS	Poly (A) Signal
PASR	Promoter Associated small RNAs
PCIF	Polytene Chromosome Immunoflourescence
Pol II	RNA polymerase II
Poly(A)	Polyadenylic acid
Poly (G)	Polyguanylic acid
Poly (U)	Polyuridylic acid
PROMPTS	Promoter Upstream Transcripts
RBPs	RNA Binding Proteins
RIP	RNA Immunoprecipitation
RNA	Ribonucleic acid
RNMT	RNA guanine-7 methyltransferase
RNP	Ribonucleoprotein

RRE	Rev Response Element
RRM	RNA Recognition Motif
sRNAs	small RNAs
SELEX	Systematic Evolution of Ligand by Exponential Enrichment
Ser-5P	Serine-5 Phosphorylation
Ser-2P	Serine-2 Phosphorylation
snRNAs	small nuclear RNAs
snRNPs	small nuclear Ribonucleoproteins
snoRNAs	small nucleolar RNAs
Su(s)	Suppressor of *sable*
TASRs	Termination Associated small RNAs
TSS	Transcription Start Site
UTR	Untranslated Region
WT	Wild Type
ZF	Zinc Fingers

CHAPTER I: INTRODUCTION

Introduction

The life cycle of a messenger RNA (mRNA) molecule begins in the nucleus where the pre-mRNA is transcribed by RNA polymerase II (Pol II) and co-transcriptionally processed. After being exported to the cytoplasm the mRNA undergoes multiple rounds of translation and is ultimately degraded. Each step in gene expression depends on the proper association of RNA binding proteins (RBPs) that facilitate these processes (1). RBPs act to modify, protect, and survey the RNA quality throughout the life cycle, thus ensuring the informational integrity of the transcriptome.

Transcription by Pol II

Pol II transcribes protein-coding genes to generate mRNAs as well as a subset of non-coding RNAs (ncRNAs), including small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs) and micro RNAs (miRNAs). Pol II is unique among polymerases in that the largest subunit contains a conserved, C-terminal domain (CTD) that consists of a hepta-peptide repeat (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) (2). The CTD exists as an extended structure that undergoes extensive post-translational modifications that are essential for the proper progression through the phases of transcription and for the recruitment of protein factors involved in RNA processing events. The post-translational modification of the CTD that has most extensively been studied is phosphorylation (3).

Pol II is recruited to the promoters of genes when the CTD is in an unphosphorylated state. Hyper-phosphorylation of the CTD at Serine 5 (Ser-5P) of the heptad repeat occurs during initiation and is thought to facilitate release of Pol II from the promoter (4). In multicellular organisms, the early elongation Pol II complex often pauses a short distance downstream of the transcription start site (TSS). Release of Pol II from this paused state is facilitated by phosphorylation of Ser-2 (Ser-2P). As Pol II escapes into the body of the gene Ser-5P is removed and Ser-2P increases. Thus, phosphorylation levels peak for Ser-5 at the 5' region of genes and at the 3' region for Ser-2, with intermediate levels of both Ser-5P and Ser2P within the body of the gene (4). Therefore this CTD code is informative of the phase of transcription of Pol II and the protein factors that may be associated with the CTD and the transcription machinery during those phases.

Co-transcriptional RNA Modifications

As mentioned above, the phosphorylation state of the CTD orchestrates the dynamic, co-transcriptional association and dissociation of proteins required to process the nascent pre-mRNA molecule to generate mRNA. During pre-mRNA maturation, the 5' end is capped, introns are removed and exon sequences are ligated together and the 3' end is generated by cleavage of the RNA followed by the addition of a protective poly adenosine (poly(A)) tail (5).

The process of modifying a nascent RNA molecule begins when the first 20-30 nucleotides have been synthesized and the growing RNA emerges from the Pol II exit channel. At this stage the CTD is hyperphosphorylated at Ser-5, which facilitates recruitment and activation of the capping enzyme (CE) and the RNA (guanine-7) methyltransferase (RNMT). Together, these enzymes generate the 7-methylguanoisine

(m^7G) cap that serves to protect the RNA from the actions of 5' to 3' exonucleases (6). The completed cap structure is cotranscriptionally bound by the cap binding complex (CBC) which aids in protecting the RNA and in the transition to productive elongation by interactions with the CTD (7).

The process of splicing, which removes introns and fuses exons together, is performed by the large multi-subunit ribonucleoprotein (RNP) complex called the spliceosome. The spliceosome is comprised of five subunits called small nuclear RNPs (snRNPs) and many protein cofactors. The RNA components (snRNAs) are non-coding and non-polyadenylated (8). snRNAs assemble with seven Sm proteins in the cytoplasm to form the various snRNP molecules. Upon reentry into the nucleus, the holo-spliceosome is formed by the consecutive association of snRNPs with active areas of transcription coordinated by interactions with the CTD (9). In particular both U1snRNP and a component of the U2 associated factors (U2AF) are believed to directly associate with the CTD (8,9). Splicing proceeds in a stepwise manner. U1snRNP recognizes the 5' splice site and CBC and SR proteins stabilize this interaction (8). U2snRNP then binds to the 3' splice site that is further defined by U2AF and SR1. Recognition of the adenosine-branch point by U2snRNP and interaction with U1snRNP define the intron and recruits U4, U5 and U6snRNP to complete the spliceosome (8). Splicing then proceeds in a two-step reaction with the initial cleavage at the 5'splice site followed by cleavage of the 3' splice site and fusion of the exons (8).

Two models have been proposed to explain how splicing might be coupled to transcription, and there is experimental support for both models (9). In the recruitment model the CTD and its phosphorylation state serves as a landing pad for the splicing

machinery and associated factors thereby facilitating splicing. The observation that intron-containing RNAs transcribed by a bacteriophage RNA polymerase are not efficiently spliced is consistent with this model and indicates that the CTD of Pol II is necessary to recruit splicing components (10). Also, antibodies directed against the CTD inhibit splicing by preventing the association of splicing components (9,10). The kinetic model states that the rate of transcription affects splicing efficiency. This is best demonstrated by alternative splicing in which certain exons are skipped or included in the mRNA. A fast Pol II elongation rate favors skipping an alternate exon, whereas a slower rate favors the inclusion of the alternate exon (11).

The final step in RNA processing involves generating a 3' end by an endonucleolytic cleavage event and the addition of a poly (A) tail. Furthermore, RNA Pol II must terminate, which requires the release of Pol II from the DNA template. The proper formation of a 3' end involves several multi-protein complexes that cooperate to recognize pre-mRNA *cis*-elements that serve to position the cleavage machinery (12). Positioning of the cleavage machinery involves a tripartite mechanism: first the recognition of the conserved (AAUAAA) poly(A) signal (PAS) located 10 to 30 nts upstream of the cleavage site (cleavage usually follows a CA dinucleotide). Secondly two downstream elements (GU-rich and U-rich) are found about 30 nt downstream of the cleavage site. Third, UGUA sequences are found 40 to 100 nts upstream of the cleavage site for several genes (13). The cleavage and polyadenylation specificity factor (CPSF) complex recognizes the PAS and a subunit of CPSF (CPSF-73) performs the endonucleolytic cleavage. The downstream elements are recognized by a component of the cleavage stimulation factor (CstF) and the upstream UGUA is recognized by a subunit of cleavage factor I (CFI). Finally, the poly(A) tail is added

by poly(A) polymerase (PAP) and its processivity is enhanced by both PABNI and CPSF, which causes the rapid addition of 200 to 300 adenosines (12). Polyadenylation is coupled to transcription as Pol II transcribes the PAS before the 3' end processing steps can occur. Interestingly, transcription of the PAS enhances processing of the terminal intron, and processing the terminal intron facilitates cleavage and poly (A) via a direct interaction of U1 and U2snRNP and CPSF (14).

Following cleavage and poly (A) synthesis, Pol II continues to transcribe, and termination, the release of Pol II from the DNA template, can occur proximal to the cleavage site or several kilobases downstream (15). There are two prevailing ideas about how termination occurs. The allosteric model suggests that conformational changes occur following the recognition of the PAS that reduces the stability of the elongation complex. Secondly, the torpedo model postulates that the 3' cleavage event exposes an unprotected 5' end of the downstream RNA; this allows entry for the 5' to 3' exonuclease Xrn2/Rat1 to degrade the RNA and catch up to Pol II, causing its release (16,17). However, it is likely that both models contribute to efficient termination. For instance, the CTD-binding cleavage factor protein, Pcf11, was shown to disassociate Pol II from the DNA template by bridging the CTD to the RNA (3). Yet, Xrn2/Rat1 has been shown to be required for efficient termination in yeast and human cells. Both Pcf11 and the Rat1 associated protein, Rtt103 interact with the Ser-2P CTD (3). Thus, allosteric rearrangements of the Pol II CTD and associated factors and degradation of the RNA generated following cleavage are both important in termination. Thus, the CTD is integral in the seamless coupling of every step in transcription and RNA processing.

Cotranscriptional Nuclear RNA Quality Control

Superimposed on the complex yet elegant coupling of transcription and RNA processing, is the need to survey the growing transcript via cotranscriptional quality control mechanisms. Nuclear RNA quality control (NRQC) mechanisms ensure rapid degradation of transcripts that are not properly processed, contain aberrant sequences, or improperly integrated into an RNP complex. This prevents the accumulation of aberrant transcripts which are potentially harmful to the cell (18). To deal with aberrant transcripts most NRQC components act at the site of transcription, while the RNA is still associated with the DNA template. NRQC employs three different methods to address faulty transcripts: degradation, nuclear retention, or down-regulation of the expressed gene (18).

The 5' m^7G cap is the first RNA processing step that subject to surveillance by NRQC mechanisms. RNA that is uncapped or improperly capped will fail to recruit CBC and this renders the transcript susceptible to the 5'-3' exonuclease activity of XRN2/Rat1 (7). Additionally, yeast Rat1 protein associates with the decapping enzymes Dox1 and Rai1 that can facilitate the removal of methylated or unmethylated caps and the hydrolysis of uncapped triphosphate ends to monophosphate, respectively, thereby providing a suitable substrate for Rat1 (19). Together these enzymes enforce a NRQC checkpoint that ensures the 5' ends of transcripts are properly capped or rapidly degraded.

Improper splicing potentially introduces errors in the sequence and structure of the mRNA, and thus surveillance of this step of RNA processing is also important. Splicing defects can cause the transcript to be retained at the site of transcription. The association of proteins required for splicing, such as SR proteins, with the nascent transcript is essential for the recruitment of factors involved in RNA export to the cytoplasm including

Mex69p (20). Also, in mammals, the cotranscriptionally deposited exon-junction complex (EJC) mediates the recruitment of the nuclear export factor, NXF1, to the mRNA, suggesting that proteins involved in RNA processing also exert NRQC as the lack of their proper association impedes downstream activities including nuclear export (21). Improperly spliced transcripts also undergo degradation by XRN2/Rat1 and the 3'-5' exonuclease, the nuclear exosome (22). The nuclear exosome is comprised of a nine-subunit core that lacks catalytic activity. The catalytic activity of the exosome is provided by the association of two RNases, RRP6, a 3'-5' exonuclease, and RRP44/DIS3, which has both endonuclease and 3'-5' exonuclease activity (23). In cells deficient in either XRN2 or the exosome, aberrantly-spliced transcripts accumulate, indicating these nucleases are fundamental components of NRQC (22).

Processing of the 3' end of pre-mRNA is another step in which NRQC mechanisms are required. One example is that failure to splice the terminal intron, which requires the exon-defining complex and CPSF, causes retention of the transcript at the gene, as only fully spliced transcripts with a poly(A) tail are released by the cleavage and polyadenylation machinery (24). Furthermore, in yeast, mutation of the poly(A) polymerase caused rapid degradation of mRNA in an RRP6 dependent manner (25). These findings demonstrate that the canonical 3' processing factors impose NRQC by tethering the transcript to the gene until a proper poly(A) tail is formed or signal for the rapid degradation via a mechanism involving the nuclear exosome component RRP6.

Cotranscriptional Nuclear RNA Quality Control of Pervasive Transcription

Initially, Pol II transcription was thought to be limited to functional coding regions. However, advances in high throughput sequencing technologies that annotated the entire

transcriptome led to the discovery that transcription is pervasive, meaning the majority of the genome, from yeast to humans, is transcribed at some level (26). Several different classes of pervasive non-coding transcripts have been identified, and these have arbitrarily been placed into two categories based on their size; small RNAs (sRNAs) are 20 nts to 200 nts in length and long RNAs (lRNAs) range from 200 nts to over 1000 nts (27). Many of the small RNAs cluster around promoters (promoter associated sRNAs (PASR)) or around the 3' end (termination-associated sRNAs (TASRs)) of genes. Interestingly, PASRs are found in both the same and divergent orientation to the promoter from which they originate and their abundance is directly proportional to the strength of the promoter. In yeast, mutation of catalytically active components of the exosome or its cofactor, the TRAMP complex, facilitated the identification of RNAs that are highly unstable and undetectable in WT cells. These RNAs were named cryptic unstable transcripts (CUTs) (28). CUTs, at 200 nts to 600 nts, are larger than PASRs, and are predominantly antisense to the promoter region with heterogenous 3' ends. The 5' ends of CUTs originate upstream of the TSS from nucleosome-depleted regions that can broadly define promoter regions. This indicates that promoters are inherently bidirectional (29). Like yeast, human cells produce unstable promoter upstream transcripts (PROMPTS) that also map to nucleosome-depleted regions. PROMPTS are transcribed from both stands but are generally antisense to the downstream promoter (30). Another type of long non-coding RNA are enhancer RNAs (eRNAs), which are generated from the transcription of enhancer elements (31). Similar to PROMPTS, eRNAs are bidirectional yet enriched for antisense transcripts (32) and rapidly degraded by the exosome (31).

The identification of protein factors involved in down regulation of pervasive transcripts has been most extensively examined in yeast. The Nrd1-Nab3-Sen1 (NNS) complex regulates CUTs. Both Nrd1 and Nab3 bind to short tetramer sequences in the nascent RNA to direct termination (33). The NNS pathway interacts with the Ser-5P form of the CTD and thus this termination pathway is limited to the region proximal to the TSS (31). Therefore a factor in the fate of CUTs is the presence of *cis*-termination sequences relative to the TSS. Sites bound by NNS are enriched in the TSS-proximal region in antisense transcripts and yet they are depleted from protein-coding mRNA transcripts (31). Thus the direct association of Nrd1 with the Ser-5P CTD of Pol II and the association of the NNS complex with the exosome, couples the transcription of CUTs to their termination and exosome mediated rapid degradation (34).

In humans, the position of transcript termination sequences in the promoter-proximal region is also involved in the regulation of antisense transcripts such as PROMPTs and eRNAs. The presence of a canonical PAS (AAUAAA) in the 5' region of the transcript invokes early termination coupled to rapid degradation. The CBC appears to interact with the nuclear-exosome-targeting (NEXT) complex to facilitate exosome-mediated degradation (35). Similar to NNS binding sites, cryptic PAS are enriched in antisense transcripts whereas protein-coding transcripts are depleted of cryptic PAS (36). Additionally, binding by U1 snRNP protects the nascent RNA by preventing the recognition of PAS that could otherwise induce early termination and thus aids in determining promoter directionality (37). Recently a role for the WD40 repeat protein, Wdr82, in preventing the readthrough of eRNAs derived from intergenic regions was defined (38).

Other proteins may also be involved in the regulation of pervasive and aberrant transcripts in metazoans.

The *Drosophila melanogaster* Protein Suppressor of Sable Negatively Regulates Aberrant Transcripts

The *D. melanogaster* protein Suppressor of sable (Su(s)) is a 140 kDa RNA binding protein that functions in NRQC. Su(s) was initially identified as a recessive sex-linked locus in which mutations suppressed spontaneous mutant alleles at a second site. Particularly, mutations at *su(s)* suppress spontaneous mutations associated with the *sable*, *purple*, *speck* and *vermilion* genes. It was determined that the spontaneous mutations at *vermilion*, *purple* and *speck* were caused by insertions of the *412* retrotransposon (39). Further analysis of the suppressible *412* induced mutations at the *vermilion* (v^1, v^2, v^k) locus revealed that *412* inserted within the 5' UTR in the opposite orientation to the *vermilion* gene (40) and that *412* sequences are removed from the primary transcript through the use of cryptic splice sites located in the long terminal repeats (LTRs) (41,42). In a WT *su(s)* background mature mutant *vermilion* mRNAs fail to accumulate, yet in a mutant *su(s)* background mature *vermilion* mRNAs accumulate and are similar in size as the WT *vermilion* despite having a 7.5kb *412* insertion (40). Intriguingly, the full *412* sequence was not required for *su(s)* mediated regulation, as a single, 480-bp LTR from *412* was sufficient (43). In a *su(s)* mutant background the single LTR also resulted in higher accumulation of *vermilion* transcripts that both retained or spliced out the single LTR (43).

In addition to *412*, insertion of a P-element in the *yellow* gene is also suppressible by mutations at the *su(s)* locus (44). Similar to *412*, the suppressible P-element inserts in an opposite orientation to the *yellow* gene and transcripts fail to accumulate in a WT *su(s)*

background. In a mutant *su(s)* background two transcripts accumulate, one containing the P-element insertion and one in which the P-element is removed from the primary transcript via recognition of cryptic donor splice sites in the 39bp-inverted repeat of the P-element and acceptor sites within the *yellow* transcript (44). These findings led to the hypothesis that the *su(s)* locus may encode a protein that is involved in preventing the recognition of cryptic splice sites, thus causing destabilization of the insertion containing transcripts (44). However, analysis of the molecular genetic interactions between *su(s)* and the suppressible alleles provided limited information about the role of the Su(s) protein product. Thus, additional targets of this regulation needed to be identified. Furthermore, molecular characterization of the *su(s)* gene and resultant protein were critical.

Targets Identified by PCIF Analysis

Polytene chromosome immunofluorescence (PCIF) studies demonstrated that Su(s) protein localized to distinct chromosomal regions (45,46). Further analysis revealed that Su(s) colocalized with the Ser5-P form of Pol II, present during initiation and the early elongation phase, but not the Ser2-P form of Pol II, which is present during active elongation, and the 3' region of genes (47). A strong signal for Su(s) was observed at the 3C locus of the X-chromosome. This region contains salivary specific intronless genes (*ng 1-4*, *Sgs4*, and *Pig1*) that lie within the introns of a larger gene. The normal expression pattern of these genes is known with *ng* genes being expressed earlier in third instar larvae. When *ng* mRNA levels decline, *Sgs4* is rapidly induced, with no overlap in the expression. However, in a mutant *su(s)* background, *ng* mRNA levels remained elevated while *Sgs4* was expressed (47). Also, *Sgs4* mRNA levels accumulated much earlier, while *ng* mRNA levels were still detectable, and maximal mRNA *Sgs4* expression was achieved at a slower rate

when compared to WT *su(s)* background. This data suggest that Su(s) plays a role in maintaining the normal temporal expression pattern for these developmental genes and the rapid induction of *Sgs4* following the decline of *ng* RNA levels (47).

PCIF analysis of the distribution of Su(s) during heat-shock revealed that Su(s) protein localized to 87C. There are four heat-shock genes at the 87C locus. Between two of the genes there is a 38 kb repetitive insert of the remnants of the retrotransposon elements, *invader1* and *Dm88*. These elements were named *alpha*/*beta* (*αβ*) elements. A subset of these elements are fused to a duplication of the *HSP70* promoter (gamma element), and these hybrid retrotransposons are referred to as *alpha-gamma* elements (47). Under mild heat-shock conditions, noncoding, polyadenylated transcripts of various lengths arise from the *αβ* RNA region (48). Northern analysis revealed that *αβ* transcripts, similar to the *vermilion* transcripts, accumulate at significantly higher levels in a mutant *su(s)* background indicating that Su(s) also regulates these aberrant transposon-containing transcripts. When Su(s) is active Hsp70-*αβ* elements generate short, unstable transcripts with heterogeneous 3' ends. Conversely, when Su(s) is inactive, longer, stable transcripts are produced that terminate following a canonical PAS (49). Unlike the antisense *412* insertion in *vermilion, purple* and *speck*, the *αβ* elements are in a sense orientation with respect to the *Hsp70* promoter (47). Analysis of the sequences required for the Su(s)-mediated negative regulation of *αβ* transcripts revealed that sequences spanning nucleotides +1 to +278 are sufficient for regulation, and *Hsp70* promoter and 5' UTR sequences (+1 to +69) do not contribute to regulation (47).

The Su(s) Protein

The *su(s)* locus is located between the cytological regions of 1B10 – C1 of the X-chromosome (50). A fortunate P-element-induced mutation in a mutant *vermilion* strain resulted in a revertant to a WT *vermilion* phenotype. The *su(s)* locus was the only known suppressor of *vermilion*, and thus the P-element insertion into *su(s)* facilitated cloning of the *su(s)* gene via transposon-tagging, leading to the identification of a 5kb transcript derived from this gene (50). Subsequent temporal analysis revealed that the *su(s)* mRNA is produced throughout the life cycle of the fly and the resultant 150 kDa protein localized to the nucleus (51). Initial examination of the protein sequence revealed that Su(s) contained a highly charged region with 28.6% sequence identity to the arginine-serine region of the Drosophila U1 70K protein that is involved in splicing and a putative C-terminal region with low similarity to a RNA recognition motif (RRM) (51). However, more extensive analysis of this putative RRM showed that this region did not fit the criteria for a true RRM and the arginine rich region of Su(s) lacked the arginine-serine repeats present in U1 70K. Thus, while Su(s) contained regions that were loosely reminiscent of proteins involved in RNA processing, the function of Su(s) had yet to be determined (52).

Su(s) is an RNA Binding Protein

To examine the biochemical functions of Su(s), our lab generated recombinant full length and truncated Su(s) protein. Full length Su(s) bound to *ftz* RNA with a high affinity and this binding could be out-competed by excess poly(U) or poly(G) (52). Furthermore, systematic evolution of ligand by exponential enrichment, SELEX, experiments identified an enriched consensus sequence, UCAGUAGUCU, which was present in many of the high affinity SELEXs RNAs. Other high affinity RNAs that lacked the SELEX consensus were

identified that were enriched for GU-dinucleotides (52). To delineate the region of Su(s) responsible for the RNA binding activity the 1325 amino acids that comprise the full length Su(s) were divided into four sections and high affinity binding was localized the N-terminal 434 amino acids (52). Immunohistochemical analysis of the endogenous Su(s) distribution pattern in embryos and the nuclei of salivary glands demonstrated that Su(s) localized to the nucleoplasm and to multiple locations on polytene chromosomes. Interestingly, in some regions, Su(s) co-localized with U1 70K, suggesting Su(s) is found at a subset of areas undergoing active transcription where pre-mRNAs are being processed (52).

Su(s) has Two Distinct RNA Binding Domains

High affinity RNA binding displayed by Su(s) had been localized to the N-terminal region (52). Sequence alignment analysis determined that this region contained two CCCH zinc-finger motifs (ZFs) (53). The CCCH-ZF motif is found in other RNA interacting proteins such as the mRNA destabilizing protein, TTP (Tis11) (54,55), the 3' end pre-mRNA processing factor subunit CPSF-30/CLIPPER (56), and the splicing factor subunit U2AF-35 (57). TTP binds to AU-rich (UUAUUUAUU) elements in the 3' UTR to destabilize the targeted mRNA (54) while U2AF-35 binds to the AG-dinucleotide at the 3'SS (57). Furthermore, *in vitro*, endonuclease activity was localized to the five CCCH-ZFs of *Drosophila* CPSF-30/CLIPPER which were required to bind and cleave RNA hairpins (58); however, the biological relevance of this activity has not been established. In addition to their RNA binding activity, the CCCH-ZFs of TTP also serve as a protein-protein interaction domain and facilitate the association with the nuclear-pore protein Nup214 (59). Based its CCCH-ZF motifs, Su(s) has been placed in an orthologous group with CPSF-30 (Yth1 in yeast) and the human protein ZC3H4 (C19orf7 or KIAA1064) (49).

Additional *in vitro* RNA binding analysis of various fragments of the N-terminal region of Su(s) localized high affinity binding to two arginine-rich motifs (ARMs) (45). ARMs have been most studied in proteins derived from viruses such as the HIV Rev and Tat proteins and the N-proteins of bacteriophages. The ARM of Rev is essential for binding to the highly structured Rev Response Element (RRE) located in the intron of the *Env* gene, and this allows incompletely spliced viral transcripts to exit the nucleus for translation (60). Tat is a trans-activator of transcription initiated in the LTR of HIV provirus. Tat binds to the *trans*-activator response (TAR) element, also a structured RNA hairpin, in the 5' UTR, to increase the rate of transcription (61). However, ARMs also recognize unstructured RNAs. For instance, the CPSF component Fip1, binds to U-rich sequences via an ARM to aid in enhancing the activity of PAP (62). Interestingly, some of the SELEX consensus RNAs bound by Su(s) were predicted to form hairpin structures, whereas SELEX RNAs that lacked the consensus were less likely to form structures (45). Furthermore, ARM1 of Su(s) was found to preferentially bind to RNAs containing the consensus, while ARM2 bound non-consensus RNAs with a higher affinity. Loss of both ARMs eliminated binding to SELEX RNAs, even though the CCCH-ZFs were present, demonstrating that, *in vitro*, the ZFs do not contribute to RNA binding (45).

The ARMs and the ZFs Contribute to Su(s) Mediated Regulation *in vivo*

The ARMs had been identified as the domain that confers high affinity binding *in vitro* and thus our lab wanted to determine if the loss of this RNA binding domain would affect Su(s) mediated regulation *in vivo*. Initial experiments using WT or ARM deletion *su(s)* cDNA transgenes that were ectopically expressed at high levels demonstrated that ubiquitous overexpression of Su(s) is embryonic lethal (45). Analysis of moderately

expressed *su(s)* transgenes in a *su(s)* null background revealed that loss of ARM1, but not ARM2, eliminated localization to polytene chromosomes under non-heat-shock conditions but did not affect nuclear localization (45). Interestingly, deletion of both ARMs increased mutant Su(s) protein and mRNA levels, suggesting that the ARMs are involved in regulating expression from these transgenes (45). Analysis of the effect of deleting of the ARMs or the ZFs on regulation of the suppressible full *412*- insertion-containing *vermilion* transcript, showed that regulation was lost only when ARM1 and ARM2 were deleted together while destabilizing ZF mutations minimally affected Su(s) mediated regulation (46). However loss of the ARMs or mutation of the ZFs greatly inhibited the ability of Su(s) to autoregulate *su(s)* genomic transgenes, expressed under the control of the endogenous promoter (46). Furthermore, loss of the ARMs or the ZFs reduced or eliminated, respectively, the ability of Su(s) to down-regulate *αβ* transcripts. The enhanced regulatory role of the ZFs with the *αβ* transcripts rather than the ARMs might be expected as mutation of the ZFs resulted in loss of localization to the 87C heat-shock locus (47). These studies suggest that different targets of Su(s) may interact with different domains by promoting various modes of binding, and, potentially different protein complexes may form in a target-dependent manner (46).

Identification of Components of the Su(s) Regulatory Pathway

Our lab determined that the nuclear exosome is a component of the Su(s) regulatory pathway. Northern analysis demonstrated that degradation of *αβ* transcripts depends on the nuclear exosome, as loss of the enzymatically active components partially restores *αβ* RNA accumulation (47).

Su(s) Interacts with Wdr82 to Regulate *Hsp70-αβ* Transcripts

Recently our lab reported that Su(s) interacts with Wdr82 and both of these proteins are required for the negative regulation of *Hsp70-αβ* transcripts (49). Wdr82 is a 35-kDa protein, containing seven WD40 repeats which typically serve as a β-propeller like platform to which proteins can bind. Wdr82 is conserved from yeast to humans. The yeast homologue, Swd2, has been shown to function in two distinct complexes; the SET1/COMPASS complex, responsible for mediating histone 3 lysine 4 (H3K4) trimethylation, and the APT complex, part of the larger holo-enzyme, cleavage and polyadenylation factor (CPF) complex, which processes the 3' end of RNA Pol II transcripts (63). 3' end polyadenylation processing by CPF does not seem to require APT(64). However, loss of Swd2 in yeast leads to improper transcription termination of some genes, particularly snoRNAs, which are not polyadenylated. Also, loss of Swd2 results in reduced global H3-K4 methylation. In human cells, Wdr82 interacts with the C-terminal domain (CTD) of RNA Pol II, which then allows recruitment of Setd1A to the promoter of genes by an interaction between Wdr82 and the RNA recognition motif (RRM) of Setd1A (65). Intriguingly, in humans cells, Wdr82 interacts with the putative human ortholog of Su(s), ZC3H4 (C19orf7, KIAA1099) (49,66). Perhaps the Su(s)-Wdr82 regulatory pathway is conserved in higher organisms, however this will require further investigation.

We demonstrated that the Su(s)-Wdr82 pathway regulates *Hsp70-αβ* transcripts via a mechanism that involves induced promoter-proximal termination and non-canonical polyadenylation (49). Furthermore, Su(s)-Wdr82 regulation depends on *cis*-sequence elements within the LTRs of *Dm88* and *invader1* that are proximal to the *Hsp70* promoter (47,49). One possible model of Su(s)-Wdr82 regulation is that Wdr82 recruits Su(s) to the

5’ region of some actively transcribed genes, explaining why Su(s) colocalized with Ser-5P Pol II (47), and thus restricting the Su(s) regulation to the promoter proximal region. Su(s) may then bind to RNA sequences in the promoter-proximal region to induce premature termination, similar to NNS pathway and the regulation of CUTS (31). However we had not previously determined if Su(s) interacts directly with *Hsp70-αβ* RNA transcripts nor had we defined biologically relevant binding sites. In the findings presented here, I demonstrate that Su(s) binds *αβ* transcripts *in vitro* and in living S2 cells. I also define sequences that are bound by Su(s) and determine that Su(s) binding sites contribute to regulation of *αβ* RNAs.

CHAPTER II: THE INHIBITION OF ABERRANT RNA ACCUMULATION BY THE SUPPRESSOR OF SABLE (SU(S)) PATHWAY DEPENDS ON THE RNA-BINDING ACTIVITY OF SU(S) AND MULTIPLE, DISTINCT REGULATORY ELEMENTS

Introduction

The maturation of eukaryotic pre-mRNAs involves several processing reactions (capping, splicing, and polyadenylation) that are intricately coupled to RNA polymerase II (Pol II) transcription. When each of these reactions is efficient, the resulting mRNAs and associated-proteins (mRNPs) are exported to the cytoplasm for translation. However, if any step in this process is impaired, e.g., because of mutations that affect key regulatory sequences, the aberrant transcripts are detected and eliminated by one of several RNA quality control (RQC) systems (18,67). Nuclear RQC (NRQC) often involves co-transcriptional degradation of defective pre-mRNAs by the 5' → 3'exonuclease XRN2/Rat1 (XRN2) or the nuclear exosome, a multi-subunit complex that includes 3' → 5' exonucleases. XRN2 has been implicated in handling improperly capped pre-mRNAs and stalled Pol II complexes, whereas nuclear exosome components degrade nascent transcripts with exposed 3' ends. The latter situation occurs when transcription elongation or 3'-processing is impaired (68,69). Although considerable progress has been made in elucidating NRQC mechanisms, much remains to be learned, especially in multicellular organisms, about how defective pre-mRNAs are identified and the regulatory proteins that participate in this process (70).

The Su(s) regulatory pathway in *D. melanogaster* functions in NRQC. Su(s) downregulates RNAs from transcription units with aberrant sequences in the 5'-transcribed region by a mechanism that involves early transcription termination and degradation of the RNAs by nuclear exosome components (47,49). Our lab recently identified the WD40 domain protein Wdr82 (yeast Swd2) as a component of this regulatory pathway (49). Wdr82 is known to function in two different promoter-proximal gene regulatory processes, i.e., regulation of histone H3 lysine 4 methylation at active genes and transcription termination at yeast genes that produce short, noncoding RNAs (64,65). Recently, an additional role for mammalian Wdr82 in promoting early transcription termination of intergenic transcripts and enhancer RNAs has been reported (38).

Several transcription units that are subject to regulation by the Drosophila Su(s) pathway have transposon sequences inserted a short distance downstream of the promoters of protein-coding genes (39). The best-studied example of this occurs within a genomic region that contains a cluster of *Hsp70* protein-coding genes and noncoding *Hsp70-αβ* elements. These elements consist of remnants of two different retrotransposons (*Dm88* and *invader1*) inserted downstream of *Hsp70* promoter/5' UTR sequences (Figure 1). During a mild heat shock induction, Su(s)-Wdr82 act to inhibit *αβ* RNA accumulation, but does not affect *Hsp70* mRNAs (47). Under these conditions, *Hsp70-αβ* elements produce short (~100 to 400 nt), unstable RNAs with heterogeneous 3' ends (Figure 1). Although the details of this regulatory process have not been fully sorted out, it appears that Su(s)-Wd82 induces transcription termination in the promoter-proximal region of *Hsp70-αβ* and the unstable RNAs are degraded by nuclear exosome components (47,49). Furthermore, the unstable *αβ* RNAs are apparently polyadenylated before being degraded,

and this process is independent of the canonical polyadenylation signal (AAUAAA) (49). In contrast, Pol II does not terminate in the promoter-proximal region when Su(s)-Wdr82 is inactive. Instead, longer and more stable *αβ* RNAs, with discrete, polyadenylated 3'-ends derived from canonical PAS, are generated under these conditions (Figure 1) (49).

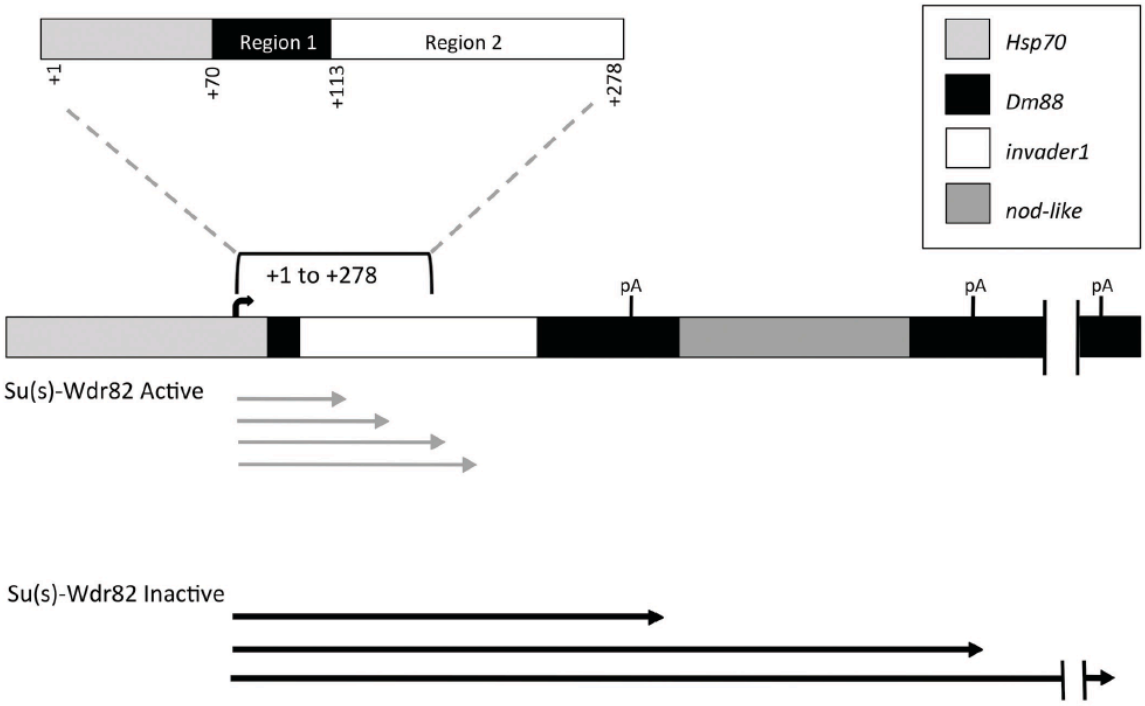


Figure 1. Schematic map of the structure and transcripts of *Hsp70-αβ* elements. *Hsp70-αβ* elements are comprised of an *Hsp70* promoter/5'UTR fragment followed by portions of the retrotransposons *Dm88* and *invader1* and an anti-sense fragment of the protein-coding *nod* gene. Sequences in the transcribed region between +70 and +278, which consists of *Dm88* and *invader1* LTR sequences, are sufficient for Su(s)-dependent downregulation. Two sub-regions within this fragment (Region 1 & Region 2) independently contribute to this regulation (49). Under heat-shock conditions the *Hsp70* promoter drives expression of *αβ* RNAs. When the Su(s)-Wdr82 regulatory pathway is active, short, unstable *αβ* RNAs with heterogeneous 3' ends are produced (grey arrows). Some of these RNAs are polyadenylated even though canonical poly(A) signals are not found in this region (49). When the Su(s)-Wdr82 pathway is inactive, several longer, stable, *Hsp70-αβ* RNAs are produced (black arrows). These RNAs end, as expected, at canonical poly (A) signals (pA).

Su(s)-dependent regulation of *αβ* RNAs depends on transposon sequences in the promoter-proximal region of *Hsp70-αβ* (+70 to +278) (47,49). Reporter gene analysis delineated two regions that contribute to this effect—a 42-nt segment of the *Dm88* long

terminal repeat (LTR) (Region 1) and a 123-nt *invader1* LTR segment (Region 2) (Figure 1). At least two elements appear to reside in Region 2 (49). Thus, it appears that multiple sequence elements mediate this regulation, presumably, by inducing transcription termination and degradation of these RNAs. The main goals of this study were to determine if Su(s) binds directly to *αβ* RNA sequences in this region and, if so, to determine if these binding sites contribute to Su(s)-mediated negative regulation.

Su(s), is a 140 kD nuclear protein that is predicted to be intrinsically disordered (49) and contains two distinct types of RNA-binding motifs (Figure 2A). Two arginine-rich motifs (ARMs) mediate the *in vitro* RNA-binding activity of Su(s) (45). In addition, two tandem CCCH zinc finger (ZF) motifs are predicted to bind to RNA, although they do not appear to mediate stable RNA binding *in vitro* (45,52). Both types of domains contribute to the regulatory activity of Su(s) *in vivo* (46,47). Whereas the RNA-binding activity of Su(s) was examined previously using substrates that are not known to be biologically relevant targets, here I have investigated the binding of Su(s) to *αβ* RNA *in vitro* and in living cells. For the first time I demonstrate that Su(s) directly binds to a relevant RNA target. Furthermore, I used a reporter gene assay to examine whether the sequences identified as binding sites are important for downregulation of *αβ* RNA. My results indicate that Su(s) recognizes sequences within Region 2 (*invader1* LTR). It binds a U-rich element within this region with a high affinity and a G-rich element with a lower affinity, and both of these elements contribute to downregulation of transposon-containing pre-mRNAs. However, a GUA-rich element in Region 1 that contributes to this regulation is not a Su(s) binding site. These findings indicate that *cis*-regulatory sequences that mediate regulation by the Su(s) pathway include binding sites for Su(s) and other regulatory proteins.

Materials and Methods

Generation and purification of MBP-Su(s) fusion proteins

The plasmid clones that were used to express MBP-Su(s) fusion proteins were generated from existing WT or mutant Su(s) cDNA clones (46). The ZF mutant construct contains three missense mutations (Cys350Gly, Cys374Gly, and His378Gly), which affect both zinc fingers. The ARM mutant construct has deletions that remove both ARMs, i.e., sequences encoding amino acids 151 to 168 and 269 to 294 (Figure 2B). The GATEWAY cloning system was used to transfer PCR-amplified fragments encoding the N-terminal region corresponding to the first 434 amino acids of WT Su(s) into the MBP-pKM596 Destination vector. In the resulting plasmid, the Su(s) coding region is positioned downstream of MBP sequences. The WT and mutant MBP-Su(s) plasmids were transformed into in a protease- deficient *E. coli* strain (ER2508). Subsequently, MBP-Su(s) fusion proteins were purified from 300 mL cultures as described in the pMAL™ Protein Fusion and Purification System (New England BioLabs) manual. Purified samples were aliquoted and stored at -70°C.

RNA binding experiments

Radioactively-labeled RNAs were synthesized by *in vitro* transcription of PCR fragments that contained the bacteriophage T7 promoter upstream of various *αβ* regions (Table A1-1). The *in vitro* transcription reactions were performed as described previously (45) with PCR amplified DNA fragments and 20 μCi of α-^{32}P-labeled CTP (800 Ci/mmol). The labeled RNAs were purified on polyacrylamide gels prior to being used in EMSA analysis. Nitrocellulose filter binding assays were performed as described previously (52). The reactions contained about 25,000 cpm of labeled RNA (0.75 nM) and various amounts

of MBP-Su(s), ranging from 1.5 nM to 100 nM in 100 µl final volume. EMSA analysis was performed as described by Ascano et al.(71). Each RNA binding reaction contained 5,000 cpm (0.75 nM) of labeled RNA and 2 nM to 2 µM protein in 20 µl final volume. Samples were analyzed on native polyacrylamide gradient gels (6%-10%).

Analysis of Su(s)-RNA interactions in cultured cells

An existing S2 cell line that had been stably transformed with an inducible FLAG-tagged Su(s) expression plasmid (Su(s)-3xF) (49) was used in these experiments. These cells were grown to 80% confluence in one to five, 15-cm cell culture plates at 22 °C, and then $CuS0_4$ was added to a final concentration of 70 µM to induce the *Mtn* promoter, which drives expression of Su(s)-3xF. After an overnight incubation, the cells were heat shocked at 37°C for 20 minutes in an air incubator. Then the cells were chilled on ice and washed with 10 ml of ice cold phosphate buffered saline (PBS). In the RNA-immunoprecipitation (RIP) experiments, the cells were harvested by centrifugation, and a nuclear fraction was prepared as described previously (49). The nuclear supernatant was transferred to a fresh tube and incubated with ANTI-FLAG® M2 Magnetic Beads (Sigma, 50 µl of packed bead volume) and placed in end-over-end rotation for 2 hours at 4°C for immuno-precipitation (IP). The beads were washed five times with 1X Tris- buffered saline (TBS). The beads were suspended in 200 µl DEPC water and boiled for 5 minutes, and the supernatant fraction was recovered (72). RNA was purified from the supernatant by extraction with 1 mL TRIzol and 200 µl chloroform according to the TRIzol instruction manual. Precipitated RNA was resuspended in 20 µl of RNase/DNase free water and stored at -70°C.

For Cross-Linking Immuno-Precipitation (CLIP) assays the cells were exposed to heat shock and then UV-irradiated at 254 nm-wavelength at 150 mJ/cm^2 as described by

Ule et al. (73). In PAR-CLIP experiments, 4-thiouridine (4SU) (74) was added to a final concentration of 100 μM for 20 minutes prior to heat shock (75), and the cells were UV-irradiation at 365 nm-wavelength at 150 mJ/cm^2 (76). IP reactions were performed as described above, and the beads were washed as described by Darnel et al. (77), except that SDS was not included in the wash buffer. The beads were sequentially washed four times with 5X TBS containing 0.5% Igepal, twice with 1X TBS containing 0.5% Igepal, and twice with 1X TBS. To isolate RNA, the beads were incubated in 200 μl of Proteinase K buffer (1X TBS, 1% SDS, 1.2 mg/ml Proteinase K, 1 mM MgCl) at 37°C for 20 minutes in a Thermomixer at 1000 rpm. Aqueous flow-through was transferred to a new tube and treated with TRIzol. RNA was precipitated as described for RIP. The purified RIP, CLIP and PAR-CLIP RNA samples were treated twice with TURBO DNase™ (Ambion), according to the protocol provided by the manufacturer.

RT-PCR was performed with gene-specific primers to determine if *αβ* RNA was present in the RIP and CLIP samples (Table A1-2). The reverse *αβ* primer (+258 to +278) was used in the cDNA synthesis reactions using AMV-RT (Promega). Subsequent PCR reactions were performed using the same reverse primer and a forward *αβ* primer (+53 to +72). The forward primer spans *Hsp70* and *αβ* sequences, and, thus, cDNAs of *Hsp70-αβ* elements were preferentially amplified by PCR. Control RT-PCR reactions were performed with *Hsp70*-specific primers (reverse, +268 to +287; forward, +29 to +49). As appropriate, PCR products were cloned into a TOPO-TA vector and sequenced.

Generation and Analysis of Reporter Gene Constructs

A recombinant PCR approach (78) was used to generate the fragments that were evaluated in the reporter gene assays. First we generated a fragment containing various

portions of *αβ* sequences in the *invader1* region of +113 to +278 (Table A1-3) or the *Dm88* region of +70 to +204 (Table A1-4). To do this we used a pair of synthetic oligomers (each up to 108 nts, SIGMA) that, together, spanned this region and contained the desired changes. The sense strand oligo consisted of 20 nts of *Hsp70-αβ* sequence (+50 to +69) followed by the 5' portion of *αβ* fragment. The antisense strand oligo overlapped the last 20 nts of the sense strand oligo and contained the remainder of the nts in 3' portion of this region. Each pair of oligos was incubated with Taq DNA polymerase for 10 PCR cycles. Subsequently, PCR was used to ligate the resulting fragment downstream of an *Hsp70* promoter fragment (-361 to + 69). The *Hsp70-αβ* fragment was cloned into a p-Pelican *LacZ* reporter plasmid as previously described (47). Transfections and *LacZ* reporter gene analysis were carried out as previously reported (49).

Results

Su(s) Binds *αβ* RNA *in vitro*

I hypothesized that Su(s)-mediated regulation of *αβ* RNA depends on Su(s) binding to specific sequences in the region between +70 to +278. To identify those sequences, I first used nitrocellulose filter binding assays to examine Su(s) binding to various *Hsp70-αβ* RNA segments. In these experiments, the N-terminal portion of Su(s) (amino acids 1-434), which exhibits the strongest RNA binding activity and encompasses the known and predicted RNA binding domains (ARMs and ZFs, respectively) (45), was expressed in *E. coli* as a maltose-binding protein fusion protein (MBP-Su(s)$_{WT}$) and affinity purified (Figure 2). In addition, mutant protein derivatives that either lacked the ARMs (MBP-Su(s)$_{\Delta ARM}$) or contained missense mutations in the ZFs (MBP-Su(s)$_{ZFmut}$) were also purified and analyzed (Figure 2).

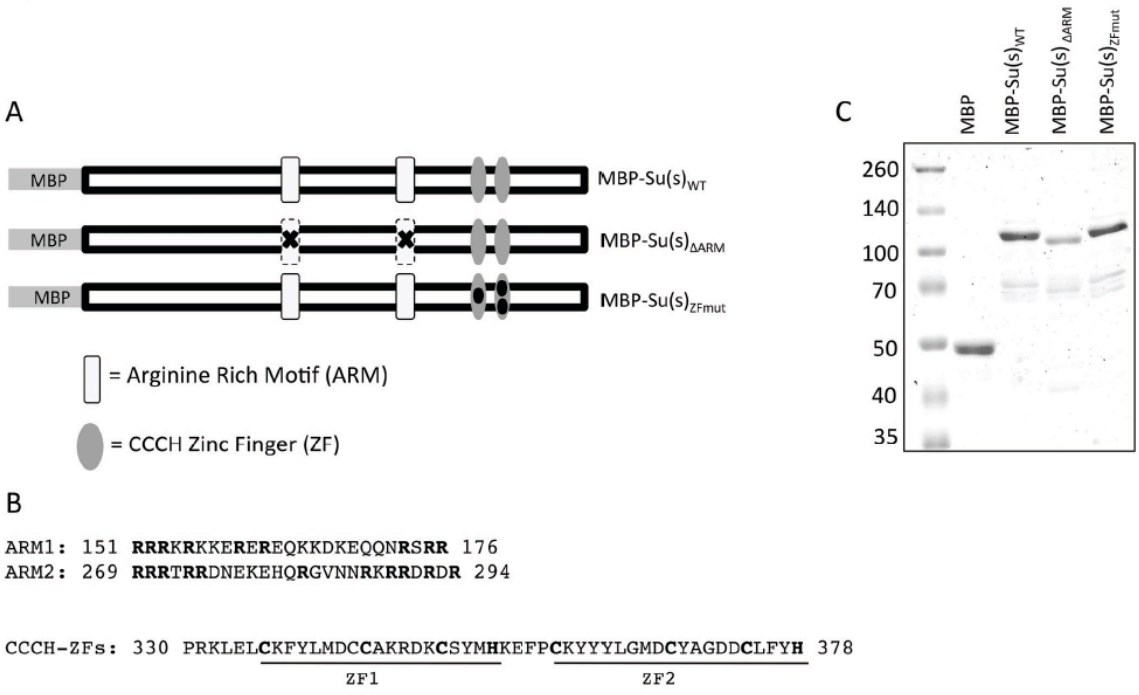


Figure 2. MBP-Su(s) fusion proteins used for *in vitro* RNA binding assays. (A) Schematic of the MBP-Su(s)$_{WT}$ fusion protein and the mutant derivatives. Su(s) has two distinct RNA binding domains. The arginine rich motifs (ARMs) are represented by light grey rectangles; the tandem CCCH zinc fingers (ZFs) are represented by dark grey ovals. The symbol X in the rectangles represents the ARM deletions in MBP-Su(s)$_{\Delta ARM}$. Dots in ovals represent point mutations in the ZF motif in MBP-Su(s)$_{ZFmut}$. (B) Wild-type amino acid sequences of the ARMs and ZFs. The arginine residues in the ARMs, and the cysteine and histidine residues in the ZFs are indicated in bold. Each ZF is underlined. (C) Coomassie-stained gel of affinity-purified, recombinant MBP or MBP fusion proteins (5 µg per well) used in the RNA binding assays. Molecular weight markers are indicated on the left side of the gel.

In the filter binding assays, I examined the binding of increasing amounts of MBP-Su(s)$_{WT}$ (1.5 nM to 100 nM) to four different ^{32}P-labeled RNA segments that span from +1 to +278 (Figure 3A). MBP-Su(s)$_{WT}$ exhibited a relatively low affinity for *Hsp70* 5' UTR sequences (*Hsp70*, +2 to +69) and the segment that includes Region 1 (*αβ*-1, +53 to +128). However, it had a higher affinity for two non-overlapping segments that span Region 2 (*αβ*-2 and *αβ*-3; +120 to +201 and +202 to +278, respectively). A longer segment that includes both Region 2 segments (*αβ*-2+3, +120 to +278), was bound with a similar affinity as *αβ*-3 alone (Figure 3B). The affinity of MBP-Su(s)$_{ZFmut}$ for *αβ*-2+3 was similar to MBP-Su(s)$_{WT}$;

however, MBP-Su(s)$_{\Delta ARM}$ did not bind to this RNA (Figure 3C). These results indicate that MBP-Su(s)$_{WT}$ preferentially binds to sequences within two regions of Region 2, and the ARMs mediate these interactions.

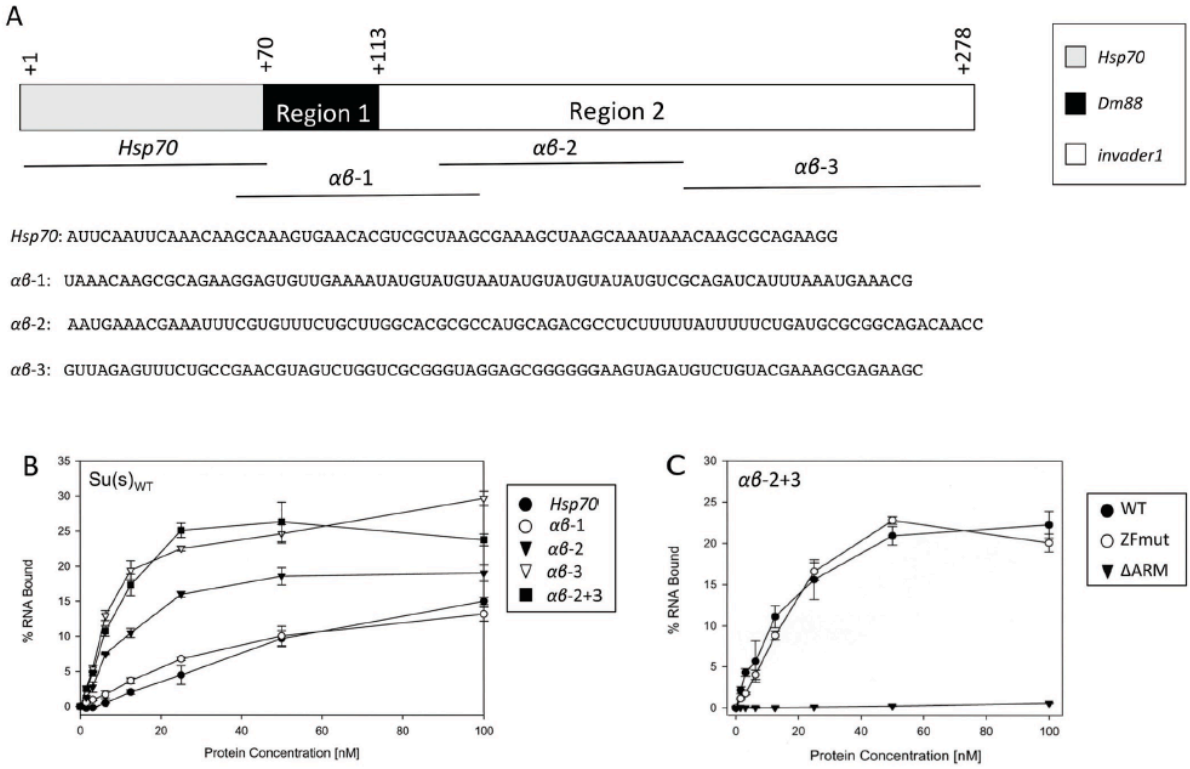


Figure 3. Nitrocellulose filter binding assays demonstrate the variable affinity of MBP-Su(s) for various *αβ* RNA fragments. (A) Region +1 to +278 was divided into four different fragments for the RNA binding assays: *Hsp70* (+2 to +69), *αβ*-1 (+53 to +128), *αβ*-2 (+120 to +201) and *αβ*-3 (+202 to +278). The RNA sequence of each fragment is shown. (B) Nitrocellulose filter binding assays were performed to examine the binding of MBP-Su(s)$_{WT}$ to each of four fragments described in **A** and to a longer fragment (*αβ*-2+3). Each binding reaction contained 25,000 counts per minute (cpm) of uniformly labeled RNA fragments and various amounts of MBP-Su(s)$_{WT}$ (1.5 nM to 100 nM). (C) Binding of MBP-Su(s)$_{WT}$, MBP-Su(s)$_{\Delta ARM}$, and MBP-Su(s)$_{ZFmut}$ to *αβ*-2+3. The binding reactions were performed as described in **B**.

I also performed electrophoretic mobility shift assays (EMSAs) (79) to examine the binding of WT and mutant MBP-Su(s) (2 nM to 500 nM) to these RNA segments. With this technique, specific protein-RNA interactions are indicated by the presence of one or more discrete bands that migrate slower in the gel than unbound RNA. As was seen in filter

binding assays, MBP-Su(s)$_{WT}$ did not bind significantly to *Hsp70* 5' UTR RNA in EMSAs, i.e., faint complexes were observed only at the highest protein concentration (Figure 4, left panel). However, MBP-Su(s)$_{WT}$ formed discrete complexes with *αβ*-1, *αβ*-2, and *αβ*-3 RNAs. Two or more complexes were observed when MBP-Su(s)$_{WT}$ was incubated with *αβ*-1 and *αβ*-3 RNAs, but only a small proportion of the total RNA was associated with these complexes, even at high protein concentrations (Figure 4, left and right panels). In contrast, a single complex was observed with MBP-Su(s)$_{WT}$ and *αβ*-2 RNA (Figure 4, right panel). This complex was detected at a 15-fold lower protein concentration than the complexes that formed with *αβ*-1 and *αβ*-3, and a larger fraction of *αβ*-2 RNA was associated with the complex. MBP alone did not bind to *αβ* RNA (Figure 5), and MBP-Su(s)$_{\Delta ARM}$ bound very weakly to *αβ* RNA (Figure 6). However, the affinity of MBP-Su(s)$_{ZFmut}$ for *αβ* RNA was similar to WT protein (Figure 6). Thus, the EMSA results indicate that MBP-Su(s)$_{WT}$ forms a discrete, high-affinity complex with *αβ*-2 RNA, and the ARMs greatly enhance the efficiency of this interaction. WT protein forms lower-affinity, discrete complexes with *αβ-1* and *αβ-3* RNAs and the ARMs are important for this interaction as well. I did not detect significant binding to the *Hsp70* 5' UTR segment.

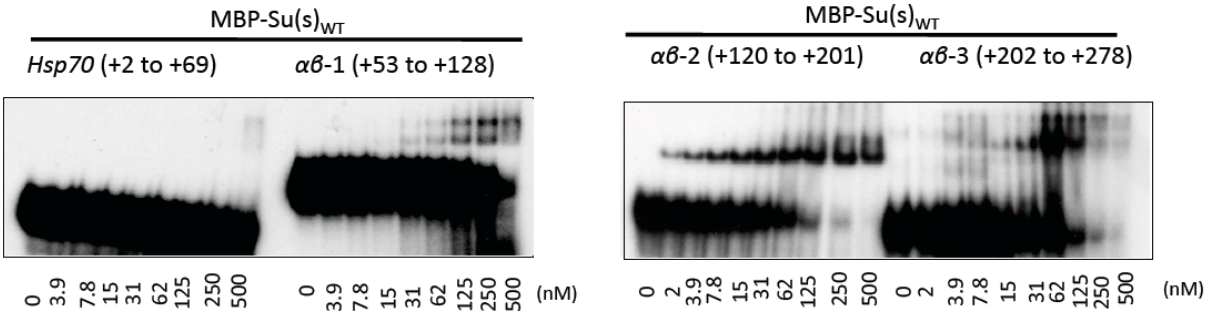


Figure 4. EMSAs reveal that MBP-Su(s)$_{WT}$ and the *αβ* RNA fragments form distinct complexes. EMSA comparing MBP-Su(s)$_{WT}$ binding to the four fragments described in Figure 3. Increasing amounts of MBP-Su(s) (2 nM to 500 nM) were incubated with 5,000 cpm of uniformly-labeled RNA prior to electrophoresis on native polyacrylamide gels.

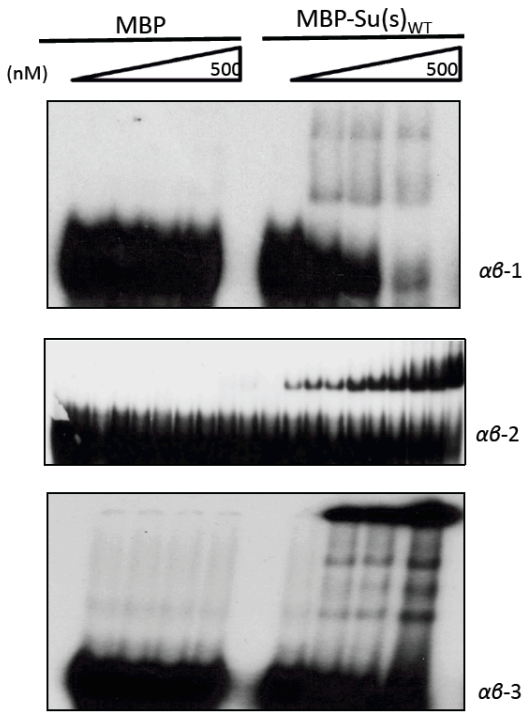


Figure 5. MBP does not form stable complexes with *αβ* RNA. Increasing concentrations of MBP or MBP-Su(s)$_{WT}$ (125 nM to 500 nM for *αβ*-1 and *αβ*-3, 2nM to 500nM for *αβ*-2) were incubated with 5000 CPM of uniformly-labeled RNA. EMSAs were performed as described above.

Although similar results were obtained with both methods, *αβ*-3 RNA behaved differently in the two assays. The filter-binding data indicate that MBP-Su(s)$_{WT}$ has a high affinity for *αβ*-3 RNA; however, a discrete, high-affinity complex was not detected in EMSAs with this RNA. The reasons for this discrepancy are unclear. Perhaps the high affinity *αβ*-3 RNA-protein complex is unstable in EMSAs. Alternatively, because I consistently observed that a significant fraction of *αβ*-3 RNA did not enter the gel (see Figures 5 and 6), it is also possible that interactions of the protein with *αβ*-3 RNA leads to aggregation. Thus, the apparent high affinity *αβ*-3 RNA-Su(s) binding that was observed in the filter binding assays might have been due to aggregation. Although the aggregate may be a nonspecific complex, it appears to depend on sequences within this RNA because this was not consistently observed with the other RNA segments.

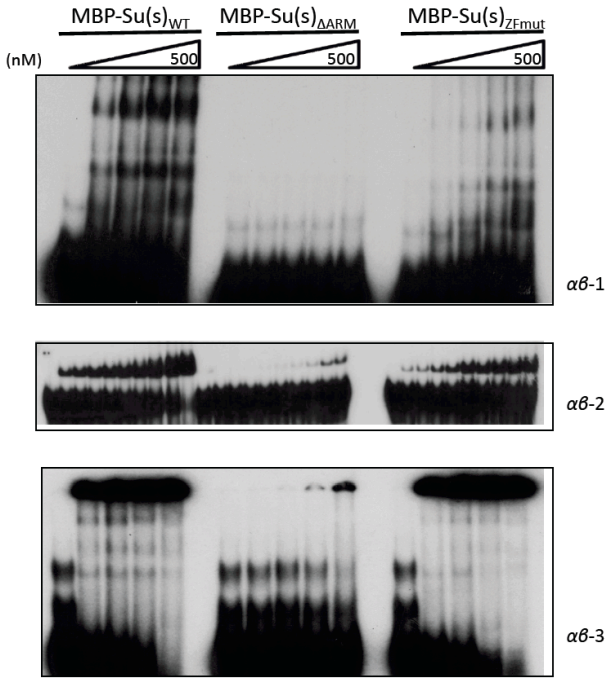


Figure 6. The ARMs mediate stable interaction with *αβ* RNA. EMSA comparing the binding of WT and mutant Su(s) proteins to *αβ*-1, *αβ*-2 or *αβ*-3.

In summary, the formation of discrete complexes with all three *αβ* RNA segments suggests that Su(s) recognizes specific sequences within these RNAs. However, its affinity for these sequences varies. The protein appears to have a higher affinity for sequences within *αβ*-2 than the other *αβ* RNA segments, and it is possible that sequences within *αβ-3* induce Su(s)-dependent aggregation. Furthermore, consistent with previous observations (45), Su(s)/*αβ* RNA interactions depend on the ARMs.

Su(s) Binds *αβ* RNAs *in vivo*

I next wanted to determine if Su(s) binds to *αβ* transcripts in the context of living cells. To do this I used existing *Drosophila melanogaster* S2 cells that had been stably transformed with a full length *su(s)* cDNA construct containing a C-terminal 3X-Flag tag (Su(s)-3xF) (Figure 7A). The inducible *Metallothionein* promoter drives transcription of

this construct. Following an overnight induction of Su(s)-3xF, the cells were heat shocked at 37°C for 20 minutes to induce expression of *αβ* RNAs, prior to the isolation of Su(s)-3xF-RNA complexes. I used two strategies to isolate Su(s)-3xF/RNA complexes, cross-linking immuno-precipitation (CLIP) (73) and RNA-immunoprecipitation (RIP) (72), followed by RT-PCR with gene-specific primers to detect RNAs associated with Su(s). For CLIP the cells were irradiated at 254 nm prior to lysis, whereas the RIP samples were not irradiated. The primers amplified the +53 to +278 region of *αβ,* which has sequences that are sufficient for regulation and contains multiple binding sites for MBP-Su(s)$_{WT}$ *in vitro*. I also used a primer pair targeting *Hsp70*, which is not bound or regulated by Su(s) (Figure 7B). This analysis indicated that Su(s)-3xF interacts stably with *αβ* transcripts with or without UV-crosslinking. In contrast, *αβ* RNAs were not detected in IPs from untransformed S2 cells, the negative control sample. Furthermore, Su(s)-3xF did not interact with *Hsp70* transcripts (Figure 7B). These results indicate that Su(s)-3xF interacts specifically with *αβ* transcripts *in vivo*.

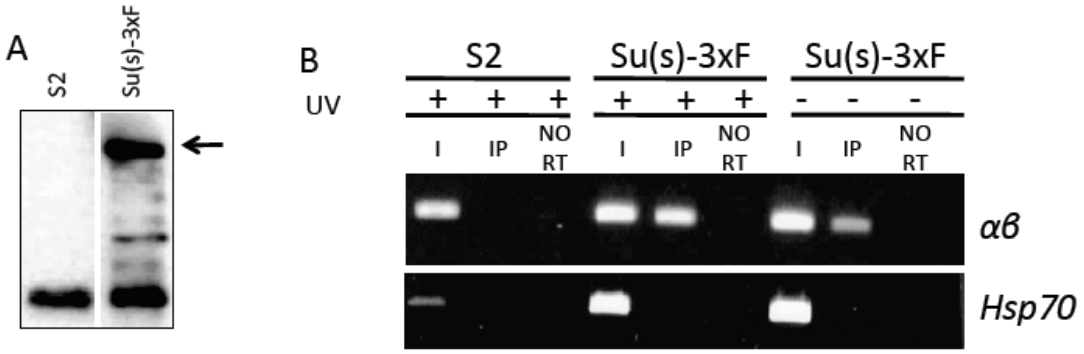


Figure 7. Su(s)-3xF binds to *αβ* transcripts in cultured Drosophila cells. (A) Western blot of Flag-IP from normal S2 cells and stably transformed Su(s)-3xF-expressing cells. The blot was probed with anti-Flag antibody. Arrow indicates Su(s)-3xF. (B) RT-PCR analysis of transcripts that co-purified with Su(s)-3xF. RNA immunoprecipitations were performed on nuclear lysates from normal S2 cells or Su(s)-3xF cells with or without UV-irradiation. Input (I) from nuclear lysates and co-purified RNAs were subject to gene specific reverse transcription (RT), followed by PCR to detect *αβ* or the

negative control *Hsp70* cDNA fragments. Additionally, a no RT reaction was performed to ensure the absence of genomic DNA.

Sequence Analysis of UV-crosslinked cDNAs Reveal Potential Sites of Contact Between Su(s) and the 209nt Region of *αβ* Transcripts

Having found that Su(s)-3xF crosslinks to *αβ* transcripts, I wanted to identify *in vivo* contact sites within the 209 nt region. To do this, I used PCR to amplify cDNAs obtained from CLIP (73)or PAR-CLIP (photoactivatable-ribonucleoside-enhanced- crosslinking and immunoprecipitation)(76). Subsequently, the PCR products were cloned and sequenced. DNA sequencing can reveal sites of protein/RNA contact because nucleotide sequence mutations are introduced at a fraction of the crosslinked sites during cDNA synthesis (80).

From six separate CLIP experiments and two separate PAR-CLIP experiments, a total of 300 clones were sequenced, and 59 of these contained mutations. In compiling the CLIP data, I only included mutations that appeared in at least two separate experiments. However, because the use of 4-thiouridine in PAR-CLIP causes a unique T to C transition (76), all mutations of this type were included.

The mutations in both CLIP and PAR-CLIP clustered in the *invader1* fragment between +113 to +278 (Figure 8A and 8B). Many of the mutations occurred in U-rich sequences located in +120 to +210, which contains the *αβ*-2 segment that MBP-Su(s)$_{WT}$ bound with high affinity in EMSA analysis. A substantial number of PAR-CLIP mutations were located within clustered triplets of U's and in a U-rich sequence that extends from +171 to +181 (UUUUUAUUUUU) (Figure 8B and 8C). Two of the CLIP identified mutations (+256 and +257) were located a short distance downstream of a consecutive run of 6 G's located at +244 to +249 (Figure 8A and 8C). These results are consistent with the EMSA analysis which indicated that Su(s) binds to the region between +113 and +278.

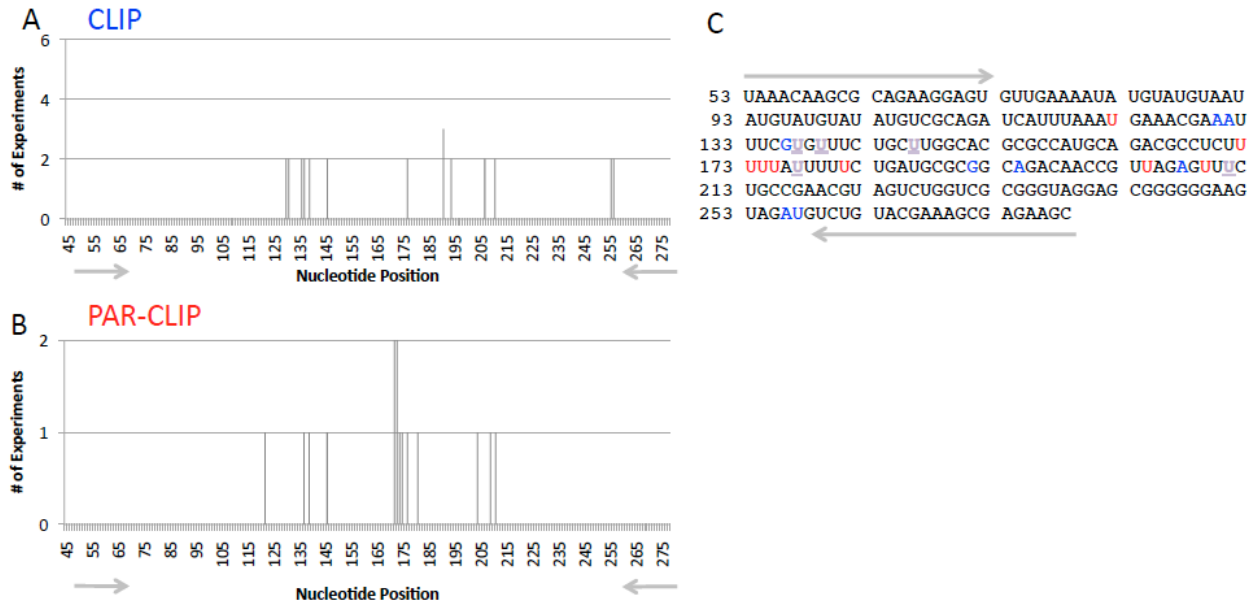
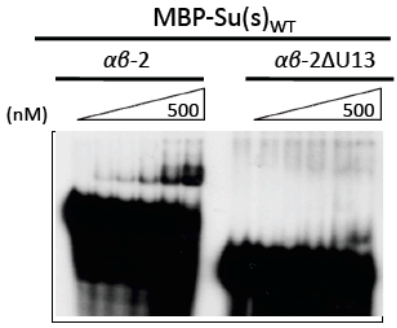


Figure 8. Sequence analysis of UV-crosslinked cDNAs**. (A)** Locations of putative Su(s)/RNA contact sites obtained from sequencing cDNA clones derived from CLIP assays. (B) Putative Su(s)/RNA contact sites interpreted as T to C mutations obtained from samples that were UV-crosslinked after 4-thiouridine incorporation (PAR-CLIP). (C) Sequence of the *αβ*-RNA region amplified during the CLIP and PAR-CLIP analysis. Blue sequences correspond to CLIP identified putative Su(s)-3xF interaction sites, and red sequences are PAR-CLIP identified putative Su(s)-3xF interaction sites. Sequences identified in both CLIP and PAR-CLIP are shown in purple and are underlined. Grey arrows indicate the locations of primers.

Loss of *In Vivo* Binding Sites Reduces Binding Affinity to *αβ* Transcripts *In Vitro*

The CLIP/PAR-CLIP experiments indicated that Su(s) binds to U-rich sequences. However, because the CLIP experiments were performed under non-denaturing conditions it is possible that Su(s) interacts with a protein that crosslinks to these regions rather than interacting directly with the RNA. Additionally, because CLIP analysis preferentially cross-links at uridines (81), low occupancy sites are sometimes overrepresented relative to high occupancy sites (82). Thus, I examined the effect of altering U-rich sequences within *invader1* on MBP-Su(s)$_{WT}$ binding *in vitro*. In these experiments I compared MBP-Su(s)$_{WT}$ binding to the WT *αβ*-2 fragment and to a derivative of this fragment containing a deletion of the U-rich region from +171 to +183 (*αβ*-2ΔU13) (Figure 9). Interestingly, the binding to

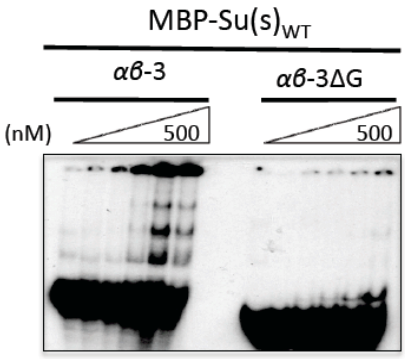
αβ-2 was eliminated when these sequences were deleted. This indicates that the sequence UUUUUAUUUUUCU, is a binding site for MBP-Su(s)$_{WT}$ (Figure 9).



αβ-2: AAUGAAACGAAAUUUCGUGUUUCUGCUUGGCACGCGCCAUGCAGACGCCUC**UUUUUAUUUUUCU**GAUGCGCGGCAGACAACC
αβ-2ΔU13: AAUGAAACGAAAUUUCGUGUUUCUGCUUGGCACGCGCCAUGCAGACGCCUCGAUGCGCGGCAGACAACC

Figure 9. Deletion of the U-rich putative Su(s)-3xF crosslinking sites affects the affinity of MBP-Su(s)$_{WT}$ for *αβ* RNA *in vitro*. EMSA comparing MBP-Su(s)$_{WT}$ binding to *αβ* RNA fragment (*αβ*-2) versus the same fragment containing the 13nt deletion of the sequence UUUUUAUUUUUCU (+171 to +183, *αβ*-2ΔU13) region identified in CLIP/PAR-CLIP analysis as a possible binding site for Su(s). The RNA sequence is shown below the EMSA with the nucleotides that were deleted shown bold. The sequence of RNA remaining after deletion of the specified region is also shown.

I also detected crosslinking sites in the vicinity of a G-rich region within *invader1*. I therefore compared MBP-Su(s)$_{WT}$ binding to WT *αβ*-3 to a deletion mutant that removes the G-rich region between +225 to +249 (*αβ*-3ΔG) (Figure 10). Deletion of the G-rich region eliminated the high molecular weight complex/aggregate and the multiple shifted bands observed with the unaltered RNA, leaving a single shifted complex (Figure 10). Thus it appears that the G-rich region also mediates binding of MBP-Su(s)$_{WT}$ to the segment. Together these EMSA results indicate that MBP-Su(s)$_{WT}$ binds to U-rich and G-rich sequences in the *invader1* LTR (Region 2).



αβ-3: GUUAGAGUUUCUGCCGAACGUAG**UCUGGUCGCGGGUAGGAGCGGGGGG**AAGUAGAUGUCUGUACGAAAGCGAGAAGC
αβ-3ΔG: GUUAGAGUUUCUGCCGAACGUAGAAGUAGAUGUCUGUACGAAAGCGAGAAGC

Figure 10. Deletion of the G-rich putative Su(s)-3xF crosslinking sites affects the affinity of MBP-Su(s)$_{WT}$ for *αβ* RNA *in vitro*. EMSA comparing MBP-Su(s)$_{WT}$ binding to *αβ*-3 versus a fragment containing a 25nt deletion of a G-rich sequence UCUGGUCGCGGGUAGGAGCGGGGGG (+225 to +249, *αβ*-3ΔG) that is located slightly upstream of two CLIP crosslinked sites. The MBP-Su(s)$_{WT}$/*αβ*-3 binding reactions were analyzed on a lower percentage gel (6%) to facilitate resolution of three complexes. The RNA sequence is shown below the EMSA with the nucleotides that were deleted shown bold. The sequence of RNA remaining after deletion of the specified region is also shown.

U-rich and G-rich Elements in *Invader1* Contribute to Regulation by Endogenous Su(s)

I used a previously established reporter gene assay (49) to determine if the U-rich and G-rich sequences within a portion of the *invader1* LTR (+113 to +278 of *Hsp70-αβ*) are relevant regulatory elements. Thus, I made a set of deletion derivatives with alterations in this DNA fragment and inserted the mutated fragments between the *Hsp70* promoter and the *LacZ* coding region (Figure 11A). The unaltered and mutant reporter plasmid constructs were transiently transfected separately into control (mock-KD) cells, which express a normal level of Su(s), or Su(s)-depleted (Su(s)-KD) cells. Subsequently, the cells were subjected to a mild heat shock, and northern blot analysis was performed to determine the relative *LacZ* mRNA level in mock-KD versus Su(s)-KD cells. If regulatory sequences are intact in the *invader1* region (+113 to +278) of a given reporter construct, the level of *LacZ* mRNA is expected to be about 5-fold lower in mock-KD cells than in Su(s)-

KD cells (Figure 11B, compare lanes 1 & 2 or lanes 9 & 10; Figure 11C). However, if a deletion removes an important regulatory element, *LacZ* mRNA is expected to accumulate at a higher-than-normal level in mock-KD cells, and the ratio of *LacZ* mRNA in mock-KD cells versus Su(s)-KD cells would increase, i.e., *LacZ* mRNA ratio >0.2. As reported previously (49) cells transfected with the control reporter plasmid γA, which contains the *Hsp70* 5' UTR region (+1 to +69) but lacks *αβ* sequences, accumulated similar levels of LacZ mRNA in mock-KD and Su(s)-KD cells (Figure 11B, compare lanes 7 & 8 or lanes 17 &18; Figure 11C). A mutant reporter construct that contained a deletion of the sequence 5'-UUUUUAUUUUUCU-3', located between +171 and +183 (Figure 11A), showed about a two-fold increase (p-value 0.0002) in the *LacZ* mRNA ratio (mock-KD/Su(s)-KD) compared to the unaltered (WT) reporter construct (ΔU13, Figure 11B and 11C).

I also analyzed three reporter constructs with deletions that affect the G-rich region of the *invader1* LTR. One of these constructs contained the 25-nt deletion that extends from +225 to +249 (Figure 11A). This deletion resulted in a four-fold increase in the *LacZ* mRNA ratio (p-value 0.00026) and, thus, eliminated most of the negative regulation (ΔG, Figure 11B & 11C). The analysis of two smaller deletions within this region (+230 to 239 and +240 to 249, Figure 11A) revealed that sequences in both of these segments contribute to this effect. Specifically, *LacZ* mRNA ratios in Δ240-249 and Δ230-239 were 3-fold (p-value 0.001) and 1.8 fold (p-value 0.025) higher than WT, respectively (Figure 11B & 11C). Thus, the G-rich region is also an essential regulatory element.

These data suggest that the U-rich element (+171 to +183) and the G-rich element (+225 to +249) within *invader1* contribute to regulation by Su(s). This finding is consistent with previous results, which indicated that regulatory sequences in *invader1* are found

upstream and downstream of +205 (49). Both of these elements can mediate Su(s) binding to RNA *in vitro*, and Su(s) makes close contact with the U-rich element *in vivo* (see Figures 8, 9 and 10). Although the EMSA analysis indicates that Su(s) has a higher affinity for the U-rich element, the G-rich element makes a larger contribution to the regulation.

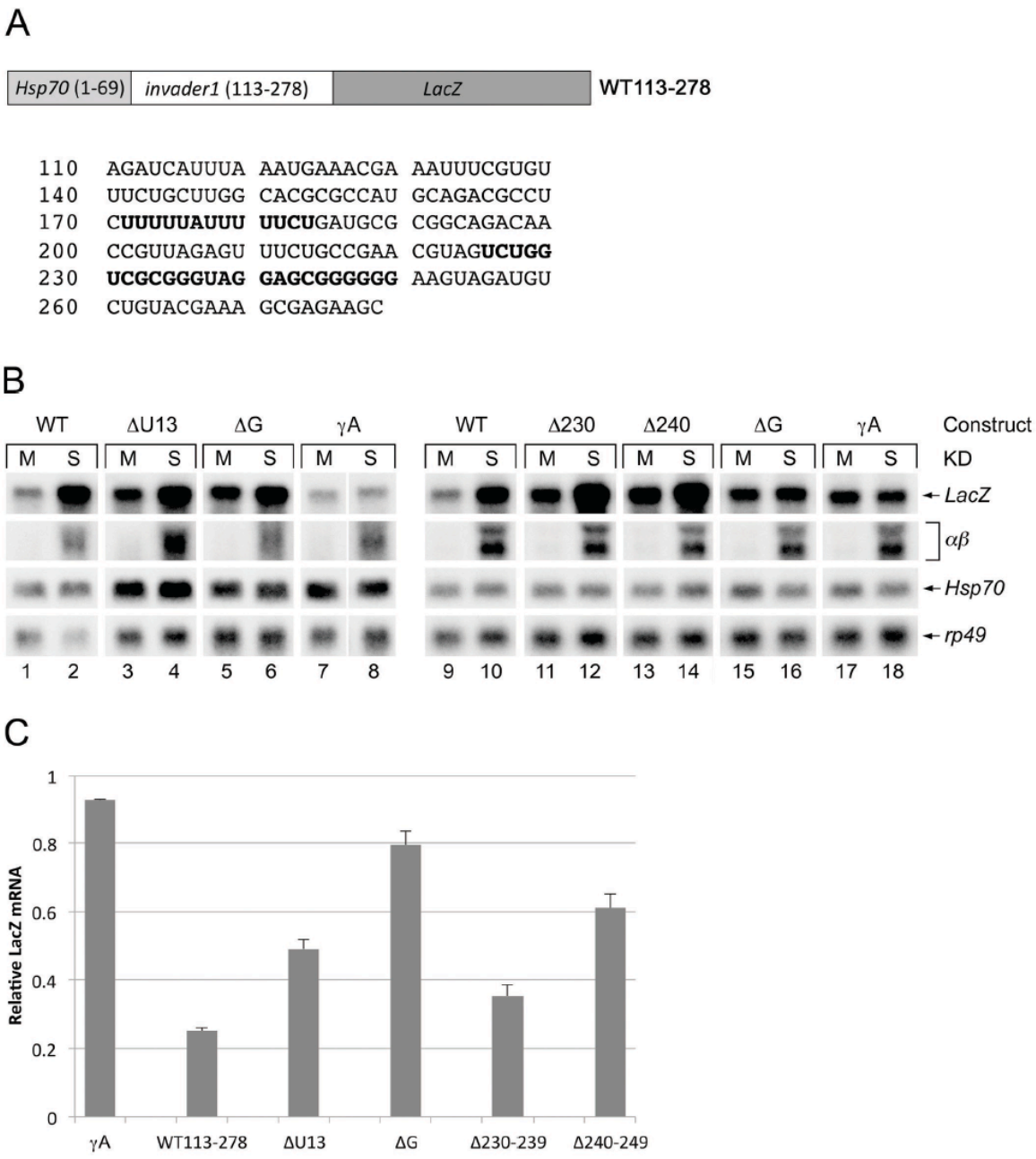


Figure 11. Deletion of Su(s) binding sites affects regulation by endogenous Su(s) in cultured cells. (A) Schematic of the *LacZ* reporter construct that was used to test the effect of deleting the U-rich and G-rich *invader1* sequences on regulation by endogenous Su(s) in S2 cells. Sequence of the *invader 1* region spanning +113 to +278 is shown beneath the schematic map. Bold sequences are nucleotides that were deleted for the reporter gene analysis. (B) Representative northern blot illustrating the level of *LacZ* mRNA generated by WT and mutant reporter constructs. The expression of each plasmid construct was examined in mock-KD (M) or Su(s)-KD cells after a brief heat shock. The mock-KD and Su(s)-KD samples are shown in the odd and even numbered lanes, respectively. The control plasmid γA, lacks αβ sequences but contains *Hsp70* 5' UTR sequences (+1 to +69) which do not contribute to regulation (49) and are not bound by Su(s) *in vitro*. Two different blots are shown in lanes 1-8 and lanes 9-18. The construct names are the same as described in the text, except that Δ230 and Δ240 are abbreviations for the10-nt deletions that extend from +230 to +239 and +240 to +249, respectively. The blots were sequentially probed to detect *LacZ*, *αβ*, *Hsp70*, and *rp49* RNAs. The latter three RNAs control for the RNAi effectiveness, heat shock, and loading, respectively. (C) Graphical representation of cumulative northern blot data. Relative *LacZ* RNA is the ratio of mock-KD *LacZ* RNA level/Su(s)-KD *LacZ* RNA level. Each construct was examined in at least three independent experiments. The error bars indicate standard deviations.

A GUA-rich Element in the *Dm88 LTR* Mediates Regulation by Su(s) but is Not a Su(s) Binding Site

We previously showed (49) that a 42-nt segment at the 5' end of the *Dm88* LTR (Region 1 of *Hsp70-αβ*, Figure 1A) mediates regulation by the Su(s) pathway, independent of *invader1* LTR sequences. Furthermore, a *LacZ* reporter construct (HDL-3.0, Figure 12A) that includes most of this segment plus other *Dm88* LTR sequences downstream of this region is strongly downregulated by Su(s) ((49); Figure 12B, compare lanes 1 & 2). The most notable feature of the 42-nt segment is a tandem 11-nt direct repeat (+79 to +100) that consists entirely of the nucleotides G, U, and A (Figure 12A). To test whether the relevant regulatory sequences lie within this repeat, we generated and analyzed reporter constructs with GU to AC substitutions in this 22-nt region and at other sites within the 134-nt *Dm88* segment. The results of this analysis are shown in Figure 12B. Mutation of all four GU dinucleotides within the direct repeat (1-4, Figure 12A) essentially eliminated

the Su(s)-dependent regulation observed in mock-KD cells (compare HDL-3.0 and -3.1, Figure 12B). A similar effect was observed when either the first two or the second two GU dinucleotides (1-2 or 3-4, Figure 12A) were altered (HDL-3.2 and -3.3, respectively, Figure 12B). In contrast, a construct with GU to AC substitutions at 12 positions outside of this region but with the wild-type sequence within the direct repeat (HDL-3.4, Figure 12B) was downregulated as efficiently as the unaltered HDL-3.0 construct. As expected, regulation was lost when all 16 GU dinucleotides, including those in the direct repeat, were altered (HDL-3.5, Figure 12B). Thus, regulation depends on sequences that include the four GU dinucleotides that are located within the direct repeat.

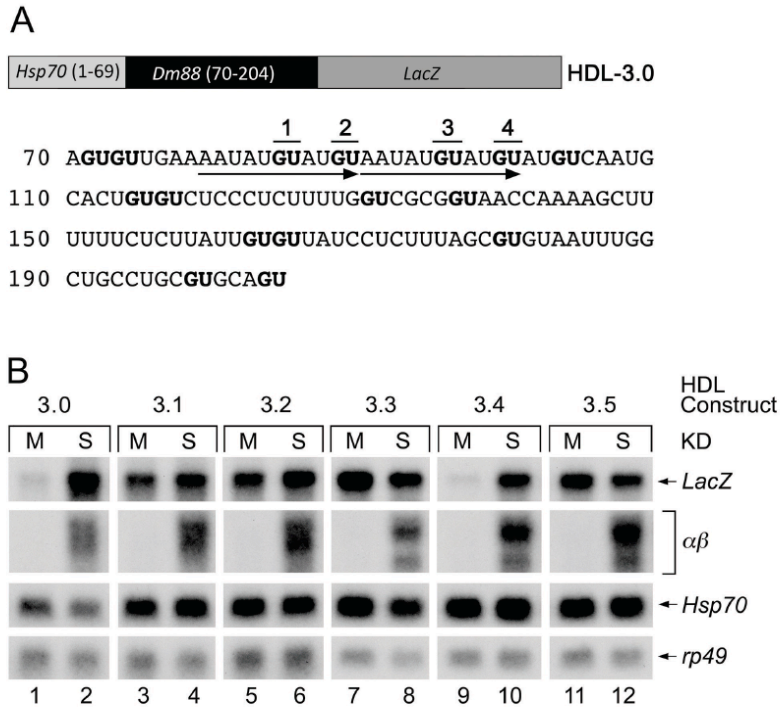


Figure 12. Nucleotide substitution mutations in the *Dm88* LTR disrupt Su(s)-mediated regulation. (A) Schematic map of the reporter construct HDL-3.0 (49), which contains a segment of the *Dm88* LTR that includes Region 1. The *Dm88* sequences are shown beneath the map. The 11-nt tandem direct repeat is indicated by the horizontal arrow. GU to AC substitutions were introduced at the positions indicated in bold. The numbers 1-4 above the sequence indicate the positions of GU dinucleotides in the direct repeat. (B) Northern blot analysis of the unaltered (HDL-3.0) and mutant reporter constructs. The mutant constructs have GU to AC substitutions at the following

positions: 1-4 (HDL-3.1), 1-2 (HDL-3.2), 3-4 (HDL-3.3), all positions indicated in bold in **A**, except those in the direct repeat (HDL-3.4), all positions indicated in bold in **A**, including those in the direct repeat (HDL-3.5). The northern blot analysis was performed as described for Figure 11.

I performed EMSA analysis to determine if the binding of MBP-Su(s)$_{WT}$ to an *Hsp70-Dm88* RNA segment, extending from +53 to +140, depends on these four GU dinucleotides (Figure 13). I found that the weak binding of MBP-Su(s)$_{WT}$ to this RNA was not affected by the nucleotide substitutions that eliminated the regulation. This suggests that the GUA-rich elements are recognized by another protein that participates in this regulatory process.

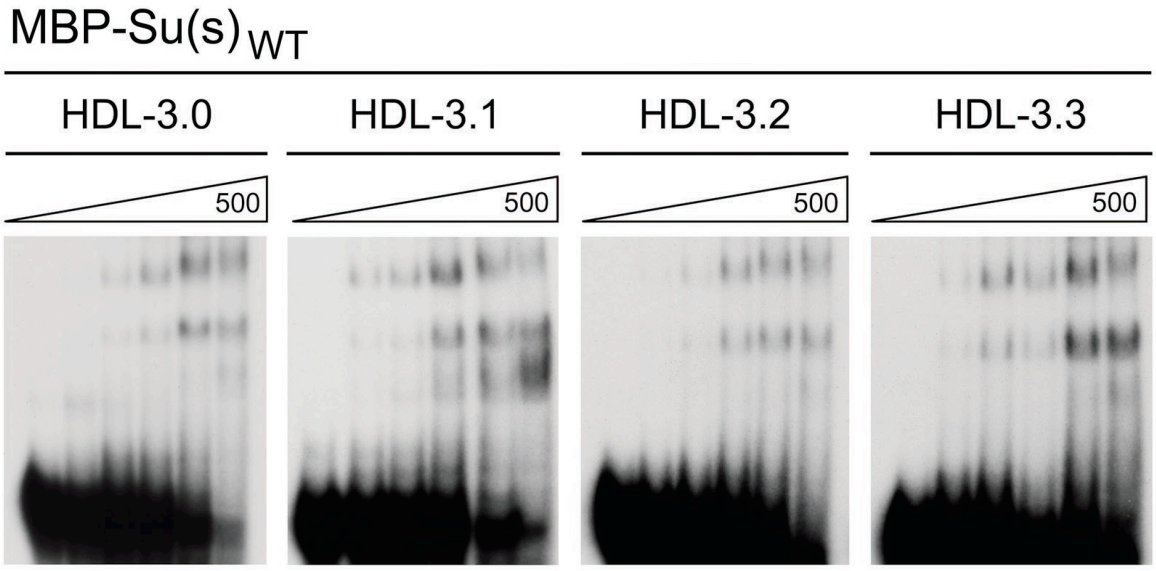


Figure 13. Nucleotide substitution mutations in the *Dm88* LTR do not affect the RNA binding of MBP-Su(s)$_{WT}$ *in vitro*. EMSA analysis of the binding of MBP-Su(s)$_{WT}$ to 89-nt segments of *Hsp70-Dm88* RNA. The RNAs extend from +53 to +140 and include either the HDL-3.0 sequences or mutant RNAs with GU to AC substitutions in the direct repeat (HDL-3.1, -3.2, and -3.3) as described in Figure 12.

Discussion

The role that the Su(s) regulatory pathway plays in down-regulating aberrant transcripts has been examined in considerable depth by the analysis of the Su(s)-mediated control of transcripts produced by *Hsp70-αβ* elements, which have retrotransposon LTR sequences inserted in the 5' transcribed region. Previous results from this lab indicate that

the Su(s)-Wdr82 complex induces transcription termination in the promoter-proximal region *Hsp70-αβ*, and the resulting RNAs are degraded by nuclear exosome components (47,49). In addition, we previously showed that down-regulation of these RNAs depends on the presence of *Dm88* and *invader1* LTR sequences, located between +70 and +278 of *αβ* (Regions 1 and 2, respectively; see Figure 1). Each of these regions can mediate regulation independently, and together their effects are additive (49).

Here, I have examined the extent to which this regulation depends on the binding of Su(s) to RNA sequences within these two regions. Several lines of evidence indicate that the binding of Su(s) to a U-rich element (+171 to +183) within Region 2, the *invader1* LTR, is a contributing factor. First, I showed that the ARMs of Su(s) mediate strong binding of MBP-Su(s)$_{WT}$ to an RNA fragment that contains this element, and deletion of the U-rich element completely eliminates this binding. Second, my CLIP/PAR-CLIP analysis indicates that Su(s)-3xF crosslinks to the U-rich element *in vivo.* Third, my reporter gene analysis indicates that deletion of the U-rich element attenuates regulation. These results are consistent with prior *in vitro* RNA-binding experiments, which indicated that Su(s) binds to poly(U) *in vitro* (52). Furthermore, previous analysis of a transgenic Su(s)$_{\Delta ARM}$ mutant derivative indicated that deletion of the ARMs partially impairs Su(s)-mediated regulation (47).

Consistent with the prediction that at least two regulatory elements lie within Region 2 (49), I also delineated a G-rich element (+225 to +249) in Region 2 that makes a strong contribution to this regulation. However, it is unclear if this is a true *in vivo* Su(s) binding site for several reasons. First, the CLIP experiments did not provide evidence that Su(s)-3xF interacts directly with the G-rich region, although two cross-links were detected

a short distance downstream of this region. Furthermore, the *in vitro* RNA-binding results were not clear-cut. In EMSAs where I examined the binding of MBP-Su(s)$_{WT}$ to a fragment that contains this element (*αβ-3*), a large RNA-protein aggregate formed that did not enter the gel. This was especially apparent at high protein concentrations (see Figures 5 and 6). The EMSA results suggest that MBP-Su(s)$_{WT}$ binds specifically, but weakly, to multiple sites on this fragment, and it is possible that this induces RNA-protein aggregation. Although the G-rich element contributes to this interaction, and Su(s) binds poly(G) *in vitro* (52), sequences of this type can form unusual structures (83,84) that might promote aggregation. Thus, these large complexes may not be specific, and, if not, high-affinity binding between MBP-Su(s)$_{WT}$ and *αβ-3* observed in the nitrocellulose filter binding assays (Figure 3B) might not be biologically relevant. However, if self-association is a property of Su(s), it is possible that this is enhanced by the interaction of Su(s) with G-rich sequences. Because of the uncertainty about whether the G-rich element is specifically recognized by Su(s), the possibility remains this sequence is not directly recognized by Su(s).

We also showed that a GUA-rich region in the *Dm88* LTR (Region 1), mediates Su(s)-dependent down-regulation of *αβ* RNA (49). Although I observed a weak interaction between MBP-Su(s)$_{WT}$ and an RNA that includes this element in EMSAs, this interaction does not depend on the GUA-rich element. Furthermore, Su(s)-3xF did not crosslink to this region *in vivo*. Thus, not all of the regulatory elements are Su(s) binding sites.

Eventually, we want to understand the mechanism by which these sequence elements interfere with the accumulation of the aberrant, transposon-containing RNAs. Our prior results indicate that the unstable *αβ* RNAs that are generated when the Su(s)-Wrd82 pathway is functional are polyadenylated at heterogeneous sites in the promoter-

proximal region before being degraded by 3' to 5' exonucleases (49). Interestingly, the 3' ends of a substantial fraction of the unstable RNAs correspond to sites in the vicinity of two of the regulatory elements that we have defined here. HDL-3.0 transcripts were polyadenylated downstream of the GUA-rich element in Region 1, and many of the *Hsp70-αβ* transcripts ended within or downstream of the G-rich region. However, only one polyadenylation event occurred in the vicinity of the U-rich element (49).

Thus, one possible model could be that the Su(s)-Wdr82 complex is recruited to the promoter of some actively transcribed genes. The presence of U-rich or G-rich Su(s) binding sites within the promoter-proximal region allows Su(s) to directly interact with the nascent RNA and this induces premature termination (Figure 14A). Furthermore, other proteins may interact with the Su(s)-Wdr82 complex and these proteins may recognize other regulatory sequences, such as the GUA-rich sequences within Region 1, and this interaction also induces premature termination (Figure 14A). However, when binding sites are deleted or mutated, the Su(s)-Wdr82 complex is no longer able to interact with the RNA, thereby preventing premature termination, resulting in stable transcripts that end following a canonical PAS (Figure 14B).

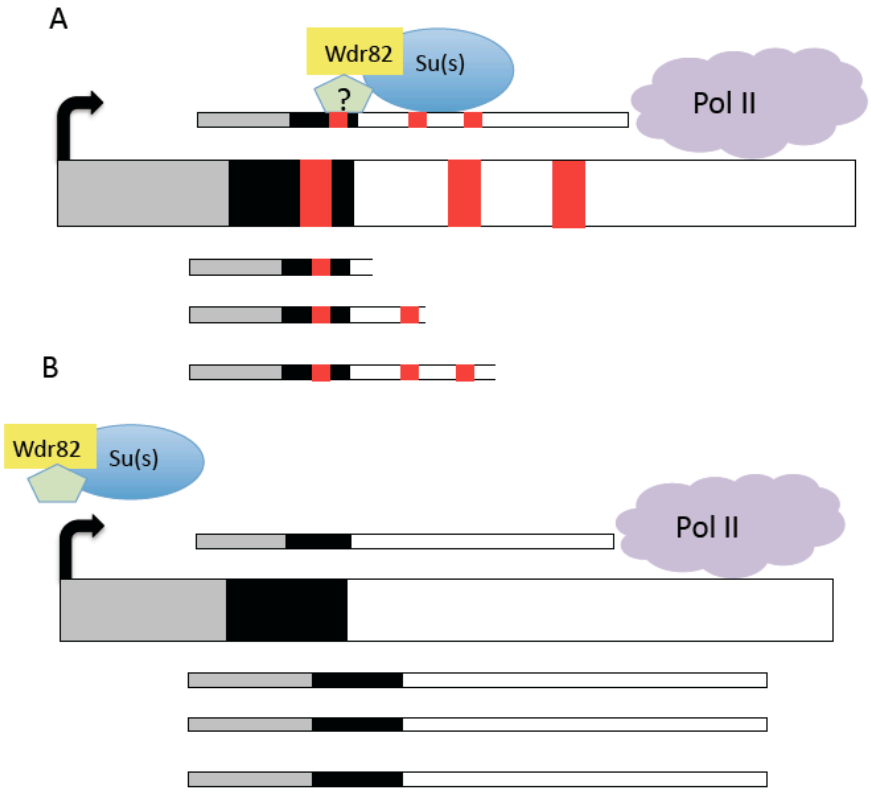


Figure 14. Putative model of the Su(s)-Wdr82 mediated regulation *αβ* transcripts. (A) Su(s) binding to regulatory sites in the promoter-proximal region induces premature termination. The Su(s)-Wdr82 complex is recruited to some actively transcribed genes. Su(s) interacts with U-rich or G-rich regulatory sites in Region 2 of the nascent RNA (indicated as red rectangles). Other proteins (green pentagon) may associate with Su(s)-Wdr82 complex and recognize GUA-rich regulatory elements in the *Dm88* sequences contained in Region 1. RNAs that have been terminated prematurely downstream of binding sites are indicated. These RNAs have heterogeneous 3' ends, are short and unstable. (B) When regulatory binding sites are removed or mutated the Su(s)-complex interaction is prevented, and transcripts end after a canonical PAS. These RNAs are longer, stable and have discrete 3' ends.

The details of the mechanism by which these unstable transcripts are polyadenylated and terminated have not yet been sorted out. However, the sequence elements defined here are likely to be involved in this process. One interesting possibility is that Su(s)-Wdr82 mediates cleavage and polyadenylation at noncanonical sites by a mechanism that involves Su(s) binding to U-rich elements and interactions between Su(s)

and cleavage and polyadenylation factors that bind to auxiliary polyadenylation elements. For example, in the canonical polyadenylation pathway, UGUA elements are recognized by a subunit of Cleavage Factor I_m (CFI_m) (13,85), and four copies of this sequence are found within the GUA-rich element. In addition, the recognition of G-rich elements by hnRNP H or CFII can stimulate polyadenylation in mammals (86,87). So perhaps, in the promoter-proximal region, these sequences induce cleavage and polyadenylation in the absence of a canonical poly(A) signal, and inefficient poly(A) tail addition leads to the degradation of these RNAs. Alternatively, since Pol II pausing and transcriptional arrest can occur at U-rich and G-rich elements (88,89), perhaps one or more of these regulatory elements interferes with transcription elongation. In this case, Pol II backtracking at stalled sites might expose the 3' end of the RNA transcript, providing an entry point for the exonuclease activity of the nuclear exosome (90). Thus, perhaps Su(s)-Wdr82 induces transcriptional stalling in the vicinity of these sequence elements, and the stalled elongation complexes are removed from the DNA in a process that involves polyadenylation and RNA degradation. These possibilities will be explored in future investigations.

CHAPTER III: DISCUSSION AND FUTURE DIRECTIONS

Through the use of *Hsp70-αβ* elements as a model for the Su(s) mediated negative regulation of transposon-containing transcripts, I have demonstrated that regulation by Su(s) involves its direct binding to sequences in the promoter-proximal region. Namely, U-rich, and G-rich sequences in the *invader1* LTR (Region 2) are bound by Su(s) *in vitro* and loss of these bindings sites negatively affects regulation. These findings have advanced the understanding of the regulation mediated by Su(s) by establishing that its RNA binding-activity is biologically relevant.

Further Analysis of the ARMs

In the context of the N-terminal portion of Su(s), *in vitro* binding and stable complex formation with *αβ* transcripts requires the ARMs. Su(s) binds to low-complexity U-rich and G-rich sequences. Previous analysis of the ARMs demonstrated that ARM1 conferred sequence specificity and bound to consensus SELEX RNAs while ARM2 bound to non-consensus RNAs (45), thus, *in vitro*, the ARMs appear to recognize different sequences. However, *in vivo*, a full-length Su(s) protein that contained a deletion of ARM1 or ARM2 maintained its ability to regulate the mutant *vermilion* transcript and loss of regulation only occurred when both ARMs were deleted (46). Perhaps, *in vivo*, the ARMs have a redundant function or other regions of the protein are involved in the regulation and compensate for the loss of a single ARM. Interestingly, a fusion construct that contained only the 25 aa ARM1 or ARM2 did not bind RNA, supporting the possibility that amino acids outside of the

ARMs contribute to stable association with RNA (45). It would be interesting to determine if ARM1 and ARM2 work together as a larger unit to recognize the U-rich and G-rich sequences or if they work independently and recognize different sequences. To examine if the ARMs work as a cohesive unit, deletion mutants of ARM1 or ARM2 could be used to determine the effect on binding affinity to *αβ* transcripts. If they work together then loss of either ARM should reduce binding affinity to both U-rich and G-rich regions of *αβ*, yet, if they work independently, then one deletion would affect U-rich binding and another would affect G-rich. Additionally, RNA protection assays (91) could be used with WT and ARM deletion mutants. The WT would protect both the U-rich and G-rich regions, while the double ARM deletion would not protect either region. If the ARMs work together then deletion of either would leave the RNA unprotected but if they work independently then the U-rich would be protected in one deletion and the G-rich region would be protected with the other deletion.

The ARMs of Su(s) could have an additional function as a protein-protein interaction domain. For instance, the high molecular weight protein/RNA complex or aggregate that formed in a G-rich sequence-dependent manner during EMSA analysis is dependent on the ARMS. Perhaps the ARMs permit self-association of Su(s) along the transcript in a sequence-dependent manner. For example, the ARM of the HIV-1 viral protein Rev is important in the Rev-Rev oligomerization that occurs to allow unspliced viral RNAs to exit the nucleus (92). However, the oligomerization of Rev is dependent on RNA binding (93), indicating that binding orients the protein in the proper position to facilitate oligomerization (94). Perhaps like Rev, when Su(s) binds to G-rich sequences a conformational change is induced that promotes ARM-dependent self-association. The

ARM deletion mutants could be used to determine if the large protein/RNA aggregate is dependent on the presence of both ARMs. Also, directly labeling the protein (95,96) could delineate if multiple proteins are present in the higher molecular-weight complexes observed in EMSAs and if ARM deletion mutants affects the formation of these higher molecular-weight complexes.

Further Analysis of the CCCH-ZFs

The CCCH-ZFs of Su(s) do not appear to contribute to stable binding and complex formation with *αβ RNAs in vitro*, yet the ZFs are important for regulation of *αβ* RNAs *in vivo* (47). PCIF analysis demonstrated that mutation of the ZFs resulted in loss of localization of Su(s) to the 87C heat-shock locus (47). In addition to binding to RNA, the CCCH-ZFs of TTP mediate interactions with various proteins including PAP, PABPN1 and Nup214 (97). Perhaps Su(s) is recruited to the *αβ* elements by a protein-protein interaction between Wdr82 and the ZFs, similar to the proposed recruitment of Setd1A to the promoter region via an interaction with Wdr82 and the RRM of Setd1A (65). To determine if the ZFs are required for the interaction with Wdr82 co-IP experiments could be performed that compare association of WT or a ZF-mutant Su(s) protein with Wdr82. If the ZFs are important for the Wdr82-Su(s) interaction then their mutation would reduce or eliminate the presence of Wdr82 in the IP.

Furthermore, the CCCH-ZFs of the putative Su(s) ortholog, CPSF-30/CLIPPER, are able to bind and cleave RNA hairpins (58) and this endonuclease activity is also observed in the *Arabidopsis* ortholog of CPSF-30 (98). Thus, another possible function of the ZFs could be to cleave RNA. Interestingly, most of the Su(s) dependent promoter-proximal premature-termination sites happened a short distance downstream of the U-rich or G-rich

binding sites (49). Perhaps binding to U-rich or G-rich sequences by the ARMs properly orients and activates the ZFs to cleave the RNA downstream of the binding site. To examine if the ZFs of Su(s) have endonuclease activity, *in vitro* nuclease activity assays could be performed (98). A highly purified N-terminal portion of Su(s) could be incubated with *αβ* RNA and a time course could be used to examine the stability of the RNA in the presence of the WT, ARM deletion or the ZF-mutant. If stable binding is required prior to cleaving the RNA, then the ARM deletion would have intact RNA and the same would be expected for the ZF-mutant. In the WT sample the RNA would show heterogeneous smaller RNA fragments being generated over time, an indication that the RNA is being cleaved.

Possible Secondary Structure within *αβ* RNAs

Su(s) binds to U-rich and G-rich sequences, however, are these sequences required in a linear context or are they present in the context of a secondary structure such as an RNA hairpin? Su(s) binding *in vitro* requires the ARMs (45) and loss of the ARMs affects regulation *in vivo* (45-47). As mentioned earlier, ARMs often recognize RNA hairpins (99). Additionally, ARMs are adaptable and display different conformations based on the RNA target (99,100), which could allow recognition of various sequences and structures. Furthermore, RNAfold (101) software predicts that Region 2 can form a stable RNA hairpin and deletion of the U-rich or G-rich region destabilizes the predicted hairpin (Figure 15). Yet, *in vitro* binding assays showed that Su(s) binds to the 5' or 3' portion of the predicted hairpin (*αβ*-2 or *αβ*-3, respectively) with a high affinity and the combination fragment that encompasses the entire predicted hairpin (*αβ*-2+3, +120 to +278) did not increase the binding affinity. Thus, a hairpin is not required for high affinity binding *in vitro*. However, previous reporter constructs demonstrated that a region that contains only the U-rich

sequences of *invader1* (+135 to +205) or only the G-rich sequences (+187 to +336) were not efficiently regulated, yet a segment that contains both sequences (+113 to +278) required to form the predicted hairpin is highly regulated (49). Also, the sequences required for the Su(s) regulation are localized to the LTR (43,47), and complex structures may be a common feature of transcripts derived from the LTRs of retroviruses or retrotransposons (102,103). Thus, while it is quite possible that Su(s) binds to multiple linear *cis*-sequence elements in the promoter-proximal region to induce termination, it is also possible that a stable hairpin in the promoter proximal region is an aberrant feature that is recognized by Su(s).

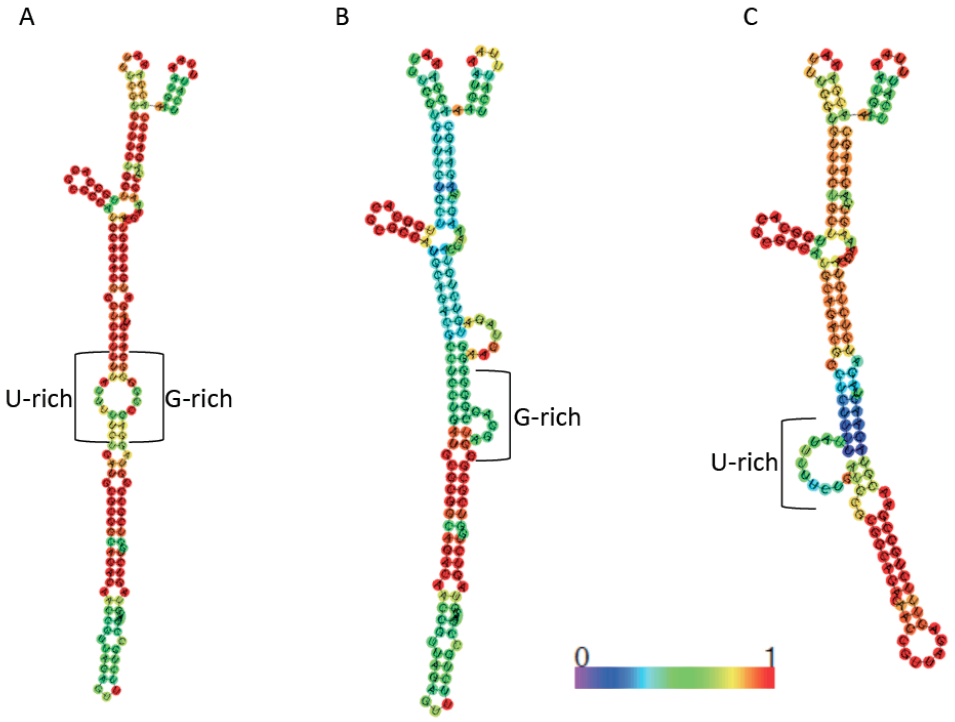


Figure 15. RNAfold analysis of the secondary structure of Region 2 of *αβ* RNA. (A) The RNAfold predicted structure of the WT *invader1* sequences (+113 to +278) contained within Region 2 of *αβ* RNA. The U-rich and G-rich Su(s) binding sites are indicated. (B) RNAfold analysis of the effect of the U-rich deletion (+171 to +183) or (C) the G-rich deletion (+225 to +249) on the predicted secondary structure in the context of Region 2. The base-pairing probability scale is shown.

To explore the possibility that *αβ* RNAs form hairpin structures *in vitro*, Selective 2' Hydroxyl Acytylation analyzed by Primer Extension (SHAPE) analysis could be utilized (104,105). SHAPE examines the local flexibility of RNA based on the reactivity of accessible 2'-OH groups to N-methylisatoic anhydride (NMIA). Regions of flexibility within the RNA structure, such as loops, bulges and junctions are more reactive thereby elucidating sites that are constrained by RNA-RNA base pairing or single-stranded, unconstrained regions (105). Additionally, SHAPE could be performed in presence of Su(s) to determine how protein binding affects the flexibility of the RNA structure and where the protein binds in the context of secondary structure (106).

Analysis of the *Dm88* Region 1

The *Dm88* (Region1) of αβ RNA is regulated by Su(s) independent of the downstream *invader1* sequences (49). On the other hand, Su(s) binds relatively weakly to this segment (*αβ*-1) *in vitro* and does not crosslink to Region1 *in vivo*. Furthermore, unlike the regulatory sequences that are Su(s) binding sites in *invader1*, regulatory sequences in Region1 do not appear to be binding sites for Su(s) as their mutation did not affect binding. Thus the regulation of Region1 must involve the interaction of Su(s) with an unidentified protein that binds to this region. To identify proteins that interact with Dm88 Region1, modified RNA pull-down assays could be used (107). In vitro transcribed wt and mutant *Dm88* transcripts could be biotinylated and incubated with nuclear lysate from S2 cells. Using Streptaviden beads, the Dm88/protein complexes could be isolated and subjected to Mass Spectrometry. Proteins present in the wt samples but absent from the mutant RNA pull down could be depleted by RNAi in S2 cells to determine if they contribute to regulation of *Dm88*.

Identification of Global Targets of Su(s)

Using CLIP and PAR-CLIP, I demonstrated that Su(s) specifically and directly binds to *αβ* transcripts *in vivo*. Having demonstrated proof of principle that we can detect Su(s)/RNA interaction *in vivo*, it would be interesting examine the global targets of Su(s) by identifying all RNAs that co-IP with Su(s). By generating a cDNA library of the precipitated RNAs it would be possible to identify sites of crosslinking, strand biases and the types of genes that are subject to Su(s) regulation (80). Given our current understanding of the sequences bound by Su(s), it would be predicted that Su(s) would crosslink to, or within close proximity to, U-rich and G-rich sequences (45). It might also be expected that Su(s) could have targets that map to both the sense and antisense strands. This could be expected as the spontaneous suppressible alleles of *vermilion*, *purple* and *speck* all contain *412* insertions that are antisense with respect to the gene to which it inserts while the *αβ* elements are sense with respect to the *Hsp70* promoter (39,47). Targets of Su(s) might include RNAs derived from the LTRs of retrotransposons or retroviruses in the promoter proximal region, or from antisense P-elements such as the element found in the *yellow* gene that is regulated by Su(s) (44). Also, normal genes such as the developmentally regulated short-intronless genes *ng1* and *sgs4* might be identified (47). However technical difficulties could be encountered as RNAs regulated by Su(s) are rapidly degraded by the nuclear exosome (47). Thus parallel experiments would need to be performed in the presence or absence of the enzymatically active nuclear exosome components to select for RNAs that are stabilized in its absence and enriched in the Su(s)-IP.

The Su(s)-Wdr82 Regulatory Pathway

Su(s) directly interacts with Wdr82 and both of these proteins are required for the regulation of *αβ* transcripts (49). Additionally, this interaction appears to be conserved as Wdr82 interacts with the putative human ortholog of Su(s), ZC3H4 (C19orf7) (66). Furthermore, the recently reported role for Wdr82 in restricting the transcription of eRNAs to the promoter-proximal region via a mechanism that involves enforced termination (38) may be similar to the induced promoter-proximal termination of *αβ* transcripts that is observed in a Su(s)-Wdr82 dependent manner (49). Su(s) and ZC3H4 are similar in size at 1325 and 1305 aa, respectively, and both proteins are predicted to be intrinsically disordered (49). If ZC3H4 is indeed an ortholog of Su(s), then it could be hypothesized that ZC3H4 would also bind and regulate its targeted RNAs. However, because the similarity is based on the ZFs, which do not mediate RNA binding in Su(s), it is also possible that ZC3H4 has a considerably different function than Su(s) and may not interact with RNA.

To examine if ZC3H4 interacts with RNA and regulates RNAs, the *αβ* RNAs could be used as a model. Using EMSA analysis and MBP fusions of different regions of ZC3H4, it would be possible to ask if ZC3H4 binds and forms stable complexes with *αβ* RNA *in vitro*. Intriguingly, ZC3H4 is enriched for RG/RGG motifs just upstream of the CCCH-ZF domains, a similar organization as the ARMs and CCCH-ZFs in Su(s) (45). RG/RGG motifs have been shown to mediate RNA binding, protein-protein interaction and nuclear localization (108). To examine regulatory role of ZC3H4 *in vivo*, previously generated *LacZ* reporter constructs could be expressed in HEK293 cells in the presence or absence of dsRNA targeting ZC3H4 to determine if it regulates *αβ* RNA, and whether U-rich and G-rich sequences are required for this regulation. Finally, to determine if ZC3H4 or Su(s) can functionally compensate for

each other, it could be asked if ZC3H4 is knocked down in HEK293 cells could Su(s) rescue the regulation of transiently expressed *αβ* RNAs? Additionally, the reciprocal experiment would ask if Su(s) is knocked down in S2-cells could ZC3H4 rescue the regulation of endogenous *αβ* RNAs? These experiments would demonstrate that Su(s) and ZC3H4 share a common biological function with respect to *αβ* RNA. (108)

APPENDIX 1

Table A1-1: Primers used to make substrates for *in vitro* RNA binding assays

Amplified Region	Primer Sequence (5' to 3')
Hsp70 (+2 to +69)	Forward TAATACGACTCACTATAGGGTTCAATTCAAACAAGCAAAGTGAAC Reverse CCTTCTGCGCTTGTTTATTTGC
αβ-1 (+53 to + 128)	Forward TAATACGACTCACTATAGGGTAAACAAGCGCAGAAGGAGTG Reverse CGTTTCATTTAAATGATCTGCGAC
αβ-2 (+120 to +201)	Forward TAATACGACTCACTATAGGGAATGAAACGAAATTTCGTGTTTCTG Reverse GGTTGTCTGCCCGCGATCAGAA
αβ-3 (+202 to +278)	Forward TAATACGACTCACTATAGGGGTTAGAGTTTCTGCCGAACGTAGTC Reverse GCTTCTCGCTTTCGTACAGAC

Table A1-2: Primers used for RIP, CLIP, and PAR-CLIP analysis

Primer Region	Primer Sequences (5' to 3')
Hsp70-αβ (+53 to +72) *Hsp70-αβ* (+258 to +278)	Forward - TAAACAAGCGCAGAAGGAGT Reverse - GCTTCTCGCTTTCGTACAGAC
Hsp70 (+29 to +49) *Hsp70* (+268 to +287)	Forward - GTCGCTAAGCGAAAGCTAAGC Reverse - CACGCAGGAGTAGGTGGTGC

Table A1-3: Synthetic oligomers used to generate the *invader1* reporter constructs

Construct	Oligomer Sequence (5' to 3')
WT (+113 to +209), ΔG, Δ230-23, Δ240-249 Forward	TCA TTT AAA TGA AAC GAA ATT TCG TGT TTC TGC TTG GCA CGC GCC ATG CAG ACG CCT CTT TTT ATT TTT CTG ATG CGC GGC AGA CAA CCG TTA GAG T
ΔU13 Forward	AAA TAA ACA AGC GCA GAA GGT CAT TTA AAT GAA ACG AAA TTT CGT GTT TCT GCT TGG CAC GCG CCA TGC AGA CGC CTC GAT GCG CGG CAG ACA ACC GTT AGA GT
WT (+184 to +278) Reverse	GC TTC TCG CTT TCG TAC AGA CAT CTA CTT CCC CCC GCT CCT ACC CGC GAC CAG ACT ACG TTC GGC AGG GGC TCT AAC GGT TGT CTG CCG CGC ATC
ΔG Reverse	G CTT CTC GCT TTC GTA CAG ACA TCT ACT TCT ACG TTC GGC AGA AAC TCT AAC GGT TGT CTG CCG
Δ230-239 Reverse	G CTT CTC GCT TTC GTA CAG ACA TCT ACT TCC CCC CGC TCC CAG ACT ACG TTC GGC AGA AAC TCT AAC GGT TGT CTG CCG
Δ240-249 Reverse	G CTT CTC GCT TTC GTA CAG ACA TCT ACT TCT ACC CGC GAC CAG ACT ACG TTC GGC AGA AAC TCT AAC GGT TGT CTG CCG

Table A1-4: Synthetic oligomers used to generate the *Dm88* reporter constructs

Construct	Oligomer Sequence (5' to 3')
HDL-3.1 Forward	GCA AAT AAA CAA GCG CAG AAG GAG TGT TGA AAA TAT ACA TAC AAT ATA CAT ACA GTC AAT GCA CTG TGT CTC CCT CTT TTG GTC GCG GTA AC
HDL-3.2 Forward	GCA AAT AAA CAA GCG CAG AAG GAG TGT TGA AAA TAT ACA TAC AAT ATG TAT GTA GTC AAT GCA CTG TGT CTC CCT CTT TTG GTC GCG GTA AC
HDL-3.3 Forward	GCA AAT AAA CAA GCG CAG AAG GAG TGT TGA AAA TAT GTA TGT AAT ATA CAT ACA GTC AAT GCA CTG TGT CTC CCT CTT TTG GTC GCG GTA AC
HDL-3.4 Forward	GCA AAT AAA CAA GCG CAG AAG GAA CAC TGA AAA TAT GTA TGT AAT ATG TAT GTA ACC AAT GCA CTA CAC CTC CCT CTT TTG ACC GCG ACA AC
HDL-3.5 Reverse	GCA AAT AAA CAA GCG CAG AAG GAA CAC TGA AAA TAT ACA TAC AAT ATA CAT ACA ACC AAT GCA CTA CAC CTC CCT CTT TTG ACC GCG ACA AC
HDL-3.1, 3.2, 3.3 Reverse	CAG GTA CCA CTG CAC GCA GGC AGC CAA ATT ACA CGC TAA AGA GGA TAA CAC AAT AAG AGA AAA AAG CTT TTG GTT ACC GCG ACC AAA AGA G
HDL-3.4, 3.5 Reverse	CAG GTA CCG TTG CGT GCA GGC AGC CAA ATT ACG TGC TAA AGA GGA TAG TGT AAT AAG AGA AAA AAG CTT TTG GTT GTC GCG GTC AAA AGA G

REFERENCES

1. Lunde, B.M., Moore, C. and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology*, **8**, 479-490.

2. Chapman, R.D., Heidemann, M., Hintermair, C. and Eick, D. (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet.*, **24**, 289-296.

3. Hsin, J.P. and Manley, J.L. (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.*, **26**, 2119-2137.

4. Allen, B.L. and Taatjes, D.J. (2015) The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology*, **16**, 155-166.

5. Licatalosi, D.D. and Darnell, R.B. (2010) RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.*, **11**, 75-87.

6. Cowling, Victoria H. (2010) Regulation of mRNA cap methylation. *Biochem. J.*, **425**, 295-302.

7. Lenasi, T., Peterlin, B.M. and Barboric, M. (2011) Cap-binding Protein Complex Links Pre-mRNA Capping to Transcription Elongation and Alternative Splicing through Positive Transcription Elongation Factor b (P-TEFb). *J. Biol. Chem.*, **286**, 22758-22768.

8. Matera, A.G. and Wang, Z. (2014) A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*, **15**, 108-121.

9. Merkhofer, E.C., Hu, P. and Johnson, T.L. (2014) Introduction to Cotranscriptional RNA Splicing. *Methods Mol Biol.*, **1126**, 83-96.

10. Hirose, Y., Tacke, R. and Manley, J. (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.*, **13**, 1234-1239.

11. Fu, X.-D. and Ares, M. (2014) Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.*, **15**, 689-701.

12. Xiang, K., Tong, L. and Manley, J.L. (2014) Delineating the Structural Blueprint of the Pre-mRNA 3'-End Processing Machinery. *Mol. Cell. Biol.*, **34**, 1894-1910.

13. Venkataraman, K., Brown, K.M. and Gilmartin, G.M. (2005) Analysis of a noncanonical polyA site reveals a tripartite mechanism for vertebrate polyA site recognition. *Genes Dev.*, **19**, 1315-1327.

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

15. Richard, P. and Manley, J.L. (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev.*, **23**, 1247-1269.

16. Perales, R. and Bentley, D. (2009) "Cotranscriptionality" - The Transcription Elongation Complex as a Nexus for Nuclear Transactions. *Mol. Cell*, **36**, 178-191.

17. West, S., Gromak, N. and Proudfoot, N.J. (2004) Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, **432**, 522-525.

18. Eberle, A.B. and Visa, N. (2014) Quality control of mRNP biogenesis: Networking at the transcription site. *Semin. Cell Dev. Biol.*, **32**, 37-46.

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

20. Hackmann, A., Wu, H., Schneider, U.-M., Meyer, K., Jung, K. and Krebber, H. (2014) Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. *Nature Communications*, **5**.

21. Herve Le Hir, D.G., Elisa Izaurralde, Melissa J. More. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *The EMBO Journal*, **20**, 4987-4997.

22. Davidson, L., Kerr, A. and West, S. (2012) Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *The EMBO Journal*, **31**, 2566-2578.

23. Kilchert, C., Wittmann, S. and Vasiljeva, L. (2016) The regulation and functions of the nuclear RNA exosome complex. *Nature Reviews Molecular Cell Biology*.

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

25. Milligan, L., Torchet, C., Allmang, C., Shipman, T. and Tollervey, D. (2005) A Nuclear Surveillance Pathway for mRNAs with Defective Polyadenylation. *Mol. Cell. Biol.*, **25**, 9996-10004.

26. Jacquier, A. (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.*, **10**, 833-844.

27. Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L. *et al.* (2007) RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science*, **316**, 1484-1488.

28. Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B. *et al.* (2005) Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. *Cell*, **121**, 725-737.

29. Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W. and Steinmetz, L.M. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature*, **457**, 1033-1037.

30. Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R. *et al.* (2013) Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. & Mol. Biol.*, **20**, 923-928.

31. Jensen, Torben H., Jacquier, A. and Libri, D. (2013) Dealing with Pervasive Transcription. *Mol. Cell*, **52**, 473-484.

32. Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182-187.

33. Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D. and Corden, J.L. (2004) Identification of cis Elements Directing Termination of Yeast Nonpolyadenylated snoRNA Transcripts. *Mol. Cell. Biol.*, **24**, 6241-6252.

34. Vasiljeva, L. and Buratowski, S. (2006) Nrd1 Interacts with the Nuclear Exosome for 3′ Processing of RNA Polymerase II Transcripts. *Mol. Cell*, **21**, 239-248.

35. Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M. *et al.* (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. & Mol. Biol.*, **20**, 1367-1376.

36. Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L. and Dreyfuss, G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, **468**, 664-668.

37. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. and Sharp, P.A. (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, **499**, 360-363.

38. Austenaa, Liv M.I., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, Britta A.M., de Pretis, S. *et al.* (2015) Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Mol. Cell*, **60**, 460-474.

39. Searles, L.L. and Voelker, R. (1986) Molecular characterization of *Drosophila* vermillion locus and its suppressible alleles. *Proc. Natl. Acad. Sci.*, **83**, 404-408.

40. Searles, L.L., Ruth, R.S., Pret, A.-M., Fridell, R.A. and Ali, A.J. (1990) Structure and Transcription of the Drosophila melanogaster Vermilion Gene and Several Mutant Alleles. *Mol. Cell. Biol.*, **10**, 1423-1431.

41. Fridell, R.A., Pret, A.M. and Searles, L.L. (1990) A retrotransposon 412 insertion within an exon of the Drosophila melanogaster vermilion gene is spliced from the precursor RNA. *Genes Dev.*, **4**, 559-566.

42. Pret, A. and Searles, L.L. (1991) Splicing of Retrotransposon Insertions from Transcripts of the Drosophila melanogaster vermilion Gene in a Revertant. *Genetics*, **129**, 1137-1145.

43. Fridell, R. and Searles, L.L. (1994) Evidence for a Role of the *Drosophila melanogaster suppressor of sable* Gene in the Pre-mRNA Splicing Pathway. *Mol. Cell. Biol.*, **14**, 859-867.

44. Geyer, P.K., Chien, A.J., Corces, V.G. and Green, M.M. (1991) Mutations in the su(s) gene affect RNA processing in Drosophila melanogaster. *Proc. Natl. Acad. Sci.*, **88**, 7116-7120.

45. Turnage, M., Brewer-Jensen, P., Bai, W. and Searles, L.L. (2000) Arginine-Rich Regions Mediate the RNA Binding and Regulatory Activities of the Protein Encoded by the *Drosophila melanogaster suppressor of sable* Gene. *Mol. Cell. Biol.*, **20**, 8198-8208.

46. Kuan, Y.S., Brewer-Jensen, P. and Searles, L.L. (2004) Suppressor of sable, a Putative RNA-Processing Protein, Functions at the Level of Transcription. *Mol. Cell. Biol.*, **24**, 3734-3746.

47. Kuan, Y.S., Brewer-Jensen, P., Bai, W., Hunter, C., Wilson, C., Bass, S., Abernethy, J., Wing, J. and Searles, L.L. (2009) Drosophila Suppressor of Sable Protein [Su(s)] Promotes Degradation of Aberrant and Transposon-Derived RNAs. *Mol. Cell. Biol.*, **29**.

48. Lis, J.T., Ish-Horowicz, D. and Pinchin, S.M. (1981) Genomic organization and transcription of the aB heat shock DNA in Drosophila melanogaster. *Nucleic Acids Res.*, **9**, 5297-5310.

49. Brewer-Jensen, P., Wilson, C.B., Abernethy, J., Mollison, L., Card, S. and Searles, L.L. (2016) Suppressor of sable [Su(s)] and Wdr82 down-regulate RNA from heat-shock-inducible repetitive elements by a mechanism that involves transcription termination. *RNA*, **22**, 139-154.

50. Chang, D.-Y., Wisely, B., Huang, S.-M. and Voelker, R. (1986) Molecular Cloning of Suppressor of Sable, a Drosophila melanogaster Transposon-Mediated Suppressor. *Mol. Cell. Biol.*, **6**, 1520-1528.

51. Voelker, R., Gibson, W., Graves, J., Sterling, J. and Eisenburg, M. (1991) The *Drosophila* Suppressor of Sable Gene Encodes a Polypeptide with Regions Similar to Those of RNA Binding Proteins. *Mol. Cell. Biol.*, 894-905.

52. Murray, M., Turnage, M., Williamson, K., Steinhauer, W. and Searles, L.L. (1997) The Drosophila Suppressor of sable Protein Binds to RNA and Associates with a Subset of Polytene Chromosome Bands. *Mol. Cell. Biol.*, **17**, 2291-2300.

53. Mellow, C., Schubert, C., Draper, B., Zhang, W., Lobel, R. and Priess, J. (1996) The PIE-1 protein and germline specification in C. elegans embryos. *Nature*, **382**, 710 -712.

54. Lai, W.S., Carballo, E., Strum, J., Kennington, E., Phillips, R. and Blackshear, P.J. (1999) Evidence that Tristetraprolin Binds to AU-Rich Elements and Promotes Deadenylation and Destabilization of Tumor Necrosis Factor Alpha mRNa. *Mol. Cell. Biol.*, **19**, 4311-4323.

55. Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Struct. & Mol. Biol.*, **11**, 257-264.

56. Bai, C. and Tolias, P. (1998) Drosophilia clipper/CPSF 30K is a post-transcriptionally regulated nuclear protein that binds RNA containing GC clusters. *Nucleic Acids Res.*, **26**, 1597-1604.

57. Kralovicova, J., Knut, M., Cross, N.C.P. and Vorechovsky, I. (2015) Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3' splice-site organization and activity of U2AF-related proteins. *Nucleic Acids Res.*, **43**, 3747-3763.

58. Bai, C. and Tolias, P. (1996) Cleavage of RNA Hairpins Mediated by a Developmentally Regulated CCCH Zinc Finger Protein. *Mol. Cell. Biol.*, **16**, 6661-6667.

59. Carman, J.A. and Nadler, S.G. (2004) Direct association of tristetraprolin with the nucleoporin CAN/Nup214. *Biochem. Biophys. Res. Commun.*, **315**, 445-449.

60. Casu, F., Duggan, Brendan M. and Hennig, M. (2013) The Arginine-Rich RNA-Binding Motif of HIV-1 Rev Is Intrinsically Disordered and Folds upon RRE Binding. *Biophys. J.*, **105**, 1004-1017.

61. Calnan, B.J., Biancalana, S., Hudson, D. and Frankel, A.D. (1991) Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes Dev.*, **5**, 201-210.

62. Kaufmann, I., Martin, G., Friedlein, A., Langen, H. and Keller, W. (2004) Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *The EMBO Journal*, **23**, 616-626.

63. Cheng, H., He, X. and Moore, C. (2004) The Essential WD Repeat Protein Swd2 Has Dual Functions in RNA Polymerase II Transcription Termination and Lysine 4 Methylation of Histone H3. *Mol. Cell. Biol.*, **24**, 2932-2943.

64. Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C. and Greenblatt, J. (2003) Organization and Function of APT, a Subcomplex of the Yeast Cleavage and Polyadenylation Factor Involved in the Formation of mRNA and Small Nucleolar RNA 3'-Ends. *J. Biol. Chem.*, **278**, 33000-33010.

65. Lee, J.H. and Skalnik, D.G. (2007) Wdr82 Is a C-Terminal Domain-Binding Protein That Recruits the Setd1A Histone H3-Lys4 Methyltransferase Complex to Transcription Start Sites of Transcribed Human Genes. *Mol. Cell. Biol.*, **28**, 609-618.

66. Lee, J.H., You, J., Dobrota, E. and Skalnik, D.G. (2010) Identification and Characterization of a Novel Human PP1 Phosphatase Complex. *J. Biol. Chem.*, **285**, 24466-24476.

67. Schmid, M. and Jensen, T.H. (2013) Transcription-associated quality control of mRNP. *Biochim. Biophys. Acta*, **1829**, 158-168.

68. Mühlemann, O. and Jensen, T.H. (2012) mRNP quality control goes regulatory. *Trends Genet.*, **28**, 70-77.

69. Kilchert, C. and Vasiljeva, L. (2013) mRNA quality control goes transcriptional. *Biochem. Soc. Trans.*, **41**, 1666-1672.

70. Doma, M.K. and Parker, R. (2007) RNA Quality Control in Eukaryotes. *Cell*, **131**, 660-668.

71. Ascano, M., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M. *et al.* (2012) FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature*, **492**, 382-386.

72. Piacentini, L., Fanti, L., Negri, R., Del Vescovo, V., Fatica, A., Altieri, F. and Pimpinelli, S. (2009) Heterochromatin Protein 1 (HP1a) Positively Regulates Euchromatic Gene Expression through RNA Transcript Association and Interaction with hnRNPs in Drosophila. *PLoS Genetics*, **5**, e1000670.

73. Ule, J., Jensen, K., Ruggiu, M., Mele, A. and Darnell, R. (2005) CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods*, **37**, 376-386.

74. Melvin, W., Milne, A., Slater, A., Allen, H. and Keir, H. (1978) Incorporation of 6-Thioguanosine and 4-Thiouridine into RNA - Application to Isolation of Newly Synthesised RNA by Affinity Chromatography. *Eur. J. Biochem.*, **92**, 373-379.

75. Dolken, L., Ruzsics, Z., Radle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P. *et al.* (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, **14**, 1959-1972.

76. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M. *et al.* (2010) Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, **141**, 129-141.

77. Darnell, R. (2012) CLIP (Cross-Linking and Immunoprecipitation) Identification of RNAs Bound by a Specific Protein. *Cold Spring Harb. Prot.*, **2012**, 1146-1160.

78. Wang, B., Steain, M.C., Dwyer, D.E., Cunningham, A.L. and Saksena, N.K. (2011) Synthetic long oligonucleotides to generate artificial templates for use as positive controls in molecular assays: drug resistance mutations in influenza virus as an example. *Virology J.*, **8**, 1-7.

79. Ryder, S.P., Recht, M.I. and Williamson, J.R. (2008) Quantitative Analysis of Protein-RNA Interactions by Gel Mobility Shift. *Methods Mol Biol.*, **488**, 99-115.

80. Wang, T., Xiao, G., Chu, Y., Zhang, M.Q., Corey, D.R. and Xie, Y. (2015) Design and bioinformatics analysis of genome-wide CLIP experiments. *Nucleic Acids Res.*, **43**, 5263-5274.

81. Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R. and Hippel, P.H.v. (1986) Laser Cross-linking of Nucleic Acids to Proteins. *J. Biol. Chem.*, **261**, 3512-3518.

82. Konig, J., Zarnack, K., Luscombe, N.M. and Ule, J. (2012) Protein-RNA interactions-new genomic technologies and perspectives. *Nat. Rev. Genet.*, **13**, 77-83.

83. Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727-4741.

84. Kostadinov, R., Malhotra, N., Viotti, M., Shine, R., D'Antonio, L. and Bagga, P.S. (2006) GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.*, **34**, D119-D124.

85. Yang, Q., Gilmartin, G.M. and Doublié, S. (2011) The structure of human Cleavage Factor Im hints at functions beyond UGUA-specific RNA binding. *RNA Biology*, **8**, 748-753.

86. Millevoi, S. and Vagner, S. (2009) Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res.*, **38**, 2757-2774.

87. Arhin, G.K., Boots, M., Bagga, P.S., Milcarek, C. and 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.*, **30**, 1842-1850.

88. Keene, R.G., Mueller, A., Landick, R. and London, L. (1999) Transcriptional pause, arrest and termination sites for RNA polymerase II in mammalian N- and c-myc genes. *Nucleic Acids Res.*, **27**, 3173-3182.

89. Gromak, N., West, S. and Proudfoot, N.J. (2006) Pause Sites Promote Transcriptional Termination of Mammalian RNA Polymerase II. *Mol. Cell. Biol.*, **26**, 3986-3996.

90. Lemay, J.-F., Larochelle, M., Marguerat, S., Atkinson, S., Bähler, J. and Bachand, F. (2014) The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat. Struct. & Mol. Biol.*, **21**, 919-926.

91. Pearson, L., Chen, C., Gaynor, R. and Sigman, D. (1994) Footprinting RNA - protein complexes following gel retardation assays- application to the R-17-procoat-RNA and tat-TAR interactions. *Nucleic Acids Res*, **22**, 2255-2263.

92. Zapp, M., Hope, T., Parslow, T. and Green, M. (1991) Oligomerization and RNA binding domains of the type1 human immunodeficiency virus Rev protein- A dual function for an arginine-rich binding motif. *Proc. Natl. Acad. Sci.*, **88**, 7734-7738.

93. Pond, S.J.K., Ridgeway, W.K., Robertson, R., Wang, J. and Millar, D.P. (2009) HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proc. Natl. Acad. Sci.*, **106**, 1404-1408.

94. Daugherty, M.D., Liu, B. and Frankel, A.D. (2010) Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. *Nat. Struct. & Mol. Biol.*, **17**, 1337-1342.

95. Daly, T., Doten, R., Rennert, P., Auer, M., Jaksche, H., Donner, A., Fisk, G. and Rusche, J.R. (1993) Biochemical Characterization of Binding of Multiple HIV-1 Rev Monomeric Proteins to the Rev Responsive Element. *Biochemistry (Mosc).* **32**, 10497-10505.

96. Cook, K.S., Fisk, G., Hauber, J., Usman, N., Daly, T.J. and Rusche, J.R. (1991) Characterization of HIV-1 REV protein - binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res*, **19**, 1577-1583.

97. Brooks, S.A. and Blackshear, P.J. (2013) Tristetraprolin (TTP): Interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim. Biophys. Acta - Gene Reg. Mech.*, **1829**, 666-679.

98. Addepalli, B. and Hunt, A.G. (2007) A novel endonuclease activity associated with the Arabidopsis ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor. *Nucleic Acids Res.*, **35**, 4453-4463.

99. Weiss, M.N., N. (1999) RNA Recognition by Arginine-Rich Petide Motifs. *Biopolymers (Nucleic Acid Science)*, **48**, 167-180.

100. Bayer, T.S. (2005) Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA*, **11**, 1848-1857.

101. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R. and Hofacker, I.L. (2008) The Vienna RNA Websuite. *Nucleic Acids Res.*, **36**, W70-W74.

102. Ahmed, N., Benachenhou, F., Jern, P., Oja, M., Sperber, G., Blikstad, V., Somervuo, P., Kaski, S. and Blomberg, J. (2009) Evolutionary Conservation of Orthoretroviral Long Terminal Repeats (LTRs) and ab initio Detection of Single LTRs in Genomic Data. *PLoS ONE*, **4**, e5179.

103. Benachenhou, F., Sperber, G., Bongcam-Rudloff, E., Andersson, G., Boeke, J. and Blomberg, J. (2013) Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). *Mobile DNA*, **4**, 1-15.

104. Sztuba-Solinska, J. and Le Grice, S.F.J. (2012) Probing Retroviral and Retrotransposon Genome Structures: The "SHAPE" of Things to Come. *Mol. Biol. Inter.*, **2012**, 1-12.

105. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2006) Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature Protocols*, **1**, 1610-1616.

106. Legiewicz, M., Badorrek, C.S., Turner, K.B., Fabris, D., Hamm, T.E., Rekosh, D., Hammarskjold, M.L. and Le Grice, S.F.J. (2008) Resistance to RevM10 inhibition reflects a conformational switch in the HIV-1 Rev response element. *Proc. Natl. Acad. Sci.*, **105**, 14365-14370.

107. Chien, H.L., Liao, C.L. and Lin, Y.L. (2011) FUSE Binding Protein 1 Interacts with Untranslated Regions of Japanese Encephalitis Virus RNA and Negatively Regulates Viral Replication. *J. Virol.*, **85**, 4698-4706.

108. Thandapani, P., O'Connor, T.R., Bailey, Timothy L. and Richard, S. (2013) Defining the RGG/RG Motif. *Mol. Cell*, **50**, 613-623.