University of Kentucky Doctoral Dissertations   Graduate School

2011

POST-TRANSCRIPTIONAL REGULATION OF AFP AND IgM GENES

Lilia M. Turcios
University of Kentucky, lturc2@uky.edu

ABSTRACT OF DISSERTATION

Lilia M. Turcios

The Graduate School

University of Kentucky

2011

POST-TRANSCRIPTIONAL REGULATION OF AFP AND IgM GENES

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Medicine at the University of Kentucky

By

Lilia M. Turcios

Director: Dr. Martha Peterson

Lexington, KY

2011

ABSTRACT OF DISSERTATION

POST-TRANSCRIPTIONAL REGULATION OF AFP AND IgM GENES

Gene expression can be regulated at multiple steps once transcription is initiated. I have studied two different gene models, the α-Fetoprotein (AFP) and the immunoglobulin heavy chain (IgM) genes, to better understand post-transcriptional gene regulation mechanisms. The AFP gene is highly expressed during fetal liver development and dramatically repressed after birth. There is a mouse strain-specific difference between adult levels of AFP, with BALB/cJ mice expressing 10 to 20-fold higher levels compared to other mouse strains. BALB/cJ mice express low levels of Zhx2 and thus incompletely repress AFP. Despite differences in steady state AFP mRNA levels in the adult liver between Balb/cJ and wild-type mice, transcription rates across this gene were similar, indicating a post-transcriptional regulatory mechanism. I found accumulated unspliced RNA across multiple AFP introns in wild-type mice where mature AFP mRNA levels are low, suggesting overall AFP splicing is inefficient in the presence of Zhx2. The IgM gene is alternative processed to produce two mRNA isoforms through a competition between cleavage/polyadenylation (µspA) and splicing reactions and the pA/splice RNA expression ratio increases during B cell maturation. Cotranscriptional cleavage (CoTC) events, driven by specific cis-acting elements, are required downstream of some poly(A) signals to terminate transcription. In some cases, a pause site can produce similar effect. I explored whether there is a CoTC-like element within the IgM gene that may contribute to developmental changes in the mRNA ratio. In both a B cell and plasma cell line there was a gradual decrease in transcripts downstream from the µspA signal, suggesting that there is not evidence for a CoTC element within the IgM gene. To examine the effect a CoTC element would have on the competition between the splice and µspA reactions, we inserted the CoTC sequence of the β-globin gene into different locations downstream of the µspA signal. While the β-globin CoTC element caused cotranscriptional cleavage in all locations, it only affected the µspA/splice ratio when located close to the µspA site. This suggests there is a

position effect of the inserted CoTC element on the competing polyadenylation and splicing reactions within the IgM transcripts.

Lilia M. Turcios

June 29, 2011

POST-TRANSCRIPTIONAL REGULATION OF AFP AND IgM GENES

By

Lilia M. Turcios

Martha L. Peterson
Director of Dissertation

Beth Garvy
Director of Graduate Studies

June 29, 2011

DISSERTATION

Lilia M. Turcios

The Graduate School

University of Kentucky

2011

POST-TRANSCRIPTIONAL REGULATION OF AFP AND IgM GENES

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Medicine at the University of Kentucky

By

Lilia M. Turcios

Director: Dr. Martha Peterson

Lexington, KY

2011

To

Nicole and Samuel

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFP	Alpha-fetoprotein
C/EBP	CAAT/Enhancer binding protein
CBC	Cap binding complex
CFI	Cleavage factors I
CFII	Cleavage factors II
CoTC	Co-transcriptional cleavage
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulatory factor
CTD	C-Terminal domain of the RNAPII largest subunit, Rbp1
DSIF	5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole sensitivity-inducing factor
E33	Exon 33 of the Fibronectin gene
EDI	Exon 33 of the Fibronectin gene
FN	Fibronectin
Foxa	Forkhead box transcription factor a family
FTF	Fetoprotein transcription factor
GTFs	General transcription factors
HCC	Hepatocellular Carcinoma
HNF1	Hepatocyte Nuclear Factor 1
HP1γ	Heterochromatin Protein 1
Igμ	Immunoglobulin M
LPS	Lipopolysaccharides
MERV	Mouse endogenous retrovirus
MSA	Serum albumin gene
NELF	Negative elongation transcription factor
NF1	Nuclear Factor 1
Nkx2.8	NK2 homeobox 8
NMD	Nonsense-mediated decay
PGC-1	Proliferator-activated receptor γ coactivator 1
PIC	Pre-Initiation Complex
Pre-mRNA	Primary RNA transcript
PRG	Primary response genes
PTC	Pre-termination codon
pTEF-b	positive transcription elongation factor b
RNAPII	RNA Polymerase II
SMA	Spinal muscular atrophy
SMN	Survival motor neuron
SnRNPs	Small ribonucleoprotein particles
TEC	Transcription elongation complex
TNF	Tumor necrosis factor
ZBTB20	Zinc finger and BTB domain containing 20

CHAPTER I

INTRODUCTION

Changes in gene expression are critical for proper cellular differentiation during development and cell survival. It also allows adequate response to different environmental conditions. Gene expression can be modulated at different levels, including transcriptional regulation, to produce a functional mRNA. Transcription of protein-encoding-genes is a highly complex process carried out by the RNA polymerase II complex (RNAPII) which includes several steps: pre-initiation complex assembly, transcription initiation, transcription elongation and transcription termination. Each of these steps are tightly coupled to multiple steps of mRNA maturation (capping, splicing and cleavage/polyadenylation), and rigorously regulated to modulate gene expression. To better understanding the impact of RNA processing on gene expression, several transcriptional steps that are targets for gene regulation will be discussed in this thesis using two different gene models: the alpha-fetoprotein (AFP) gene to study the mechanism of post-natal repression mediated by splicing regulation and the immunoglobulin M (μ) gene to investigate the effect of co-transcriptional cleavage (CoTC) on developmental RNA processing and transcription termination. The introduction will consist of a general overview of transcription by RNAPII and how it is coupled to RNA processing followed by specific background information regarding the two genes used in this study, the AFP and the mouse μ genes.

CO-TRANSCRIPTIONAL RNA PROCESSING

The reactions involved in the maturation of mRNA are now known to often occur co-transcriptionally as the RNA is extruded from the actively transcribing RNAPII. This observation has changed the old view of eukaryotic gene expression in which the mRNA processing reactions were thought to occur post-transcriptionally, one independent of the other. The new model is more dynamic and integrated, where a cross-talk between factors involved in each of these reactions ensure accurate processing in a spatial-temporal manner which also allows for complex regulation (Bentley, 2005; Kornblihtt et al., 2004).

Coupling of transcription with RNA processing is mediated by a unique feature of the RNAPII, the C-terminal domain (CTD) of its largest subunit, Rbp1. The CTD contains multiple heptapeptide repeats (26 in yeast and 52 in mammals) with the

consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$, which is the target of regulatory phosphorylation at serines 2, 5 and 7. The phosphorylation state of the CTD ("CTD code") is a dynamic process that changes as transcription across the gene proceeds, and the CTD is thought to act as a landing pad that orchestrates the recruitment of factors involved in transcription elongation and mRNA processing. CTD phosphorylation at Ser5 occurs at early stages in transcription and promotes association with capping enzymes and promoter release. In contrast, CTD phosphorylation at Ser2 increases as RNAPII progresses through the body of the gene with a maximum peak at the 3'-end, and it is correlated with transcription elongation, cleavage and polyadenylation and transcription termination (Bentley, 2005; Kornblihtt et al., 2004)

During transcription initiation, RNAPII binds to the promoter along with the general transcription factors (GTFs) to form the Pre-Initiation Complex (PIC) (Figure 1-1, Step 1). Soon after, CTD phosphorylation at Ser5 by the GTF TFIIH takes place. Recruitment of 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole sensitivity-inducing factor (DSIF) and negative elongation transcription factor (NELF) to the transcription complex results in arrest of productive elongation (Figure 1-1, step 2). This pausing allows enough time for recruitment of capping enzymes and assembly of the transcription elongation complex (TEC). Addition of the 5'-cap, a methylated guanosine attached through an unusual 5' to 5' triphosphate linkage, is stimulated by Ser5 CTD phosphorylation and it occurs as the RNA emerges from the RNAPII when it is about 25 bases long. The 5'-cap is recognized by the cap binding complex (CBC) that in turn protects the nascent transcript from degradation by 5'-3' exonucleases and enhances mRNA translation. Promoter-proximal pausing is relieved by phosphorylation of the CTD Ser2 mediated by positive transcription elongation factor b (pTEFb), which causes dissociation of NELF and thus, allows transcription elongation to resume (Figure 1-1, step 3) (Bentley, 2002; Bres et al., 2008; Mandal et al., 2004; Proudfoot et al., 2002; Sims et al., 2004a). Once the RNAPII complex transitions to productive elongation, it is very stable and able to transcribe up to hundreds of kilobases without dissociating from the DNA template (Nechaev and Adelman, 2011). Phosphorylation of CTD at Ser2 also serves as platform for binding of factors involved in regulation of transcription elongation, RNA processing and termination as described latter.

Transcriptional pausing at the promoter-proximal region induced in the early transcriptional elongation phase, may serve as an important control for gene regulation by at least three different ways (Chiba et al., 2010): First, it may act as rate-limiting step

2

to reduce transcriptional level. Second, it contributes to the maintenance of a "poised" state where specific-gene activation and transcription initiation processes have been completed, allowing a rapid response to extracellular signals. It also may favor an open chromatin structure by preventing nucleosome formation and finally, it allows functional coupling between transcription and RNA processing ensuring that RNA capping has been completed before engaging in a productive elongation (Chiba et al., 2010; Nechaev and Adelman, 2011).

SPLICING REGULATION

Typical mammalian genes are comprised of multiple short exons (<300 bp) which are interrupted by long intervening sequences (introns). Introns must be removed from the primary transcript (pre-mRNA) to generate mRNA by a process known as splicing. About 90% of human genes undergo alternative splicing to give rise more than one mRNA isoform. Therefore, alternative splicing is an important mechanism to expand the proteomic diversity from a limited number of genes encoded by the genome. Indeed, inclusion/exclusion of alternative exons from a primary transcript may result in mRNAs that encode for protein products with different function, subcellular localization or catalytic activity, which are important for tissue specific and developmental processes (Hu and Fu, 2007; Karni et al., 2007; Luco et al., 2011). Moreover, it may introduce alternative exons or intron retention that results in production of truncated proteins or in mRNA that are target for degradation by the nonsense-mediated decay (NMD) pathway (Tazi et al., 2009).

The importance of alternative splicing regulation has been increased because mis-regulation of mRNA splicing may be involved in human disease, including cancer. Many genes involved in apoptosis are alternative spliced and the different isoforms can generate proteins with opposite function. For example, alternative usage of a 5´splice site coded within exon 2 of the *Bcl-x* gene can lead to the synthesis of either the pro-apoptotic isoform Bcl-x_S or the anti-apoptotic isoform Bcl-x_L (Mercatante et al., 2001). Additionally, the function of several tumor suppressor genes (*Brca1* and *Mdm2*) and oncogenes (*Kras* and *Wnt*), are also regulated by alternative splicing. Hence, generation of splice variants from these genes that encode an inactive tumor suppressor or an active oncogene play a role in cancer (Mercatante et al., 2001; Venables, 2004). Others genes regulated by alternative splicing include those involved in angiogenesis (*VEGF*) and cell adhesion (*CD44*) whose splicing pattern changes during tumorigenesis (Zerbe

et al., 2004). Another example is the spinal muscular atrophy (SMA) disorder caused by the absence of or mutation in the *Survival Motor Neuron 1* (*SMN1*) gene, which encodes for an essential protein involved in assembly of the spliceosome (Singh et al., 2004; Sumner, 2006). The SMN2 gene, which differs by a nucleotide within exon 7, fails to compensate loss of SMN1 gene, because the nucleotide substitution induces exon 7 skipping and production of a truncated and unstable protein (Singh et al., 2004). Increased SMN expression levels could be achieved either by modulating splicing of the SMN2 pre-mRNA, induce stabilization of the SMN protein or SMN1 gene replacement (Sumner, 2006).

Joining of exons is mediated by the spliceosome which is composed of five uridine-rich small nuclear ribonucleoprotein particles (snRNPs) called U1, U2, U4, U5 and U6, and multiple non-snRNPs splicing factors. Splicing involves the recognition of consensus sequences near the intron-exon boundaries and conserved elements in the introns (Figure 1-2A). The early steps of the spliceosome assembly involve recognition of the consensus elements at both ends of the intron to form the early (E) complex: U1 snRNP binds to the 5´splice site, the branchpoint-binding protein (BBP) binds to the branchpoint site, while the heterodimeric U2 snRNP auxiliary factor (U2AF) recognizes the polypyrimidine tract and the 3´ AG (Figure 1-2B). Next, U2 snRNP is recruited to the branch site aided by the U2AF which results in the displacement of BBP to form the A complex. The tri-snRNP particle, comprised of U4/U6/U5, then joins the complex, bringing together the two splice sites that lead to the B complex. Once the assembly of the spliceosome occurs, there is a dynamic rearrangement of interactions that promotes release of U1 and U4 snRNPs. This results in the formation of the catalytically active spliceosome known as C complex which carries out the two trans-esterification reactions leading to intron excision (Cartegni et al., 2002; Dou et al., 2006; Sanford et al., 2005; Smith and Valcarcel, 2000; Watson, 2004). Although these conserved sequences are required for splicing, they are not sufficient to determine the correct exonic boundaries. Regulatory elements such as splicing enhancers and silencers located within either exons or introns act in both constitutive and regulated splicing. These elements called exonic or intronic splicing enhancers (ESE/ISE) and exonic or intronic splicing silencers (ESS/ISS) consist of short highly degenerate sequences that are generally bound by positive or negative trans-acting splicing factors often serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNP), respectively. Binding of these regulators to their corresponding cis-acting elements favor or inhibit recruitment of early

spliceosome components. This plays an important role in the decision of splice site selection that ultimately determines whether an exon will be included or excluded in the mRNA (Baralle and Baralle, 2005; Pagani and Baralle, 2004). Indeed, different tissues show not only differences in relative concentration but also a unique pattern of SR proteins and hnRNPs. These differences can affect splice choice in multiple RNAs and contribute to cell-specific gene expression (Smith and Valcarcel, 2000; Zahler et al., 1993).

Alternative splicing is not only affected by the relative abundance of different splicing factors at the time a gene is expressed. Extensive experimental evidence indicates that the elongation rate, promoter structure, transcription regulators, chromatin structure and RNA secondary structure also can affect alternative splicing.

A slowed transcriptional elongation rate or increased pausing of the RNAPII led to enhanced inclusion of exons with weak splice sites, presumably because it gives more time for the recruitment of splicing factors before the downstream competing splice site was transcribed (Kornblihtt, 2006; Luco et al., 2011). This was neatly demonstrated by utilizing reporter minigenes containing the exon 33 (E33 or EDI) of the *Fibronectin* (*FN*) gene and a wild-type or slow mutant of RNAPII. It was observed that the slow mutant RNAPII stimulates inclusion of the E33 by about 3-fold compared to wild-type RNAPII (de la Mata et al., 2003; Kornblihtt et al., 2004). Similarly, pause sites can regulate alternative splicing by temporarily arresting or slowing RNAPII. Insertion of the MAZ pause element in the tropomyosin gene, between the splice junction and its regulatory element, enhanced inclusion of tropomyosin exon 3 (Luco et al., 2011; Roberts et al., 1998)

How promoter structure influences alternative splicing was first discovered through studies with reporter minigenes containing the alternative exon 33 (E33) of the *FN* gene, driven by different RNAPII promoters. It was observed that inclusion of the E33 exon was about 10 times higher when transcription was driven by the FN or the CMV promoters than when it was driven by the weaker α-globin promoter (Cramer et al., 1999; Cramer et al., 1997; Kornblihtt, 2005, 2006). The effect of the promoter on alternative splicing can be mediated by differences in promoter occupancy by transcription factors which then, may impact transcription elongation rates or splicing factor recruitment, directly or indirectly. For example, the thermogenic coactivator PGC-1, is recruited at promoters with DR-1 elements bound by PPARγ. PGC-1 in turn recruits SRp40, resulting in EDII exon skipping (Kornblihtt, 2005).

Studies with the *FN* reporter minigene also revealed that chromatin structure can impact alternative splicing. It was observed that increased histone acetylation, induced by the histone deacetylation inhibitor trichostatin A, enhanced skipping of the E33 exon (Nogues et al., 2002). Moreover, genome-wide analysis has shown that nucleosomes are strikingly enriched on exons as compared to introns, suggesting they may participate in the exon definition process (Schwartz et al., 2009). The average exon size is 140-150 nt, compared to the nucleosome occupancy of 147 nt of DNA. In addition, nucleosomes are even more prominent in exons with weaker splice sites, suggesting that they may regulate alternative splicing. Furthermore, some histone modifications are specifically enriched in exonic nucleosomes compared to their flanking intronic regions; these modifications possibly mark the nucleosomes for binding to RNA-processing factors (Luco et al., 2011; Perales and Bentley, 2009). In addition, histone modifications can function as docking sites for chromatin binding factors. A nice example of the correlation between histone modifications and alternative splicing was recently reported by Saint-André et al. (2011). They demonstrated that H3K9 trimethylation is specifically enriched throughout the alternative exons of the CD44 gene. This histone mark functioned as binding sites for phosphorylated HP1γ, which mediated the inclusion of CD44 alternative exons. They noticed that HP1γ induces accumulation of RNAPII on the alternative exon region and therefore suggested a decrease in the transcription elongation rate. In addition, HP1γ also stabilizes association between the nascent RNA with the chromatin which may also enhance the inclusion of the CD44 alternative exons (Saint-Andre et al., 2011).

CO-TRANSCRIPTIONAL SPLICING AND TRANSCRIPTION COUPLING

Initial evidence that splicing is co-transcriptional came from analyses of nascent transcripts of *Drosophila* embryo by electron microscopy. These data visualized ribonucleoprotein assembly and intron looping on nascent transcripts occurs still attached to RNAPII (Beyer and Osheim, 1988). A recent analysis of the *c-Src* and the *FN* pre-mRNAs from chromatin-associated and nucleoplasmic RNA fractions revealed that most introns are removed before the transcript is released from the chromatin (Pandya-Jones and Black, 2009). However, introns flanking constitutive exons were removed in a 5'-3' fashion, while introns neighboring alternative exons showed differences in excision efficiency, indicating they were not removed in a strict 5'-3' direction, likely allowing for regulation.

This and numerous other studies establish that the splicing machinery is recruited and introns are committed to splice co-transcriptionally, indicating that the decision of whether an exon will be included or skipped is generally determined before the nascent transcript is released. However, this does not rule out that splicing of some introns will occur post-transcriptionally. Additional evidence that splicing is coupled to transcription is the interaction between splicing factors including SR proteins and U1 snRNP with the RNAPII (Luco et al., 2011).

ALTERNATIVE SPLICING AND THE NONSENSE-MEDIATED DECAY PATHWAY

Nonsense-mediated mRNA decay (NMD) is a conserved surveillance pathway present in eukaryotes for degradation of transcripts containing premature termination codons (PTC). This prevents translation of mRNA that would produce deleterious truncated proteins. Indeed, about 30 % of inherited genetic disorders would be generated as result of nonsense mutations or frameshifts that introduce nonsense codons and degradation of those transcripts by NMD, has been shown to ameliorate the disease phenotype. The intriguing conservation of NMD among the eukaryotes has suggested other functions for NMD pathway. Bioinformatic analyses point out that about 10 % of naturally occurring transcripts would produce alternative splice variants containing a PTC. Thi suggests a mechanism that couples alternative splicing with NMD to achieve a suitable regulation of gene expression by modulating the stability of target transcripts in a given developmental or physiological condition. By means of this mechanism, the NMD pathway plays an important role in regulating diverse biological processes such as cell proliferation, cell cycle, metabolism, transcription and telomere maintenance (Behm-Ansmant and Izaurralde, 2006; Boelz et al., 2006; Frischmeyer and Dietz, 1999; Mendell et al., 2004).

Activation of the NMD pathway relies on protein complexes deposited 20-24 nucleotides upstream of exon-exon boundaries. These multiprotein complexes, referred to as exon junction complexes (EJCs), serves as platform for assembly of the NMD effectors (Upf1, Upf2 and Upf3) that lead to mRNA degradation when premature termination codons are recognized (Behm-Ansmant and Izaurralde, 2006). A simple model of NMD pathway activation is depicted in Figure 1-3. In this model, Upf3 is recruited to the mRNA during splicing by interactions with the EJC, while Upf2 may join the complex after mRNA export. During the first round of translation, Upf2, Upf3 and other components of the EJC are removed as ribosomes trasverse the mRNA, while

stop codons are recognized by the eukaryotic release factors (eRF1 and eRF3). Detection of a PTC, which is defined as a stop codon located more than 50 nucleotides upstream of an exon-exon junction, will avoid complete removal of downstream Upf proteins and therefore, recruitment of Upf1. This triggers phosphorylation of Upf1 and rapid degradation of mRNA by decapping enzymes and 5'-3' exonucleases (Behm-Ansmant and Izaurralde, 2006; Chang et al., 2007).

CLEAVAGE/POLYADENYLATION AND TRANSCRIPTION TERMINATION by RNAPII

Transcription termination is described as the process by which RNA polymerase ceases transcription and releases the nascent RNA and DNA template (Figure 1-1, step 4). RNAPII release generally occurs downstream of the poly(A) site, which allows polymerase recycling and prevents interference at downstream promoters. Although transcription termination by the RNAPII is not completely understood, it is known that it requires a functional poly(A) site and is coupled to 3'-end mRNA processing (Dye and Proudfoot, 2001; Gromak et al., 2006). Poly(A) signals of protein-coding genes contain a conserved hexanucleotide sequence, AAUAAA and a G/U-rich or U-rich element located 10-50 nucleotides downstream. The 3'-end processing of most mRNAs is a two-step reaction involving the cleavage at a CA between the AAUAAA and the G/U-rich elements and the addition of about 200 adenosines at the 3'-end of the mRNA (Figure 1-4A). The cleavage and polyadenylation machinery includes cleavage and polyadenylation specificity factor (CPSF) that binds to the AAUAAA sequence. It is stabilized by binding of cleavage stimulatory factor (CstF) at the G/U-rich element, followed by the cleavage factors I and II (CFI/CFII) (Figure 1-B). Rapid addition of the poly(A) tail at the exposed 3'-OH is carried out by the poly(A) polymerase (PAP) and thus, the mRNA is stabilized. The downstream cleavage product, having an uncapped 5'-end is, instead, highly unstable and rapidly degraded (Bentley, 2005; Gromak et al., 2006; Perales and Bentley, 2009; Rosonina et al., 2006).

Recruitment of 3'-end processing factors start at the promoter of protein-encoding genes. CPSF is recruited to the pre-initiation complex by the general transcription factor TFIID. Once transcription initiates, CPSF dissociates from TFIID and associates with the transcription elongation complex (TEC), presumably by binding to the RNAPII body (Dantonel et al., 1997). The CstF complex, composed of three subunits (CstF50, CstF64 and CstF77), assembles during transcription. This is based on the finding that Cst64 localizes close to the transcription start site while CstF77 is

8

detected downstream of the poly(A) signal in *MYC, GAPDH* and *p21* genes (Glover-Cutter et al., 2008). CPSF binds to CstF and RNAPII in a mutually exclusive manner. This suggests that the formation of a functional CPSF and CstF complex may be modulated by transference of CPSF from the RNAPII body to CstF at the CTD, and subsequently to the nascent RNA after the poly(A) signal is synthesized (Glover-Cutter et al., 2008; Nag et al., 2007; Perales and Bentley, 2009).

RNAPII pausing generally occurs within 0.5 and 2 Kb downstream of the poly(A) site. Pausing at the 3'-end of genes correlates with maximal CTD Ser2 phosphorylation and high level of 3'-processing factors recruitment. It is thought that pausing at the 3'-end coordinates poly(A) site processing and transcription termination by an uncertain mechanism. Indeed, many proteins involved in cleavage/polyadenylation are also required for transcriptional termination, such as Pcf11 and Rat1 (Xrn2 in mammals). The Pcf11 subunit of CFII is required for both processes. Beside its role in pre-mRNA cleavage, Pcf11 has been suggested to act as a terminator factor by directly binding to Ser2 CTD and destabilizing the TEC (Zhang and Gilmour, 2006). The yeast 5'-3' RNA exonucleases Rat1 is recruited to the CTD during transcription and is implicated to promote recruitment of cleavage/polyadenylation factors essential in termination (Luo et al., 2006). It also has been suggested that upon 3'-end RNA cleavage, Rat1/Xrn2 is responsible for 5'-3' degradation of the nascent transcript and possibly, aids in the RNAPII release (Bentley, 2005; Luo et al., 2006; Rosonina et al., 2006; West et al., 2004).

Some poly(A) signals have the intrinsic ability to pause RNAPII and direct transcription termination in model constructs (Orozco et al., 2002). In these cases, termination occurs as a gradual decrease in RNAPII loading over hundreds to thousands bases downstream of the poly(A) signal, depending on the poly(A) strength (Orozco et al., 2002). However, additional downstream elements have been reported to increase termination efficiency in several genes. These sequences may be critical for closely spaced genes to avoid transcription into promoters of downstream genes. Such is the case of the human complement *C2* gene, spaced just 421 bp from the *Factor B* gene. It was demonstrated that transcription termination in the *C2* gene occurs just downstream of the poly(A) site, mediated by a 160 bp G-rich sequence that may bind the zinc finger protein MAZ. MAZ binding induced pausing of the RNAPII and additionally, DNA bending that may favor disassembling of the TEC (Ashfield et al., 1991; Ashfield et al., 1994).

Pause sites are also reported to aid in transcription termination, which is the case in the human $^{G}\gamma$-*globin* (Plant et al., 2005) and *α2-globin* genes (Enriquez-Harris et al., 1991). The human $^{G}\gamma$-*globin* termination element consist of multiple weak pause sites and a strong poly(A) signal. In contrast, the pause site in the *α2-globin* gene consists of a 92 bp long sequence combined with a weaker poly(A) signal. The *α2-globin* pause site contains a nearly perfect $(CAAAA)_6$ repeat proposed to either form a specific structure to the DNA template or binding site for a protein binding to the nascent RNA (Enriquez-Harris et al., 1991).

More recently, another elment required for efficient termination was identified in the 3' flanking region of the human *β-globin* gene. This element is positioned between 900 and 1600 bp downstream of the poly(A) site and contains the unique ability to undergo cotranscriptional cleavage (CoTC) by a poorly understood mechanism (Dye et al., 2006; Dye and Proudfoot, 2001; Teixeira et al., 2004). This element was uncovered by performing a run-on assay with nascent transcripts selected using an antisense biotinylated RNA probe targeting the poly(A) region (Dye and Proudfoot, 2001). Transcripts containing the sequences downstream of the CoTC region were not contiguous with the selected RNA, suggesting that they were cleaved. This was further confirmed by reverse transcription followed by PCR (RT-PCR) using primers spanning the region downstream of the poly(A) site. Deletion of the CoTC sequence eliminated transcription termination but did not affect 3'-end processing. The absence of 3'-end processing, by mutating the poly(A) signal, also abolished termination but did not affect cleavage within the CoTC region. Therefore, both 3' end processing and CoTC were required for transcription termination (Dye and Proudfoot, 2001). Gromak et al. (2006) also demonstrated that the β-globin CoTC element can be replaced by four tandem MAZ pause elements to support transcription termination in a poly(A)-dependent manner in this transcripiton unit.

Similar to the β-globin CoTC element, sequences undergoing cleavage downstream of a poly(A) site have been identified in the *ε-globin* gene (Dye and Proudfoot, 2001) and more recently in the mouse serum albumin gene (*MSA*) (West et al., 2006). However, they differ from the β-globin CoTC element in various characteristics. The ε-globin termination elements are more diffuse and involve long tracts of the 3' flaking region, which corresponds to differences in transcription termination profiles; transcription termination in the *β-globin* gene occurs abruptly within

the 3' flanking region while transcription termination in the *ε-globin* gene occurs gradually over about 1.65 kb downstream of the poly(A) site (Dye and Proudfoot, 2001). Unlike the MSA CoTC element, the human β-globin CoTC is A/U-rich (West et al., 2006).

As decribed above, there is a considerable sequence heterogeneity among the termination elements identified, making their existence difficult to predict without a functional assay. But their effect on transcription temination, in all the cases, is dependent on a poly(A) signal.

The current model of transcription termination postulates that there is a cooperative association of Rat1 and cleavage/polyadenylation factors, such as Pcf11, with the Ser2 CTD. When the poly(A) signal is transcribed and recognized by these factors, this may induce allosteric changes in the RNAPII, making it prone to termination. This may be enhanced by additional sequences that act to pause or slow RNAPII transcriptional elongation. Cleavage of nascent RNA at the poly(A) site or at a CoTC element leads to two RNA products; the mRNA is rapidly polyadenylated, while the 3'-cleavage product which remains attached to the RNAPII is targeted for degradation by 5'-3' exonucleases Rat1/Xrn2. Rat1/Xrn2-induced degradation may aid to release RNAPII (Gromak et al., 2006; Luo et al., 2006).

As has been discussed, regulation of gene expression is a very complex process that involves not only control of transcription initiation, transcription elongation and transcription termination, but also it can be modulated at different RNA processing steps. For instance, regulation of alternative RNA splicing may lead expression of proteins with different functions or subcellular localization, or even mRNAs that differ in stability. Better understanding of the mechanism by which genes are tightly regulated may generate therapeutic tools that could be used for amelioration of human diseases such as spinal muscular atrophy and cancer. Here, I will discuss two different gene models, the AFP and the IgM (μ) genes, that we study to gain insight into mechanisms coupling transcription with RNA processing. First, I will give some background information on what is known about the AFP and μ gene regulation.

ALFHA-FETOPROTEIN FUNCTION AND DEVELOPMENTAL GENE EXPRESSION

The α-fetoprotein (AFP) gene consists of 15 exons that encode a glycoprotein of about 70 KDa with a carbohydrate (glycan) moiety (Mizejewski, 2004; Peterson et al., 2011). The AFP gene is member of the serum albumin gene family which consists of five members, Albumin (*Alb*), vitamin D-binding protein (*DBP*), α-fetoprotein (AFP), α-

albumin or afamin (α-*Alb* or *Afm*) and AFP-related gene (*Arg*) protein, that all evolved from a common ancestor gene by a series of duplication events (Mizejewski, 2011; Peterson et al., 2011). Although AFP function remains poorly understood, it has been reported to be involved in the regulation of growth and differentiation during development and it binds and transports many ligands including bilirubin, fatty acids, hormones, retinoids, steroids, heavy metals, dyes, dioxins and organic drugs (Mizejewski, 2004).

AFP is expressed at high levels and secreted by the yolk sac, liver and, to a lesser extent, the gastrointestinal tract during fetal development. After birth AFP expression gradually decreases about 10^4-10^5-fold and remains at a low basal level throughout adult life (Belayew and Tilghman, 1982; Olsson et al., 1977). However, increased serum AFP is associated with liver and germ cell tumors and therefore, is used as biomarker for detection of these abnormalities. In pregnant women, AFP levels are also measured using maternal blood or amniotic fluid to screen for neural tube defects and Down syndrome (Spear et al., 2006).

Much of the work to understand AFP expression and the dramatic decrease in AFP expression after birth has been done in mice (Belayew and Tilghman, 1982; Olsson et al., 1977). Early studies to understand the mechanisms of postnatal AFP silencing were performed by measuring the serum AFP levels in adult mice of different inbred strains. BALB/cJ mice were found to express up to 20-fold higher serum AFP levels compared to other mouse strains. The post-natal persistence of AFP expression is a recessive trait controlled by a gene called *raf* (regulation of AFP) and later renamed *Afr1* (Alpha-fetoprotein regulator 1) (Olsson et al., 1977). Subsequently, it was shown that the higher levels of AFP in BALB/cJ correlated with higher steady-state liver AFP mRNA levels. AFP expression also is reactivated in adult liver during liver regeneration but this induction is controlled by another gene named *rif* (regulation of induction of AFP) or *Afr2* (Belayew and Tilghman, 1982).

Linkage analysis initially mapped the *Afr1* gene to mouse chromosome 15 (Blankenhorn et al., 1988). Then, performing a positional cloning using a high-resolution mapping, *Afr1* was identified as the Zinc-fingers and homeoboxes 2 (*Zhx2*) gene (Perincheri et al., 2005). This analysis revealed that the BALB/cJ Zhx2 allele contains an ETnIIα endogenous retroviral element insertion within the intron 1. While normal transcription initiation of the Zhx2 gene in BALB/cJ mice occurs, most pre-mRNAs splice to the retroviral element and a very low fraction of the mRNA is properly spliced to encode full length Zhx2 mRNA. This hypomorphic mutation dramatically reduces Zhx2

expression in BALB/cJ adult mice compared to other mouse strains expressing the wild-type Zhx2 allele (Perincheri et al., 2008). Lower Zhx2 expression in adult BALB/cJ mice correlates with increased accumulation of liver AFP and H19 mRNA levels, indicating the role for Zhx2 in repressing expression of both genes in perinatal liver. In fact, expression of a Zhx2 transgene driven by a liver-specific transthyretin enhancer/promoter in BALB/cJ mice led to complete repression of AFP and H19 in adult liver, confirming that Zhx2 is responsible for the Afr1 phenotype (Perincheri et al., 2005). Moreover, levels of Zhx2 are observed to increase after birth as AFP levels decrease, consistent with the role of Zhx2 in AFP repression.

The H19 gene was isolated as another target of Zhx2 through a screen for murine fetal liver cDNA clones that had decreased expression after birth (Pachnis et al., 1984). The H19 gene product is a long non-coding RNA, 2.4 kb in length, which has been extensively studied as model of genomic imprinting (Brannan et al., 1990; Pachnis et al., 1988). More recently, the cell-surface proteoglycan 3 (Gpc3), which belong to a family of proteins associated with cell growth, development and growth factor responses, was also identified to be regulated by Zhx2 (Morford et al., 2007). Deregulation of Gpc3 expression has been associated with Simpson-Golabi-Behmel syndrome characterized by general overgrowth as well with different types of tumors. Zhx2 may also control many other genes based on recent work in which a quantitative trait locus (QTL) on chromosome 15, called *Hyperlipidemia 2* (*Hyplip2*), involved in regulation of cholesterol and triglyceride levels, was identified to be Zhx2. BALB/cJ mice, which express very low levels of Zhx2, showed reduced serum cholesterol and triglyceride levels compared to other mouse strains. Additionally, complementation with a Zhx2 transgene resulted in increased-serum cholesterol and triglyceride levels as compared to other mouse strains, indicating Zhx2 is responsible for this trait. Comparison of global gene expression analysis from livers of BALB/cJ and Zhx2-expressing mice indicated that about a thousand genes were differentially expressed, some of which are known to be involved in liver function and plasma lipid metabolism. From this study, Lipoprotein lipase (Lpl) was reported as a new target of Zhx2 which showed a similar pattern of developmental expression like AFP, H19 and Gpc3. However, the Ear11 gene, a member of the eosinophil-associated ribonuclease family, showed the opposite expression pattern, suggesting that Zhx2 function may act in a context-dependent manner (Gargalovic et al., 2010). Interestingly, known targets of Zhx2 such as AFP, H19, Gpc3 and Lpl are generally elevated in liver tumors (Gargalovic et al., 2010; Spear et al., 2006).

ZINC FINGER AND HOMEOBOXES (ZHX) FAMILY

The Zinc fingers and homeoboxes (ZHX) family consists of three members, ZHX1, ZHX2 and ZHX3. Human ZHX1 was identified as a NF-YA interacting protein by screening a human liver cDNA library using a yeast two-hybrid system (Yamada et al., 1999). ZHX1 was independently isolated by immunoscreening of a bone marrow stromal cell cDNA library (Barthelemy et al., 1996; Yamada et al., 1999). Human ZHX2 and ZHX3 were identified later as ZHX1-interacting proteins by screening rat liver and ovarian granulosa cell cDNA libraries using ZHX1-fused to the GAL4-DBD as bait in a yeast two-hybrid system. ZHX proteins are reported to form homodimers and heterodimers with each other as well with NF-YA (Kawata et al., 2003a; Kawata et al., 2003b; Yamada et al., 2003).

All ZHX proteins contain two C_2H_2-type zinc finger domains and 4 (ZHX2 and ZHX3) to 5 (ZHX1) homeodomains. They also share similar gene structure, consisting of 4 exons where the proteins are encoded by an unusual long exon (exon 3) with the exception of the ZHX3, in which the last 8 nt are encoded by the last exon. ZHX1 is linked to ZHX2 in chromosome 8 (Chromosome 15 in mouse) while ZHX3 is located on chromosome 20 (Chromosome 2 in mouse). Based on the sequence similarities, gene structure and phylogenetic analysis of mouse ZHX proteins, it was suggested that ZHX genes were originated from a common ancestral gene by two different duplication events (Kawata et al., 2003a; Spear et al., 2006).

Analysis of tissue distribution of human ZHX mRNAs by Northern blot indicated that they are ubiquitously expressed, although transcript levels varied among the tissues examined (Kawata et al., 2003b; Yamada et al., 2003; Yamada et al., 1999). ZHX1 mRNA showed two major transcripts of 4.5 and 5 kb, Zhx2 is expressed as a single transcript of 4.4 kb, while ZHX3 is expressed as multiple mRNAs of 9.4, 7.3, 5.0 and 4.6 kb. Similar analysis of Zhx2 mRNA from mouse tissues indicated that a single 4.4 kb Zhx2 transcript is also ubiquitously expressed, although levels are lower in liver and in testis than the other tissues assayed (Perincheri, 2005). ZHX genes are also expressed in all vertebrates analyzed including human, rodents, dogs and zebrafish but homologues are not found in non-vertebrates such as *Drosophila*, *Caenorhabditis elegans* and yeast, suggesting that ZHX proteins are restricted to the vertebrate lineage (Spear et al., 2006).

ZHX2 has been called a transcriptional repressor based on a cell culture system utilizing a luciferase reporter gene driven by 5 copies of the GAL_4-DNA binding domain.

Co-transfecting HEK293 cells with a plasmid expressing ZHX2 fused to the GAL4 DNA-binding domain reduced luciferase activity by a maximum of 25% in a dose-dependent manner. Additionally, ZHX2 was shown to interact with NF-YA in vitro and in vivo, suggesting that ZHX2 may regulate genes that are regulated by the NF-Y transcription factor. In this regard, a luciferase reporter gene driven by the promoter region of the *cdc25C* gene responsive to NF-Y was used to co-transfect Drosophila SL2 cells. Co-transfection of NF-Y expression plasmids activated luciferase expression in these cells, and this expression was reduced about 30% by co-transfection with a ZHX2-expressing plasmid in a dose-dependent manner (Kawata et al., 2003b). Similar assays were performed with other promoters responsive to NF-Y in order to determine the repressor activity of all three ZHX proteins. The M2-type pyruvate kinase (MPK) promoter was modestly repressed by ZHX1 and ZHX3, the Hexokinase II (HKII) promoter was repressed by all ZHX proteins, and the cdc25C promoter was repressed by ZHX2 and ZHX3. These results indicated that members of the ZHX family negatively regulate expression of genes stimulated by NF-Y, although with differences in activity and specificity (Yamada et al., 2009).

The list of physiological functions of Zhx proteins has increased over the recent years, and all indicate they have important roles in cell differentiation and in disease. Briefly, ZHX proteins are reported to modulate podocyte gene expression during development of nephritic syndrome, where down-regulation of ZHX3 preceded proteinuria, a frequent complication in primary podocyte disorders (Liu et al., 2006). In another study, it was shown that ZHX2 is specifically expressed in neural progenitor cells during cortical neurogenesis. Its interaction with the cytoplasmic fragment of Ephrin-B in the nucleus regulates transcription and subsequent inhibition of cell differentiation required for neural progenitor cell maintenance (Wu et al., 2009). ZHX2 also was down-regulated in patients with hereditary persistence of fetal hemoglobin (HPFH) and $\delta\beta$-thalasemia. The genomic region where ZHX2 lies on chromosome 8 is associated with a quantitative trait locus that influences the switch from fetal to adult hemoglobin, and found to be responsible for HPFH (De Andrade et al., 2010). Finally, a gene expression profile performed to determine potential genes involved in clinical behavior of multiple myeloma (MM) indicated that high levels of ZHX2 expression correlated with better response and survival after high-dose therapy (Armellini et al., 2008). All the examples listed above suggest that ZHX proteins are involved in the control of genes that are involved early in development.

Other studies have suggested that ZHX2 is deregulated during liver cancer progression. Lv et al. reported that silencing of ZHX2 by promoter hypermethylation was observed in about half of the Hepatocellular Carcinoma (HCC) samples analyzed compared to control samples (no tumor). ZHX2 mRNA levels in these HCC samples were inversely correlated with serum levels of AFP (Lv et al., 2006). However, in another report, ZHX2 protein expression was analyzed by immunohistochemistry of tissue microarrays from liver tissues obtained from cholangitis, cirrhosis, adjacent non-tumor tissues, primary HCC tissues and matched metastatic lesions, and found that ZHX2 expression was detected only in HCC tissues. Higher expression of ZHX2 correlated with HCC clinical stage III and IV compared to stage I and II, and in metastatic lesions compared to primary HCC lesions (Hu et al., 2007). Because these studies come to opposite conclusions regarding a potential role for ZHX2 in HCC progression, further studies are warranted in this area.

AFP REGULATORY ELEMENTS

Regulatory elements that drive mouse AFP expression are contained within a 7.6 kb DNA region upstream of the AFP transcription start site and include the promoter, a repressor and three enhancer elements. The AFP promoter consists of 250 bp directly adjacent to exon 1 and its activity is restricted to the liver. It contains multiple binding sites for liver-enriched and ubiquitous transcription factors including the Hepatocyte Nuclear Factor 1 (HNF1), Nuclear Factor 1 (NF1), CAAT/Enhancer binding protein (C/EBP), Fetoprotein transcription factor (FTF) and Nkx2.8. An overlapping site for HNF1 and NF1 binding located at -120 from the AFP transcription start site has been reported to contribute to the promoter activity (Feuerman et al., 1989; Spear, 1999). In fact, it has been reported that a single nucleotide substitution in human at two different HNF1 binding sites within the AFP promoter (-60 and -120) are associated with hereditary persistence of AFP (Alj et al., 2004; Blesa et al., 2003). More recently it was reported that the Zinc finger and BTB domain containing 20 (ZBTB20) is involved in transcriptional repression of AFP in perinatal liver by binding to the region between -108 and -53 of the mouse the AFP promoter (Xie et al., 2008).

A repressor region located between the -250 and -838 bp region was reported to be required for postnatal AFP repression, since an AFP transgene lacking this region continued to be expressed in adult liver (Vacher and Tilghman, 1990). Additional binding sites for Foxa and p53 family members, p53 and p73, are located around the -

850 bp region. Interference of Foxa binding at this site by p53 is suggested to contribute to post-natal AFP repression (Peterson et al., 2011; Xie et al., 2008). However, in other studies, the 250 bp promoter without this repressor region was sufficient to repress a heterologous reporter gene in transgenic mice (Peyton et al., 2000a)

The three enhancers, referred to as E1, E2 and E3, are positioned at about -2.5, -5.0 and -6.5 kb, respectively, and are each about 200-300 bp (Godbout and Tilghman, 1988). E1 and E2 enhancers showed higher similarity suggesting they arose from a duplication event; moreover, E1 is present in rodent but it is absent in other species analyzed (Long et al., 2004). Studies in transgenic mice indicated that the three enhancers are active in both fetal and adult mouse livers (Ramesh et al., 1995). In the adult liver, E1 and E2 are active in all hepatocytes while E3 is active only in hepatocytes around the central vein due to continuous repression of this enhancer in non-pericentral hepatocytes (Peyton et al., 2000b).

REGULATION OF AFP POST-NATAL REPRESSION BY ZHX2

The AFP cis-acting elements required for Zhx2 responsiveness have been studied using mouse models. Although the cis-acting elements required for transcriptional activation of AFP expression, such as the enhancer and promoter regions, have been well characterized in cultured cells, the developmental silencing observed in vivo has not been reproduced in these in vitro systems. The BALB/cJ mouse strain, containing the Zhx2 mutation that results in incomplete post-natal AFP repression has been an important tool for studying this process. Because the mouse AFP gene consists of 15 exons and is about 22 kb, the use of AFP minigenes allows ease of manipulation for deletion and mutation of sequences that may impact AFP regulation. Minigenes also can be distinguished from the endogenous AFP gene for measuring gene expression. To study Zhx2-mediated regulation, transgenic mice expressing AFP minigenes were generated in a mouse strain expressing Zhx2 and backcrossed twice to BALB/cJ. F2 offspring were genotyped to screen for transgenic mice heterozygous or homozygous for the recessive Zhx2 mutant allele. Initial studies in cultured cells and transgenic mice utilized an AFP minigene that consisted of the two first and three last exons linked to the 7.6 kb 5'-flanking region of AFP gene. This minigene was developmentally regulated and expressed in the same tissue-restricted manner as the endogenous AFP gene (Krumlauf et al., 1985).

Chimeric transgenes, in which the repressor/promoter regions and the five-exon AFP minigene were linked to the Albumin enhancer, were regulated by Zhx2 whereas transgenes containing the Albumin promoter and Albumin minigene linked to the AFP enhancers were not. These results indicated that AFP enhancers do not direct AFP post-natal repression mediated by Zhx2 and, rather, elements contained within the 1 kb 5'-flanking region or within AFP structural gene were required (Camper and Tilghman, 1989). To explore whether sequences within the AFP structural gene are required for Zhx2 regulation, transgenic mice expressing the AFP 5'-flanking region containing the E1 and the repressor/promoter regions linked to the mouse major histocompatibility complex (MHC) class I *H-2D^d* (AFP-D^d), were generated. It was found that the unrelated gene linked to AFP 5'-flanking region was expressed in the same tissues and postnatally repressed like the endogenous AFP (Spear, 1994). Moreover, a similar chimeric construct in which the D^d gene was linked to the 250 bp AFP promoter and E2 enhancer was not only postnatally repressed but was also responsive to Zhx2 in a similar manner to endogenous AFP. This indicated that the AFP promoter is sufficient for conferring Zhx2 regulation (Peyton et al., 2000b).

The studies described above indicated that Zhx2 regulates AFP and H19 genes through the promoter. However, nuclear run on assays showed that the transcription rates of H19 gene and AFP transgene were similar between mice that did or did not express Zhx2. Thus, transcriptional difference can not account for the 10 to 20-fold difference in adult liver H19 and AFP mRNAs levels. Therefore, Zhx2 was proposed to regulate target genes at the post-transcriptional level (Vacher et al., 1992). A more detailed nuclear run-on experiment was performed in our lab using single strand probes derived from M13 clones. These probes covered the 5'-end, middle and 3'-end of the mouse AFP and H19 genes in the sense and anti-sense orientations. Similar transcription rates were observed for both genes in the presence or absence of Zhx2. Low levels of antisense transcription were detected, but it was not different between mice and does not contribute to Zhx2 regulation. Also, there was no significant loss of signal between the 5'-end probe and 3'-end probes that would indicate early termination of transcription. These result confirmed the post-transcriptional regulation of AFP and H19 genes and additionally, demonstrated that neither anti-sense transcription nor transcription termination are involved in Zhx2 regulation. Based on data described above, it is proposed that Zhx2 controls target genes by a mechanism that couples transcriptional and post-transcriptional steps of gene expression. It is well documented

using artificially constructed minigenes that promoters influence transcription elongation and regulate post-transcriptional events such as splicing. Zhx2 regulation of AFP and H19 may be a natural example of where these events interact to regulate endogenous gene expression.

More recently, the effect of ZHX2 on AFP expression was studied in human hepatoma cell lines. HepG2 and HepG2215 cell lines express low levels of ZHX2 mRNA which correlated with high level of secreted AFP, while SMMC771 and LO2 cell lines express higher levels of ZHX2 mRNA and lower level of secreted AFP. Overexpression of ZHX2 in HepG2 cells reduced AFP protein levels in a dose-dependent manner by about 25%, whereas, down-regulation of ZHX2 by siRNA in SMMC771 and LO2 cell lines increased levels of AFP about two-fold without affecting Albumin levels. These results suggested AFP secretion in different hepatoma cell lines is inversely correlated with expression of ZHX2. Additionally, HepG2 cells were co-transfected with a ZHX2-expressing plasmid and a reporter construct in which the luciferase gene was linked to the 269 bp human AFP promoter. Luciferase activity was reduced by about two-fold upon increased expression of ZHX2. This indicated that ZHX2 can also repress AFP expression mediated by the AFP promoter in HepG2 cells (Shen et al., 2008). Similar results were observed with the mouse AFP promoter linked to luciferase (Yamada, 2009). Both cell culture and transgenic mice have provided insight into the AFP cis-acting sequences and factors required for expression and developmental silencing. However, it is important to note that in cell culture systems the repression of AFP observed is, in most cases, only two-fold, while in the mouse model, the presence or absence of Zhx2 results in about 10 to 20-fold difference in AFP expression. More research must be done to improve the cell culture system to study AFP regulation by Zhx2 and thus, the mouse model remains an essential tool for studying developmental repression by Zhx2.

B CELL DEVELOPMENT

B lymphocytes derive from a common lymphoid progenitor cell in the bone marrow, where resident stromal cells provide the chemokines and cytokines as well intercellular contact required for survival and proper early stage development of B cells. This involves productive DNA arrangements of the $V_H D_H J_H$ segments of the immunoglobulin (Ig) heavy chain locus and the $V_L J_L$ segments of the κ or λ light chain loci, allowing the expression of a functional B cell receptor (BCR) complex on the B cell

surface. The Ig receptor consists of two identical heavy chains and two identical light chains held together through disulfide bonds to form a membrane-bound IgM (μm), and together with the Igα and Igβ heterodimer on the B cell surface, delivers intracellular signaling important for survival and proliferation. Once the IgM is expressed on the B cell surface, the immature B cells are selected for central tolerance, so B cells that recognize self-antigens are removed by apoptosis or undergo further light chain rearrangements, mediated by receptor editing, to reduce potential auto-reactivity. Immature B cells start to express new adhesion molecules and homing receptors, becoming a transitional type 1 B cell (T1 B cell). This allows transition of the developing B cell to the periarteriolar lymphoid sheath in the spleen where local production of B cell lymphocyte activating factor induces survival and transition to a T2 B cell. At this stage, B cells migrate to other secondary lymphoid tissues where they become mature naïve B cells and co-express IgM and IgD on their surface. These B cells are ready to encounter antigen which will trigger clonal expansion and the differentiation of the B cell into an Ig-secreting plasma cell (Kurosaki et al., 2010).

Upon encounter with antigen, B cells can also undergo isotype switching, a process in which expression of IgM and IgD switch to other Ig classes, IgA, IgG or IgE, without a change in the antigenic specificity. Isotype switching is regulated by external signals such as cytokines released by T lymphocytes and it involves irreversible nonhomologous DNA recombination guided by stretches of repetitive DNA sequences known as switch regions (S_H). The S_H regions lie in the intron between the J_H gene segments and the *Cμ* gene, and upstream of the genes for each other heavy chain isotype with the exception of the *Cδ* gene. Unlike IgA, IgG and IgE, co-expression of IgD along with IgM does not involve DNA rearrangement, instead it involves alternative splicing of the variable region ($V_H D_H J_H$) to the Cδ1 exon, resulting in the removal of the Cμ region in the pre-mRNA (Janeway, 2005; Peterson, 1994b; Peterson, 2007).

Ig heavy chain genes encoding IgM, IgD, IgA, IgG and IgE, have similar gene structure. They all encode both membrane-bound and secreted Ig isoforms and therefore, undergo regulated RNA processing at the 3' end of the transcript as B cells mature to plasma cells. In naïve B cells and memory B cells, mRNA encoding the membrane-associated isoform predominates, while in plasma cells, the mRNA encoding the secreted form is preferentially expressed. The exclusive expression of the Ig on the B cell surface, and Ig secretion in plasma cells, is due in part to differences in the relative expression of the mRNA isoforms but also, to the contribution of additional post-

translational mechanisms, such as differential intracellular retention and protein degradation (Peterson, 1994b; Peterson, 2007; Peterson, 2011).

REGULATION OF IgM EXPRESSION

The shift between the membrane-associated (μm) and the secreted (μs) IgM isoforms observed as B cells differentiate into plasma cells occurs by alternative RNA processing at the 3' end of a common precursor RNA (Figure 1-5). The μs mRNA is produced by cleavage and polyadenylation of the transcript at the proximal (μs) poly(A) site, while the μm mRNA is produced by splicing between the Cμ4 and M1 exons followed by cleavage and polyadenylation at the distal (μm) poly(A) site. B cell lines express roughly similar levels of μs and μm mRNA while plasma cell lines express 10-20 fold more μs mRNA than μm mRNA, so the μs/μm mRNA ratio increases as B cells progress to plasma cells (Peterson, 2007).

The IgM gene has been extensively studied as a model to better understand the mechanisms involved in the regulated RNA processing observed during B cell development. For this purpose, plasmids expressing intact or modified sequences of the mouse *Cμ* gene driven by a heterologous promoter have been transiently or stably transfected into suitable mouse B cell and plasma cell lines, such as M12 and S194, respectively (Peterson, 2007; Peterson, 2011). The murine B cell line, M12, does not express endogenous immunoglobulin, while the murine plasmacytoma cell line S194, expresses only the endogenous IgA gene. The absence of endogenous IgM expression in these cell lines make them ideal for expression of transfected IgM constructs (Kim et al., 1979; Hyman et al., 1972). Additionally, when either the wild-type or chimeric IgM genes are transfected into B cells and plasma cells, the shift in the μs/μm mRNA ratio, from 1-2 to >10, resembles that observed in the endogenous μ gene when resting B cells are stimulated into Ig-secreting cells. These results validate this experimental system and it has been a valuable tool for understanding the elements required for the regulated shift in the μs/μm mRNA ratio observed during B cell differentiation to plasma cells, and it is extensively used in the present study (Peterson, 2007).

Based on deletion, substitution and mutations of other portions of the μ gene using the approach mentioned above, it was concluded that the strength of the two competing reactions, the cleavage and polyadenylation at the μs poly(A) site and the Cμ4-M1 splice, has to be balanced in order for regulation to occur. In the wild-type IgM gene both the μs poly(A) site and the splice site are suboptimal. Substitution of the μs

poly(A) site by a stronger poly(A) site and maintaining the suboptimal Cµ4 5' splice site predominantly favored µs mRNA in both B cell and plasma cells, thus the regulation was disrupted. Substitution of the Cµ4 5' splice site by a consensus splice signal while maintaining the suboptimal µs poly(A) site exclusively favored the production of µm mRNA, even in plasma cells. Only when both sequences were replaced with strong signals was the µs/µm mRNA ratio developmentally regulated (Peterson, 1992; Peterson and Perry, 1989).

Deletion of a 158 bp *Not*I-*Hind*III fragment 50 bp downstream of the µs poly(A) site decreased the µs/µm mRNA ratio by 3- to 8-fold in B, plasma and non-lymphoid cells (Peterson et al., 2002). A detailed analysis of the region uncovered the existence of an RNAPII pause site, with the following features: the fragment enhanced the use of a weaker upstream poly(A) site when placed between tandem poly(A) sites; it could be replaced by other known pause sites such as the H3.3 and the MAZ elements and run-on assays showed an increase in RNAPII loading about 50 bp downstream of the µs poly(A) site even when the µs poly(A) was inactivated (Peterson et al., 2002). Additionally, the µ pause site effect on µs poly(A) site use was lost when it is moved over 100 bp downstream from its original position, although it still induced RNAPII pausing (Burnside et al., 2011; Peterson et al., 2002).

Other parameters that impact the µs/µm mRNA ratio, without affecting the developmental regulation between B cells and plasma cells, are the Cµ4-M1 intron size (Peterson and Perry, 1986; Tsurushita and Korn, 1987) and the Cµ4 exon size (Peterson et al., 1994). Shortening either Cµ4-M1 intron or the Cµ4 exon favored the splicing reaction so more µm mRNA was made. Deletion and substitution of different portions of the µ gene also revealed that there are no gene-specific sequences required for the regulated µs/µm mRNA ratio shift between B cells and plasma cells. Moreover, non-immunoglobulin genes (*β-globin* and *Major Histocompatibility Complex H-2D^d*) that were modified by inserting a intronic poly(A) signal, so they could be alternative spliced or cleaved and polyadenylated at the inserted poly(A) site, were regulated in B cell and plasma cells like the Ig genes (Peterson, 1994a; Peterson, 2011). This was further tested in resting and stimulated B lymphocytes derived from transgenic mice expressing the modified D^d gene. Similarly to the endogenous µ gene, the pA/spliced D^d mRNA ratio increased as the B lymphocytes were differentiated (Peterson, 2011; Seipelt et al., 1998).

These studies indicate that the balanced efficiencies of the competing splice and cleavage-polyadenylation reactions are critical features for the developmental Ig RNA processing, and the Cμ4-M1 intron and Cμ4 exon sizes have an effect on the efficiency of these competing reactions. Consistent with these facts, other immunoglobulin genes that undergo developmental expression regulation have similar gene structures, where poly(A) strength and the intron and exon sizes are well conserved (Peterson, 2011).

Because the Ig gene structure rather than specific sequences are important for regulation, it is likely that the trans-acting factors that regulate the alternative RNA processing reactions will be involved in regulating the general splicing and/or cleavage-polyadenylation reactions. In fact, there is evidence that both splicing and cleavage/polyadenylation activities are different between B cells and plasma cells. To test whether differences in splicing environment alter μs/splice mRNA ratio, μ gene constructs were transiently co-transfected with different SR-encoding plasmids in B cell (M12) and plasma cell (S194) lines. Expression of several SR proteins, SF2/ASF, SRp30c and SRp55, decreased the μs/splice mRNA ratio, suggesting they enhance the Cμ4-M1 splice reaction (Bruce et al., 2003). On the other hand, differences in cleavage and polyadenylation activity were also observed when constructs with tandem poly(A) sites were expressed in B cell and plasma cell lines. It is also reported that cleavage and polyadenylation efficiency increases as B cell differentiate to plasma cells (Peterson et al., 1991). Therefore, several trans-acting factors involved in either splicing or cleavage and polyadenylation may be implicated in the μ RNA processing switch. Some candidates which relative levels differ between B cells and plasma cells have been studied, Cst64, hnRNP H/H'/F family and U1 snRNP, however their roles in regulating the μs/splice mRNA ratio are not fully understood (Peterson, 2007; Peterson, 2011). More recently, the elongation factor ELL2 was reported to be an important factor in shifting the Ig heavy-chain (IgH) pre-mRNA processing from the membrane associated form to the secreted mRNA form (Martincic et al., 2009). ELL2, which increases during B cell differentiation to plasma cells was shown to bind to the IgH promoter and favor the recruitment of Cst64 to the RNAPII near the 5' end of the gene. ELL2 expression was associated with enhanced polyadenylation at the promoter proximal poly(A) site in a construct with tandem poly(A) sites and also increased exon skipping of the Ed1 splicing reporter gene. Therefore, it was suggested that ELL2 influences the IgH pre-mRNA processing by increasing the polyadenylation at the promoter-proximal poly(A) site and reducing the efficiency of splicing (exon skipping).

IgM TRANSCRIPTION TERMINATION

Several studies have investigated the transcription termination profiles in the μ-δ locus by nuclear run-on analysis, using either resting and stimulated B cells or cells lines that represent different B cell developmental stages. These studies have shown the transcription unit is gradually foreshortened as B cells progress to plasma cells. In IgM secreting cells, transcription is terminated predominantly downstream of the μm poly(A), while in surface-IgM expressing cells, transcription proceeds throughout the δ locus. The gradual foreshortening of transcription termination through the μ-δ locus observed in plasma cells, is responsible for the lack of IgD expression. However, it is not likely to be involved in the increased expression of μs mRNA because most transcripts contain the Cμ4-M1 splice and thus are able to produce μm mRNA. Based on the fact that a functional poly(A) site is required for efficient transcription termination to occur, the earlier transcription termination observed in plasma cells may be explained by a developmental increase in efficiency of the μs poly(A) usage (Kelley and Perry, 1986; Peterson, 2007; Weiss et al., 1989; Yuan and Tucker, 1984).

1. Transcription pre-initiation

- Assembly of general transcription factors onto the promoter region and recruitment of RNAPII
- RNAPII CTD is unphosphorylated

2. Transcription initiation

- RNAPII CTD is phosphorylated at Ser-5 (yellow stars)
- RNAPII escapes from promoter and starts RNA synthesis
  - Capping enzymes associate and add the 7-methyl-guanoisne cap to the nascent RNA

DSIF and NELF binding induces promoter-proximal pausing which serves as a checkpoint for proper 5'-end RNA processing and a control point for gene expression regulation.

3. Transcription elongation

- p-TEFb recruitment induces phosphorylation of DSIF and RNAPII CTD at Ser-2 (red stars). This releases RNAPII from promoter-proximal pause to begin elongation
  - Factors required for RNA splicing and cleavage-polyadenylation are recruited

The elongation rate may affect alternative splicing outcome and serve as a point for gene expression regulation. Factor recruitment can also be regulated.

4. Transcription termination

- RNAPII disengages from the transcript and DNA template
- Nascent RNA is cleaved and rapidly polyadenylated.

mRNA is exported and translated

RNAPII is recycled

Pausing occurs at the end of the gene which may serve as checkpoint to ensure co-transcriptional splicing as well recruitment of additional factors involved in cleavage-polyadenylation and transcription termination.

Figure 1-1. Transcription coupling with RNA processing. Transcription is a step-wise process (black print) in which each step is tightly associated with different RNA processing stage (blue print) as described. Checkpoints at each step (red print) ensure proper transcriptional regulation and RNA processing. Additional detail is explained in the text.
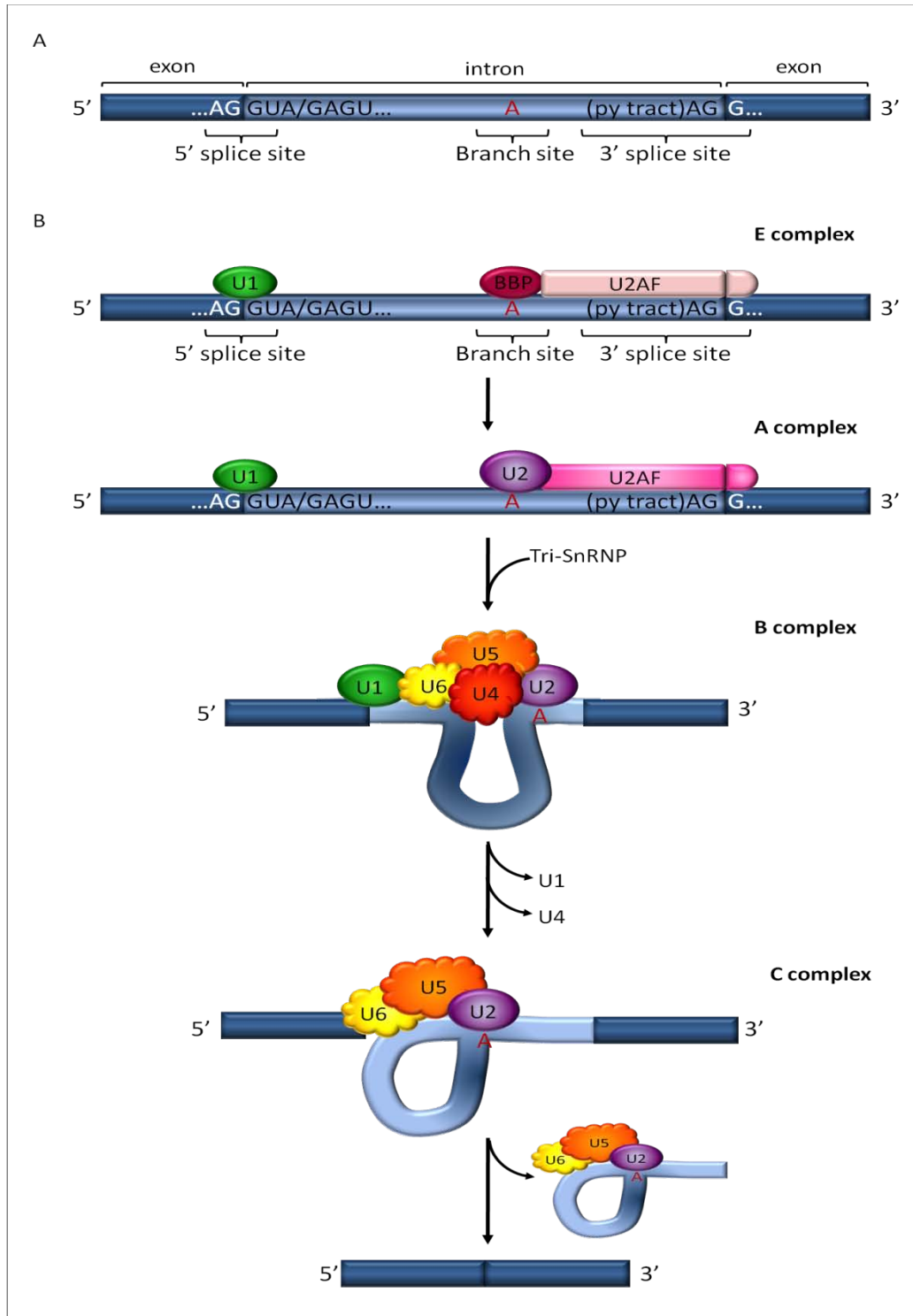
Figure 1-2. Spliceosome assembly. (A) Consensus sequences at the intron-exon boundaries. (B) Steps involved in spliceosome assembly.

26
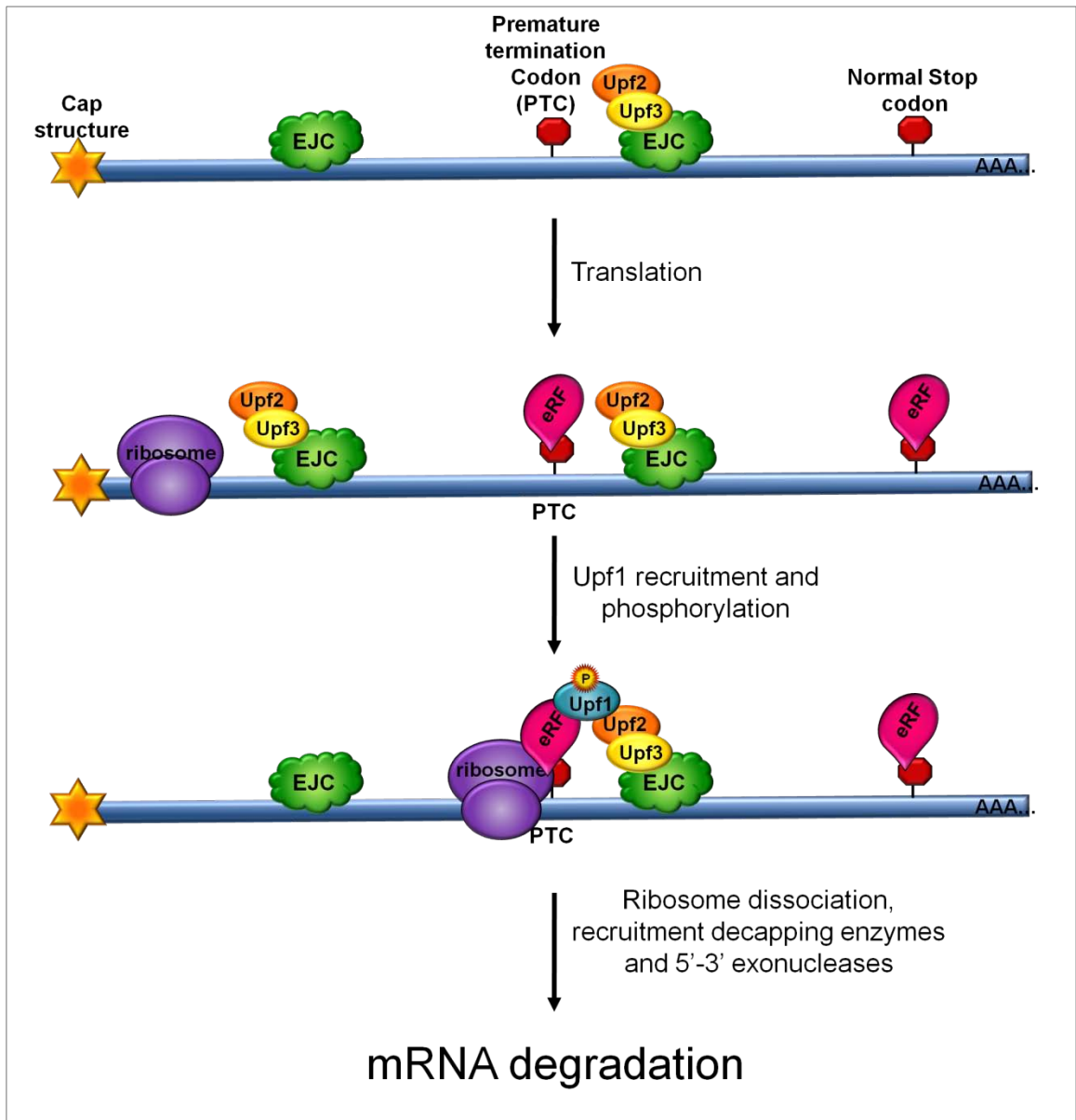
Figure 1-3. Simplified model of the nonsense mediated decay pathway activation.
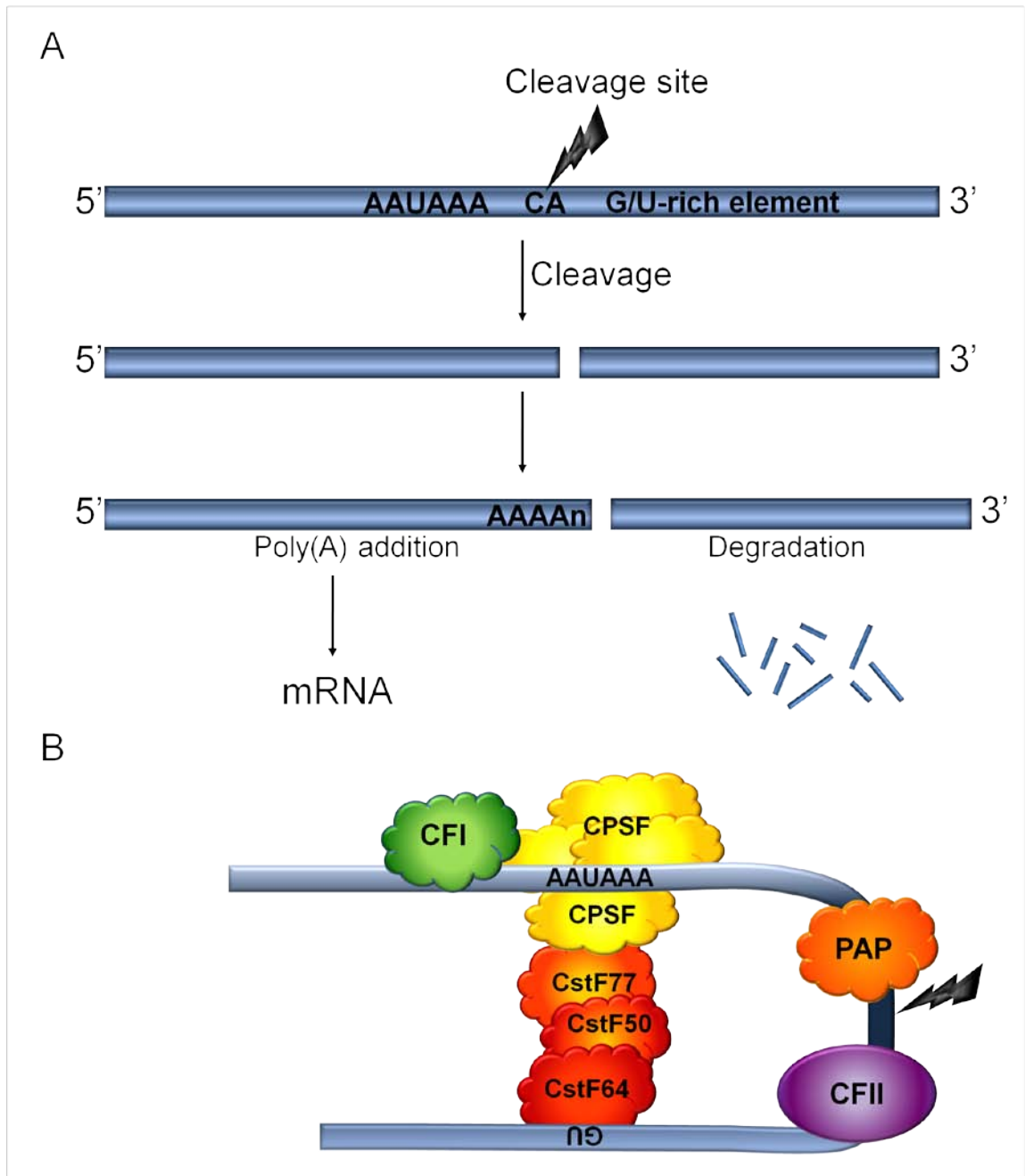
Figure 1-4. Cleavage and polyadenylation process. (A) Cleavage and polyadenylation scheme. (B) Sequences involved in cleavage and polyadenylation factor recognition.
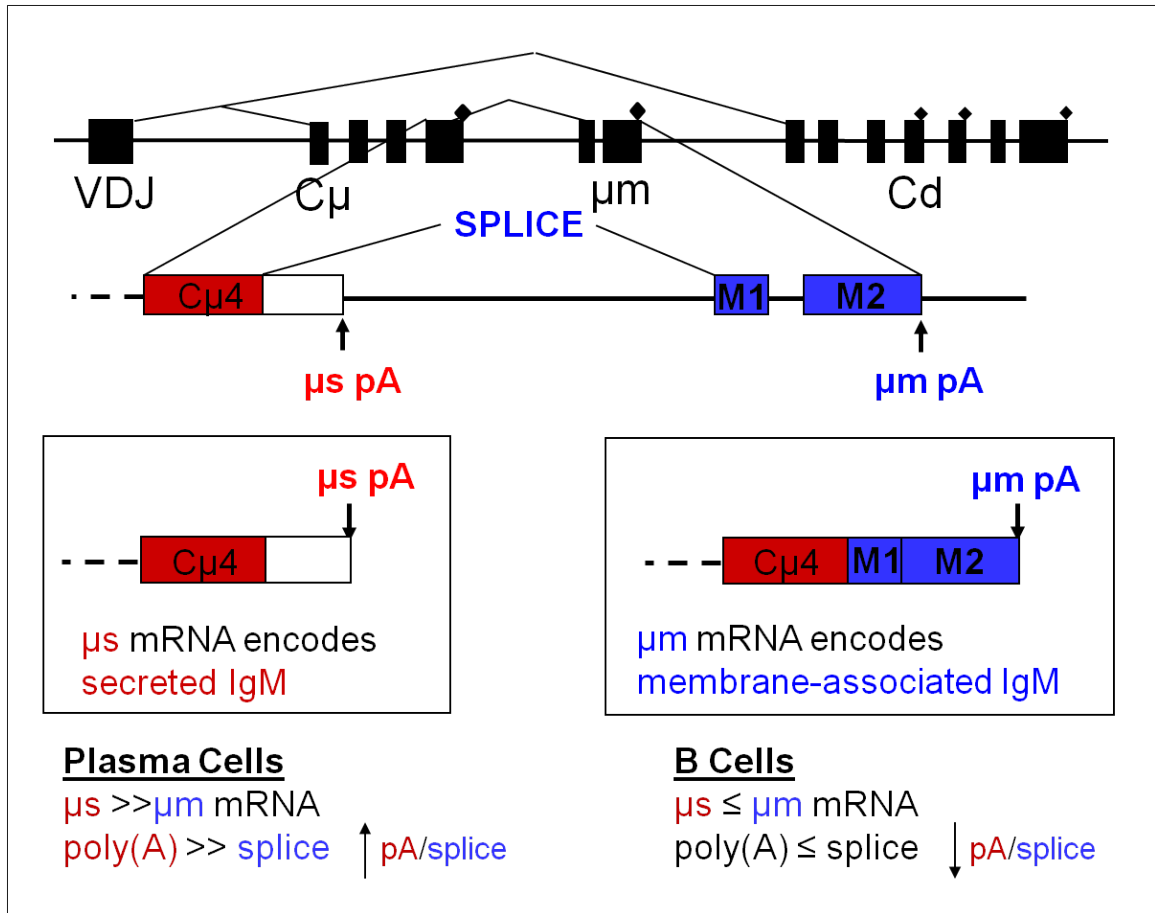
Figure 1-5. IgM gene structure and alternative RNA processing.

CHAPTER II

MATERIALS AND METHODS

PLASMID CONSTRUCTION

pGEM-T-CoTC was made by cloning the β-globin CoTC element into the pGEM-T easy vector (Promega). The full β-globin CoTC fragment was PCR-amplified with *Asp*718 ends from βΔ5-7p(A)mut (Dye and Proudfoot, 2001) using the primers CoTC-Fc/CoTC-R800. The 800bp-PCR product was *Dpn*I-treated and cloned into pGEM-T Easy vector following manufacturer instructions.

Cμ/CoTC(+), Cμ/CoTC(-), ΔNH/CoTC(+), ΔNH/CoTC(-), ΔAR/CoTC(+) and ΔAR/CoTC(-) (Table 2-1) were made to determine the effect of having a CoTC element at the *Kpn*I site within the Cμ4-M1 intron on the μm/μs mRNA expression ratio. The 800 bp *Asp*718-digested fragment from pGEM-T-CoTC was cloned into IgCμ(WT), IgCμ(ΔNH) and Cμ(ΔAR), that had been digested with *Asp*718 and treated with SAP, to obtain Cμ/CoTC, ΔNH/CoTC and ΔAR/CoTC, respectively. Cμ(WT) refers to the wild-type μ construct; Cμ(ΔNH) has a 158 bp deletion between the *Not*I and *Hind*III sites within the Cμ4-M1 intron which contains a RNAPII pause site element (Peterson et al., 2002); and Cμ(ΔAR) contains a 797 bp deletion between the *Acc*I and *EcoR*V sites in the middle of the Cμ4-M1 intron (Peterson and Perry, 1986). For each construct, clones containing the CoTC sequence in either orientation were selected, denoted as (+) when β-globin CoTC is in 5´-3´orientation and (-) when the β-globin CoTC is in 3´-5´orientation, to obtain Cμ/CoTC(+), Cμ/CoTC(-), ΔNH/CoTC(+), ΔNH/CoTC(-), ΔAR/CoTC(+) and ΔAR/CoTC(-).

Cμ/CoTC(+)-H and ΔNH/CoTC(+)-H (Table 2-1) were made to determine the effect of having a CoTC element at the *Hind*III site on the μm/μs mRNA expression ratio in presence or absence of the RNAPII pause site element. Cloning the β-globin CoTC at the *Hind*III or *Not*I-*Hind*III (referred as ΔNH) sites of Cμ required three sequential cloning steps. The 800 bp *Asp*718-digested fragment from pGEM-T/CoTC was Klenow treated to make blunt ends and inserted at either the *Hind*III or *Not*I-*Hind*III sites of pUCPst, also Klenow-treated after restriction enzyme digestion. pUCPst contains the 2318 bp-*Pst*I fragment of μ gene spanning the region between the Cμ4 and M2 exons cloned into pUC9 (Peterson and Perry, 1989). The *Apa*I-*Kpn*I fragment containing the modified

sequences was cloned into the *Apa*I-*Kpn*I-cut pSV2Cμ which contains the 6 Kb *Bgl*II fragment comprising the μ constant region (Peterson and Perry, 1986). The last cloning step transferred the *BamH*I-*Bgl*II fragment of the modified pSV2Cμ into the *BamH*I-cut pIgCμ where an intact μ gene containing the modification could be expressed stably in lymphoid cell lines such as M12 and S194 cells (Peterson and Perry, 1989).

Cμ/rpI900-H and ΔNH/rpI900-H (Table 2-1) were made to use as intron size controls by inserting an unrelated sequence with no known CoTC activity, into Cμ. The rpI900 fragment is a 900-bp *BamH*I-*Sma*I fragment from the second intron of the S16 ribosomal protein gene. This fragment was Klenow-treated to make blunt ends and inserted into the *Hind*III or *Hind*III-*Not*I sites of the μ gene, following the thee-step cloning procedure described above.

Other plasmid constructs described in previous studies also were used as controls. These constructs differ in the Cμ4-M1 intron length by deleting or inserting unrelated sequences (Table 2-1). The ΔNR contains a 351 bp deletion between the *Nco*I and *EcoR*V sites, ΔAN has a deletion of 446 bp between *Acc*I and *Nco*I sites, AR/rpI900 has 797 bp deleted between the *Acc*I and *EcoR*V sites and the rpI900 fragment inserted at the *Kpn*I site, Cμ/rpI900 has the rpI900 fragment inserted at the *Kpn*I site in the wild-type Cμ gene (Peterson and Perry, 1986).

CELL LINES AND CULTURE CONDITIONS

<u>B Cells</u>:

M12 murine B cells, a lymphoma-derived cell line deficient in IgM expression when established in culture (Kim et al., 1979), were grown in RPMI 1640 medium supplemented with 10% heat-inactivated fetal bovine serum, 50 U/ml penicillin/Streptomycin and 50 μM β-mercaptoethanol. The 38C-13 cell line is derived from a C3H/eB mouse-B cell lymphoma and expresses roughly similar amount of μm and μs mRNA (Nelson et al., 1983). This cell line was grown in the same medium as M12 cells.

<u>Plasma Cell lines</u>

The S194 murine plasmacytoma cell line, a myeloid-derived cell line expressing IgA but defective in IgA secretion (Hyman et al., 1972), was obtained from the American Type Culture Collection. S194 cells were maintained in Dulbelcco´s modified Eagles medium (DMEM) supplemented with 10% heat-inactivated horse serum and 50 U/ml

penicillin/Streptomycin. The D2 cell line is a 38C-13-derived hybridoma cell line actively secreting IgM (μs mRNA >> μm mRNA) (Nelson et al., 1983) and was propaged in RPMI 1640 medium supplemented with 20% fetal bovine serum.

<u>HEK293 cells</u>

HEK293 cells were obtained from the American Type Culture Collection and maintained in Dulbelcco´s modified Eagles medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum and 50 U/ml penicillin/streptomycin.

DNA PREPARATIONS, PRIMERS AND SEQUENCING

Plasmid DNA used for sequencing or transfections was prepared by cesium chloride protocol as described by Ausubel et al. (1987). DNA concentration was quantified by optical density at 260 nm. Primers were purchased from Integrated DNA Technologies, Inc. (IDT; Coraville, IA).   DNA sequencing was performed by MWG Biotech (High Point, NC) for sequencing.

DNA TRANSFECTIONS

<u>Stable transfections</u>

M12 B cells and S194 plasma cells were stably transfected by electroporation as described by Bruce et al. (2003). About 20 μg of *Pvu*I-linearized plasmid DNA were used for transfecting 5×10^6 cells that had been washed and resuspended in 0.8 ml of Opti-MEM-I+GlutaMAX-I (GIBCO).   Cells were electroporated at 250 volts and a capacitance of 960 (BioRad) and allowed to recover overnight before being selected with G418 at 300 μg/ml for M12 and 600 μg/ml for S194. Cells were collected after at least 2-3 weeks of G418 selection and RNA was extracted.

<u>Calcium phosphate transient transfections</u>

HEK293 cells were transiently transfected by the calcium phosphate protocol described by Spear and Tilghman (Spear and Tilghman, 1990).   Briefly, two 80% confluent plates of cells were split into 6 plates the day before transfection. 15 μg of plasmid DNA resuspended in 450 μl of 250 mM $CaCl_2$ were added dropwise with slow bubbling to 500 μl of 2X HBS (280mM NaCl, 50 mM HEPES, 1.5 mM $NaPO_4$, pH 7.1). The precipitate was allowed to sit for 30 minutes at room temperature and then added to

a plate of HEK293 cells. After 6 hours, the cells were rinsed and incubated with fresh media for 48 hours before being collected for nuclear extract preparation.

RNA PREPARATIONS

Total RNA extraction

Total RNA from 1×10^7 stably transfected M12 B cells or S194 plasma cells was extracted with Trizol reagent following manufacturer instructions (Invitrogen). Total RNA from mouse liver was obtained by homogenization of about 100 mg of liver sample followed by two sequential Trizol reagent extractions. The RNA concentration was quantified by optical density at 260 nm.

Nuclear RNA extraction from B cells and plasma cells

Nuclear RNA from 38C-13 and D2 cells was isolated by the hot phenol method from citric acid-purified nuclei (Schibler et al., 1975). Briefly, 1-2 $\times10^8$ cells were resuspended in 10 ml of cold 5% citric acid, homogenized and then centrifuged over 10 ml of 30% sucrose/5% citric acid cushion at 1000×g for 10 minutes at 4°C. The pellet was resuspended in 10 ml of cold 10mM sodium acetate/10 mM NaCl and centrifuged at 2000 rpm for 5 minutes. The pellet was resuspended in 1 ml of 50 mM sodium acetate and RNA was extracted by the hot phenol method. Briefly, 10 ml of 50 mM sodium acetate/1% SDS was added while gently vortexing the sample. The suspension was extracted with an equal volume of hot phenol saturated with 50 mM sodium acetate, incubated for 15 minutes at 68°C with intermittent shaking and then chilled on ice for at least 5 minutes. After centrifugation, a second extraction was performed, followed by a final extraction with chloroform:isoamylic alcohol (24:1). RNA was precipitated with 1/10 volume of 3 M sodium acetate and 2.5 volumes of ethanol. For RT-PCR analysis, 100-200 µg of nuclear RNA was further treated with Trizol reagent in order to eliminate remnant DNA contamination.

Cellular RNA fractions from mouse liver

Mouse liver was homogenized in 10 ml of Buffer I (0.25 M sucrose, 60 mM KCl, 15 mM NaCl, 15 mM Hepes pH7.4, 2 mM EDTA, 0.5 mM EGTA, 0.15 mM spermine, 0.5 mM spermidine and 14 mM β-mercaptoethanol) in a Dounce homogenizer using 18 strokes with pestle A and 18 strokes with pestle B. Homogenate was layered on 10 ml of Buffer II (same as buffer I but containing 0.75 M sucrose) and spun at 1000 ×g for 10

minutes at 4 ºC. Cytoplasmic fraction corresponding to the pinkish top faction was saved for cytoplasmic RNA extraction, while the interface phase was discarded. The nuclear pellet was resuspended in 3.2 ml of Buffer III (same as buffer I but containing 2 M sucrose) and layered over 1 ml cushion of Buffer III, then spun at 33000 rpm for 1 h at 4 ºC in a Beckman Coulter "Optima" ultracentrifuge (SW55 Ti rotor) for nuclei purification. Chromatin-soluble and chromatin-insoluble fractions were obtained from purified nuclei as described by Pandya-Jones and Black (Pandya-Jones and Black, 2009). Briefly, nuclei were resuspended in 400 µl of glycerol buffer (20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, 50% glycerol and proteinase inhibitors) and 400 µl of urea lysis buffer (20 mM HEPES pH7.6, 1 mM DTT, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 0.3M NaCl, 1M Urea, 1% NP-40) was added, mixed and incubated for 10 minutes on ice. The nuclear soluble fraction is recovered from the supernatant by centrifugation at 14000 rpm in a microcentrifuge for 3 minutes at 4ºC, and used for RNA extraction. The insoluble pellet was washed with 400 µl 1X PBS/1mM EDTA and resuspended in 100 µl DNase buffer containing 10 U of RNase-free DNase (Promega). After 5 minutes at 37ºC, RNA was extracted with Trizol reagent with following modifications: the DNase-treated insoluble fraction was incubated with Trizol reagent at 50ºC for about 5 minutes, 200 µl of chloroform was added and the samples were centrifuged 15 minutes at 11500 rpm in a microcentrifuge. The supernatant was transferred to a new tube and extracted with phenol/chloroform and ethanol precipitated. Cytoplasmic and nuclear soluble fractions were treated with proteinase K prior to RNA extraction with Trizol as described for the nuclear insoluble RNA fraction.

Cytoplasmic A$^+$/A$^-$ RNA extraction

Cytoplasmic RNA was diluted to <500 µg/ml in sterile buffer E (0.1 M Tris, 1 M EDTA and 0.1% SDS) and passed over oligo-dT columns. The A$^-$ RNA fraction was recovered from the flow-through by ethanol precipitation. The A$^+$ RNA fraction was eluted with Buffer A (0.1 M Tris, 1 mM EDTA, 0.1% SDS and 0.2 M NaCl), ethanol precipitated and resuspended to 0.5 µg/µl..

RNA ANALYSIS

S1 nuclease protection assay

Expression of the exogenous IgM gene was analyzed by the S1 nuclease protection assay utilizing a 3' end-labeled probe that differentiates cleaved and

polyadenylated µs mRNA from spliced µm mRNA (Peterson et al., 2002; Peterson and Perry, 1989). The probe corresponds to a 640 bp region containing the Cµ4 exon and part of the intron downstream of the µspA, as well about 30 bp of unrelated sequence (see Figure 3.2). The 640 bp probe was PCR amplified from 100 ng of any Cµ plasmid DNA with primers "Pst-T" and "Hind" (Table 2.2) for 30 cycles of 1 minute at 94°C, 2 minutes at 55°C and 2 minutes at 72°C. The PCR product was purified using the QIAEX II desalting and concentrating DNA protocol following manufacturer instructions, and 200 ng of *Pst*I-digested fragment was 3'-end labeled using the Klenow fragment of DNA polymerase I and [α-^{32}P]-dCTP. The 3´end-label PCR product was purified through a Centri-Spin-20 column (Princeton Separation, N.J) and its specific activity (cpm) was determined in a scintillation counter.

A total of 100 µg of RNA containing 10 µg of either M12 or S194 RNA and carrier RNA were precipitated along with 20,000 cpm of 3´-end-labeled probe. The pellet was resuspended in 50 µl of hybridation solution (80% formamide, 0.6 M NaCl, 10 mM PIPES, 1mM EDTA) and incubated overnight at 50°C after denaturation for 15 minutes at 80°C. The S1 nuclease digestion was carried out by adding 450 µl of S1 buffer (3 mM Zinc acetate, 30 mM sodium acetate and 250 mM NaCl) and 60 units of S1 nuclease (Invitrogen), incubating at 37°C for 30 minutes. Protected fragments were extracted with chloropane, precipitated with ethanol and analyzed on a 6 % denaturing polyacrylamide gel (6 % acrylamide/bis-acrylamide, 7 M urea) at 350 V for 1 hour. The dried gel was quantitated using Phosphorimage analysis and ImageQuant 5.0 (Molecular Dynamics).

In Vitro transcription

To synthesize a control RNA for reverse transcriptase (RT) reactions, a 2976-bp EcoRV fragment of µ gene was subcloned into pGEM4. This fragment contains about 594-bp upstream of the M1 exon (including the *Kpn*I site) and 1900-bp downstream of µm pA. This fragment was *in vitro* transcribed from pGEM-RV using SP6 RNA polymerase (Promega) and the Ampliscribe transcription kit (Epicentre) following the manufacturer instructions. Briefly, 10 µg of pGEM-RV was linearized with *BamH*I followed by chloropane extraction and ethanol precipitation. The pellet was resuspended in water at 1 µg/µl and 2 µg of the *BamH*I-linearized pGEM-RV was *in vitro* transcribed by incubation at 37°C for 90 minutes in presence of 10 mM DTT, 80 U of RNaseout (Invitrogen), 0.5 mM of each rNTP and 17 U of SP6 (Promega). The sample was treated with 2.5 U of RQ1 DNase I (Promega) for 15 minutes at 37°C follow

chloropane extraction and ethanol precipitation. The pellet was resuspended in 6 μl of water and 2 μl was electrophoresed on a 1 % agarose formaldehyde gel and visualized by ethidium bromide staining to ensure it was intact. One μl of a 1:1000 dilution of the *in vitro* transcribed RNA was used for the reverse transcription reactions with a μ sequence specific primer and 1 μl of the 1:10 diluted RT reaction was used for PCR reactions.

Reverse Transcriptase-PCR (RT-PCR)

One to two μg of total RNA was reverse transcribed with hexamer random primers (Invitrogen) or a gene-specific primer using the Omniscript RT Kit (QIAGEN) following manufacturer instructions. cDNA was PCR amplified using 1-5 μl of the (undiluted or 1:5 diluted) RT reaction. To ensure the reactions were within the linear amplification range, PCR products were collected at different cycle numbers. PCR products were analyzed on agarose gels containing ethidium bromide, photographed and when required, specific PCR product bands were quantified from the digital images using ImageQuant 5.0 (Molecular Dynamics). All PCR reactions were performed with controls in which the reverse transcriptase was omitted from the reaction; all the RT(-) reactions were blank, indicating there was no DNA contamination in the RNA samples (data not shown). The images were often inverted for presentation in the figures.

For radioactive PCR, 40 ng of reverse primer was 5'-end labelled with γ-^{32}P-ATP and polynucleotide kinase. The 5´end-labeled primer was eluted through a Centri-Spin-10 column (Princeton Separation, N.J) and its specific activity (cpm) was determined in a scintillation counter. PCR reactions were performed as usual but in 25 μl final volume and containing 0.5 μl of 5'-end labelled primer and 5 μl of RT reaction. PCR products were ethanol precipitated, resuspended in 5 μl of S1 loading buffer (80% formamide, 1 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol) and analyzed on a 6% or 8% denaturing polyacrylamide gel (6-8% acrylamide/bis-acrylamide, 7 M urea). The PCR product bands were visualized from the dried gels using phosphorimage analysis.

3´-Rapid amplification of cDNA-end by PCR (3'-RACE)

One μg of total RNA was reverse transcribed with a tagged-poly-dT primer (RT-RACE primer) using the Omniscript RT Kit (QIAGEN) following manufacturer instructions. Five μl of synthesized cDNA was PCR amplified using a reverse primer targeting the 3´-end tag (3´RACE-1 primer) and a forward primer complementary to the Cμ3 exon (Cμ3T primer) for 35 cycles of 94°C for 1 minute, 56°C for 1 minute and 72°C

for 2 minutes. 5 µl of the PCR product was submitted to a nested PCR with the "3´RACE-1" primer and a forward primer complementary to the Cµ4 exon (Apa primer) for 35 cycles of 94°C for 1 minute, 52°C for 1 minute and 72°C for 2 minutes. Ten to twelve µl of PCR product were digested with 5 units of either *Nhe*I (Invitrogen) or *Spe*I (NEB) overnight at 37°C and digested and undigested samples were visualized by ethidium bromide staining in either 1.8% agarose gels or 8% non-denaturing polyacrylamide gels.

TRANSGENIC MICE

Transgenic mice expressing the TTR-Flag-Zhx2 were generated in a (C57BL/6 × C3H) background by the University of Kentucky Transgenic Mouse Facility. Transgenic founders were bred to BALB/cJ mice and offspring were screened by PCR using DNA extracted from tail biopsies for the presence of the TTR-Flag-Zhx2 transgene and primers TTR and H1R. The endogenous Zhx2 allele was assessed with primers Afr1, Afr2 and ETn3Flank. Nuclear extracts or RNA were prepared from liver of 3 week old or older littermate with or without the Flag-Zhx2 transgene.

Transgenic mice expressing the AFP minigenes were generated in a (C57BL/6 × C3H) background by the University of Kentucky Transgenic Mouse Facility. AFP cDNA and genomic minigenes were jointly constructed in the Peterson and Spear labs. The AFP minigenes consist of the AFP exon 1 to exon 3 fused to the last two exons, exon 14 and exon 15, in its genomic (intron containing) or cDNA (no introns) forms. An 18 bp oligonucleotide was inserted into the *Bsu*36I site in exon 15 to distinguish the transgene form the endogenous gene. The minigenes were linked to the AFP enhancer II and either the AFP 250 bp promoter or 1 kb AFP promoter/repressor regions. Independent founders were crossed to BALB/cJ and two F1 from each founder were backcrossed to BALB/cJ. F2 offspring were screened by PCR using DNA extracted from tail biopsies. Presence of 250-AFP minigenes were analyzed with primers AFP120+165R and AFPE2, the 1kb-AFP minigenes with primers E3F and E15R and presence of the mutant or wild-type Zhx2 allele with primers Afr1, Afr2 and ETn3Flank. To determine Zhx2 regulation of AFP minigenes, liver RNA was prepared from F2 transgenic mice and analyzed by RT-PCR with primers E3F or E14E15-F and E15R.

PROTEIN PREPARATION AND ANALYSIS

Nuclear extract preparation from mouse liver

Nuclei from mouse liver were purified following the two sequential sucrose cushion procedure as described in the cellular RNA extraction section. Purified nuclei obtained from two mouse livers were resuspended in 900 µl of low-salt buffer (20 mM HEPES pH 7.9, 25% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT and proteinase inhibitors) using a 7 ml Dounce homogenizer and pestle B. Equal volume of high-salt buffer (20 mM HEPES pH 7.9, 800 mM KCl, 25% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT and proteinase inhibitors) was added in a dropwise fashion to the nuclei suspension while mixing with the pestle B. The extract was incubated for 30 minutes at 4ºC with gentle mixing and transferred to a 2 ml microcentrifuge tube. Nuclear extract was obtained by centrifugation at 21,000×g for 30 minutes at 4ºC. Aliquot of supernatants (NE) were quickly frozen in liquid nitrogen and storage at -80ºC. Pellets were resuspended in 900 µl of 8M urea and 900 µl of 2X loading buffer (125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 0.02% bromophenol blue, 0.2 M DTT) and incubated for 4-5 minutes in a boiling water bath. Total nuclei suspension in low-salt buffer, nuclear extract and suspended pellet were analyzed by 8% SDS-PAGE, transferred to PVDF membrane and blotted with anti-Zhx2 and/or anti-Flag antibodies.

Nuclear extract preparation from HEK293 cells

HEK293 cells transiently transfected with a Flag-Zhx2-expressing plasmid were harvested, washed with PBS (4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) and prepared nuclear extract using the NE-PER nuclear and cytoplasmic extraction reagent (Thermo Scientific) following the instruction manual. Total protein was determined by BCA protein assay and 2.5 to 5 µg of total protein was used as positive control in Western blot analysis with anti-Flag or anti-Zhx2 antibodies.

Flag-tagged Zhx2 immunoprecipitation

Immunoprecipitation was performed by batch absorption following the instruction manual of the anti-Flag M2 affinity gel (Sigma). Nuclear extracts (100 µg) were resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) up to a final volume of 500 µl and incubated overnight with 10 µl of packed volume of beads previously equilibrated with TBS. The beads were washed three to four times with 500 µl of TBS (50 mM Tris-HCl pH 7.4, 150 mM NaCl) and elute with 20 µl of

2X loading buffer incubated 4 minutes in boiling water bath. Beads were collected by centrifugation at 7000×g for 1 minute and the eluate was analyzed by Western blot.

Western blotting

Protein samples from total nuclei suspension, nuclear extracts or insoluble pellets were fractionated on 8% SDS-PAGE and electro-blotted to PVDF (Polyvinylidene fluoride) membrane (Immobilon-P, Milipore) as described by Ausbel et al. (Ausubel et al., 1987). Filters were blocked 2 hours or overnight with blocking buffer (20 mM Tris-HCl pH 7.6, 140 mM NaCl, 0.1 % Tween-20 and 5% non-fat dry milk). Membranes were was incubated 2 hours at room temperature or overnight at 4ºC with primary antibody diluted in blocking buffer and then washed 4 times with TBST (20 mM Tris-HCl pH 7.6, 140 mM NaCl, 0.1 % Tween-20). Membranes were incubated with secondary antibody 2 hours at room temperature, washed 4 times with TBST and bands were visualized by enhanced chemiluminescence (Pierce). Primary antibodies used: monoclonal mouse anti-Flag M2 (Sigma) 1:500 dilution; polyclonal rabbit Anti-Zhx2 (Bethyl laboratories) 1:2000 dilution. Secondary antibodies used: goat anti-mouse IgG-HRP (sc-2031) 1:4000 dilution and goat anti-rabbit IgG-HRP (sc-2004) 1:8000 dilution.

BCA protein assay

Total protein concentration from nuclear extracts was determined by BCA (PIERCE), following the instruction manual.

Table 2-1: Plasmid constructs used in Cµ study.

Plasmid	Construct
pGEM-T-CoTC[a]	β-globin CoTC (800 bp) PCR product cloned into pGEM-T
Cµ [c]	Intact Ig µ gene in the pR-SP6 plasmid backbone
ΔNH [d]	Intact Ig µ gene containing a 158 bp deletion of the pause element between *Not*I and *Hind*III sites
ΔAR [b]	Intact Ig µ gene containing a 797 bp deletion between *Acc*I and *EcoR*V sites
ΔNR [b]	Intact Ig µ gene containing a 351 bp deletion between *Nco*I and *EcoR*V sites
ΔAN [b]	Intact Ig µ gene containing a 446 bp deletion between *Acc*I and *Nco*I sites
ΔAR/pBR403 [b]	403 bp fragment from pBR322 inserted in the *Kpn*I site of ΔAR
ΔAR/rpl900 [b]	900 bp fragment from the intron of the S16 ribosomal protein gene (rpl900) inserted at the *Kpn*I site of ΔAR
Cµ/rpl900 [b]	rpl900 fragment inserted at the *Kpn*I site of Cµ
Cµ/CoTC(+)[a]	β-globin CoTC inserted in 5'-3' orientation 1200 bp downstream of µspA (*Kpn*I site) of Cµ
Cµ/CoTC(-)[a]	β-globin CoTC inserted in 3'-5' orientation 1200 bp downstream of µspA (*Kpn*I site) of Cµ
ΔNH/CoTC(+)[a]	β-globin CoTC inserted in 5'-3' orientation 1042 bp downstream of µspA (*Kpn*I site) of ΔNH
ΔNH/CoTC(-)[a]	β-globin CoTC inserted in 3'-5' orientation 1042 bp downstream of µspA (*Kpn*I site) of ΔNH
ΔAR/CoTC(+)[a]	β-globin CoTC inserted in 5'-3' orientation 403 bp downstream of µspA (*Kpn*I site) of ΔAR
ΔAR/CoTC(-)[a]	β-globin CoTC inserted in 3'-5' orientation 403 bp downstream of µspA (*Kpn*I site) of ΔAR
Cµ/CoTC-H[a]	β-globin CoTC inserted in 5'-3' orientation 200 bp downstream of µspA (*Hind*III site) of Cµ
ΔNH/CoTC-H [a]	β-globin CoTC inserted in 3'-5' orientation 50 bp downstream of µspA (*Hind*III site) of ΔNH
Cµ/rpl900-H [a]	rpl900 fragment inserted 200 bp downstream of µspA (*Hind*III site) of Cµ
ΔNH/rpl900-H [a]	rpl900 fragment inserted 50 bp downstream of µspA (*Hind*III site) ΔNH

[a] This work
[b] Peterson and Perry (1986)
[c] Peterson and Perry (1989)
[d] Peterson et al (2002)

Table 2-2: Oligos used for reverse transcriptase and PCR reactions in Cμ study.

CoTC-Fc (#342): 5' ACTTAGGTACCATAGTGTTACCATCAACCACC 3'
CoTC-R800 (#344): 5' ACAATGGTACCTCCCACCCCCAACCCCT 3'
Hind (#4): 5' ATATGTGCCTGAATGCTGCC 3'
Pst-T (#264): 5' TGGTGAAGGGCTTCTCTCCTG 3'
RT-RACE (#48): 5' AGAGAATTCACCGGATCCTACCCGGGTTTTTTT 3'
3´RACE-1(#51): 5' GAGAATTCACCGGATCCTAC 3'
Apa (# 7): 5' TATGTGACCAGTGCCCCGAT 3'
Cμ3T (#86): 5' TGACTCACAGGGATCTGCCT 3'
C4B (#274): 5' CATTCCTCCTCTGTCACAGTCAGG 3'
CμpA-T (#242): 5' CCTGGGTGTCCAGTTGCTCTGT 3'
432AS (#246): 5' TGTACCTTTGTCTGTGTGTCCATGA 3'
1528AS (#248): 5' GGCAGACTCAGGAACTGTGGCA 3'
Kpn-T (#265): 5' GATCAGAGATCCCAATAAATGCCC 3'
Kpn-B (#278): 5' AAGGACTCGGACATCTCATGACTTT 3'
M1-B (#277): 5' CAAAGCCTTCCTCCTCAGCATT 3'
M2-T(#279): 5' TGTACAATCTGAAGCAATGTCTGGC 3'
M2-B (#280): 5' CAGATAATCCTTGGGCTGTCAACA 3'
469-T (#281): 5' TGGGTGTGCCTTTGGGGAGT 3'
469-B (#282): 5' GCCTTTATCTGTCCTCGAGTGACT3'
473AS (#546): 5' AGGGACCATCACCAAGGTGC 3'
982AS (#548): 5' TAGCTCTGAGGTCATGGCGGGT 3'

Table 2-3: Oligos used for PCR reactions in Zhx2 study.

TTR (#110S): 5' TGTTCCAGAGTCTATCACCG 3'
B-ActF (#131): 5' GTGGGCCGCTCTAGGCACCA 3'
B-ActR (#132): 5' CGGTTGGCCTTAGGGTTCAGGGGGG 3'
L30R (#138): 5' CCTCAAAGCTGGACAGTTGTTGGCA 3'
AFPE12A (#267): 5' TGAGCTTGGCACAGATCCTTGTGGA 3'
E1E5H19-F (#322): 5' TCCAGAGTCCGTGGCCAAGGAGG 3'
E1E5H19-R (#323): 5' TGGATTCTCAGGGGTGGGTGGGT 3'
E2E3-F (#324): 5' TCCACGTTAGATTCCTCCCAGTGCG 3'
E2E3-R (#325): 5' TGTTCACTTCCTCCTCGGTGGCTTC 3'
E5E6-R (#327): 5' TTGCCTGGAGGTTTCGGGATCC 3'
E10E11-F (#330): 5' TGCAGAAACACATCGAGGAGAGCCA 3'
E10E11-R (#331): 5' CGTGGAGGCAATGCTCACCATCTTC 3'
E14E15-F (#334): 5' GGTCCAAAGTTGATTTCCAAAACTC 3'
E13E15-F (#337): 5' GCCAAAAGGCTCACACCAAAGAGTC 3'
E11E12F (#373): 5' TTATTGGTTACACGAGGAAAGCCC 3'
E12E13-FB (#377): 5' GCCGACATTTTCATTGGACATTTGT 3'
E12E13-RB (#378): 5' CAGACTTCCTGGTCCTGGGCTTT 3'
Afp3'-F (#396): 5' TTGATTTCCAAAACTCGTGATGC 3'
Afp3'-R1 (#397): 5' GAAATCTCACATGGACATCTTCACC 3'
Afp3'-R2 (#398): 5' TCAGAGGAGAAGTGGAATGGTGA 3'
H1R (#405S): 5' CCCATAGATTTCACCTCAACC 3'
Gpc3 U-2 (#423): 5' TGACCGTGACCTTGAGCGTC 3'
nest12-13BF (#432): 5' TTCTCCCAGTTTTCTTCACCCCAG 3'

nest12-13BR (#433): 5' AACCCTAGCAGCCATCCTCAG 3'
Gpc3E7-U (#438): 5' ACGGGATGGTGAAAGTGAAGAAT 3'
Gpc3E8-L (#440): 5' CGCTGTGAGAGGTGGTGA 3'
H19Ex2-L (#443): 5' AAGAAGGCTGGATGACTG 3'
S-H19 (#463S): 5' CGTTCTGAATCAAGAAGATGCTGC 3'
AS-H19 (#464S): 5' TTTGAGTCTCTCAAGCAAGGAAGG 3'
E1F (#499): 5' GTGAAGGAACCAGCAGCCATGA 3'
E3F (#500) 5' AATGACTAGCGATGTGTTGGCTGC 3'
E14R (#501): CGCCCAAAGCATCACGAGTTT 3'
E15R (#502) 5' TTCCTAAGGCAAGCTTCGAGGCT 3'
U6F (#579): 5' GTGCTCGCTTCGGCAGCACATATA 3'
U6R (#580): CGAATTTGCGTGTCATCCTTGCG 3'
L30F2 (#581): 5' TGTCCCGCACCTAAGGCAGGAAGAT 3'
i9-F (#596): 5' AAGGCAAAATCTCCTGTCCAGTGG 3'
i10-R (#597) 5' GGTTTGGGGAACCCGAAGTCAAT 3'
i10-F (#598) 5' TCAAGGAAGAGCTAAGGTTGGGAGG 3'
i11-R (#599) 5' TTTGAGAAGGACAGAACGCCTGGC 3'
i2-F (#600) 5' CTGAGTGACTCGAACCCAGTTTCCC 3'
i2-R (#601) 5' CTGGAACTCCAGTCAATGTCAAGGC 3'
i3-R (#602) 5' GGCGGGATTCCAGAACTAATCG 3'
i5-R (#605) 5' CCCACAGGAGGCACATTTCTGC 3'
i1-F (#607) 5' TGTGAGGGAGCCAAGTAGTAAGGG 3'
E5-F (#608) 5' TGCCCCAGCCATTCTGTCCT 3'
Z2HD1-F (#1017S): 5' TCCCCGTCCCGCTGAATACT 3'
Z2HD1-R (#1018S): 5' TGTGCTCCTCTGGGTGTTTG 3'
Gpc3URT (1031S) 5' CTCCCAGCAACGCCAATATAGATC 3'
Gpc3LRT (1032S) 5' CTGATTCTTCATCCCGTTCCTTGC 3'
H19Ex2-U (#1128S) 5' ACAGAAGGGCAGTCATCCAG 3'
H19Ex3-L2 (#1129S) 5' CCAGAGAGCAGCAGAGAAGT 3'
H19Ex3-L3 (#1131S) 5' GTTCAAGGTAGGGGGAGGAG 3'
Afr1 (#1143S) 5' AACCCAAGTAGCAAGTGAGTGGTG 3'
Afr2 (#1144S) 5' ATCTCTGGTCCAGTAGAAAGGTGC 3'
ETn3Flank (#1259S) 5' CCTATCTTGGTACCCTGATCAAGC 3'
AFP120+165R (#1276S) 5' CCTTACGTTGAAGTTAGGCAA 3'
AFPE2 (#1280S) 5' GATCTGGATCCGGGGAAATAATCT 3'

CHAPTER III

AFP SPLICING REPRESSION BY ZHX2

INTRODUCTION

That transcription and post-transcriptional events are closely associated is suggested by interactions between components involved in transcription initiation, elongation and termination with those involved in RNA processing (capping, splicing and cleavage and polyadenylation). Coupling transcription with RNA processing is mediated, at least in part, by the C-terminal domain (CTD) of the RNAPII largest subunit, which contains multiple heptapeptide repeats that are targets of regulated phosphorylation. Dynamic changes in CTD phosphorylation are involved in modulating recruitment of factors required for elongation and RNA processing in a temporal-spatial manner. During transcription initiation, the RNAPII CTD is phosphorylated at serine 5 of the heptapeptide (S5), resulting in recruitment of DSIF and NELF to the transcription complex (Figure 1-1). This induces a promoter proximal pausing that allows binding of capping enzymes. Pausing is relieved by phosphorylation of the RNAPII CTD at serine 2 (S2), mediated by recruitment of pTEFb to the elongation complex. This causes dissociation of NELF and transcription elongation proceeds (Bentley, 2002; Bres et al., 2008; Mandal et al., 2004; Sims et al., 2004b). Although under certain conditions, it has been reported that RNAPII elongation could proceed in the absence of pTEFb, the coupling of transcription with RNA processing is disrupted, giving rise to full-length RNA transcripts that are quickly degraded (Hargreaves et al., 2009). Such is the case of primary response genes in unstimulated macrophages. Soon after activation, pTEFb is recruited to the elongation complex and RNA processing is reestablished to allow rapid production of mature mRNA (Hargreaves et al., 2009). Additionally, promoter structure has also been implicated in coupling transcriptional and post-transcriptional events to control gene expression. Studies with reporter minigenes containing the alternative exon (E33 or EDI) of the *fibronectin* (*FN*) gene showed that FN and CMV promoters induced 10 times higher inclusion of the exon E33 compared to the α-globin promoter (Cramer et al., 1999; Cramer et al., 1997; Kornblihtt, 2005, 2006). Therefore, post-transcriptional events can be modulated by altering the elongation properties of the RNAPII in the pre-initiation complex. This may result in changes in elongation rates or

43

recruitment of RNA processing factors and ultimately may have effect on the fate of the nascent transcript and regulate gene expression (Kornblihtt, 2005).

Expression of the AFP gene is developmentally regulated. It is expressed at high levels in fetal liver and dramatically repressed after birth. Interestingly, AFP expression is reactivated in hepatocellular carcinomas and during liver regeneration. Because of this characteristic pattern of AFP gene expression, it has become an important model for liver-specific gene expression during liver development, liver regeneration and disease (Spear et al., 2006). The first information about the mechanism of AFP repression was obtained by screening adult serum AFP levels of different inbred mouse strains (Olsson et al., 1977). All the strains expressed low serum levels of AFP with the exception of BALB/cJ mice that expressed about 10 to 20-fold higher levels than the other mouse strains analyzed. Post-natal persistence of AFP expression in BALB/cJ mice is due to a mutation in the *Zinc-fingers and homeoboxes 2* (*Zhx2*) gene (Blankenhorn et al., 1988; Olsson et al., 1977; Perincheri et al., 2005), a member of a small family of proteins that contains two zinc fingers and four or five homeodomains (Spear et al., 2006). Analysis of the BALB/cJ Zhx2 allele revealed the presence of a mouse endogenous retrovirus (MERV) element that interferes with proper splicing of the Zhx2 transcript, downregulating its expression in this mouse strain (Perincheri et al., 2008). Lower Zhx2 levels in adult BALB/cJ liver correlate with increased accumulation of AFP and H19 mRNA levels, indicating the role of Zhx2 in repressing expression of these genes after birth. Other Zhx2 target genes identified are Gpc3 (Morford et al., 2007) and Lpl (Gargalovic et al., 2010) Both showed developmental expression patterns similar to the AFP and H19 genes, high in fetal liver, repressed after birth and reactivated in liver cancer and during regeneration (Gargalovic et al., 2010; Morford et al., 2007; Spear et al., 2006).

Transgenic mouse studies using minigenes or other reporter genes linked to different portions of the AFP 5' flanking sequences have been used as a strategy to better understand the mechanism of AFP repression mediated by Zhx2. The 250 bp AFP promoter was sufficient to confer Zhx2 regulation to a linked reporter gene (*H-2D^d*) suggesting that Zhx2 acts at transcriptional level (Peyton et al., 2000a). However, transcriptional rates on the AFP and H19 genes, measured by nuclear run-on assays, were the same in the presence or absence of Zhx2 expression, indicating that Zhx2 may act at post-transcriptional level (Morford; Vacher et al., 1992). Based on these results, it was proposed that Zhx2 controls accumulation of mRNA levels of targets genes by a

mechanism that couples transcription and post-transcriptional events (Peyton et al., 2000a). Therefore, we explored splicing as possible post-transcriptional step involved in Zhx2 regulation and indeed, we found that splicing of AFP was repressed in the presence of Zhx2. We further investigated whether introns are required for splicing regulation and whether alternative splicing plays a role in AFP post-natal repression mediated by Zhx2. We also analyzed the splicing efficiency of the H19 and Gpc3 genes, two other Zhx2 target genes. Preliminary results found no evidence of splicing inhibition in these genes in response to Zhx2. To investigate possible factors controlled by Zhx2 that are involved in coupling transcription and RNA processing, we intend to purify and identify proteins that interact with Zhx2. Toward this end, we initiated optimization of nuclear extracts preparation and immunoprecipitation assays that eventually will lead to purification of the Flag-tagged Zhx2 and its interacting proteins.

RESULTS

AFP splicing repression mediated by Zhx2

Based on the preliminary data described above, it was proposed that Zhx2 regulates AFP expression through a post-transcriptional mechanism controlled by the AFP promoter. To determine what post-transcriptional step is involved in that regulation, splicing efficiency across the AFP gene was analyzed first. Total liver RNA from adult littermate mice with or without the Flag-tagged Zhx2 transgene was utilized for cDNA preparation using random hexamers. PCR amplification was performed with sets of primers spanning several exon-exon boundaries throughout the AFP gene (Figure 3-1). Data from the PCR reactions diagrammed in Figure 3-1, along with the expected sizes for the unspliced and/or spliced AFP transcripts are shown in Figure 3-2. Ribosomal protein L30 (rpL30) was used as an expression control. Negative controls without RT enzyme were performed and no signal corresponding to PCR products for L30, H19 or AFP E10-E11 were observed, indicating there was no DNA contamination (Figure 3-2). When primers targeting several exon-exon regions was performed (Figure 3-2, top panels: AFP E2-E3, AFP E5-E6, AFP E10-E11, AFP E12-E13, AFP E14-E15), samples from animals expressing Flag-Zhx2 showed lower accumulation of spliced mRNA than that observed from samples of animals not expressing Flag-Zhx2. This result is consistent with previous data that showed AFP and H19 mRNA accumulated to higher levels in BALB/cJ mice (no Zhx2 expression). Interestingly, PCR analysis of AFP E10-E11, AFP E12-E13 and E14-E15 (Figure 3-2, top panels of AFP E10-E11, AFP E12-E13

45

and AFP E14-E15) showed that in the presence of Zhx2, bands corresponding to unspliced RNA were detected with higher intensity than those observed in the absence of Zhx2. This suggested that AFP splicing efficiency across multiple exons within the 3'-end of the AFP gene was substantially reduced in Zhx2-expressing animals. To verify differences in unspliced RNA levels between mice with or without Zhx2, PCR reactions with primers targeting exon-intron or intron-intron boundaries were performed. This strategy will only detect unprocessed RNA. In all the cases assessed (Figure 3-2, middle and bottom panels of AFP E2-E3, AFP E5-E6, AFP E10-E11, AFP E12-E13, AFP E14-E15), bands corresponding to the unspliced RNA showed higher intensity in the presence of Zhx2 than in the absence of Zhx2. This strategy allowed detection of PCR product due to unprocessed RNA in E2-E3 and E5-E6 and indicated that AFP splicing repression mediated by Zhx2 is extended towards the 5'-end. Therefore, we conclude that Zhx2 has a general effect on the splicing efficiency throughout the AFP gene which would explain the reduced accumulation of AFP mRNA in adult BALB/cJ mice.

Analysis of cellular RNA fractions

Preliminary data demonstrated that Zhx2 was enriched in the chromatin-associated (insoluble) fraction, which is consistent with the idea that Zhx2 may be involved in a co-transcriptional regulation step. Therefore, the unprocessed RNA observed in the presence of Zhx2 may be more abundant in the nuclear fraction. To explore the cellular localization of the unspliced RNA, liver tissue from littermate mice with or without the Flag-tagged Zhx2 transgene was fractionated into cytoplasmic, chromatin soluble (N_{Sol}) and chromatin-associated (insoluble, N_{Ins}) fractions (Pandya-Jones and Black, 2009). RNA was extracted from each fraction, cDNA was made with random hexamers and AFP expression was analyzed by PCR using primers spanning AFP E10-E11 (Figure 3-3). β-actin and rpL30 were utilized as loading controls. U6 RNA, which is found exclusively in the nucleus, was used as a fractionation control. A U6 RNA PCR product was observed only in the nuclear soluble and nuclear insoluble fractions as expected, indicating that cytoplasmic fraction was clean. A PCR signal was not detected in the nuclear insoluble fraction using primers targeting the AFP E10-E11 or L30, suggesting that the spliced RNAs must be rapidly released from the chromatin and unspliced AFP RNA is not retained in this fraction. However, PCR products from the β-actin and U6 reactions were visible, indicating that the RNA in this fraction was not degraded and thus, the absence of AFP E10-E11 PCR product was due to low

abundance of the AFP transcript in this fraction. In the nuclear soluble fraction, both spliced and unspliced AFP E10-E11 was detected and relatively more of the PCR band corresponding to the spliced RNA was seen in absence of Zhx2. Thus, the Zhx2 regulation of mature mRNA abundance is seen in the nucleus, consistent with this being due to a nuclear regulatory event. However, there wasn't an obvious difference in the unspliced PCR product in the presence or absence of Zhx2. Surprisingly, a dramatic difference was observed in the cytoplasmic fraction in which high accumulation of the unspliced RNA was observed in presence of Zhx2 with concomitant lower levels of spliced RNA, compared to that observed in absence of Zhx2. Negative controls without RT enzyme were performed and no signal corresponding to PCR products for L30, β-actin or AFP E10-E11 were observed, indicating there was no DNA contamination (data not shown). This experiment was repeated with the cytoplasmic RNA fraction from other independent sets of littermate mice and the same result was observed. Additional RT-PCR reactions were performed with primer pairs spanning different exon-exon and exon-intron boundaries (as shown in Figure 3-1) using the cytoplasmic RNA fractions (results not shown). In all cases, higher levels of unspliced product was observed in presence of Zhx2, indicating that unprocessed AFP transcripts are present in the cytoplasm and thus have escaped from nuclear surveillance processes.

The cytoplasmic fraction was further characterized by analyzing poly-A$^+$ and poly-A$^-$ RNA fractions by RT-PCR. A$^+$ and A$^-$ RNA fractions were obtained from liver cytoplasmic RNAs by separation through oligo-dT columns. cDNA was made from each fraction utilizing random hexamers and PCR reactions were performed with primers targeting AFP E10-E11, E14-E15 and L30 as an internal control (Figure 3-4). In the A$^+$ fraction, the AFP E10-E11 spliced PCR product was mainly detected in absence of Zhx2, consistent with the known mRNA expression differences seen in the presence or absence of Zhx2. The unspliced product was not detected in either of the A$^+$ RNA fractions, indicating that the unspliced RNA is not polyadenylated. Indeed, the main PCR product observed in the A$^-$ fraction corresponded to the unspliced RNA, which was more abundant in presence of Zhx2. Similar results in both the A$^+$ and A$^-$ fractions were seen with the AFP E14-E15 primers. Although it was unexpected to find unspliced RNA in the cytoplasmic fraction, there are some reports in the literature that unspliced RNAs are found in the cytoplasm when splicing is inhibited with small molecules or mutations in the core splicing machinery (Kaida et al., 2007; Legrain and Rosbash, 1989; O'Brien

et al., 2008). Therefore, the data discussed here is consistent with global repression of AFP splicing mediated by Zhx2.

Splicing analysis of other known targets of Zhx2

Other known targets of Zhx2 are the H19 and Gpc3 genes. We proceeded to investigate whether Zhx2 may regulate expression of these genes by repressing pre-mRNA splicing as was observed with AFP. Conventional RT-PCR was used to analyze Gpc3 splicing in the presence and absence of Zhx2. Both cytoplasmic and nuclear soluble fractions were used for cDNA preparation. PCR reactions were performed with pairs of primers targeting exon and intron boundaries across the Gpc3 gene (Figure 3-5). Higher accumulation of spliced Gpc3 mRNA (PCR E4-E6 and PCR E7-E8) was observed in absence of Zhx2 in both cytoplasmic and nuclear soluble fractions, though the difference in abundance was not as obvious as that observed in the AFP gene as seen previously (Morford et al., 2007). Unspliced product (Gpc3 I5-E6) was not detected in the cytoplasmic fraction and little difference in accumulation was seen in the nuclear soluble fraction. PCR product signals for the unprocessed RNA were significantly weaker than those observed for the spliced RNA, indicating that unprocessed transcripts are present in lower abundance. To examine H19 splicing, primers in Exons 4 and 5 were used (Figure 3-5). No unspliced RNA was detected, but again, the repression due to Zhx2 was seen in both the cytoplasmic and nuclear fractions. The fact that the Zhx2 regulation was seen in the nuclear RNA fraction is consistent with this being due to a nuclear regulatory event, but it is not clear whether splicing is the regulated event for the Gpc3 and H19 genes.

Preliminary data had indicated that unspliced H19 RNA could not be detected by conventional RT-PCR. Therefore, PCR amplification was performed using a 5'-end radiolabeled primer to increase sensitivity. Splicing analysis across the H19 gene was performed by using pairs of primers that target different exons of the H19 gene, diagrammed in Figure 3-6A. The cytoplasmic RNA fraction was used for cDNA preparation because this fraction showed the higher difference in unprocessed AFP RNA. In all the cases, higher accumulation of mRNA was observed in the absence of Zhx2, which is consistent with previous Northern blot data (Figure 3-6B). However, unspliced H19 RNA could not be detected by this procedure; genomic DNA was used as a control to identify the correct size of unspliced RNA. There are several possible explanations for this outcome. One possibility is that splicing of H19 transcripts may be

repressed but the unprocessed RNAs are rapidly turned over before they can reach the cytoplasm. In fact, unspliced RNAs are usually rapidly degraded in the nucleus, so this result with the H19 RNA may be expected. To detect splicing repression, it may be better to examine nuclear RNA fractions. On the other hand, the H19 gene is small, about 3.6 kb, compared to the AFP gene that is 20 kb, and consists of five exons separated by unusually short introns (~79-87 bp) which may favor efficient splicing. Thus, it is also possible that H19 is post-transcriptionally regulated at a step later than splicing repression

Are introns required for Zhx2 regulation of target genes?

As discussed previously, the transcriptional rate across the AFP and H19 genes are similar in spite of Zhx2 expression. However, in the presence of Zhx2 there is a decrease of mRNA levels compared to that observed in absence of Zhx2. This difference in accumulated mRNA can be explained by general splicing inhibition of AFP pre-mRNA mediated by Zhx2. Early transgenic mouse studies utilized the AFP promoter to drive expression of AFP minigenes or the MHC class I *H-2D^d* (250AFP-D^d) gene, both containing introns. These transgenes were developmentally repressed and regulated by Zhx2 in a pattern similar to the endogenous AFP gene (Peyton et al., 2000a). However, when Zhx2 was transiently co-transfected into liver cell lines with intron-less luciferase reporter genes driven by the AFP promoter, only about a 2-fold repression was observed (Shen et al., 2008). It is possible that the lower fold repression observed with the luciferase reporter gene is because introns may be required for or contribute to Zhx2 regulation. This hypothesis was tested using AFP minigenes that differ only by the presence or absence of introns in littermate mice that either express the endogenous Zhx2 gene or have the mutant BALB/cJ allele. If introns are required for Zhx2 regulation, Zhx2 should regulate expression of the genomic AFP minigene whereas the cDNA AFP minigene expression should not be regulated.

To test the prediction that introns are required for Zhx2 regulation, the AFP minigenes diagrammed in Figure 3-7 were made, based on those used in previous studies that are known to be repressed at birth and responsive to Zhx2 expression (Peyton et al., 2000a; Vacher et al., 1992). These AFP mini-genes consist of the AFP enhancer 2 and the 250 bp AFP promoter linked to AFP exon 1 through 3 fused to the last two AFP exons (250AFP-MG), exon 14 and 15, either containing (genomic) or lacking introns (cDNA). Additionally, an insertion of 18 bp was included within exon 15

49

to allow exclusive detection of the AFP transgene. The AFP minigenes were sequenced and their expression was tested in HepG2 cells before being introduced into mice. Both genomic and cDNA AFP minigenes, were expressed at similar levels in HepG2 cells (A. Ribble, data not shown). The AFP transgenes were introduced into mice expressing the wild-type Zhx2 allele. Only one founder from each line was obtained and each backcrossed twice with BALB/cJ mice to determine whether these transgenes were responsive to Zhx2 regulation. Total RNA from adult liver of littermate mice from each line was analyzed by RT-PCR using primers targeting exon 3 and the 18-bp sequence inserted within exon 15 (Figure 3-7). Endogenous AFP and H19 RNA levels were analyzed by RT-PCR as controls for Zhx2 responsiveness and L30 was used as an expression control. In both the cDNA and genomic AFP minigene transgenic lines, lower levels of endogenous AFP and H19 mRNA were observed in transgenic mice expressing Zhx2 wild-type allele (Zhx2+), as was unspliced AFP RNA. In contrast, higher levels of endogenous AFP and H19 mRNA were observed in absence of Zhx2 expression, indicating that the endogenous Zhx2 targets genes were regulated in both mouse lines, as expected.

Analysis of RNA from five transgenic littermate mice expressing the AFP cDNA minigene showed that this transgene was responsive to Zhx2 regulation; lower minigene mRNA levels were detected in mice that expressed the Zhx2 wild-type allele (Figure 3-7, left panel; Zhx2+) compared to mice with the BALB/cJ allele (Figure 3-7, left panel; Zhx2-). The estimated fold-repression of the AFP cDNA transgene was about 3 to 4-fold while the endogenous H19 and AFP genes were repressed about 15-23-fold and 10-30-fold, respectively. This indicated that the cDNA AFP transgene was not repressed to the same extent as the endogenous AFP and H19 genes. In contrast, analysis of the AFP genomic transgene, showed that it was not well expressed among the littermate mice tested (Figure 3-7, right panel) and therefore, it was difficult to assess whether this transgene was regulated. There was high expression variability among littermates that did not correlate with the presence or absence of Zhx2. Similar results were obtained from analysis of two independent litters from this AFP genomic transgenic line (data not shown). Thus, while the cDNA transgene does appear to respond to Zhx2, the response is not as robust as the endogenous gene, which may suggest that introns are not required for Zhx2 regulation, but that they may contribute to the regulatory mechanism. However, since expression of the AFP genomic transgene is so low, we can not use it as

a positive control for this experiment to know whether the mutated transgene response is due to the lack of introns.

Because we only obtained one founder animal from each of the transgenes containing the AFP250 bp promoter and because the genomic minigene did not express well, additional mouse lines expressing the AFP cDNA or genomic minigene under control of the 1 kb AFP promoter and AFP E2 were generated (1kb AFP-MG) to test the contributions introns make to Zhx2 regulation. It was anticipated that the 1 kb promoter may provide more robust transgene expression. Five founder mice were obtained from each construct and named A-E for the cDNA minigene and G-K for the genomic minigene (Figure 3-8). Founders were crossed to BALB/cJ mice and at least two F1 mice from each founder were backcrossed to BALB/cJ. F2 mice were obtained from all founders except for the H and I of the AFP genomic minigene line that were sterile. RNA from adult liver of F2 littermate mice from A (Figure 3-9), E (Figure 3-10) and G (Figure 3-11 and Figure 3-12) founders were analyzed by RT-PCR as described above. In all cases, the endogenous AFP and H19 mRNA was regulated by Zhx2. However, there was variability in expression of AFP transgenes among the lines. The A2 AFP cDNA-minigene line did not express the transgene well, based on the very low AFP E3-E15* PCR signal (Figure 3-9). The AFP promoter-regulated transgenes should be repressed by four weeks after birth, so to see whether transgene expression in the A2 line might be higher at three weeks rather than five weeks after birth, multiple litters were collected at various ages. However, this did not improve detection of this transgene. In all the other cases, E2 AFP cDNA minigene and G1 and G2 AFP genomic minigene, AFP E3-E15* PCR signals were detected but unexpectedly, no significant differences in the transgene expression was observed in spite of the Zhx2 expression levels. These results indicated that the transgenes were not strongly regulated by Zhx2. Other strategies were utilized to analyze the AFP transgene expression, such as Northern blot and S1 nuclease (results not shown). No bands for AFP minigene were visualized in Northern blot from total RNA. Previous studies utilized A$^+$ RNA because AFP is expressed in low levels in adult liver. In general, the AFP transgenes were not well-expressed because about 2.5-fold more cDNA was required in the PCR reactions to get a signal intensity similar to the endogenous AFP gene. Additionally, low intensity bands corresponding to the AFP minigene could be detected by S1 nuclease protection assays but it required over 75 µg of total RNA.   Based on the published literature, we fully expected that the AFP genomic minigenes would be regulated by Zhx2; it is not clear why this is not the case.

Alternative splicing across the AFP gene

Based on results described above, we conclude that AFP splicing efficiency is dramatically reduced in Zhx2-expressing mice. To determine whether the presence of Zhx2 also caused any alternative splicing of the AFP RNA, we searched for alternatively spliced RNAs by RT-PCR using primers that spanned across different AFP exons (Figure 3-13A). Total RNA from two pairs of littermate mice with or without the Flag-Zhx2 transgene was analyzed. Under these conditions, AFP E1-E6 and AFP E5-E11 PCR yielded only the products expected for fully spliced AFP mRNA and these products were more abundant in the absence of Zhx2. The fully spliced mRNA also was observed for the AFP E10-E15 PCR, but in this case a weak smaller band of about 430 bp was also observed. The E10-E15 region was further analyzed to determine whether this band corresponded to an alternatively spliced AFP mRNA (Figure 3-13B). AFP E10-E12, AFP E10-E13 and AFP E11-14 PCRs yielded PCR products with the expected sizes for the fully spliced AFP mRNA. In addition, the AFP E10-E13 PCR also produced PCR products with smaller sizes. These other PCR products may be derived from alternatively processed RNAs or they may be artifacts from the PCR amplification. However, because the overlapping E10-E12 and E11-E14 PCR reactions did not detect any smaller products, it is more likely that the products in E10-E13 are PCR artifacts. Also, the abundance of these products was the same in the presence or absence of Zhx2. Thus, we did not pursue this further and conclude that there is no obvious alternative splicing mediated by Zhx2.

Nuclear extract preparation from mouse liver and Flag-tagged Zhx2 Immunoprecipitation

As discussed previously, Zhx2 regulates AFP expression post-transcriptionally by repressing AFP splicing in a promoter dependent manner. It is possible that Zhx2 may interfere with factors that directly or indirectly affect the pre-initiation complex. Alternatively, Zhx2 may affect steps of RNAPII elongation that impact post-transcriptional steps such as RNA splicing, RNA export and/or RNA degradation. To gain more insight into the mechanism of Zhx2 regulation, we intend to identify proteins that interact with Zhx2 using a proteomics approach. This strategy requires co-purification of Flag-tagged Zhx2 and its interacting proteins from liver nuclear extracts by immunoprecipitation with anti-Flag M2 agarose beads, followed by mass spectrometry. Transgenic mice that express Flag-tagged Zhx2 in the liver will be used in these experiments. Analysis of liver tissue fractions from Flag-Zhx2 transgenic mice by

Western blot demonstrated that the Flag-Zhx2 protein appeared mainly in the chromatin-associated fraction with the hyperphosphorylated RNAPII (Morford, L., unpublished data). It was also shown that Flag-Zhx2 was expressed at similar levels to the endogenous liver Zhx2 in wild-type mice and thus, similar to physiological levels. Moreover, Flag-Zhx2 expression in transgenic mice also complemented the Afr1 phenotype, so the AFP and H19 mRNAs were repressed after birth to the same extent as observed in mice that express wild-type Zhx2. Therefore, livers of these transgenic mice are ideal starting material for achieving the co-purification of Flag-Zhx2 and *in vivo* partners for the proteomic study.

Because Zhx2 was found to be tightly associated with the chromatin-insoluble fraction, the first step was to find a nuclear extract (NE) protocol that allowed recovery of Flag-Zhx2 without disrupting protein-protein interactions. As a control for unspecific interactions, littermate mice that did not express the Flag-Zhx2 transgene were included. NE were prepared by lysis of purified nuclei obtained from liver of adult mice with or without the Flag-Zhx2 transgene. Lysis of nuclei were performed based on three different protocols (Aygun et al., 2008; Dignam et al., 1983; Wuarin and Schibler, 1994) and analyzed by Western blot with anti-Flag or anti-Zhx2 antibodies. The best results were obtained using a Dignam-based protocol (Ausubel et al., 1987). NE preparation was optimized using buffers with different concentration of NaCl and KCl. The best Flag-Zhx2 recovery was obtained with 300 mM to 400 mM KCl (results not shown). Preparation of NE using this methodology was reproducible and at least half of the Flag-Zhx2 protein was recovered from the chromatin-insoluble pellet. Figure 4-14A shows a representative Western blot with anti-Zhx2 antibody for one NE preparation using 400 mM KCl in lysis buffer. As a positive control, NE from transfected HEK-293 cells with a plasmid expressing the Flag-tagged Zhx2 was prepared; Western blot analysis with anti-Zhx2 detects a band of about 98 KDa, which is the expected size for the Flag-Zhx2 protein product (Figure 3-14, Lanes 1 and 8). Similar results were observed with the anti-Flag antibody (results not shown). Analysis by Western blot of the NE prepared from the mice that expressed the Flag-Zhx2 transgene (NE+) also showed the 98 KDa band recognized by the anti-Zhx2 antibody that is not observed in the NE from littermate mice not expressing the transgene. The next step was optimizing the Flag-Zhx2 immunoprecipitation using the anti-Flag agarose beads. As an initial trial, immunoprecipitation was performed following the procedure indicated in the instruction manual. NE input and eluate from the immunoprecipitation were analyzed by Western

blot as shown in Figure 3-14B. The band of 98 KDa is observed in the eluate from NE of mice that expressed the Flag-Zhx2 (lanes 9 and 10, Figure 3-14B) that is not observed in the eluate from NE of transgene-negative mice (lanes 11 and 12, Figure 3-14B), indicating that this procedure could specifically pull down the Flag-tagged Zhx2 protein. However, one concern was that the buffer used in this immunoprecipitation assay contained 1% Triton X-100, which may be too harsh for maintaining some protein-protein interactions and is known to interfere with mass spectrometry. Because NF-YA was reported to interact with ZHX2 in HEK-293 cells (Kawata et al., 2003b), it could be used as a positive control for protein co-immunoprecipitation. Other IP trials were performed with reduced the detergent concentration, but a precipitate formed during the incubation of the NE with the agarose beads, and the eluates showed no specificity.

DISCUSSION

The results presented here indicate that Zhx2 dramatically reduces splicing efficiency of the AFP transcript. This is consistent with the reduced steady state AFP mRNA levels and increased pre-mRNA accumulation observed in the presence of Zhx2. Interestingly, the repression of splicing was observed across all the introns analyzed indicating that the effect on splicing is global rather than being specific for certain introns. Surprisingly, analysis of the cellular RNA fractions demonstrated that the higher accumulation of fully unspliced RNA was observed in the cytoplasmic RNA fraction. Moreover, analysis of the cytoplasmic A^+ and A^- RNA fractions showed that the unspliced AFP RNA is mainly in the A^- RNA fraction. This result suggested that the unspliced AFP RNA found in the cytoplasm is not polyadenylated or that the poly-A tail is too short to bind to the oligo-dT column. Therefore, the unprocessed AFP RNA is escaping the nuclear surveillance mechanisms and being exported to the cytoplasm where it is likely deadenylated. Although these findings were unexpected, there are some reports in the literature indicating that pre-mRNA could leak into the cytoplasm and even be translated when splicing is repressed by small molecules or mutations that target spliceosome components (Kaida et al., 2007; Legrain and Rosbash, 1989; Lo et al., 2007; O'Brien et al., 2008). Thus, the presence of unspliced AFP transcripts in the cytoplasm is consistent with the idea that Zhx2 globally represses splicing of AFP pre-mRNA.

Zhx2 also decreased mRNA accumulation of other two target genes, H19 and Gpc3, and at least H19 was shown by nuclear run-on to be transcribed at the same rate

in the presence or absence of Zhx2. It may be expected that Zhx2 will affect expression of these genes by a similar mechanism. Therefore, we investigated whether splicing of these genes was repressed similarly to that observed for AFP gene. Radiolabeled RT-PCR reactions were performed to detect unspliced H19 RNA in the cytoplasmic RNA fraction to increase the sensitivity of the assay compared to standard RT-PCR. With these assays, a lower accumulation of PCR product corresponding to the mRNA was observed in the presence of Zhx2, but PCR products corresponding to unspliced RNA were not detected across the multiple exon-exon regions of the H19 gene analyzed (Figure 3-5). One possible explanation for this result is that Zhx2 inhibits H19 splicing but unprocessed H19 RNA is rapidly turned over in the nucleus, as normally occurs with aberrant spliced transcripts. Additional experiments should be done analyzing nuclear RNA and primers that target exon-intron or intron-intron regions as another way to determine whether there are differences in accumulation of unspliced H19 RNA in the presence of Zhx2. In the case of Gpc3, there was a lower accumulation of spliced RNA in presence of Zhx2 in both nuclear and cytoplasmic RNA fractions, but the difference was not as marked as that observed in the H19 and AFP genes. Unspliced Gpc3 RNA product was not detected in the cytoplasmic fraction and no remarkable difference was observed in the nuclear soluble fraction between the presence or absence of Zhx2. The fact that Gpc3 pre-mRNA is not detected in the cytoplasm suggested that any unspliced Gpc3 transcripts must be rapidly processed or degraded. Together, these splicing results raise the question of what accounts for the unexpected stability of unprocessed AFP RNA compared to the transcripts of other Zhx2 target genes? In fact, it is perhaps fortunate that unprocessed AFP RNA is stabilized so that it can be detected as it provides important clues to the Zhx2 regulatory mechanism. It is not clear at this point whether Zhx2 affects splicing of target genes in the same way as the AFP gene or whether other post-transcriptional mechanism may be involved. Further analysis of the nascent RNA splicing, in chromatin-associated RNA fractions, may provide information about Zhx2 regulation of the H19 and Gpc3 genes.

Analysis of AFP intron requirement for Zhx2 regulation in transgenic mice

Because Zhx2 affected the splicing of AFP pre-mRNA, whether AFP introns were required for Zhx2 mediated regulation also was investigated. For this purpose transgenic mice that expressed AFP minigenes in its cDNA or genomic (intron containing) forms were generated. These minigenes were based on other constructs

that were reported to be repressed after birth and responsive to Zhx2. The cDNA AFP minigene driven by 250 bp AFP promoter was regulated in response to Zhx2 but the fold-change between mice that did or did not express Zhx2 was smaller than the difference in the endogenous AFP and H19 genes. However, the genomic AFP minigene driven by the 250 bp AFP promoter was not expressed well in the transgenic mice and therefore, its regulation by Zhx2 could not be assessed. Transgenic mice that expressed the cDNA or the genomic AFP minigenes driven by the 1 kb AFP repressor/promoter region were also generated. Surprisingly, these AFP minigenes were not strongly regulated in response to Zhx2 in the F2 offspring from E (cDNA) and G (genomic) founders, while the cDNA AFP minigene was not expressed well in the F2 offspring from the A founder. In contrast, endogenous AFP and H19 genes were regulated by Zhx2 as expected in all the cases. No obvious Zhx2 regulation was observed either by S1 analysis or by semi-quantitative RT-PCR analysis. F2 littermate liver RNA from three other cDNA transgene founders and two other genomic AFP transgene founders remain to be analyzed. At this point, why these minigenes are not regulated in response to Zhx2 is not known.

Mechanism of AFP splicing repression by Zhx2

Transport of pre-mRNA to the cytoplasm has been observed by interfering with early steps in the spliceosome assembly. This includes mutations in U1 snRNA or mutations in either the 5' splice site or the branchpoint sequence (Legrain and Rosbash, 1989). Accumulation of pre-mRNAs in the cytoplasm also was induced with spliceostatin A, an inhibitor of splicing, which target the SF3b subcomplex of the U2 snRNP (Kaida et al., 2007). Because SF3b interacts with components of the retention and splicing (RES) complex in yeast, it was suggested that SF3b is involved in recognizing intron-containing transcripts for nuclear retention. Based on these studies it can be proposed that Zhx2 might interfere with recruitment of one or more of the trans-acting factors involved in early spliceosome assembly and subsequently, avoid the nuclear retention and degradation. If this is the mechanism by which Zhx2 regulates its target genes, it would require introns for regulation. Because preliminary results suggest that a cDNA transgene is at least partially regulated, it is possible that introns contribute to the Zhx2 regulation but they can not be required.

Only a few other cases of gene regulation by intron retention have been observed, one being in the posttranscriptional regulation of the TNF gene expression

(Yang et al., 1998). TNF pre-mRNA was found to accumulate in the nucleus in naïve T cells. Upon T cell activation, TNF pre-mRNA splicing was quickly induced, resulting in high levels of TNF mRNA and protein. The mechanism by which TNF pre-mRNA splicing is repressed in naïve T cells and then induced soon after T cell activation, remains unknown. However, it was suggested that a negative regulator may block TNF pre-mRNA processing and be released upon T cell activation to allow proper TNF pre-mRNA processing. Another example of regulation through intron retention is observed in the primary response genes (PRGs), which produce full-length transcripts, but no mRNA, in unstimulated macrophage (Hargreaves, 2009). Promoters of PRGs are pre-associated with serine 5-phosphorylated RNAPII prior LPS stimulation, allowing transcription of full-length pre-mRNA that are quickly degraded rather than spliced. Uninduced promoters do not recruit pTEF-b to phosphorylate RNAPII on serine 2, which is required for co-transcriptional RNA processing, because the promoters are bound by co-repressor molecules.   Upon LPS stimulation, pTEF-b is recruited through the adaptor protein Brd4, inducing RNAPII phosphorylation at serine 2 and generation of mature RNA.

By analogy to these situations, it is possible that Zhx2 interferes with recruitment of factors involved in coupling transcription with RNA processing factors (Figure 3-15), for example by blocking RNAPII phosphorylation. This would result in AFP pre-mRNA splicing inhibition and concomitant reduction of AFP mRNA. These factors may also impact stability of the AFP pre-mRNA and avoid nuclear retention. Therefore, AFP pre-mRNA could escape from the nuclear surveillance mechanisms and be transported to the cytoplasm. Insight into the mechanism of Zhx2 regulation could be obtained by proteomic analysis of Zhx2-interacting proteins.
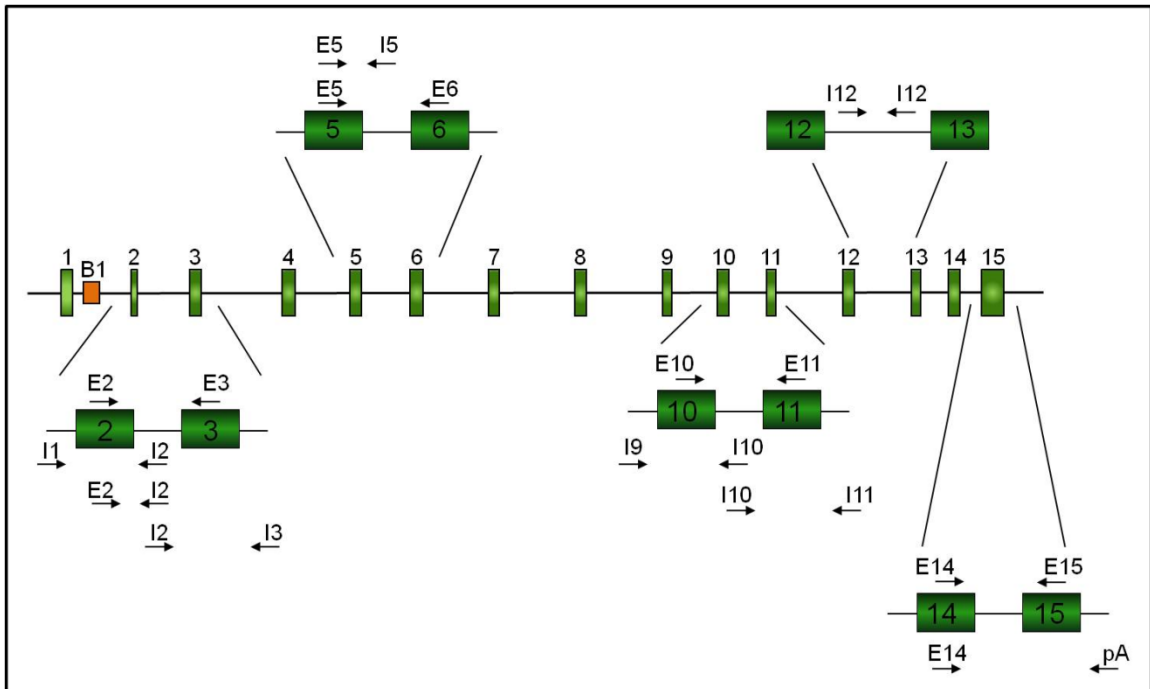
Figure 3-1. Scheme of PCR reactions performed for analysis of AFP splicing.
Diagram of the 15-exon AFP structural gene and location of the PCR primer pairs
utilized for AFP splicing analysis. B1 is a repetitive element within intron 1 known to be
transcribed by RNA polymerase III from the opposite strand to that encoding AFP RNA
(Godbout et al., 1986). AFP E2-E3, primers E2E3-F/E2E3-R; AFP I1-I2, primers i1-F/i2-
R; AFP E2-I2, primers E2E3-F /i2-R; AFP I2-I3, primers i2-F/i3-R; AFP E5-E6, primers
E5-F/E5E6-R; AFP E5-I5, primers E5-F/i5-R; AFP E14-E15, primers Afp3'-F/Afp3'-R1;
AFP E14-pA, primers Afp3'-F/Afp3'-R2; AFP I9-I10, primers i9-F/i10-R; AFP I10-I11,
primers i10-F/ i11-R; AFP E10-E11, primers E10E11-F/ E10E11-R, AFP E12-E13,
primers E12E13-FB/E12E13-RB, AFP I12-I12, primers nest12-13BF/nest12-13BR. All
primer sequences are listed in Table 2-3.
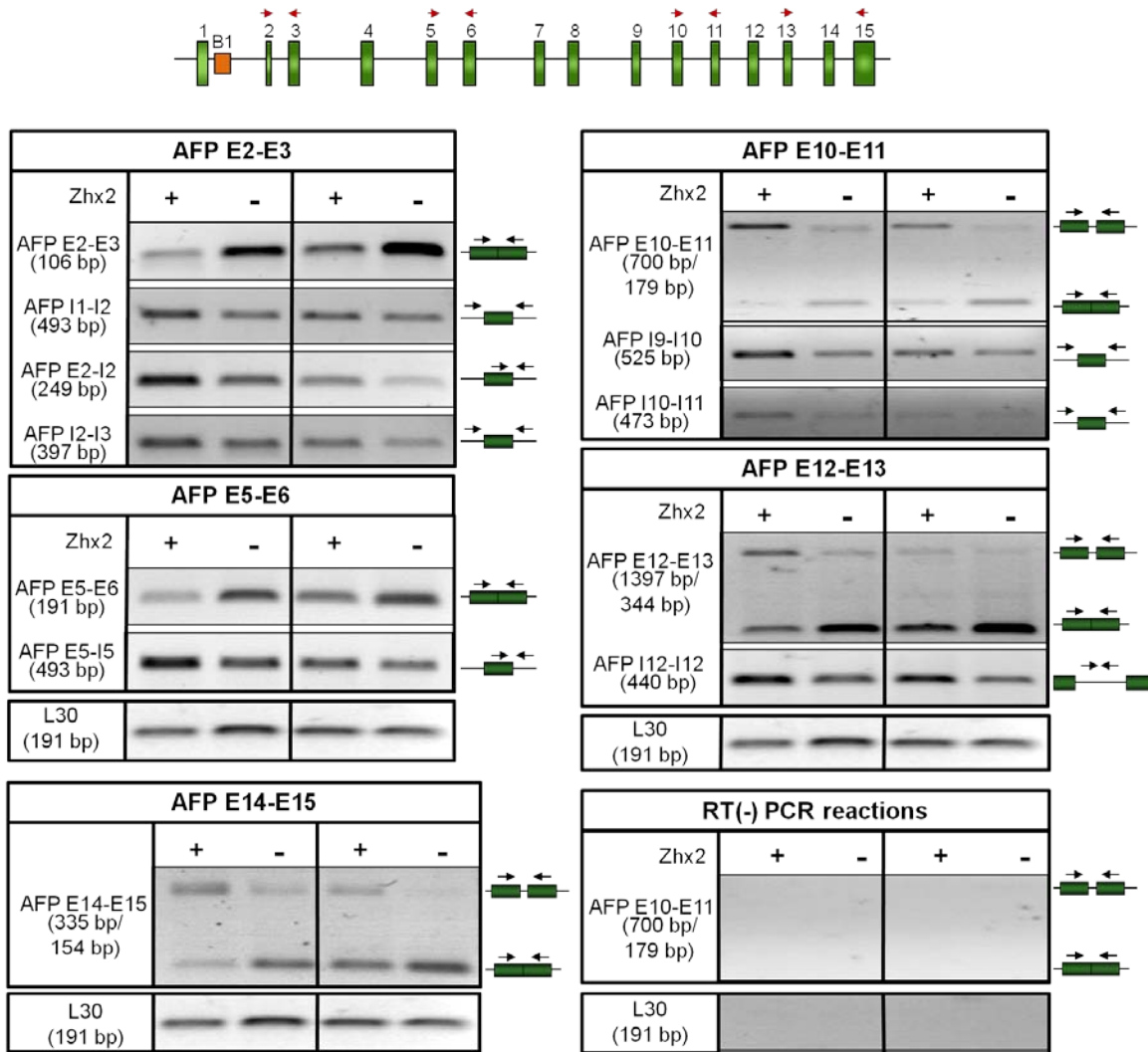
Figure 3-2. AFP pre-mRNA splicing is repressed in the presence of Zhx2. RT-PCR reactions using total RNA extracted from adult liver of two independent sets of littermate mice with (Zhx2+) or without (Zhx2-) the Flag-Zhx2 transgene. cDNA was made from 1 μg of total RNA and PCR reactions were performed with 1 to 5 μl of cDNA diluted 1:5. The expected sizes of the PCR products are shown on the left; a diagram of the location of each primer pair is shown on the right. PCR products were collected at different cycles to ensure a linear PCR amplification range (only one set of reactions is shown). rpL30 is used as loading control (23 cycles); AFP E2-E3, AFP I1-I2, AFP E2-I2, AFP I2-I3, AFP E5-E6, AFP E5-I5, AFP E14-E15, AFP E14-pA (33 cycles); AFP I9-I10, AFP I10-I11 (30 cycles); AFP E10-E11, AFP E12-E13, AFP I12-I12 (31 cycles). To check for DNA contamination, negative controls without RT enzyme were performed followed by PCR amplification with primers for L30 and AFP E10-E11 as described above.
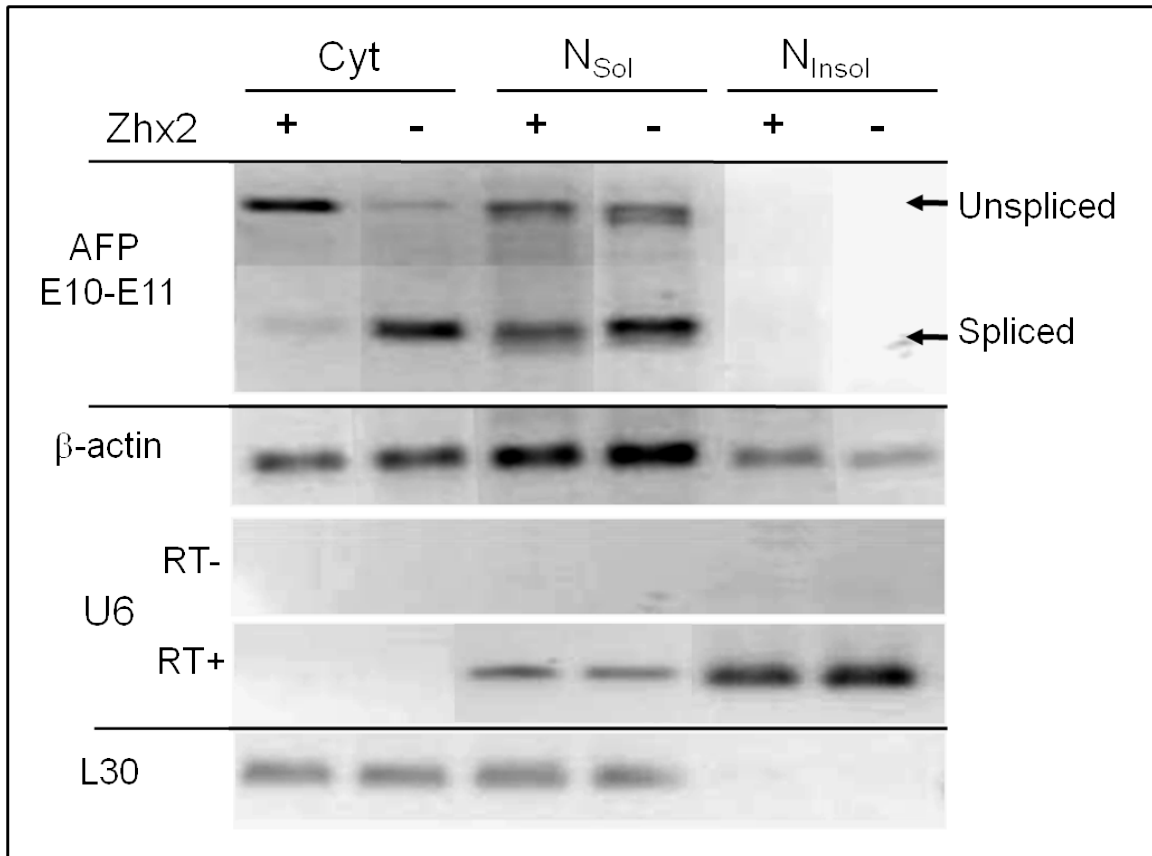
Figure 3-3. Splicing analysis of AFP transcripts in cellular RNA fractions from liver of littermate mice with or without the Flag-Zhx2 transgene. cDNA was made from 1 μg of the indicated RNA fraction. PCR reactions were performed with 1 to 5 μl of cDNA diluted 1:5 from cytoplasmic RNA fraction and undiluted cDNA from nuclear soluble or insoluble RNA fractions. PCR products were collected at different cycles to ensure a linear PCR amplification range (only one set of reactions is shown). rpL30 (L30R/L30F2) was used as loading control (27 cycles); AFP E10/E11 PCR (E10E11-F/ E10E11-R; 31 cycles), β actin PCR (B-ActF/ B-ActR; 26 cycles cytoplasmic fraction and 29 cycles nuclear soluble and insoluble fractions), U6 PCR (U6F/U6R; 26 cycles). U6 PCRs were performed from cDNAs prepared with 1 μg of RNA fractions without dilution, so that undetectable levels of U6 in the cytoplasmic fractions was not due to a dilution factor. U6 PCR for nuclear soluble fractions were performed under same conditions but at 27 cycles. The AFP and β actin lanes shown were assembled from non-adjacent lanes of the same gel.
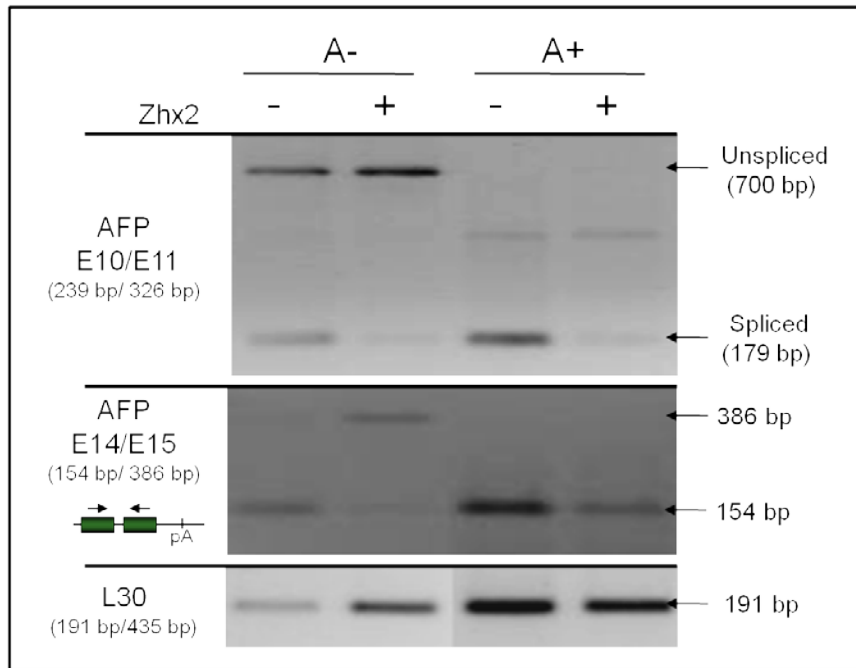
Figure 3-4. Splicing analysis of cytoplasmic A⁺ and A⁻ RNA fractions. Liver cytoplasmic RNA from littermate mice with or without the Flag-Zhx2 transgene were fractionated into A+ and A- RNA fractions using oligo-dT columns. cDNA was made using 1 µg of A⁻ RNA fraction or 12 ng of A⁺ RNA fraction with random hexamers. PCR reactions were performed with 0.5 µl of cDNA for L30 PCR or 2 µl of cDNA for AFP E10/E11 PCR. PCR products were collected at different cycles to ensure a linear PCR amplification range (only one set of reactions is shown). rpL30 (primers L30R/L30F2) were used as loading control (23 cycles); AFP E10/E11 PCR, primers E10E11-F/ E10E11-R (31 cycles); AFP E14/E15 PCR, primers E14E15-F/Afp3'-R1 (33 cycles). All primer sequences are listed in Table 2-3.
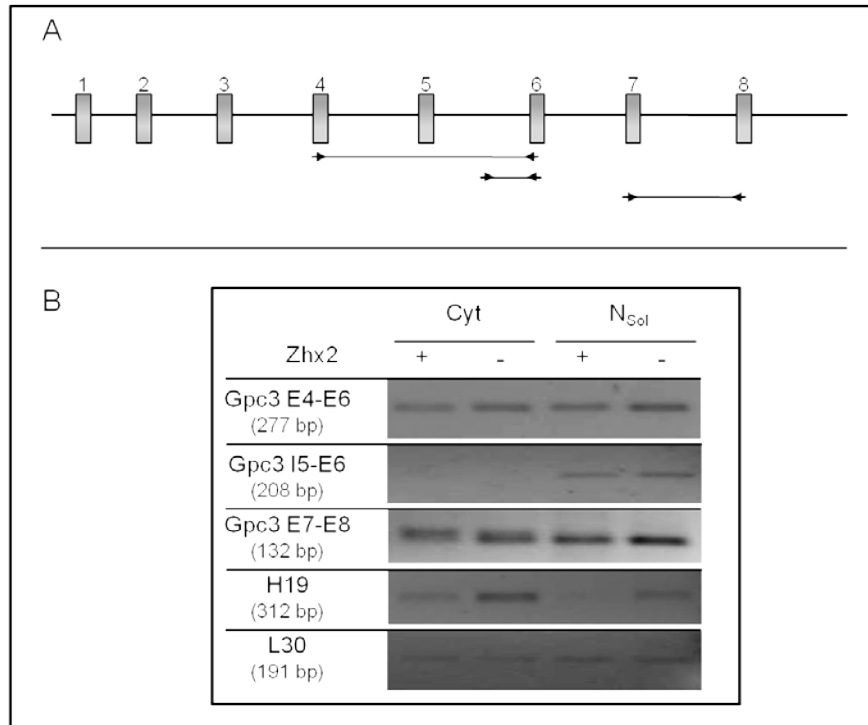
Figure 3-5. Splicing analysis of Gpc3 RNA in cytoplasmic and nuclear soluble RNA liver fractions. RNA was extracted from cytoplasmic and nuclear soluble fractions of liver tissue obtained from littermate mice with or without the Flag-Zhx2 transgene. cDNA was made with 1 μg of the indicated RNA fractions. PCR reactions performed with 1-2 μl of cDNA diluted 1:5 from the cytoplasmic RNA fraction and cDNA undiluted from the nuclear soluble fractions. PCR products were collected at different cycles to ensure a linear PCR amplification range. rpL30 (L30R/L30F2) was used as loading control (22 cycles); Gpc3 E4/E6 PCR (Gpc3URT/Gpc3LRT, 33 cycles); Gpc3 I5/E6 (Gpc3 U-2/Gpc3LRT, 32 cycles); Gpc3 E7/E8 PCR (Gpc3E7-U/Gpc3E8-L, 33 cycles); H19 PCR Exons 4 and 5 (S-H19/AS-H19, 30 cycles).
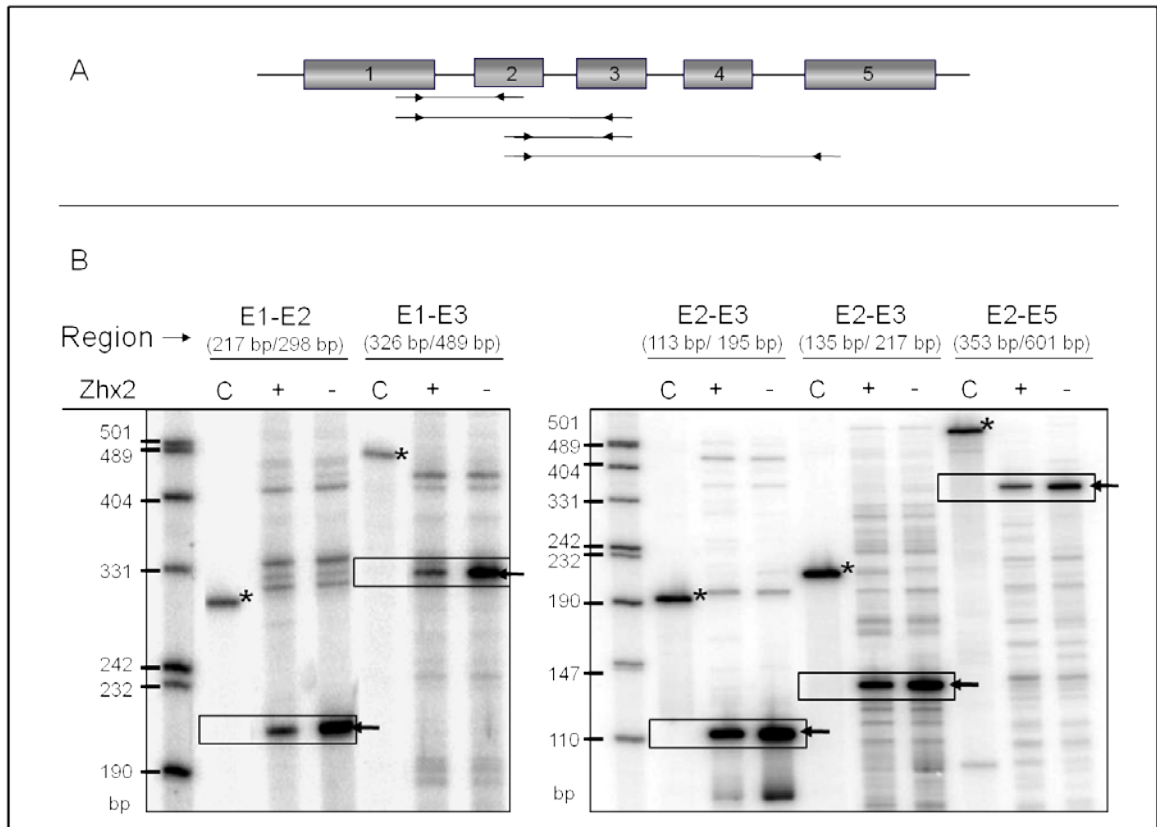
Figure 3-6. Splicing analysis of H19 RNA in cytoplasmic RNA liver fractions. 1 µg of cytoplasmic RNA fractions from littermate mice with or without the Flag-Zhx2 transgene were used for cDNA preparation with random hexamers. PCR amplification was performed with 5 µl of cDNA diluted 1:5 and primers targeting different exon-exon regions of H19 gene represented in (A). All PCR reactions were carried out for 29 cycles at 55 °C with the reverse primer radiolabeled with polynucleotide kinase (PNK) and $[\gamma$-^{32}P]-ATP. PCR products were analyzed on 6% or 8% denaturing polyacrylamide gels and detected by phosphorimaging (B). Boxes indicated PCR products corresponding to spliced mRNA. "*" indicates the PCR product from genomic DNA used as positive control (C) for expected size of the unspliced RNA. From left to right: PCR E1-E2, primers E1E5H19-F/ H19Ex2-L; PCR E1-E3, primers E1E5H19-F/H19Ex3-L3; PCR E2-E3, primers H19Ex2-U/H19Ex3-L2; PCR E2-E3, primers H19Ex2-U/H19Ex3-L3; PCR E2-E5, primers H19Ex2-U/E1E5H19-R.
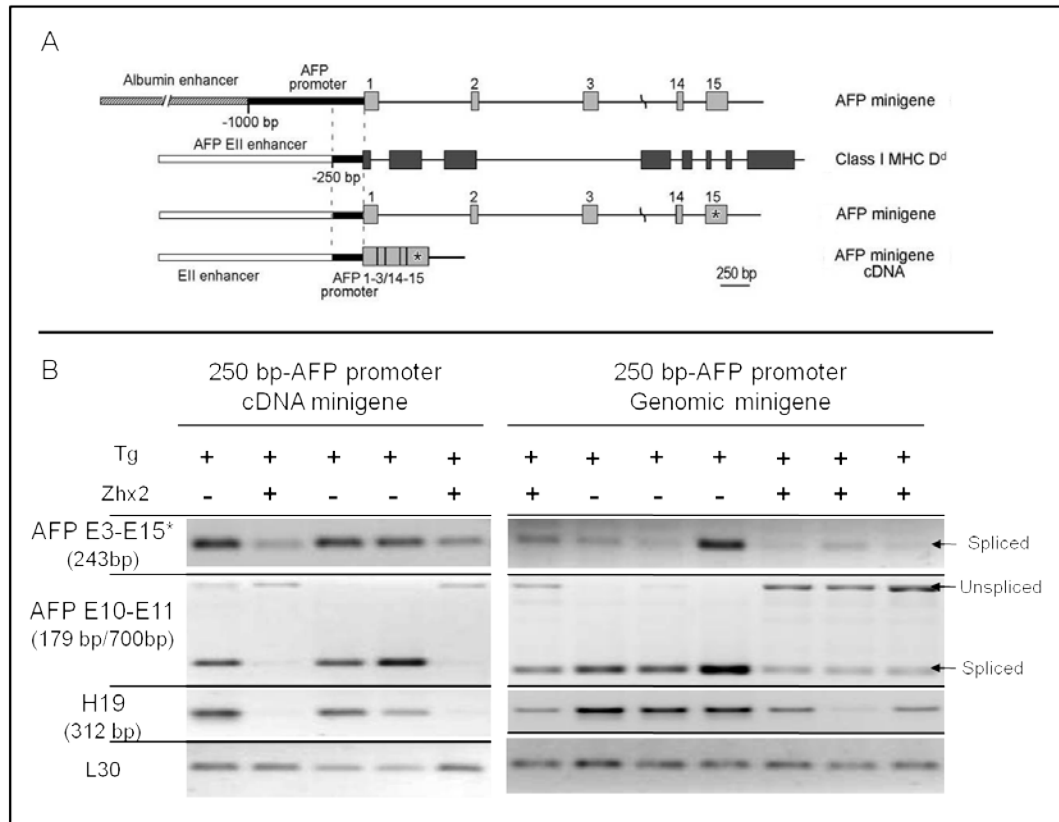
Figure 3-7. Splicing regulation by Zhx2 of AFP minigenes driven by the 250 bp AFP promoter region. (A) AFP minigenes used in previous studies contained the albumin enhancer and 1kb promoter to drive expression of a short version of the AFP gene where the internal sequence from exon 4 through exon 13 was deleted (Vacher et al., 1992). The second construct contains the AFP enhancer II and AFP 250 bp promoter to drive expression of the class I MHC D^d gene (Peyton et al., 2000a). Both construct were repressed after birth and responsive to Zhx2 regulation. Based on these transgenes, the AFP minigenes in the genomic or cDNA forms were designed. These AFP genomic and cDNA minigenes were driven by the AFP enhancer II and AFP 250 bp promoter. Additionally, an 18 bp sequence was inserted within exon 15 (*) to detect the transgene by RT-PCR. To study regulation of transgene by Zhx2, at least two offspring from each founder were backcrossed to BALB/cJ. RNA was extracted from livers of 3-4 week old F2 littermate mice and analyzed by RT-PCR (B). cDNA was prepared from 1 µg of RNA and 1-2 µl of 1:5 diluted cDNA was used for PCR amplification. AFP transgene, AFP E3-E15 PCR, primers E3F/E15R (2 µL RT 1:5 diluted, 33 cycles); endogenous AFP and H19 expression was analyzed as control for Zhx2 regulation: AFP E10-E11 PCR, primers E10E11-F/ E10E11-R (2 µL RT 1:5 diluted, 31 cycles). H19 PCR, primers S-H19/AS-H19 (1 µL RT 1:5 diluted, 28 cycles cDNA transgene and 29 cycles genomic transgene samples). rpL30 was used as loading control, primers L30R/L30F2 (1 µL RT 1:5 diluted, 23 cycles cDNA transgene and 22 cycles genomic transgene samples).
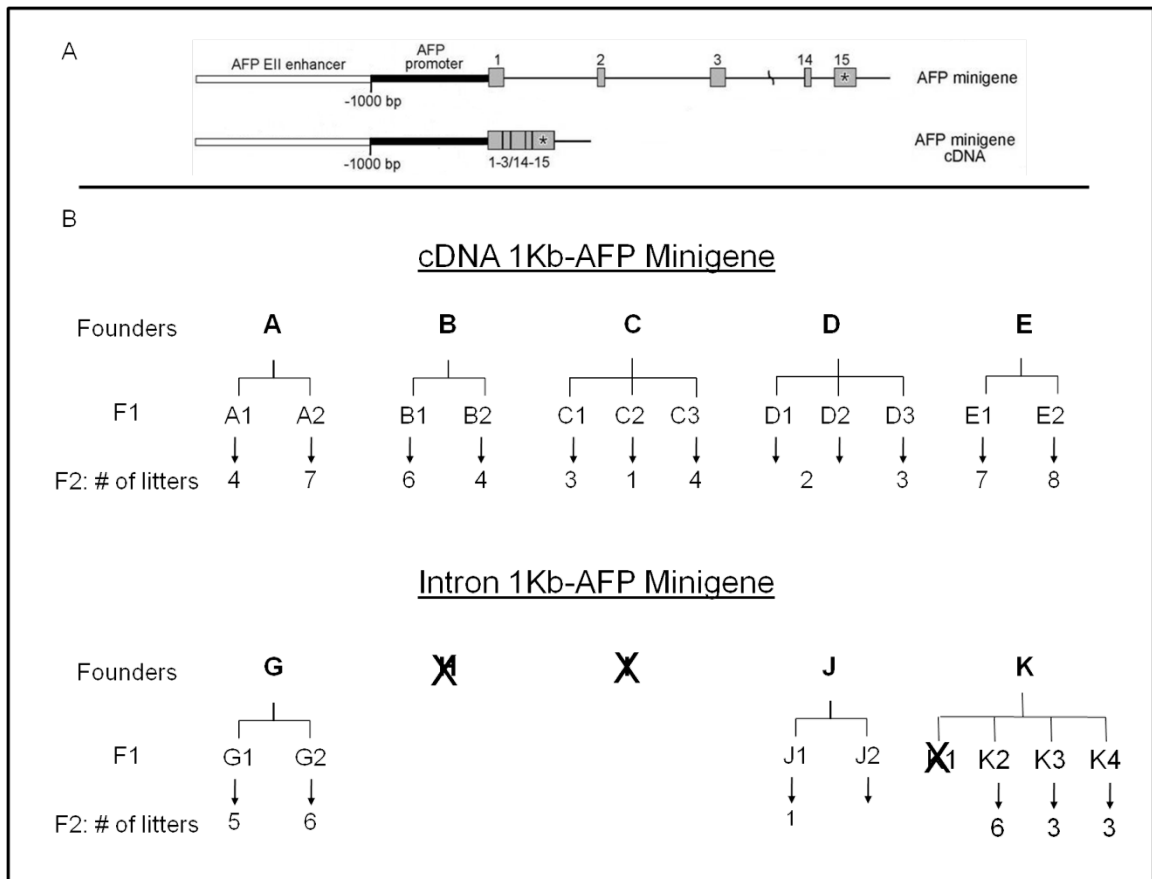
Figure 3-8. Scheme of transgenic founder mice expressing the cDNA and genomic AFP minigenes driven by the 1kb AFP promoter region. (A) AFP cDNA and genomic minigene as described in figure 4-7 (A) but driven by the AFP enhancer II and AFP 1kb promoter. (B) Founders expressing the AFP cDNA (A-E) and genomic (G-K) minigenes. Founder animals were crossed to BALB/cJ mice. To study regulation of the transgene by Zhx2, at least two offspring from each founder were backcrossed to BALB/cJ. The numbers indicate the number of litters harvested for RNA analysis. The H and I founders were sterile and K1 did not segregated the transgene; these are indicated by an "X".   Not all of the litters have been analyzed yet.

Figure 3-9. Analysis of splicing regulation by Zhx2 of the AFP cDNA minigene driven by the 1kb AFP promoter region, line A2. 1 µg of liver RNA from 3-5 week old littermate mice was used for cDNA preparation and 1-5 µl of 1:5 diluted cDNA was used for PCR amplification. AFP transgene expression was determined by AFP E3-E15 PCR (primers E3F/E15R, 5 µL RT 1:5 dilution and 33 cycles) or AFP E14-E15 (primers E14E15-F/E15R, 5 µL RT 1:5 dilution and 31 cycles); endogenous AFP and H19 expression was analyzed as controls for Zhx2 regulation by primer targeting the AFP E10-E11 (E10E11-F/E10E11-R, 2 µL RT 1:5 dilution and 31 cycles) and the H19 gene (S-H19/AS-H19, 2 µL RT 1:5 dilution and 30 cycles). Zhx2 expression levels were determined with primers Z2HD1-F/ Z2HD1-R (3 µL RT 1:5 dilution and 30 cycles). rpL30 (L30R/L30F2, 1 µL RT 1:5 dilution and 25 cycles) was used as loading control. The numbers below the figure are the sample tracking numbers of the animals analyzed.
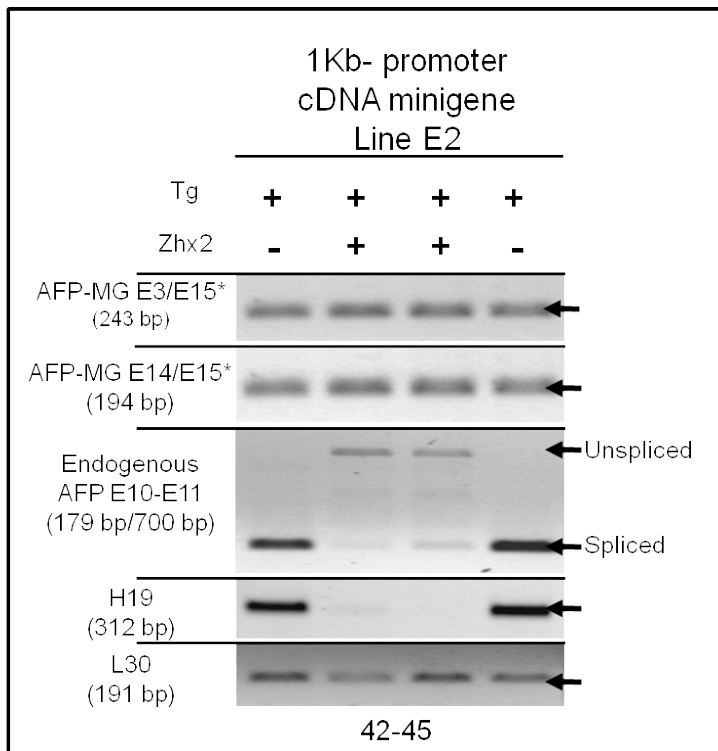
Figure 3-10. Analysis of splicing regulation by Zhx2 of AFP cDNA minigene driven by the 1kb AFP promoter region, line E2. 1 µg of liver RNA from 5 weeks littermate mice was used for cDNA preparation and 1-5 µl of 1:5 diluted cDNA was used for PCR amplification. AFP transgene expression was determined by AFP E3-E15 PCR (primers E3F/E15R, 5 µL RT 1:5 dilution and 31 cycles) or AFP E14-E15 (primers E14E15-F/E15R, 5 µL RT 1:5 dilution and 31 cycles); endogenous AFP and H19 expression was analyzed as controls for Zhx2 regulation by primers targeting AFP E10-E11 (E10E11-F/E10E11-R, 2 µL RT 1:5 dilution and 31 cycles) and the H19 gene (S-H19/AS-H19, 2 µL RT 1:5 dilution and 30 cycles). rpL30 (L30R/L30F2, 1 µL RT 1:5 dilution and 25 cycles) was used as loading control. The numbers below the figure are the sample tracking numbers of the animals analyzed.
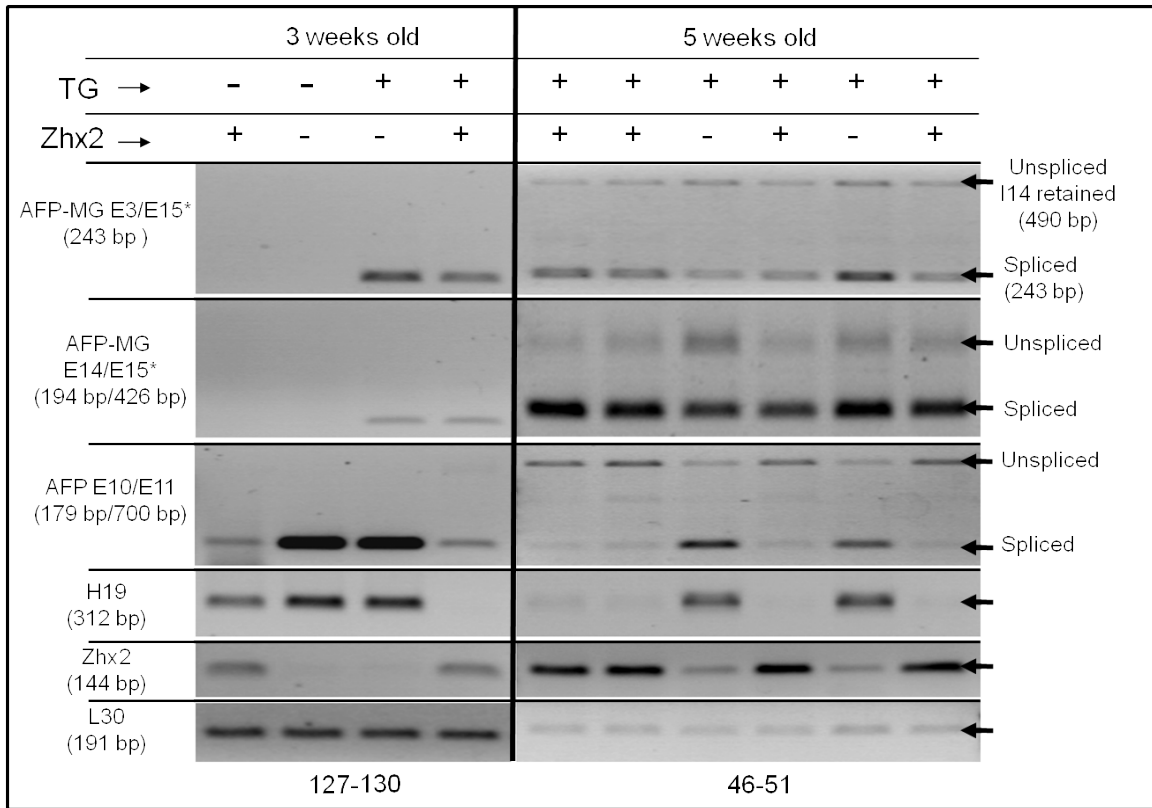
Figure 3-11. Analysis of splicing regulation by Zhx2 of AFP genomic minigene driven by the 1kb AFP promoter region, line G1. 1 µg of liver RNA from 3 or 5 week old littermate mice was used for cDNA preparation and 1-5 µl of 1:5 diluted cDNA was used for PCR amplification. AFP transgene expression was determined by AFP E3-E15 PCR (primers E3F/E15R, 5 µL RT 1:5 dilution, 30 cycles samples 127-130 and 31 cycles samples 46-51) or AFP E14-E15 (primers E14E15-F/E15R, 5 µL RT 1:5 dilution, 31 cycles); 490 bp band in AFP E3-E15 PCR correspond to the expected size of RNA transcripts with intron 14 retention. Endogenous AFP and H19 expression was analyzed as controls for Zhx2 regulation by primers targeting AFP E10-E11 (E10E11-F/E10E11-R, 2 µL RT 1:5 dilution, 31 cycles) and the H19 gene (S-H19/AS-H19, 2 µL RT 1:5 dilution, 30 cycles). Zhx2 expression levels were determined with primers Z2HD1-F/ Z2HD1-R (3 µL RT 1:5 dilution, 31 cycles samples 127-130 and 33 cycles samples 46-51). rpL30 (L30R/L30F2, 1 µL RT 1:5 dilution, 25 cycles) was used as loading control. The numbers below the figure are the sample tracking numbers of the animals analyzed.
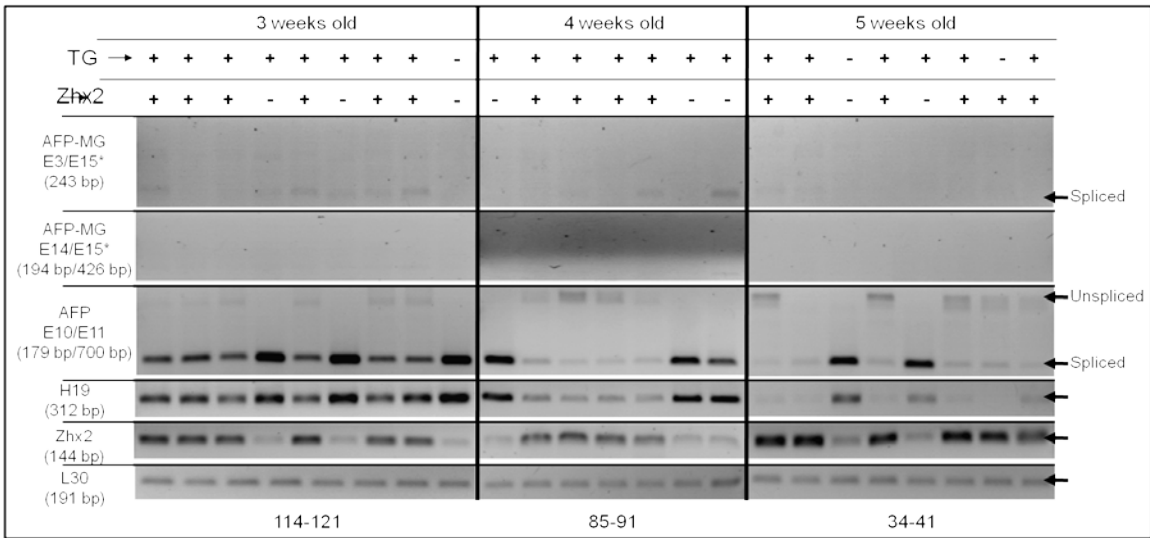
68
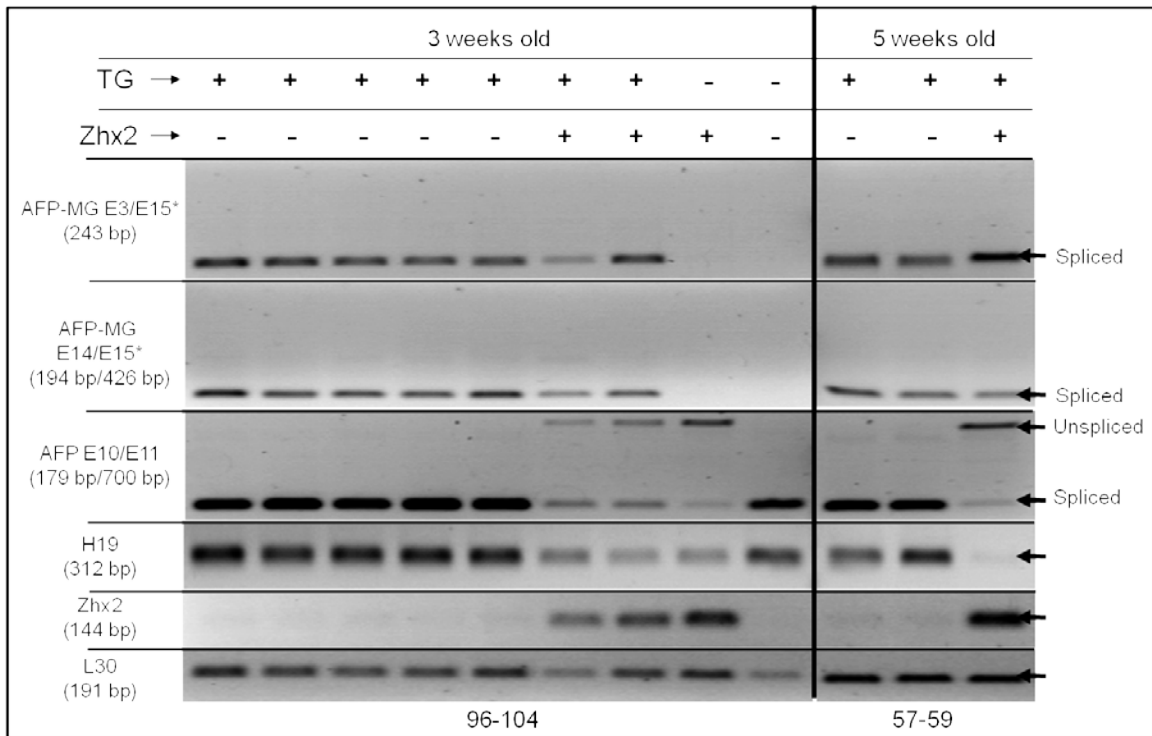
Figure 3-12. Analysis of splicing regulation by Zhx2 of AFP genomic minigene driven by the 1kb AFP promoter region, line G2. 1 µg of liver RNA from 3 or 5 week old littermate mice was used for cDNA preparation and 1-5 µl of 1:5 diluted cDNA was used for PCR amplification. AFP transgene expression was determined by AFP E3-E15 PCR (primers E3F/E15R, 5 µL RT 1:5 dilution, 30 cycles) or AFP E14-E15 (primers E14E15-F/E15R, 5 µL RT 1:5 dilution, 31 cycles); endogenous AFP and H19 expression was analyzed as controls for Zhx2 regulation by primers targeting AFP E10-E11 (E10E11-F/E10E11-R, 2 µL RT 1:5 dilution, 31 cycles) and the H19 gene (S-H19/AS-H19, 2 µL RT 1:5 dilution, 30 cycles). Zhx2 expression levels were determined with primers Z2HD1-F/ Z2HD1-R (3 µL RT 1:5 dilution, 31 cycles). rpL30 (L30R/L30F2, 1 µL RT 1:5 dilution, 25 cycles) was used as loading control. The numbers below the figure are the sample tracking numbers of the animals analyzed.
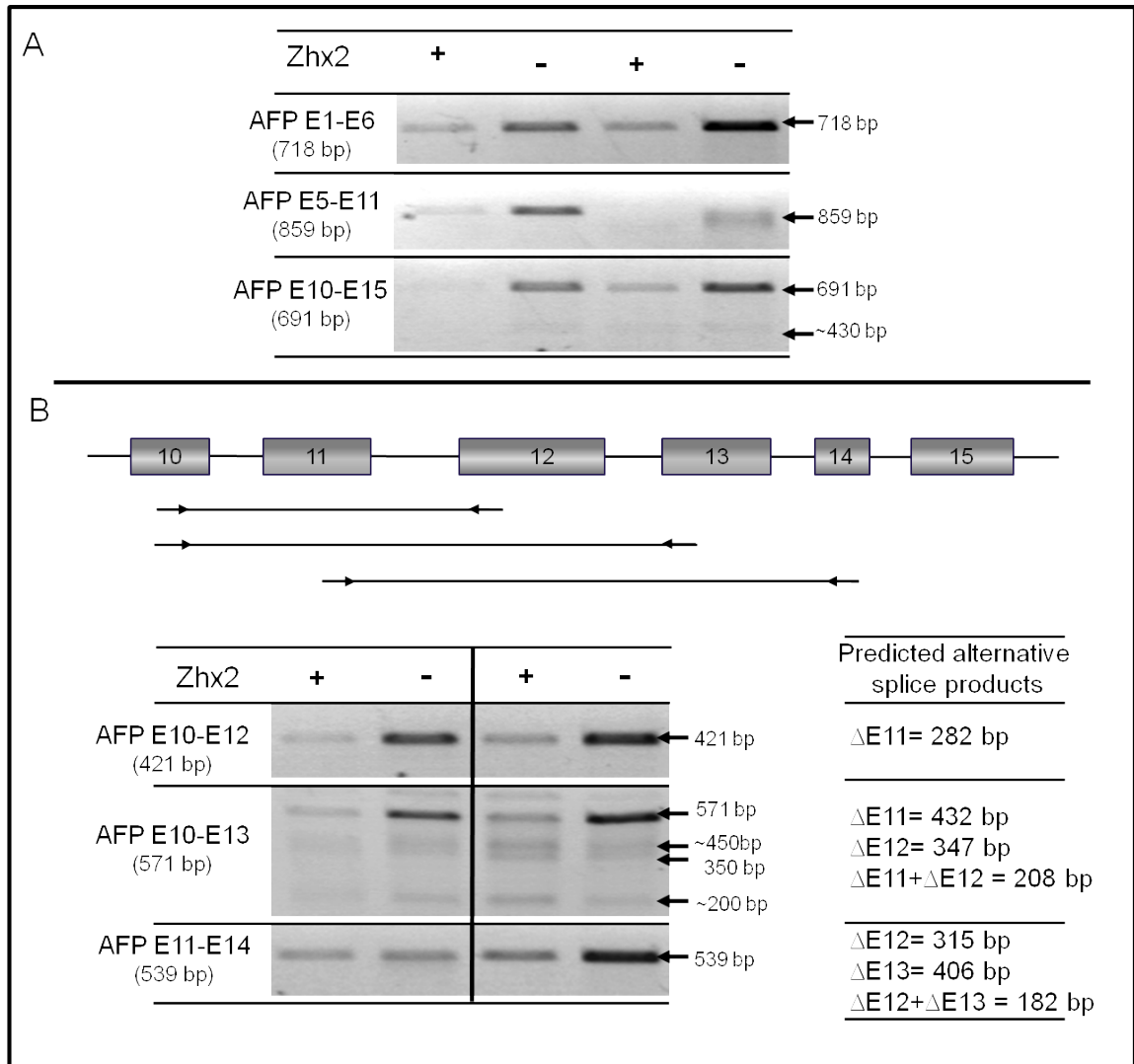
Figure 3-13. Analysis of alternative splicing across the AFP gene. cDNA was made from 1 µg of total RNA from liver of two sets of littermate mice with or without the Flag-Zhx2 transgene. (A) One to five µl of 1:5 diluted cDNA was used for PCR amplification of the AFP gene regions indicated on the left. The expected sizes of the PCR products corresponding to the fully spliced mRNAs are shown on the left. To further examine the ~430 bp product in the AFP E10-E15 reaction, the additional PCR reactions in (B) were performed. On the right are shown the predicted PCR product sizes for alternatively spliced RNAs if specific exons are skipped (identified by Δ). rpL30 (primers L30R/L30F2) was used as a loading control and is shown in Figure 4-2. PCR E1-E6, primers E1F and E5E6R; PCR E5-E11, primers E5-F and E10E11-R; PCR E10-E15, primers E10E11-F and E13E15-R; PCR E10-E12, primers E10E11-F and AFPE12A; PCR E10-E13, primers E10E11-F and E12E13-RB; PCR E11-E14, primers E14R and E11E12F.
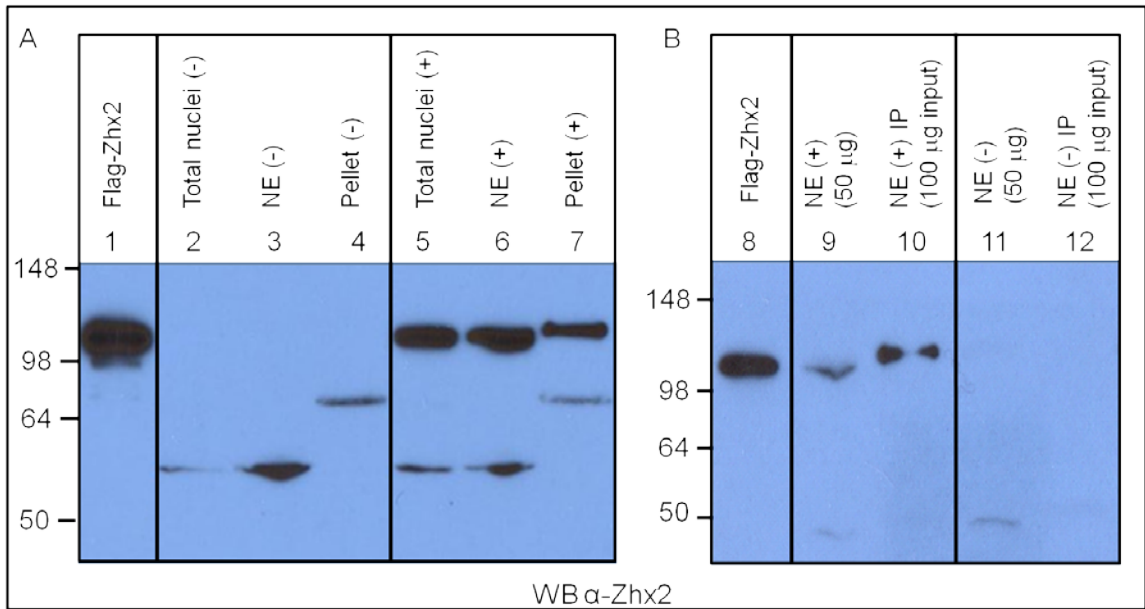
Figure 3-14. Nuclear extract preparation from liver of Flag-Zhx2 transgenic mice and anti-Flag-epitope immunoprecipitation. (A) Total nuclei (lanes 2 and 5, 10 µl), nuclear extract (lanes 3 and 6, 20 µl) and the insoluble pellet (lanes 4 and 7, 20 µl) fractions from liver of mice with (+) or without (-) the Flag-Zhx2 transgene were analyzed by Western blot with a rabbit polyclonal anti-Zhx2 antibody. (B) 100 µg of nuclear extracts from an independent preparation to that shown in (A) was used for immunoprecipitation using the M2 anti-flag agarose (Sigma). NE input (lanes 9 and 11, 50 µg) and respective eluates (lanes 10, and 12). NE from transiently transfected HEK293 cells were used as positive control and molecular size reference for Flag-Zhx2 (lanes 1 and 8, 2.5 µg total protein per line).
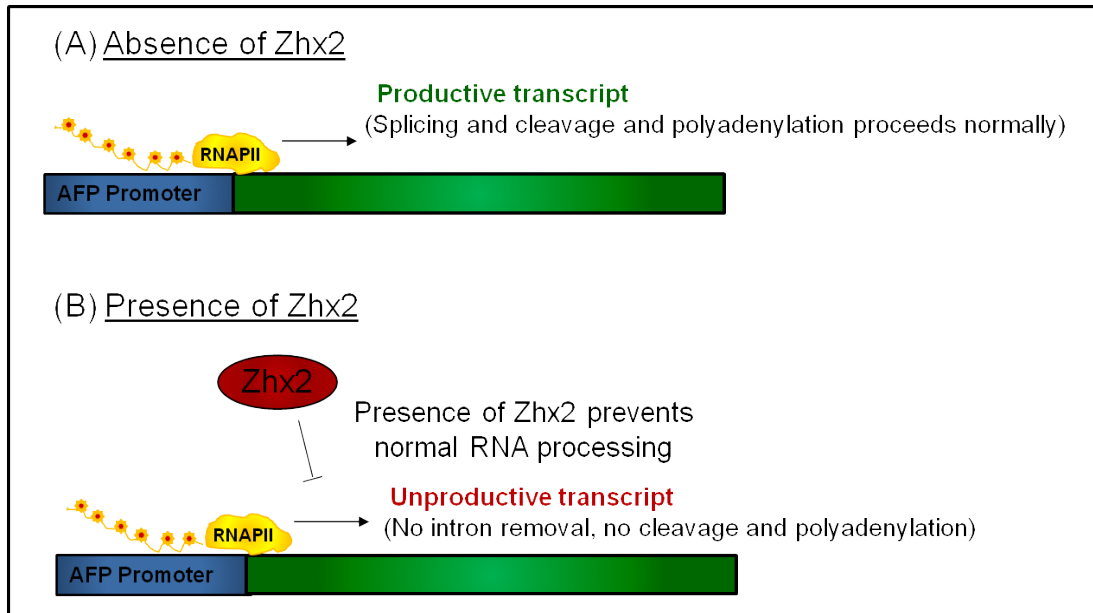
Figure 3-15. Proposed model for Zhx2-mediated splicing repression of AFP gene. (A) Before birth levels of Zhx2 are low consistent with high expression of AFP during fetal development. In the absence of Zhx2 expression, AFP RNA transcripts are fully spliced resulting in high levels of mRNA accumulation and AFP protein expression. (B) After birth levels of Zhx2 increase and AFP expression is repressed. Our data suggest that Zhx2 interferes with normal splicing so that AFP mRNA levels are low. Possible mechanisms include blocking recruitment of splicing factors or altering RNA polymerase II phosphorylation. Zhx2 may also be acting indirectly.

CHAPTER IV
TRANSCRIPTION TERMINATION IN THE Cμ GENE

INTRODUCTION

Termination of RNAPII transcription occurs at variable distances downstream from functional cleavage-polyadenylation [poly(A)] signals to recycle RNAPII complexes and prevent read-through transcription into adjoining genes. Termination has also been shown to enhance protein and mRNA expression from genes transiently transfected into cells (West and Proudfoot, 2009). While a functional poly(A) signal is essential for termination, additional sequences downstream from poly(A) signals are also required in some genes (Richard and Manley, 2009). For example, RNAPII pause sites have been identified downstream of multiple poly(A) signals (Ashfield et al., 1991; Enriquez-Harris et al., 1991; Gromak et al., 2006; Peterson et al., 2002). Also, an element positioned between 800 and 1600 nt downstream from the human β-globin poly(A) signal, which was shown to have co-transcriptional cleavage (CoTC) activity, is required for β-globin gene termination (Dye and Proudfoot, 2001). This CoTC cleavage element also functions to direct transcription termination when it is moved to a position about 200 nt from the poly(A) signal (Dye and Proudfoot, 2001). The β-globin CoTC element can also be replaced by the well-characterized MAZ4 RNAPII pause element or a pause element found downstream from the human β-actin gene (Gromak et al., 2006). The human ε-globin gene and the mouse serum albumin gene were also found to have a termination region that contained CoTC activity (Dye and Proudfoot, 2001; West et al., 2006), but CoTC activity is not found in all terminator regions examined (Gromak et al., 2006; Plant et al., 2005).

The current model for how RNAPII terminates downstream from protein-coding genes is a hybrid version of two original models (Bentley, 2002). Transcription through a poly(A) signal seems to make the RNAPII complex "termination competent"; a functional poly(A) signal is always required. Once the nascent transcript is cleaved, either by the cleavage-polyadenylation machinery acting on the poly(A) signal or by a CoTC element, the exposed 5' end of the 3' product becomes an entry point for an exonuclease to degrade the downstream RNA. This is required, but not sufficient, for RNAPII to release the template DNA (West et al., 2008). RNAPII is often seen to pause downstream from poly(A) signals, in some cases, it is due to the presence of specific sequences (Dye and Proudfoot, 2001; Fong and Bentley, 2001; Peterson et al., 2002) and in others, the

pausing is directed by the poly(A) signal itself (Orozco et al., 2002). While this model for the events that occur after RNAPII transcribes a poly(A) signal provides a framework for understanding the termination process, not all the details of this process have been established. In addition, it is not yet clear how termination downstream of more complex transcription units, such as genes containing multiple poly(A) signals, is regulated or whether termination is a part of the regulatory choice among multiple poly(A) signals.

The immunoglobulin M (μ) gene is a complex transcription unit that contains two poly(A) signals, μs and μm. The upstream μs poly(A) signal is within the Cμ4-M1 intron and the relative strengths of the competing cleavage-polyadenylation and splicing reactions is a major determinant of which mRNAs are produced. In B cells, the μs poly(A) signal is recognized less efficiently so more spliced μm mRNA is made whereas in plasma cells, use of the μs poly(A) signal predominates over Cμ4-M1 splicing, resulting in mostly μs mRNA being produced. Transcription termination in the μ gene has been well-studied and was shown to change during B lymphocyte development; the transcription unit is gradually foreshortened as B cells mature to plasma cells (Kelley and Perry, 1986; Peterson, 2007; Weiss et al., 1989; Yuan and Tucker, 1984). In B cells, transcription terminates far downstream of the distal μm poly(A) signal while in plasma cells, transcription terminates in the vicinity of this μm poly(A) signal, about 2 kb downstream from the upstream μs poly(A) signal. It is possible that the developmental differences in the extent of transcription may be due to changes in the use of the poly(A) signals in this gene. A RNAPII pause element was identified in the region 50 – 200 nt downstream of the μs poly(A) signal (Peterson et al., 2002). When this element was deleted, the amount of RNA cleaved and polyadenylated at the μs poly(A) site decreased while the RNA spliced between the Cμ4 and M1 exons was increased, in both B cells and plasma cells. Thus, this element contributes to the balance in efficiencies between the competing splice and cleavage-polyadenylation reactions, but does not seem to be directly involved in the RNA processing regulation. This pause element can be replaced by pause elements found downstream of other poly(A) signals (Peterson et al., 2002) and can cause RNAPII pausing in heterologous contexts (Peterson et al., 2002; Robson-Dixon and Garcia-Blanco, 2004). Whether any co-transcriptional cleavage activity exists within the μ gene locus that may contribute to the developmental termination has not been investigated.

Much of our knowledge of transcription termination and how CoTC and/or pause elements are involved in this process have been established using a model substrate,

where the human β-globin gene driven by the highly active HIV promoter was transiently transfected into HeLa cells. While this model system has discovered mechanistic details that would have been difficult to find in less robustly expressed genes, it is an artificial system, which raises the question of how broadly the details apply to natural transcription units. Thus, to examine the effects of the CoTC element in another context and determine how it affects transcription termination and alternative RNA processing choices of a more complex transcription unit, we have introduced the CoTC element into the well-studied mouse μ gene and explored the effects in stably transfected lymphocyte cell lines. While the β-globin CoTC element has been shown to promote termination downstream from poly(A) signals, it did not disrupt splicing when placed within the 850 nt β globin second intron (Dye et al., 2006). However, placement of the β-globin CoTC element in the 1862 nt Cμ4-M1 intron may have several possible outcomes. It was possible that splicing of the longer Cμ4-M1 intron would be disrupted, depending on how fast cleavage occurs, relative to transcription of the complete intron. On the other hand, CoTC elements and pause sites were functionally interchangeable downstream from the β globin poly(A) signal (Gromak et al., 2006). Therefore, it was possible that the CoTC element would be able to replace the activity of the μ pause element, or since pause and CoTC elements act in different ways, it was also possible that these elements would not be interchangeable in the μ gene context. Thus, to examine how the CoTC element will affect the delicate balance between the competing poly(A) signal and splice reaction, we placed the CoTC element in several locations within the Cμ4-M1 intron, in the presence or absence of the μ pause element. It is known that changes in the length of the Cμ4-M1 intron affect the pA/splice mRNA ratio, so expression of the constructs containing the CoTC element were compared to constructs in which the Cμ4-M1 intron size was modified with fragments not known to contain pause or CoTC activity. We find that the CoTC located within the Cμ4-M1 intron near, but not far, from the μs poly(A) signal affected μs/μm mRNA expression. In addition, we find no evidence for natural CoTC activity in the μ gene.

RESULTS

The β globin CoTC at a distal position does not affect μ mRNA processing

To examine the how a CoTC element may effect μs/μm processing, we placed the 800-bp β globin CoTC element at the *Kpn*I site within the Cμ4-M1 intron (1200 bp from the μs poly(A) and 500 nt upstream from the M1 3' splice site), in the context of an

intact μ gene, containing the Ig promoter and enhancer, a rearranged VDJ region, the complete IgM constant region and about 2 kb of sequence downstream of the μm poly(A) signal ((Peterson and Perry, 1989); Cμ/CoTC (+), Figure 4-1). Because the CoTC and pause elements had similar effects on transcription termination in other contexts (Gromak et al., 2006), we also inserted the CoTC into the μ gene from which the pause had been deletedΔNH/CoTC(+), Fig ure 4-1).   We have shown previously that the size of the Cμ4-M1 intron affects the μs/μm mRNA expression ratio; a shorter intron decreases the pA/splice ratio while a longer intron increases it (Peterson, 1992; Peterson, 1994b; Peterson et al., 1994; Peterson and Perry, 1986, 1989; Seipelt and Peterson, 1995). Thus, to detect any effects specific to the CoTC rather than just changes to the intron size, we made multiple intron size control constructs. In one, we inserted the 800 bp CoTC into the Cμ4-M1 intron from which an 800 bp fragment had been deleted (ΔAR/CoTC(+), Fig ure 4-1), which places the CoTC element about 400 nt downstream of the μs poly(A) signal. We also isolated CoTC element insertions in the opposite, presumably inactive, orientation for these three constructs (CoTC(-) versions, Figure 4-1). We also used other constructs containing various intron sizes described in previous studies as controls (Table 2-1, Figure 4-1). These includeΔNR, containing a 351 bp deletion between the *Nco*I and *EcoR*V sites; ΔAN, containing a 446 bp deletion between the *Acc*I and *Nco*I sites; ΔAR, containing a 797 bp deletion between the *Acc*I and *EcoR*V sites; ΔAR/pBR403 and ΔAR/rpl900 containing a 403 bp or 900 bp fragment inserted at the *Kpn*I site of ΔAR; and Cμ/rpl900, containing the 900 bp fragment inserted at the *Kpn*I site of Cμ (Peterson and Perry, 1986). All of these genes were stably introduced into M12 B cells and S194 plasma cells, both of which do not express their endogenous μ genes, and total RNA was analyzed by S1 nuclease mapping to measure the relative expression of μs (pA) and μm (splice) mRNA (Figure 4-2). Since previous placement of the CoTC into the β-globin intron did not affect expression, it was possible that it would have no effect on μ gene expression. However, because this intron is involved in the cis-competition with the μs poly(A) signal, it was possible that the CoTC could affect the balance between these reactions; this cis-competition is sensitive so that minor changes can be detected. Also, since this intron contains a pause element, we could determine whether the CoTC element was able to replace this pause element as it could in other genes.

Predictably, because inserting the CoTC element lengthened the Cμ4-M1 intron, the pA/splice expression ratio increased for all CoTC(+) constructs relative to their

starting constructs and this was observed in both B cells and plasma cells [Cμ vs. Cμ/CoTC(+), ΔNH vs. ΔNH/CoTC(+) and ΔAR vs. ΔAR/CoTC(+), Figures 4-1 and 4-2]. The one exception was Cμ and Cμ/CoTC in plasma cells, which expressed similarly high pA/splice mRNA ratios. To determine whether the CoTC element could replace the pause element, we compared the Cμ and ΔNH constructs to these genes that contain the inserted CoTC element (Figures 4-1 and 4-2). In B cells, the pA/splice ratio of ΔNH is about 3-fold lower than Cμ. A similar difference in the pA/splice ratio is also seen between Cμ/CoTC(+) and ΔNH/CoTC(+). This suggests the CoTC is not able to replace the activity of the pause site when inserted at this location about 1 kb downstream from where the pause site is normally located. It is interesting to note that various pause sites inserted at this *Kpn*I site also did not replace the activity lost by deletion of the NH pause element (Peterson et al., 2002).

To better visualize the expression changes caused by the CoTC elements, relative to the intron size controls, we plotted intron size vs the pA/splice expression ratio (Figure 4-3). The ΔNH constructs, which lack the pause element have, by definition, an effect that is greater than the size of the element and thus have not been included in this plot. The trend of the greater intron size causing an increase in the pA/splice ratio can clearly be seen from the size control constructs. While there is some scatter from a straight line, which indicates that primary sequence or RNA structure also contributes to the pA/splice ratio, the strong trend indicates that intron size is a major determinant of the expression ratio. In general, the CoTC(+) constructs fall within the range expected based on intron size, rather than having an increased pA/splice ratio that would indicate a specific effect of the CoTC element. The constructs ΔAR/CoTC(+), ΔAR/rpI900 and Cμ, which all have Cμ4-M1 introns of a similar size, all have similar expression ratios in B cells (Figures 4-1 and 4-2), which is also consistent with the CoTC element not affecting pA/splice ratio from the *Kpn*I site location in the Cμ4-M1 intron.

Curiously, all the constructs with the CoTC element in the opposite orientation had much lower expression ratios than expected based on the intron size (Figure 4-1, Figure 4-2 and "x" in plots in Figure 4-3). It appears that the CoTC(-) fragment contains activity that greatly enhances Cμ4-M1 splicing relative to cleavage-polyadenylation at the μs poly(A) signal. Because expression is measured by S1 nuclease mapping, which would not detect any aberrant splicing caused by this fragment, we performed multiple RT-PCR reactions, as well as 3' RACE, to determine whether the Cμ4 5' splice site was being spliced to a new 3' splice junction. However, only the expected products were

detected (data not shown). We propose that the enhanced splicing activity may be due to a fortuitous splice enhancer sequence introduced with the anti-sense β-globin CoTC, but we did not test this hypothesis further.

Because the expression data indicated that the CoTC element did not affect Ig expression, we needed to determine whether this was because the CoTC element was not being cleaved in the μ gene context. To do this, we performed semi-quantitative RT-PCR using primers that flank the CoTC element using RNA extracted from M12 cells expressing the constructs with the CoTC in both orientations (Figure 4-4, PCR 2). Because the CoTC in the antisense orientation should not be cleaved, the RT-PCR signal obtained for the CoTC(-) constructs will measure the amount of uncleaved intronic RNA. If the CoTC element in the CoTC(+) constructs has been cleaved, an RT-PCR product will not be amplified. Primers that amplify the region between the Cμ3 and Cμ4 exons were used to ensure similar expression levels among the constructs (Figure 4-4, PCR 1). Because the signal across the CoTC insertion site was consistently higher in the CoTC(-) constructs than the CoTC(+) constructs, we conclude that the CoTC element is being cleaved in a major portion of the transcripts. However, because this activity does not affect the pA/splice expression ratio, cleavage is likely occurring after a commitment to one of the two competing reactions has been made.

The β-globin CoTC at a proximal position does affect μ mRNA processing

Because the CoTC element placed at the *Kpn*I site in the Cμ4-M1 intron did not affect μ alternative RNA processing choices and did not replace the activity of the NH pause fragment, we determined whether this element could affect μ mRNA expression from a location closer to the μs poly(A) signal. We cloned the CoTC element into the *Hind*III site in the presence or absence of the NH pause element, so that the CoTC is either 200 bp or 50 bp, respectively, downstream of the μs poly(A) signal (Cμ/CoTC-H and ΔNH/CoTC-H, Figure 4-5A). This also places the CoTC element 1475 nt upstream from the M1 3' splice site. As controls for intron size, we also inserted the rpI900 fragment into the same locations (Cμ/rpI900-H and ΔNH/rpI900-H, Figure 4-5A). RNA from M12 cells stably expressing these constructs was analyzed by S1 nuclease analysis (Figure 4-5B). Compared to Cμ and ΔNH, the constructs containing the CoTC element had dramatically increased pA/splice mRNA ratios. These ratios were about 3-fold higher than the intron size controls, which indicates that the CoTC element is affecting μ mRNA expression at this location. The pA/splice ratios for the constructs

carrying the CoTC and rpI900 fragments in either the *Kpn*I or *Hind*III sites are compared in Figure 4-5D. It is interesting that the rpI900 fragment at either site leads to a similar pA/splice ratio, indicating this ratio is mainly due to the intron size. However, while the construct with the CoTC element at the *Kpn*I site expresses a pA/splice ratio similar to the constructs containing the rpI900 element, when the CoTC is placed in the *Hind*III site, cleavage-polyadenylation at the µs poly(A) signal is greatly enhanced. Cµ/CoTC-H and ΔNH/CoTC-H expressed similar pA/splice ratios, suggesting that the CoTC element activity either replaced or dominated over the lost pause site activity in the ΔNH construct. Thus, these experiments show that the CoTC, when placed close to the µs poly(A) signal does affect the pA/splice ratio.

The CoTC element in the β globin gene is required for transcription termination (Dye and Proudfoot, 2001), did not disrupt splicing of an 850 bp intron (Dye et al., 2006). Thus, the high pA/splice expression ratio observed when the CoTC was located near the µs poly(A) signal could be due to CoTC-induced early termination so that the complete Cµ4-M1 intron wasn't synthesized efficiently. It could do this either by enhancing use of the µs poly(A) signal or by creating cleaved ends that are entry sites for 5'-3' exonucleases. Alternatively, CoTC cleavage could interfere with Cµ4-M1 intron splicing; this intron is about 1 kb longer than the β globin intron that was not affected. To examine the fate of RNA downstream of the CoTC element, we performed semi-quantitative RT-PCR on total M12 RNA, using primers flanking the *Hind*III insertion site and primers across the M1 3' splice junction in both the CoTC-containing constructs and the rpI900 size controls (PCR 2 and PCR 3, Figure 4-5C). As an expression control, each sample was analyzed with primers in the Cµ3 and Cµ4 exons, which showed a similar level of expression in all four constructs (PCR 1, Figure 4-5C). The RT-PCR reaction across the CoTC or rpI900 fragment insertion site clearly showed a decrease in RNA from the CoTC-containing constructs compared to the rpI900-containing constructs; the product sizes vary due to the presence or absence of the NH pause fragment and the difference in size between the CoTC and rpI900 fragments. This difference in RNA abundance, seen with multiple dilutions of the RT reaction, suggests that the CoTC element is being cleaved in a substantial population of transcripts. The PCR reaction detecting RNA across the 3' M1 splice junction shows an even greater difference between the CoTC-containing constructs compared to the rpI900-containing constructs. A signal for the CoTC-containing constructs can be seen only with the undiluted RT reaction, when the signal for the rpI900-containing constructs is over-

amplified (left panel of Figure 4-5C). This lack of a downstream RNA signal suggests that either the CoTC element is causing transcription to terminate before the M1 exons or, if this RNA is transcribed, that it is rapidly degraded. Thus, while these results do not distinguish between the possible mechanisms of CoTC actions, they are fully consistent with previous results that demonstrated enhanced expression and termination due to the CoTC (West and Proudfoot, 2009).

CoTC like-elements are not found in the μ gene

Transcription termination in the μ gene has been measured in both primary B-lymphocytes and various lymphoid cell lines; in all cases, the transcription termination region was further upstream in plasma cells or stimulated B cells than in B cell lines or resting B cells. It is not clear whether this change in termination pattern is due to changes in the use of different poly(A) signals or whether termination sequences, such as a CoTC element, that are more active in the plasma cells may be involved. Thus, we examined the termination region of the μ gene for evidence of a CoTC element that may contribute to termination, possibly in a developmental stage-dependent manner. CoTC elements were identified previously using a nuclear run-on hybrid selection technique that requires a high level of gene expression. The cleavage activity was confirmed by RT-PCR reactions, using primers that spanned the predicted CoTC region (Dye and Proudfoot, 2001; Yuan and Tucker, 1984); PCR products were produced when the RNA was contiguous, but where the RNA had been cleaved, no PCR product was detected. We used an RT-PCR approach to screen the μ gene sequences for CoTC activity because expression levels are not sufficient for the hybrid selection technique. We isolated nuclear RNA, to enrich for nascent RNA, from two lymphoid cell lines expressing the endogenous μ gene, 38C-13, a B cell line and D2, a 38C-13-derived hybridoma cell line (Nelson et al., 1983). These two cell lines contain the same productively rearranged μ gene allele expressed in different developmental stages to produce different ratios of pA/splice μ mRNA. These cell lines also have different transcription termination profiles by nuclear run-on assays (Kelley and Perry, 1986). D2 terminates downstream of the μm poly(A) signal whereas 38C-13 transcription continues over this region (Kelley and Perry, 1986). We designed sets of primers to amplify overlapping regions of the μ gene spanning about 3.7 kb of sequence downstream of the μs poly(A) signal (Figure 4-6A). As a positive control for RT-PCR amplification over the sequences farther downstream from the μs poly(A) signal, we used an *in vitro*

transcribed RNA, containing about 600 nt upstream of the M1 exon and 1900-bp downstream of the µm poly(A) site (denoted as the RV RNA, Figure 4-6A). PCR reaction "a" was used as the quantitation control to compare the signals between the 38C-13 and D2 RNAs and the amount of RNA used in the RT reactions was adjusted so the signal intensity was similar. This adjusted amount of RNA (1 µg of 38C-13 and 2 µg D2 RNA) was used in all the RT reactions. The primers a – e were used to make cDNA from the two RNA samples. The cDNA was then amplified with the primers shown (Figures 4-6A and 4-6B). If a CoTC existed within the µ gene, we would expect to see an abrupt loss in PCR product from the RT primer that was downstream of the cleavage site. If there were a CoTC element that was active selectively in one cell type, we would only see this abrupt loss in one RNA sample. In fact, we observe a gradual decline in PCR signal as we use primers further downstream from the µs poly(A) signal; the D2 signal declines relative to 38C-13 mostly over primers c and d (Figure 4-6C). This is in good agreement with the nuclear run-on data that shows D2 termination downstream of the µm poly(A) signal (Kelley and Perry, 1986). Thus, we find no evidence for a CoTC element that contributes to transcription termination in the µ gene.

DISCUSSION

Transcription termination has been extensively studied in the IgM/IgD gene (Peterson, 1994a). While B cells express both IgM and IgD and transcribe the entire locus, transcription terminates between IgM and IgD in plasma cells so that IgD is no longer expressed. However, termination is not required for the RNA processing shift seen during B lymphocyte development because cells that express high pA/splice µ mRNA ratios often terminate transcription downstream of the µm poly(A) signal. While there are clear differences in the transcription termination regions in B cells compared to plasma cells, little is known about the mechanism driving this, although changes in the use of the µs poly(A) signal is a possible regulatory mechanism. Here, we explored whether additional sequences such as CoTC elements, located within the Cµ4-M1 intron or in sequences downstream of the µm poly(A) signal, are involved in transcription termination in B cell and plasma cells. Using semi-quantitative RT-PCR and primers spanning the region 3.7 kb downstream from the Cµ4 exon, we explored whether nascent µ transcripts are cleaved co-transcriptionally. This event would result in an abrupt loss of the PCR signal downstream of the cleavage site due to the loss of continuity within the nascent RNA. However, this is not what we observed. Instead, we

measured a gradual decrease in PCR signal across the region in both the B cell line (38C-13) and the 38C-13-derived hybridoma cell line (D2). Termination was observed to occur further upstream in the D2 cells, compared to 38C-13 cells, which is consistent with run-on data from these cell lines (Kelley and Perry, 1986). Therefore, we conclude that a CoTC element is not detectable in the μ gene. Thus, the most likely explanation for transcription termination profile differences between B cells and plasma cells is that changes in the μs poly(A) signal use drive differential termination (Peterson, 2007).

Because alternative RNA processing in the μ gene relies on a balanced competition between a splice and a cleavage-polyadenylation reaction, subtle changes that affect the efficiencies of either reaction can be sensitively detected in this gene. We have shown that the μ gene contains an RNA pol II pause element between 50 and 200 nt downstream of the μs poly(A) signal that contributes to the efficient use of this poly(A) signal in both B cells and plasma cells (Peterson et al., 2002). To gain a better understanding of the CoTC element from the β globin gene, and how it may function in place of the μ pause site, we placed the 800 bp CoTC element within the Cμ4-M1 intron at several different positions downstream of the μs poly(A) signal in the presence or absence of the μ pause element. Because we knew that changes in the Cμ4-M1 intron size would affect the pA/splice RNA ratio, we used a series of constructs containing Cμ4-M1 introns of different lengths as controls for intron size. When the CoTC element was inserted in the *Kpn*I site in the active orientation (+), there was no effect on the pA/splice mRNA ratio over and above that expected due to the intron length, whether the pause site was present or not. In this set of constructs, the distance between the CoTC element and the μs poly(A) signal varied from 400 nt ΔAR/CoTC(+) to 1200 nt in Cμ/CoTC(+), but the distance between the CoTC element and the M1 3' splice junction was a constant 500 nt. In contrast, when the CoTC element was placed in the *Hind*III site, use of the μs poly(A) signal was greatly enhanced. In these constructs, the distance between the CoTC element and the μs poly(A) signal was either 200 nt in Cμ/CoTC-H or 50 nt in ΔNH/CoTC -H whereas the distance between the CoTC and the M1 3' splice junction is 1475 nt. We showed that, in both locations, the CoTC element was causing RNA cleavage, as seen by a decrease in RT-PCR signal across the cleavage element compared to a similarly placed fragment without cleavage activity. Thus, the difference between the CoTC affecting μ processing from one intronic location but not another must be due to its position within the intron rather than differences in the fragment activity. There are several possible explanations for this position effect, based

on whether the spacing between the µs poly(A) signal and CoTC or the spacing between the CoTC and the M1 3' splice junction is more critical.

Previous studies in our lab showed that the µ pause region increased usage of the µs poly(A) site when it was positioned up to 100 bp downstream of its normal location. However, if the pause element was moved further away, effects on the pA/splice RNA expression were lost (Burnside et al., 2011). It is possible that we are detecting a similar strict distance effect with the CoTC at 200 nt from the µs poly(A) signal being able to affect the pA/splice RNA expression whereas all effects are lost at a position 400 nt from the µs poly(A) signal. The basis for the limited spacing between the µs poly(A) signal and pause elements is not yet clear, but it suggests that a poly(A) signal is able to respond to pausing over a limited distance or a limited time after being synthesized. Although it hasn't been systematically examined, other studies that have measured the effects of pause sites and CoTC elements on transcription termination have not described a distance effect. Indeed, the β globin CoTC is naturally located 800 nt from the β globin poly(A) signal although it continued to direct transcription termination when it was relocated to 200 nt from the poly(A) signal (Dye and Proudfoot, 2001). A major difference between these assays is that in the µ gene we measure affects on alternative RNA processing choices of the nascent RNA, not just transcription termination.

While it is possible that the poly(A) signal – CoTC spacing affects whether the CoTC impacts µ RNA processing, perhaps a more likely interpretation is that the distance between the CoTC and the M1 3' splice site is the more important variable that determines whether the CoTC affects µ mRNA processing or not. We have previously explained the effect of intron size on the pA/splice RNA ratio in terms of the relative timing between when the µs poly(A) signal and the complete competing intron are synthesized; a smaller intron would allow less time for the µs poly(A) signal to be recognized before the splice reaction could begin to compete. This same relative timing explanation may be used to interpret the position effect of the CoTC element. In the constructs in which the CoTC does affect the pA/splice RNA ratio, this distance is 1475 nt, whereas it is 500 nt in the constructs in which the CoTC has no effect. Thus, with only 500 nt to be synthesized, the complete intron will be present in the nascent transcript sooner, so the CoTC cleavage reaction likely occurs after the nascent RNA is committed to one of the two competing reactions. However, significant cleavage of the CoTC may occur before the M1 3' splice junction is synthesized when this distance is

1475 nt and this cleavage may drive transcription termination. Indeed, we have shown that there was a dramatic decrease in nascent RNA over the M1 exon in constructs containing the CoTC element in this location, consistent with the idea that CoTC cleavage is causing transcription termination so that the complete Cμ4-M1 intron was not synthesized efficiently to compete with the μs poly(A) signal. This interpretation, based on timing, also is consistent with previous studies that introduced the β globin CoTC, a hammerhead ribozyme or the hepatitis δ ribozyme within the introns of the β globin gene. The fast-cleaving hepatitis δ ribozyme, but not the CoTC or hammerhead ribozyme affected expression of β globin mRNA (Dye et al., 2006; Fong et al., 2009). This suggested that RNA processing was disrupted by the faster cutting hepatitis δ ribozyme, which did not allow time for co-transcriptional spliceosome assembly before the nascent RNA was cleaved. However, the slower cleaving CoTC element and hammerhead ribozyme did allow the spliceosome to commit the pre-mRNA to be spliced before the intron was cleaved and thus did not affect overall mRNA expression. The lengths of the β globin introns are 130 and 850 nt, so the distances between the CoTC and the splice junctions are, in general smaller than in the μ gene. It is possible that a systematic study of the distances over which effects can be detected could be used to estimate how rapidly these various RNA cleavage activities occur in vivo, relative to transcription elongation which has been measured to be about 3.8 kb min^{-1} (Singh and Padgett, 2009).

Previous studies on transcription termination have shown that pol II pause sites and CoTC elements are interchangeable to facilitate termination downstream from functional poly(A) signals (Gromak et al., 2006; West and Proudfoot, 2009; West et al., 2008). When we placed the CoTC element in the HindIII site of the Cμ4-M1 intron, directly downstream of the μs poly(A) signal in either the presence or absence of the μ pause site region, we saw a dramatic increase in the pA/splice RNA ratio, from 2.4 in the wild-type gene to 23-24, but we no longer detected an expression difference due to the pause site. This suggests that the CoTC element activity dominates in this location. When other pause sites were substituted for the μ pause element, several were shown to be stronger because they caused an increase the pA/splice ratio compared to the wild-type gene, from 2.4 to 6-8 (Burnside et al., 2011). Interestingly, transcription terminated further upstream in the μ genes with the stronger pause sites. Thus, it seems that the pause and CoTC elements may be interchangeable for directing transcription termination in this μ gene context; cleavage at the CoTC element drove early termination

and thus a high pA/splice ratio whereas strong pause sites increased use of the μs poly(A) signal which then drove early termination. An increase in the pA/splice expression ratio was seen in each case. We did not determine whether the constructs containing the CoTC at the *Hind*III site were regulated between B cells and plasma cells. Since the pA/splice was already very extreme in B cells, we would have been able to measure only a 2-3-fold increase in plasma cells before we reached the limits of our assay. Thus, the regulatory mechanism is likely to have been blunted due to the strong effect of the CoTC element.

While it is important to study transcription termination in simple transcription units to establish the connections between poly(A) signals and downstream termination elements, it is also valuable to consider these events in a more complex transcription units such as the μ gene where poly(A) site use, pause sites and transcription termination are developmentally regulated to alter RNA processing reactions. A major conclusion from studying termination in β globin constructs was that poly(A) signal recognition, not cleavage, is sufficient for termination in the presence of a terminator such as the CoTC element (West and Proudfoot, 2008). However, simply having a poly(A) signal appear in the nascent transcript can not be sufficient for termination in the μ gene because the μs poly(A) signal, the first of two poly(A) signals, is not recognized efficiently in B cells and transcription does not terminate until after the downstream μm poly(A) signal is used. In plasma cells, use of the first μs poly(A) signal predominates and termination occurs sooner than in B cells. Thus, there must be a mechanism to regulate when a poly(A) signal is recognized to trigger termination or to make the pol II "termination prone" and when it is skipped so transcription continues. Because the μ transcription unit is both complex in structure and is regulated during B lymphocyte development, it is a unique model system to understand how the basic steps of transcription termination can be modulated to lead to different gene expression outcomes.
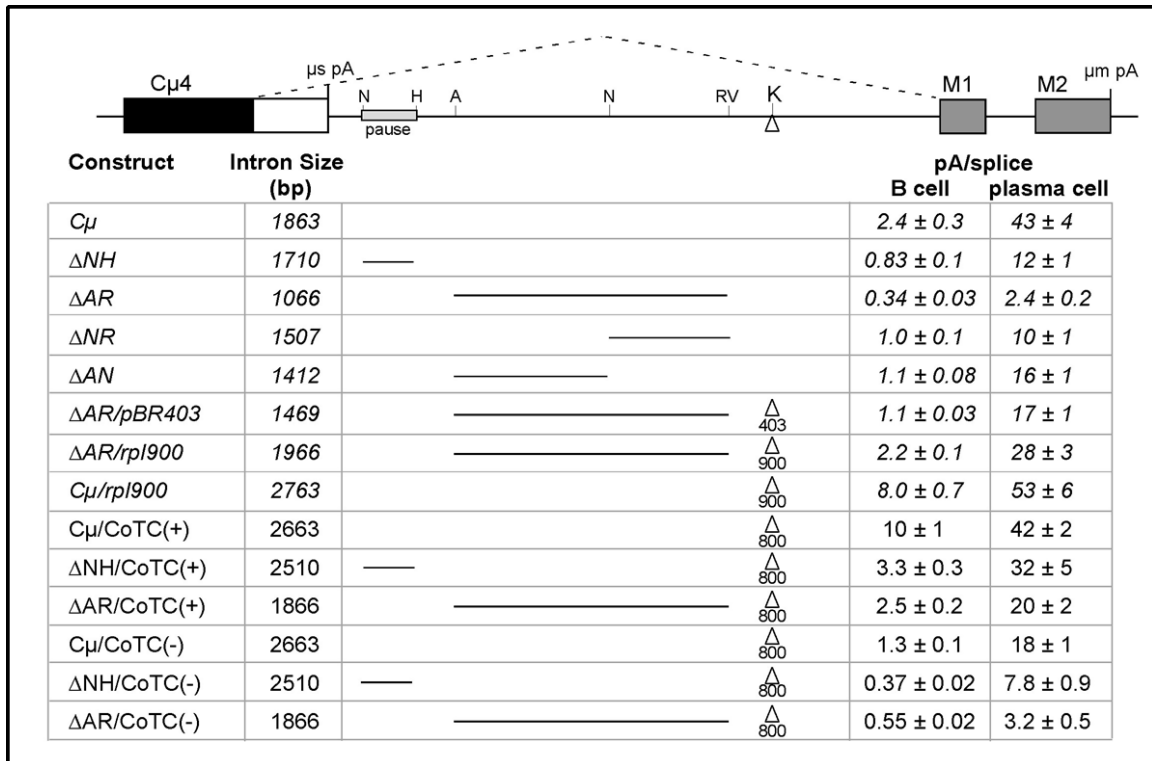
Construct	Intron Size (bp)		pA/splice B cell	plasma cell
Cµ	*1863*		*2.4 ± 0.3*	*43 ± 4*
ΔNH	*1710*		*0.83 ± 0.1*	*12 ± 1*
ΔAR	*1066*		*0.34 ± 0.03*	*2.4 ± 0.2*
ΔNR	*1507*		*1.0 ± 0.1*	*10 ± 1*
ΔAN	*1412*		*1.1 ± 0.08*	*16 ± 1*
ΔAR/pBR403	*1469*	△403	*1.1 ± 0.03*	*17 ± 1*
ΔAR/rpl900	*1966*	△900	*2.2 ± 0.1*	*28 ± 3*
Cµ/rpl900	*2763*	△900	*8.0 ± 0.7*	*53 ± 6*
Cµ/CoTC(+)	2663	△800	10 ± 1	42 ± 2
ΔNH/CoTC(+)	2510	△800	3.3 ± 0.3	32 ± 5
ΔAR/CoTC(+)	1866	△800	2.5 ± 0.2	20 ± 2
Cµ/CoTC(-)	2663	△800	1.3 ± 0.1	18 ± 1
ΔNH/CoTC(-)	2510	△800	0.37 ± 0.02	7.8 ± 0.9
ΔAR/CoTC(-)	1866	△800	0.55 ± 0.02	3.2 ± 0.5

Figure 4-1. The CoTC element inserted in the Cµ4-M1 intron at the KpnI site does not affect the pA/splice mRNA expression ratio besides that expected due to changes in intron size. The map of the µ gene 3' region containing the competing µs poly(A) signal (µs pA) and the Cµ4-M1 splice (dotted line) is shown at the top. The Cµ4 exon common to both µs and µm mRNA is shown as a black box, the µs-specific exon is shown as a white box and the µm-specific exons M1 and M2 are shown as grey boxes. The previously described pol II pause site region is denoted by the small light grey box. Restriction sites used in cloning are shown: N, *Not*I; H, *Hin*dIII;, A, *Acc*I; N, *Nco*I; RV, *EcoR*V; K, *Kpn*I. Listed below the map are the CoTC insertion constructs and the intron size controls (in italics), the size of each Cµ4-M1 intron, a diagram of the intron structure with a bar denoting deleted sequence and a triangle denoting an inserted fragment and the pA/splice expression ratio for each construct in M12 B cells and S194 plasma cells. Expression was measured by S1 nuclease analysis from cell lines stably transfected with each construct (see Figure 4-2); the average and standard deviation of at least three analyses are shown.
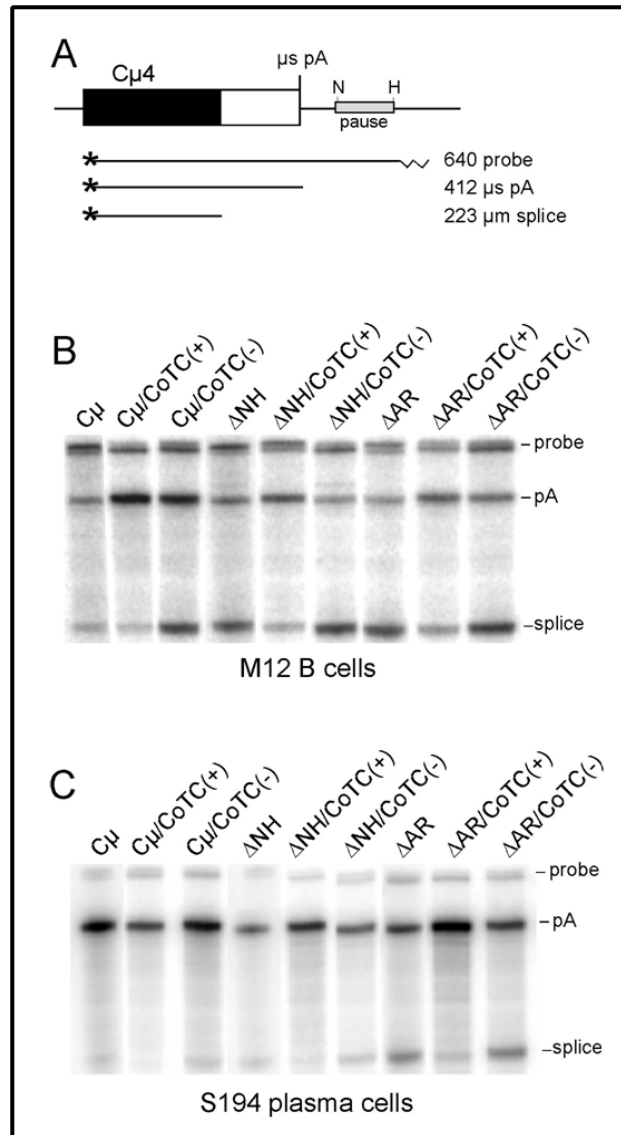
Figure 4-2. S1 nuclease analysis of the CoTC constructs and controls. **(A)** Diagram of the end-labelled probe used to distinguish mRNA spliced at the Cµ4 5' splice junction from mRNA that is cleaved and polyadenylated at the µs poly(A) signal. **(B)** Representative S1 analyses for the constructs shown above the lanes stably expressed in M12 B cells; the bands for probe, mRNA that is cleaved and polyadenylated at the µs poly(A) signal (pA) and mRNA spliced at the Cµ4 5' splice junction (splice) are denoted on the right. **(C)** Representative S1 analyses for the constructs shown above the lanes stably expressed in S194 plasma cells; the bands for probe, mRNA that is cleaved and polyadenylated at the µs poly(A) signal (pA) and mRNA spliced at the Cµ4 5' splice junction (splice) are denoted on the right.
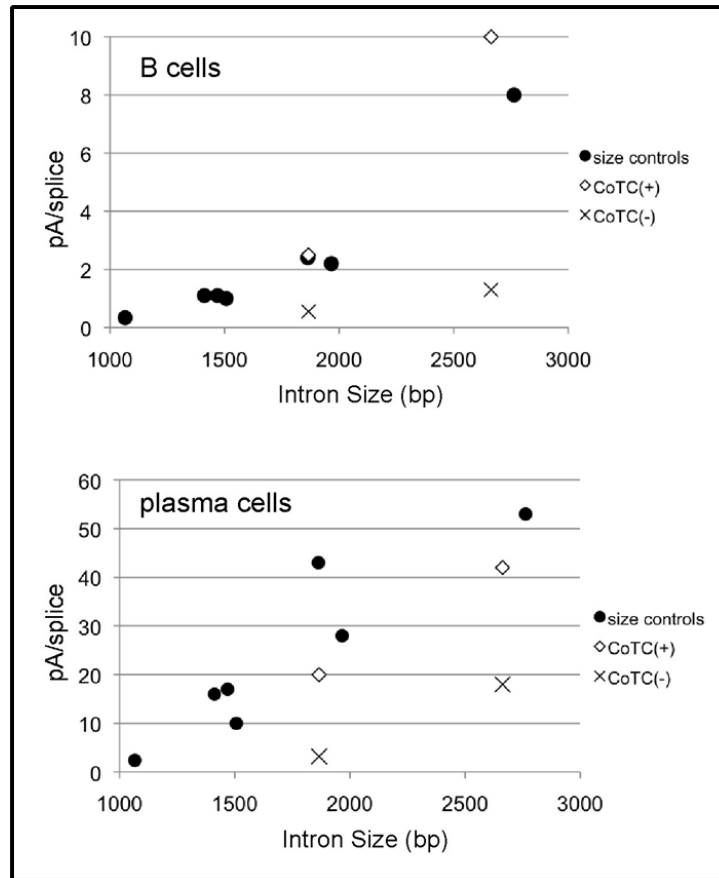
Figure 4-3. The pA/splice expression ratio increases with increasing intron size.
The pA/splice expression ratios, in both B cells (top panel) and plasma cells (lower panel), for the CoTC(+) and CoTC(-) constructs and intron size control constructs are plotted versus the intron size. The data are taken from Figure 4-1.
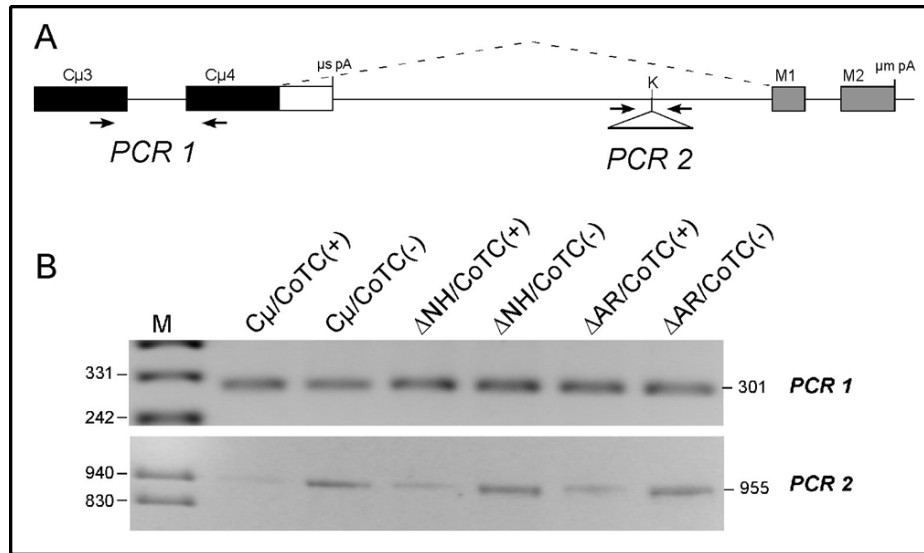
Figure 4-4. The CoTC element inserted at the *Kpn*I site is being co-transcriptionally cleaved. (A) Map of the μ gene 3' region, as in Figure 4-1, showing the locations of the two sets of PCR primers used to analyze RNA from cells stably expressing μ genes containing the CoTC element. PCR 1 (primers Cμ3T and C4B, Tabla 2-2) spans exons Cμ3 and Cμ4 and will detect all spliced μ mRNA. PCR 2 (Kpn-T and Kpn-B, Table 2-2) spans the KpnI site that contains the CoTC(+) and CoTC(-) elements. **(B)** RT-PCR reactions using RNA from M12 B cells stably expressing the constructs shown above each lane. The sizes of the marker lanes (M) and the size of the PCR products are shown. RT reactions in which the reverse transcriptase enzyme was omitted were used in the PCR reactions to ensure there was no DNA contamination in the RNA samples; all reactions were blank (data not shown).
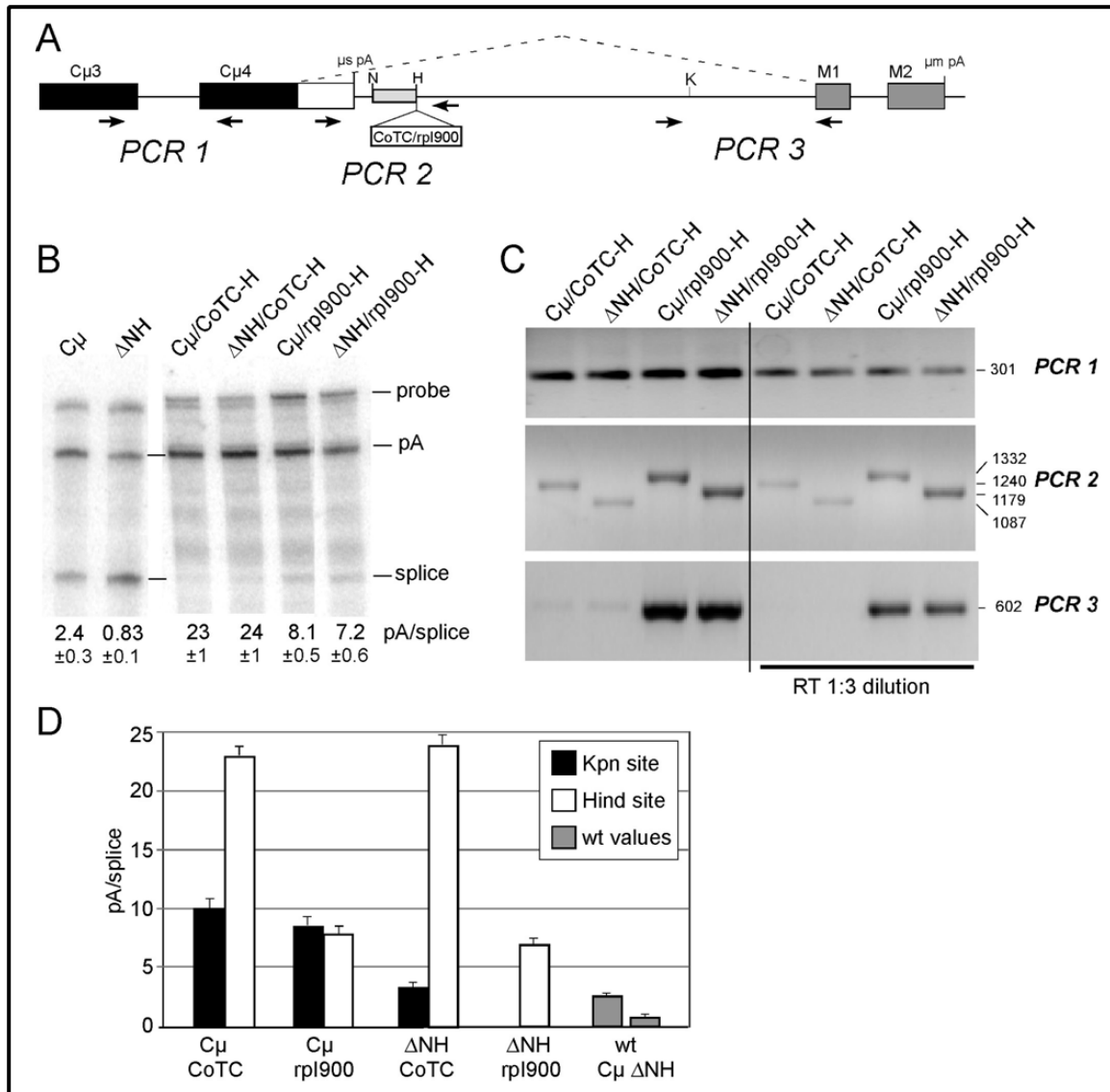
Figure 4-5. The CoTC element inserted at the *Hind*III site affects expression of the µ gene. **(A)** Map of the µ gene 3' region, as in Figure 4-1, showing the locations of the three sets of PCR primers used to analyze RNA from cells stably expressing µ genes containing the CoTC element or the rpl900 size control fragment. The site of fragment insertion is shown. PCR 1 (primers Cµ3T and C4B, Table 2-2) spans exons Cµ3 and Cµ4 and will detect all spliced µ mRNA. PCR 2 (Apa and Hind, Table 2-2) spans the region that contains the CoTC and rpl900 fragments. PCR 3 (Kpn-T and M1B, Table 2-2) spans the M1 3' splice junction. **(B)** Representative S1 analyses for the constructs shown above the lanes stably expressed in M12 B cells; the bands for probe, mRNA that is cleaved and polyadenylated at the µs poly(A) signal (pA) and mRNA spliced at the Cµ4 5' splice junction (splice) are denoted on the right. The average and standard deviation of at least three independent analyses are shown below each lane. This figure was assembled from multiple gels. **(C)** RT-PCR reactions using RNA from M12 B cells stably expressing the constructs shown above each lane. The sizes of the marker lanes

(M) and the size of the PCR products are shown. The panel on the left used the undiluted RT reaction in the PCR. The RT reactions were diluted 1:3 for the PCR reactions in the right panel. The PCR 2 products vary in size because of differences in the size of the CoTC and rpl900 fragments and because of the presence or absence (in ΔNH) of the NH pause region. **(D)** Graph of the pA/splice expression ratio for the constructs shown along the bottom, with the CoTC element or rpl900 fragment in either the *Kpn*I site (black bars) or the *Hind*III site (white bars). The control Cμ and ΔNH constructs are shown for comparison (grey bars).
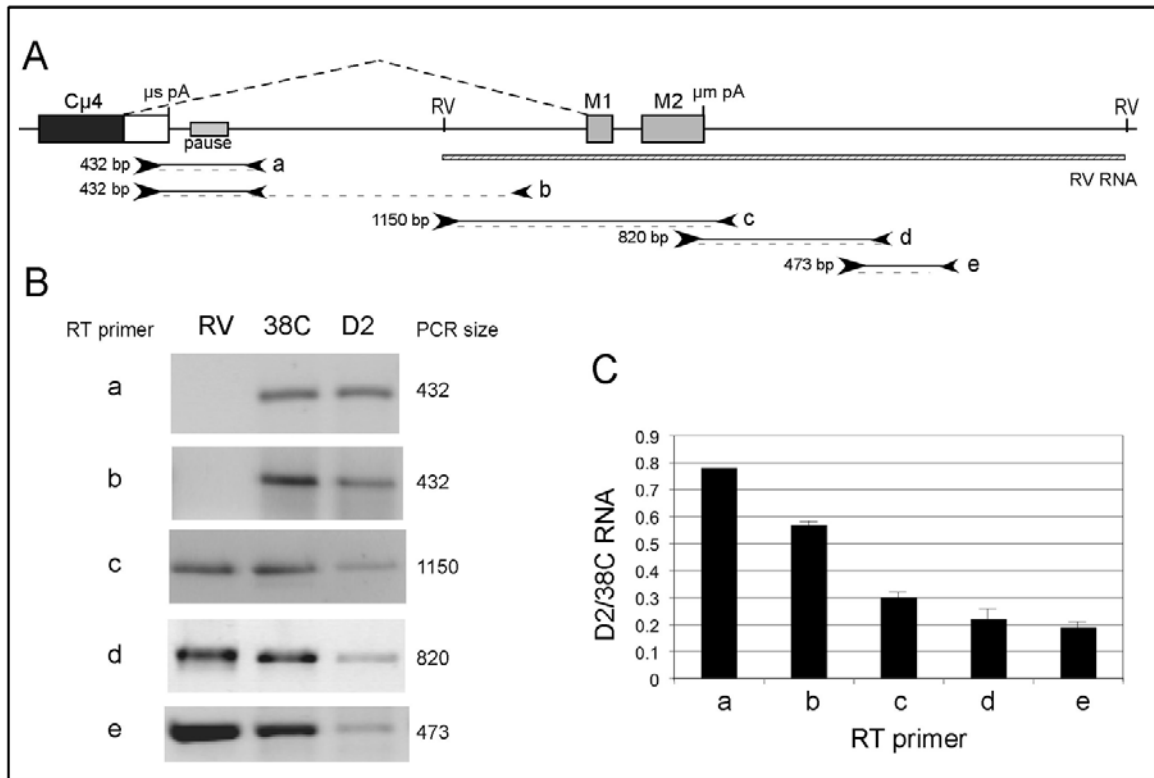
Figure 4-6. There is no evidence for a natural CoTC element in the μ gene. (A) Map of the μ gene 3' region, as in Figure 4-1, but extending downstream to the *EcoRV* site, about 400 nt upstream of the first IgD exon. (Cδ1, cross-hatched box). Below in the hatched slim box is shown the region in vitro transcribed as the RV RNA used as a positive control. Below the map is the scheme used to explore the μ gene for CoTC elements: the arrows marked "a" – e" are the RT primers used to make cDNA (denoted by dashed line) from the 38C-13 B cell line and D2 hyridoma cell line nuclear RNAs and the arrows connected by a solid line are the primers to make the PCR products from each RT reaction. The sizes of the PCR products are indicated on the left. **(B)** RT-PCR reactions from nuclear RNA from the 38C-13 (38C) B cell line or its hybridoma partner D2 or the in vitro transcribed RNA RV. The RT primers used are shown on the left and correspond to the map in A and the PCR product sizes are shown on the right. One μg of 38C-13 RNA and 2 μg of D2 RNA were used in each RT reaction to generate similar product intensity in RT-PCR "a". The primers used are (Table 2-2): "a", 432AS for RT and CμpA-T and 432AS for PCR; "b", 1528AS for RT and CμpA-T and 432AS for PCR; "c", M2-B for RT and Kpn-T and M2-B for PCR; "d", 469-T for RT and M2-T and 469-T for PCR; "e", 473AS for RT and 469-T and 473AS for PCR. **(C)** The ratio of the D2 to 38C PCR signals for each RT-PCR reaction were quantitated from two to four independent RT-PCR reactions and graphed.

CHAPTER V

CONCLUSIONS AND STATEMENT OF SIGNIFICANCE

Regulation of gene expression occurs at multiple levels and it is now recognized that modulation of RNA processing during transcription contributes to the complexity of how and when a gene may be expressed. Many human diseases have been attributed to misregulation of gene expression where RNA processing steps such as splicing, are involved. Therefore, a better understanding of the mechanisms by which transcription is coupled to RNA processing could be used to develop therapeutic tools to improve human health. Here, I studied two different model genes: the AFP and the μ genes to investigate different post-transcriptional events where a change in RNA processing is responsible for developmental changes in gene expression.

Early studies identified that AFP developmental repression by Zhx2 not only requires the AFP promoter (Peyton et al., 2000a) but it also involves post-transcriptional events (Morford, Unpublished; Vacher et al., 1992). These data suggested that Zhx2 may control accumulation of AFP mRNA levels by a mechanism that couples transcription and posttranscriptional events (Figure 1-1) (Peyton et al., 2000a). In this work, we described one of the post-transcriptional steps involved in AFP developmental repression by Zhx2. We observed that splicing efficiency of AFP transcripts is greatly reduced in the presence of Zhx2. The global repression of AFP pre-mRNA splicing may explain the reduced levels of AFP expression in adult mouse liver, when Zhx2 levels are higher than in pre-natal stage. Moreover, inhibition of AFP pre-mRNA splicing may also explain the presence of the unprocessed AFP RNA in the cytoplasm, maybe due to an increase in AFP pre-mRNA stability. It is possible that the extraordinary stability of AFP transcripts may be attributed to the absence of exon junction complex (EJC) complexes that are normally deposited during splicing. This in turn would avoid recruitment of NMD core effectors, allowing the unprocessed AFP transcripts escape from this surveillance mechanism.

One hypothesis about the mechanism of AFP repression by Zhx2 is that Zhx2 may interfere with recruitment of some factors required for coupling transcription with RNA processing. Because the promoter is required for AFP regulation by Zhx2, recruitment of these factors may occur in a promoter dependent-manner during pre-initiation complex formation, transcription initiation and/or early transcription elongation (Figure 1-1, steps 1 through 3). There are only a few reported examples where

uncoupling of transcription with RNA processing leads to full length unprocessed RNA. Two of these examples are the TNF gene in naïve T cells and the primary response genes (PRGs) in unstimulated macrophages (described in discussion of chapter III). In the case of naïve T cells, TNF pre-mRNA splicing is blocked. Soon after T cell activation, the accumulated TNF pre-mRNA is properly spliced, allowing immediate TNF protein expression. The mechanism by which TNF pre-mRNA splicing is co-transcriptionally repressed and then post-transcriptionally activated is not yet known. In the case of unstimulated macrophages, the promoters of the PRGs are pre-associated with S5-phosphorylated RNAPII before LPS stimulation, allowing production of full-length pre-mRNA that is quickly degraded. Upon LPS stimulation, p-TEFb is recruited, through Brd4, to phosphorylate RNAPII at S2, which results in the generation of mature RNA and expression of PRGs. These examples clearly indicate that regulation of RNA processing can affect gene expression as a mechanism that allows quick response to stimuli. In addition to p-TEFb, the Spt5 subunit of DSIF is known to be recruited during transcription initiation and accompanies the RNAPII during transcription elongation (Sims et al., 2004b). The Spt5 subunit has been shown to participate in different RNAPII protein complexes and to interact with RNA processing factors such as capping enzymes (Lindstrom et al., 2003; Sims et al., 2004b). Moreover, mutations in Spt5 cause splicing defects in yeast, indicating it is required for coupling transcription and mRNA maturation (Lindstrom et al., 2003). Based on all these data, factors involved in regulation of RNAPII CTD phosphorylation and recruitment of RNA processing factors, such as P-TEFb and Spt5, may be potential candidates to participate in Zhx2 regulation. To investigate whether the presence of Zhx2 alters the phosphorylation pattern of RNAPII CTD across the AFP gene, ChIP assays could be performed using antibodies that specifically recognize either S5- or S2-phosphorylated forms of the RNAPII CTD. We may expect a reduction in S2- CTD phosphorylation toward the end of the AFP gene if Zhx2 interferes with this process. Recruitment of factors such as pTEF-b and Spt5 also could be determined by ChIP assays. Additionally, identification of Zhx2-interacting proteins by proteomic analysis should provide significant clues to the post-transcriptional steps involved in Zhx2 regulation. Investigation of the AFP splicing repression mechanism by Zhx2 may increase our understanding of how gene expression could be regulated at early transcription steps (Figure 1-1, steps 1 through 3). This knowledge may be extended to other gene systems and used as tool for amelioration of human

diseases. This information also can be used to optimize cell culture systems so that they will reproduce the AFP similar repression levels observed in the mouse model.

In addition to AFP regulation, Zhx2 seems to have an important role in regulating expression of other genes, such as those involved in lipid metabolism. For instance, BALB/cJ mice, which express low levels of Zhx2, have reduced triglyceride levels compared to BALB/cJ mice expressing a Zhx2 transgene in the liver, when fed with a fat diet (Gargalovic et al., 2010). It was determined that reduced triglyceride levels in BALB/cJ mice are, in part, due to the enhanced lipoprotein lipase (Lpl)-mediated lipolysis and plasma clearance of triglyceride-rich plasma lipoproteins. Since Zhx2 negatively impacts expression of genes involved in lipid metabolism such as *Lpl*, low expression of Zhx2 may have a protective effect on atherosclerosis and heart diseases (Gargalovic et al., 2010). Additionally, the role of Zhx2 in cancer is not well established because RNA expression data and protein array are not consistent to each other (Hu et al., 2007; Lv et al., 2006). It is possible, that subcellular localization of Zhx2 is perturbed in disease states. If that were the case, studies on Zhx2 post-translational modifications may be performed and contribute to the overall understanding of the Zhx2 regulatory mechanism.

In contrast to the AFP gene, where regulation of gene expression occurs at an early step before pre-mRNA enters the splicing pathway, the μ gene developmental expression is due to regulated changes in cleavage/polyadenylation and splicing that ultimately affects transcription termination (Figure 1-1, steps 3 and 4 and Figure 1-5). Because CoTC elements are known to drive transcription termination in some genes such as the human *β-globin*, the human *ε-globin* and the mouse albumin genes, we decided to study whether the CoTC elements may drive transcription termination differences in the μ gene that occur during B cell differentiation to plasma cell. We did not find evidence of a CoTC element in the Cμ4-M1 intron or further downstream of the M2 exon. Therefore, we utilized the complex gene structure of the μ gene to better understanding the effect of cleaving the pre-mRNA within the Cμ4-M1 intron on the μ gene RNA processing. We did show that a cleavage induced by the β-globin CoTC element can affect the pA/splice mRNA ratio in certain intronic positions. Inserting the β-globin CoTC within the Cμ4-M1 intron induced cleavage of the RNA transcript independent of where it was located, but only insertion of the CoTC element downstream of the μspA dramatically affected the balance of the two competing reactions to enhance usage of the μspA. We explain this positional effect based on

timing (Figure 5-1). The presence of the CoTC toward the end of the Cμ4-M1 intron (*Kpn*I site) may allow enough time for proper co-transcriptional spliceosome assembly before cleavage occurs and thus splicing could compete with cleavage/polyadenylation at μspA (Figure 5-1A). However, when the CoTC element is present directly downstream of the μspA, cleavage may occur before spliceosome assembly so that splicing can not outcompete cleavage/polyadenylation at μspA (Figure 5-1B). In addition, the CoTC element drives early transcription termination. Several studies have addressed the effect on splicing by inducing cleavage within an intron (Dye et al., 2006; Fong et al., 2009). It was observed that a fast-cutting ribozyme such as the δ ribozyme impairs splicing, while insertion of a slower-cutting ribozyme or the β-globin CoTC element did not have the same effect. It was concluded that splicing is affected when cleavage within the intron is fast enough to interfere with the co-transcriptional spliceosome assembly (Fong et al., 2009). Here, we extended these findings using the μ gene where a splicing reaction is in competition with a cleavage/polyadenylation reaction. We observed that indeed, the CoTC element could affect splicing when inserted close to the 5' splice site and in this case, it enhances the cleavage at the μspA which results in higher μs mRNA production, rather than leads to an unproductive transcript. These results can also support evidence about microRNA (miRNA) processing. MiRNAs are noncoding RNA species of ~22 nt that are involved in post-transcriptional gene silencing (Kim and Kim, 2007). These RNAs are encoded not only within intergenic regions but also within introns of protein-coding transcrips. MiRNA processing includes cleavage by Drosha to release ~70 nt pre-miRNAs that are then processed by Dicer to generate mature ~22 nt miRNAs. This raises the question of how intronic miRNA are processed without disrupting proper RNA processing of mRNAs (Kim and Kim, 2007). Our studies together with those mentioned above support the idea that Drosha may excise intronic miRNA after splicing is committed (Fong et al., 2009; Kim and Kim, 2007).

Better understanding of μ gene RNA processing regulation could improve our knowledge of the cis- and trans-acting factors involved in alternative RNA processing and the mechanism involved in that regulation. The μ gene with its complex gene structure can serve as a model for studying splicing, cleavage/polyadenylation and transcription termination. The information gained about the μ gene regulatory mechanisms could be extended to other gene systems to modulate gene expression under pathological conditions.

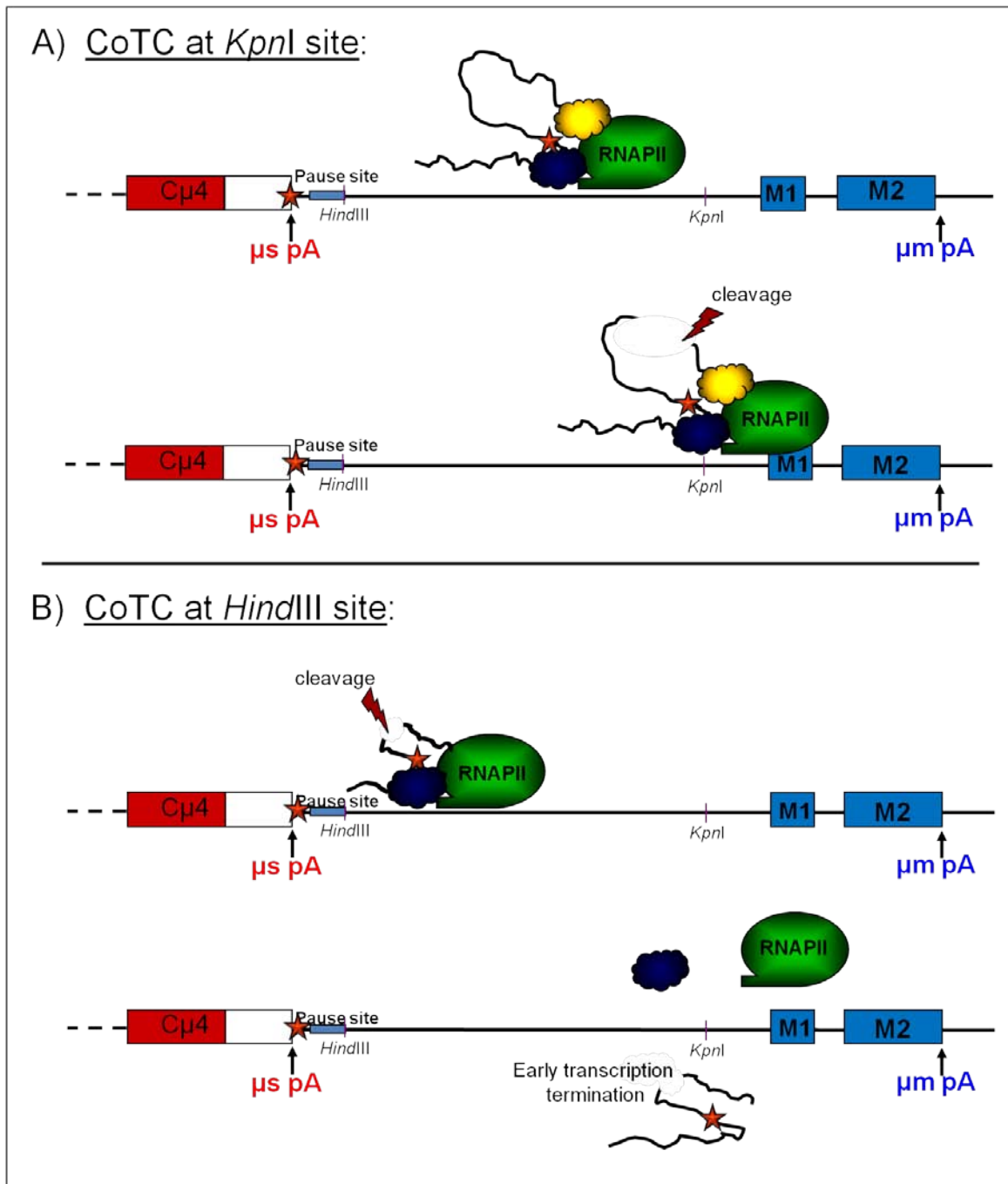**Figure 5-1. Model of the β-globin CoTC element insertion effect on μ gene RNA processing.** (A) Insertion of CoTC element at the *Kpn*I site with in the Cμ4-M1 intron. (B) Insertion of CoTC element at the *Hind*III site with in the Cμ4-M1 intron.

REFERENCES

- Alj, Y., Georgiakaki, M., Savouret, J.F., Mal, F., Attali, P., Pelletier, G., Fourre, C., Milgrom, E., Buffet, C., Guiochon-Mantel, A.*, et al.* (2004). Hereditary persistence of alpha-fetoprotein is due to both proximal and distal hepatocyte nuclear factor-1 site mutations. Gastroenterology *126*, 308-317.

- Armellini, A., Sarasquete, M.E., Garcia-Sanz, R., Chillon, M.C., Balanzategui, A., Alcoceba, M., Fuertes, M., Lopez, R., Hernandez, J.M., Fernandez-Calvo, J.*, et al.* (2008). Low expression of ZHX2, but not RCBTB2 or RAN, is associated with poor outcome in multiple myeloma. Br J Haematol *141*, 212-215.

- Ashfield, R., Enriquez-Harris, P., and Proudfoot, N.J. (1991). Transcriptional termination between the closely linked human complement genes C2 and factor B: common termination factor for C2 and c-myc? EMBO J *10*, 4197-4207.

- Ashfield, R., Patel, A.J., Bossone, S.A., Brown, H., Campbell, R.D., Marcu, K.B., and Proudfoot, N.J. (1994). MAZ-dependent termination between closely spaced human complement genes. EMBO J *13*, 5656-5667.

- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., Albright, L.M., Coen, D.M., and Varki, A., eds. (1987). Current protocols in molecular biology (Brooklyn, Greene Publishing Associates).

- Aygun, O., Svejstrup, J., and Liu, Y. (2008). A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. Proc Natl Acad Sci U S A *105*, 8580-8584.

- Baralle, D., and Baralle, M. (2005). Splicing in action: assessing disease causing sequence changes. J Med Genet *42*, 737-748.

- Barthelemy, I., Carramolino, L., Gutierrez, J., Barbero, J.L., Marquez, G., and Zaballos, A. (1996). zhx-1: A novel homeodomain protein containing two zinc-fingers and five homeodomains. Biochem Biophys Res Comm *224*, 870-876.

- Behm-Ansmant, I., and Izaurralde, E. (2006). Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. Genes Dev *20*, 391-398.

- Belayew, A., and Tilghman, S.M. (1982). Genetic analysis of alpha-fetoprotein synthesis in mice. Mol Cell Biol *2*, 1427-1435.

- Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol *14*, 336-342.

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

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

- Blankenhorn, E.P., Duncan, R., Huppi, K., and Potter, M. (1988). Chromosomal location of the regulator of mouse alpha-fetoprotein, Afr-1. Genetics *119*, 687-691.

- Blesa, J.R., Giner-Duran, R., Vidal, J., Lacalle, M.L., Catalan, I., Bixquert, M., Igual, L., and Hernandez-Yago, J. (2003). Report of hereditary persistence of alpha-fetoprotein in a Spanish family: molecular basis and clinical concerns. J Hepatol *38,* 541-544.

- Boelz, S., Neu-Yilik, G., Gehring, N.H., Hentze, M.W., and Kulozik, A.E. (2006). A chemiluminescence-based reporter system to monitor nonsense-mediated mRNA decay. Biochem Biophys Res Commun *349*, 186-191.
- Brannan, C.I., Dees, E.C., Ingram, R.S., and Tilghman, S.M. (1990). The product of the H19 may function as an RNA. Mol Cell Biol *10*, 28-36.
- Bres, V., Yoh, S.M., and Jones, K.A. (2008). The multi-tasking P-TEFb complex. Curr Opin Cell Biol *20*, 334-340.
- Bruce, S.R., Dingle, R.W.C., and Peterson, M.L. (2003). B-cell and plasma-cell splicing differences: A potential role in regulated immunoglobulin RNA processing. RNA *9*, 1264-1273.
- Burnside, R.D., Ribble, A., and Peterson, M.L. (2011). The spatial relationship between RNA polymerase II pause sites and poly(A) signals in the IgM gene. submitted.
- Camper, S.A., and Tilghman, S.M. (1989). Postnatal repression of the α-fetoprotein gene is enhancer independent. Genes Dev *3*, 537-546.
- Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.
- Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem *76*, 51-74.
- Chiba, K., Yamamoto, J., Yamaguchi, Y., and Handa, H. (2010). Promoter-proximal pausing and its release: molecular mechanisms and physiological functions. Exp Cell Res *316*, 2723-2730.
- Cramer, P., Caceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E., and Kornblihtt, A.R. (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. Mol Cell *4*, 251-258.
- Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci USA *94*, 11456-11460.
- Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature *389*, 399-402.
- De Andrade, T., Moreira, L., Duarte, A., Lanaro, C., De Albuquerque, D., Saad, S., and Costa, F. (2010). Expression of new red cell-related genes in erythroid differentiation. Biochem Genet *48*, 164-171.
- de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell *12*, 525-532.
- Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

- Dou, Y., Fox-Walsh, K.L., Baldi, P.F., and Hertel, K.J. (2006). Genomic splice-site analysis reveals frequent alternative splicing close to the dominant splice site. RNA *12*, 2047-2056.
- Dye, M.J., Gromak, N., and Proudfoot, N.J. (2006). Exon tethering in transcription by RNA polymerase II. Mol Cell *21*, 849-859.
- Dye, M.J., and Proudfoot, N.J. (2001). Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. Cell *105*, 669-681.
- Enriquez-Harris, P., Levitt, N., Briggs, D., and Proudfoot, N.J. (1991). A pause site for RNA polymerase II is associated with termination of transcription. EMBO J *10*, 1833-1842.
- Feuerman, M.H., Godbout, R., Ingram, R.S., and Tilghman, S.M. (1989). Tissue-specific transcription of the mouse α-fetoprotein gene promoter is dependent on HNF-1. Mol Cell Biol *9*, 4204-4212.
- Fong, N., and Bentley, D.L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev *15*, 1783-1795.
- Fong, N., Ohman, M., and Bentley, D.L. (2009). Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. Nat Struct Mol Biol *16*, 916-922.
- Frischmeyer, P.A., and Dietz, H.C. (1999). Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet *8*, 1893-1900.
- Gargalovic, P.S., Erbilgin, A., Kohannim, O., Pagnon, J., Wang, X., Castellani, L., LeBoeuf, R., Peterson, M.L., Spear, B.T., and Lusis, A.J. (2010). Quantitative trait locus mapping and identification of Zhx2 as a novel regulator of plasma lipid metabolism. Circ Cardiovasc Genet *3*, 60-67.
- Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat Struct Mol Biol *15*, 71-78.
- Godbout, R., and Tilghman, S.M. (1988). Configuration of the alpha-fetoprotein regulatory domain during development. Genes Dev *2*, 949-956.
- 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.
- Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. Cell *138*, 129-145.
- Hu, A., and Fu, X.D. (2007). Splicing oncogenes. Nat Struct Mol Biol *14*, 174-175.
- Hu, S., Zhang, M., Lv, Z., Bi, J., Dong, Y., and Wen, J. (2007). Expression of zinc-fingers and homeoboxes 2 in hepatocellular carcinogenesis: a tissue microarray and clinicopathological analysis. Neoplasma *54*, 207-211.
- Hyman, R., Ralph, P., and Sarkar, S. (1972). Cell-specific antigens and immunoglobulin synthesis of murine myeloma cells and their variants. J Natl Cancer Inst *48*, 173-184.
- Janeway, C., Travers, P., Walport, M., and Shlomchik, M. (2005). Immunobiology: The immune system in health and disease, 6 edn (New York, Garland Science).

- Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H*., et al.* (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol *3*, 576-583.
- Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol *14*, 185-193.
- Kawata, H., Yamada, D., Shou, Z., Mizutani, T., and Miyamoto, K. (2003a). The mouse zinc-fingers and homeoboxes (ZHX) family; ZHX2 forms a heterdimer with ZHX3. Gene *323*, 1330140.
- Kawata, H., Yamada, K., Shou, Z., Mizutani, T., Yazawa, T., Yoshino, M., Sekiguchi, T., Kajitani, T., and Miyamoto, K. (2003b). Zinc-fingers and homeoboxes (ZHX) 2, a novel member of the ZHX family, functions as a transcriptional repressor. Biochem J *373*, 747-757.
- Kelley, D.E., and Perry, R.P. (1986). Transcriptional and post-transcriptional control of immunoglobulin mRNA production during B lymphocyte development. Nucleic Acids Res *14*, 5431-5447.
- Kim, K.J., Kanellopoulos-Langevin, C., Merwin, R.M., Sachs, D.H., and Asofsky, R. (1979). Establishment and characterization of BALB/c lymphoma lines with B cell properties. J Immunol *122*, 549-553.
- Kim, Y.K., and Kim, V.N. (2007). Processing of intronic microRNAs. EMBO J *26*, 775-783.
- Kornblihtt, A.R. (2005). Promoter usage and alternative splicing. Curr Opin Cell Biol *17*, 262-268.
- Kornblihtt, A.R. (2006). Chromatin, transcript elongation and alternative splicing. Nat Struct Mol Biol *13*, 5-7.
- Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J., and Nogues, G. (2004). Multiple links between transcription and splicing. RNA *10*, 1489-1498.
- Krumlauf, R., Hammer, R.E., Tilghman, S.M., and Brinster, R.L. (1985). Developmental regulation of α-fetoprotein genes in transgenic mice. Mol Cell Biol *5*, 1639-1648.
- Kurosaki, T., Shinohara, H., and Baba, Y. (2010). B Cell Signaling and Fate Decision. Annual Review of Immunology *28*, 21-55.
- Legrain, P., and Rosbash, M. (1989). Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. Cell *57*, 573-583.
- Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., 3rd, and Hartzog, G.A. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. Mol Cell Biol *23*, 1368-1378.
- Liu, G., Clement, L.C., Kanwar, Y.S., Avila-Casado, C., and Chugh, S.S. (2006). ZHX proteins regulate podocyte gene expression during the development of nephrotic syndrome. J Biol Chem *281*, 39681-39692.

- Lo, C.W., Kaida, D., Nishimura, S., Matsuyama, A., Yashiroda, Y., Taoka, H., Ishigami, K., Watanabe, H., Nakajima, H., Tani, T.*, et al.* (2007). Inhibition of splicing and nuclear retention of pre-mRNA by spliceostatin A in fission yeast. Biochem Biophys Res Commun *364*, 573-577.
- Long, L., Davidson, J.N., and Spear, B.T. (2004). Striking differences between the mouse and human a-fetoprotein enhancers. Genomics *83*, 694-705.
- Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell *144*, 16-26.
- Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Dev *20*, 954-965.
- Lv, Z., Zhang, M., Bi, J., Xu, F., Hu, S., and Wen, J. (2006). Promoter hypermethylation of a novel gene, ZHX2, in hepatocellular carcinoma. Am J Clin Pathol *125*, 740-746.
- Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J., and Reinberg, D. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. Proc Natl Acad Sci U S A *101*, 7572-7577.
- Martincic, K., Alkan, S.A., Cheatle, A., Borghesi, L., and Milcarek, C. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nat Immunol *10*, 1102-1109.
- Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet *36*, 1073-1078.
- Mercatante, D.R., Bortner, C.D., Cidlowski, J.A., and Kole, R. (2001). Modification of alternative splicing of Bcl-x pre-mRNA in prostate and breast cancer cells. analysis of apoptosis and cell death. J Biol Chem *276*, 16411-16417.
- Mizejewski, G.J. (2004). Biological roles of alpha-fetoprotein during pregnancy and perinatal development. Exp Biol Med (Maywood) *229*, 439-463.
- Mizejewski, G.J. (2011). Review of the putative cell-surface receptors for alpha-fetoprotein: identification of a candidate receptor protein family. Tumour Biol *32*, 241-258.
- Morford, L. (Unpublished).
- Morford, L., Davis, C., Jin, L., Dobierzewska, A., Peterson, M.L., and Spear, B.T. (2007). The oncofetal gene Glypican 3 is regulated in postnatal liver by Zhx2 and regenerating liver by Afr2. Hepatology *46*, 1541-1547.
- Nag, A., Narsinh, K., and Martinson, H.G. (2007). The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat Struct Mol Biol *14*, 662-669.
- Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochim Biophys Acta *1809*, 34-45.

- Nelson, K.J., Haimovich, J., and Perry, R.P. (1983). Characterization of productive and sterile transcripts from the immunoglobulin heavy-chain locus: processing of μm and μs mRNA. Mol Cell Biol *3*, 1317-1332.
- Nogues, G., Kadener, S., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2002). Transcriptional activators differ in their abilities to control alternative splicing. J Biol Chem *277*, 43110-43114.
- O'Brien, K., Matlin, A.J., Lowell, A.M., and Moore, M.J. (2008). The biflavonoid isoginkgetin is a general inhibitor of Pre-mRNA splicing. J Biol Chem *283*, 33147-33154.
- Olsson, M., Lindahl, G., and Ruoslahti, E. (1977). Genetic control of alpha-fetoprotein synthesis in the mouse. J Exp Med *145*, 819-827.
- Orozco, I.J., Kim, S.J., and Martinson, H.G. (2002). The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. J Biol Chem *277*, 42899-42911.
- Pachnis, V., Belayew, A., and Tilghman, S.M. (1984). Locus unlinked to alpha-fetoprotein under the control of the murine raf and Rif genes. Proc Natl Acad Sci U S A *81*, 5523-5527.
- Pachnis, V., Brannan, C.I., and Tilghman, S.M. (1988). The structure and expression of a novel gene activated early in mouse embryogenesis. EMBO J *7*, 673-681.
- Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet *5*, 389-396.
- Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA *15*, 1896-1908.
- Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell *36*, 178-191.
- Perincheri, S., Dingle, R.W., Peterson, M.L., and Spear, B.T. (2005). Hereditary persistence of alpha-fetoprotein and H19 expression in liver of BALB/cJ mice is due to a retrovirus insertion in the Zhx2 gene. Proc Natl Acad Sci U S A *102*, 396-401.
- Perincheri, S., Peyton, D.K., Glenn, M., Peterson, M.L., and Spear, B.T. (2008). Characterization of the ETnII-alpha endogenous retroviral element in the BALB/cJ Zhx2 (Afr1) allele. Mamm Genome *19*, 26-31.
- Peterson, M.L. (1992). Balanced efficiencies of splicing and cleavage-polyadenylation are required for μs and μm mRNA regulation. Gene Expr *2*, 319-327.
- Peterson, M.L. (1994a). Regulated immunoglobulin (Ig) RNA processing does not require specific cis-acting sequences: Non-Ig genes can be alternatively processed in B cells and plasma cells. Mol Cell Biol *14*, 7891-7898.
- Peterson, M.L. (1994b). RNA processing and expression of immunoglobulin genes. In Handbook of B and T Lymphocytes, E.C. Snow, ed. (San Diego, Academic Press), pp. 321-342.
- Peterson, M.L. (2007). Mechanisms controlling production of membrane and secreted immunoglobulin during B cell development. Immunologic Res *37*, 33-46.

- Peterson, M.L. (2011). Immunoglobulin heavy chain gene regulation through polyadenylation and splicing competition. Wiley Interdisciplinary Reviews: RNA *2*, 92-105.
- Peterson, M.L., Bertolino, S., and Davis, F. (2002). An RNA polymerase pause site is associated with the immunoglobulin µs poly(A) site. Mol Cell Biol *22*, 5606-5615.
- Peterson, M.L., Bryman, M.B., Peiter, M., and Cowan, C. (1994). Exon size affects competition between splicing and cleavage-polyadenylation in the immunoglobulin µ gene. Mol Cell Biol *14*, 77-86.
- Peterson, M.L., Gimmi, E.R., and Perry, R.P. (1991). The developmentally regulated shift from membrane to secreted mu mRNA production is accompanied by an increase in cleavage-polyadenylation efficiency but no measurable change in splicing efficiency. Mol Cell Biol *11*, 2324-2327.
- Peterson, M.L., Ma, C., and Spear, B.T. (2011). Zhx2 and Zbtb20: novel regulators of postnatal alpha-fetoprotein repression and their potential role in gene reactivation during liver cancer. Semin Cancer Biol *21*, 21-27.
- Peterson, M.L., and Perry, R.P. (1986). Regulated production of µm and µs mRNA requires linkage of the poly(A) addition sites and is dependent on the length of the µs-µm intron. Proc Natl Acad Sci USA *83*, 8883-8887.
- Peterson, M.L., and Perry, R.P. (1989). The regulated production of µm and µs mRNA is dependent on the relative efficiencies of µs poly(A) site usage and the Cµ4-to-M1 splice. Mol Cell Biol *9*, 726-738.
- Peyton, D.K., Huang, M.-C., Giglia, M.A., Hughes, N.K., and Spear, B.T. (2000a). The alpha-fetoprotein promoter is the target of Afr1-mediated postnatal repression. Genomics *63*, 173-180.
- Peyton, D.K., Ramesh, T., and Spear, B.T. (2000b). Position-dependent activity of α-fetoprotein enhancer element III in the adult liver is due to negative regulation. ProcNatAcadSci,USA *97*, 10890-10894.
- Plant, K.E., Dye, M.J., Lafaille, C., and Proudfoot, N.J. (2005). Strong polyadenylation and weak pausing combine to cause efficient termination of transcription in the human Ggamma-globin gene. Mol Cell Biol *25*, 3276-3285.
- Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501-512.
- Ramesh, T.M., Ellis, A.W., and Spear, B.T. (1995). Individual mouse alpha-fetoprotein enhancer elements exhibit different patterns of tissue-specific and hepatic position-dependent activities. Mol Cell Biol *15*, 4947-4955.
- Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes Dev *23*, 1247-1269.
- Roberts, G.C., Gooding, C., Mak, H.Y., Proudfoot, N.J., and Smith, C.W.J. (1998). Co-transcriptional commitment to alternative splice site selection. Nucleic Acids Res *26*, 5568-5572.
- Robson-Dixon, N.D., and Garcia-Blanco, M.A. (2004). MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. J Biol Chem *279*, 29075-29084.

- Rosonina, E., Kaneko, S., and Manley, J.L. (2006). Terminating the transcript: breaking up is hard to do. Genes Dev *20*, 1050-1056.
- Saint-Andre, V., Batsche, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. Nat Struct Mol Biol *18*, 337-344.
- Sanford, J.R., Ellis, J., and Caceres, J.F. (2005). Multiple roles of arginine/serine-rich splicing factors in RNA processing. Biochem Soc Trans *33*, 443-446.
- Schibler, U., Wyler, T., and Hagenbüchle, O. (1975). Changes in size and secondary structure of the ribosomal transcription unit during vertebrate evolution. Journal of Molecular Biology *94*, 503-510.
- Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat Struct Mol Biol *16*, 990-995.
- Seipelt, R.L., and Peterson, M.L. (1995). Alternative RNA processing of IgA pre-mRNA responds like IgM to alterations in the efficiency of the competing splice and cleavage-polyadenylation reactions. Mol Immunol *32*, 277-285.
- Seipelt, R.L., Spear, B.T., Snow, E.C., and Peterson, M.L. (1998). A nonimmunoglobulin transgene and the endogenous immunoglobulin μ gene are coordinately regulated by alternative RNA processing during B-cell maturation. Mol Cell Biol *18*, 1042-1048.
- Shen, H., Luan, F., Liu, H., Gao, L., Liang, X., Zhang, L., Sun, W., and Ma, C. (2008). ZHX2 is a repressor of alpha-fetoprotein expression in human hepatoma cell lines. J Cell Mol Med *12*, 2772-2780.
- Sims, R.J., 3rd, Mandal, S.S., and Reinberg, D. (2004a). Recent highlights of RNA-polymerase-II-mediated transcription. Curr Opin Cell Biol *16*, 263-271.
- Sims, R.J.I., Belotserkovskaya, R., and Reinberg, D. (2004b). Elongation by RNA polymerase II: the short and long of it. Genes Dev *18*, 2437-2468.
- Singh, N.N., Androphy, E.J., and Singh, R.N. (2004). The regulation and regulatory activities of alternative splicing of the SMN gene. Crit Rev Eukaryot Gene Expr *14*, 271-285.
- Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci *25*, 381-388.
- Spear, B.T. (1994). Mouse alpha-fetoprotein gene 5' regulatory elements are required for postnatal regulation by raf and Rif. Mol Cell Biol *14*, 6497-6505.
- Spear, B.T. (1999). Alpha-fetoprotein gene regulation: Lessons from transgenic mice. Seminars in Cancer Biology *9*, 109-116.
- Spear, B.T., Jin, L., Ramasamy, S., and Dobierzewska, A. (2006). Transcriptional control in the mammalian liver: liver development, perinatal repression, and zonal gene regulation. Cell Mol Life Sci *63*, 2922-2938.
- Spear, B.T., and Tilghman, S.M. (1990). Role of α-fetoprotein regulatory elements in transcriptional activation in transient heterokaryons. Mol Cell Biol *10*, 5047-5054.
- Sumner, C.J. (2006). Therapeutics development for spinal muscular atrophy. NeuroRx *3*, 235-245.

- Tazi, J., Bakkour, N., and Stamm, S. (2009). Alternative splicing and disease. Biochim Biophys Acta *1792*, 14-26.
- Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Dye, M., James, W., Proudfoot, N., and Akoulitchev, A. (2004). Autocatalytic RNA cleavage in the human ß-globin gene transcript promotes transcriptional termination. Nature *432*, 526-530.
- Tsurushita, N., and Korn, L.J. (1987). Effects of intron length on differential processing of mouse µ heavy-chain mRNA. Mol Cell Biol *7*, 2602-2605.
- Vacher, J., Camper, S.A., Krumlauf, R., Compton, R.S., and Tilghman, S.M. (1992). *raf* regulates the postnatal repression of the mouse α-fetoprotein gene at the posttranscriptional level. Mol Cell Biol *12*, 856-864.
- Vacher, J., and Tilghman, S.M. (1990). Dominant negative regulation of the mouse α-fetoprotein gene in adult liver. Science *250*, 1732-1735.
- Venables, J.P. (2004). Aberrant and alternative splicing in cancer. Cancer Res *64*, 7647-7654.
- Watson, J.D., Baker T.A., Bell, S.P., Gann, A., Levine, M., Losick, R., ed. (2004). Molecular Biology of the Gene, 5th edn (San Francisco, CA, Pearson Education, Inc).
- Weiss, E.A., Michael, A., and Yuan, D. (1989). Role of transcriptional termination in the regulation of µ mRNA expression in B lymphocytes. J Immunol *143*, 1046-1052.
- West, S., Gromak, N., and Proudfoot, N. (2004). Human 5'->3' exonuclease Xrn2 promotes transcriptional termination at co-transcriptional cleavage sites. Nature *432*, 522-525.
- West, S., and Proudfoot, N.J. (2008). Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. Nucleic Acids Res *36*, 905-914.
- West, S., and Proudfoot, N.J. (2009). Transcriptional termination enhances protein expression in human cells. Mol Cell *33*, 354-364.
- West, S., Proudfoot, N.J., and Dye, M.J. (2008). Molecular dissection of mammalian RNA polymerase II transcriptional termination. Mol Cell *29*, 600-610.
- West, S., Zaret, K., and Proudfoot, N.J. (2006). Transcriptional termination sequences in the mouse serum albumin gene. RNA *12*, 655-665.
- Wu, C., Qiu, R., Wang, J., Zhang, H., Murai, K., and Lu, Q. (2009). ZHX2 Interacts with Ephrin-B and regulates neural progenitor maintenance in the developing cerebral cortex. J Neurosci *29*, 7404-7412.
- Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol Cell Biol *14*, 7219-7225.
- Xie, Z., Zhang, H., Tsai, W., Zhang, Y., Du, Y., Zhong, J., Szpirer, C., Zhu, M., Cao, X., Barton, M.C.*, et al.* (2008). Zinc finger protein ZBTB20 is a key repressor of alpha-fetoprotein gene transcription in liver. Proc Natl Acad Sci U S A *105*, 10859-10864.
- Yamada, K., Kawata, H., Shou, Z., Hirano, S., Mizutani, T., Yazawa, T., Sekiguchi, T., Yoshino, M., Kajitani, T., and Miyamoto, K. (2003). Analysis of zinc-fingers and

homeoboxes (ZHX)-1-interacting proteins: molecular cloning and characterization of a member of the ZHX family, ZHX3. Biochem J *373*, 167-178.

- Yamada, K., Ogata-Kawata, H., Matsuura, K., Kagawa, N., Takagi, K., Asano, K., Haneishi, A., and Miyamoto, K. (2009). ZHX2 and ZHX3 repress cancer markers in normal hepatocytes. Front Biosci *14*, 3724-3732.
- Yamada, K., Printz, R.L., Osawa, H., and Granner, D.K. (1999). Human ZHX1: Cloning, chromosomal location, and interaction with transcription factor NF-Y. Biochem Biophys Res Comm *261*, 614-621.
- Yang, Y., Chang, J.F., Parnes, J.R., and Fathman, C.G. (1998). T cell receptor (TCR) engagement leads to activation-induced splicing of tumor necrosis factor (TNF) nuclear pre-mRNA. J Exp Med *188*, 247-254.
- Yuan, D., and Tucker, P.W. (1984). Transcriptional regulation of μ-δ heavy chain locus in normal murine B lymphocytes. J Exp Med *160*, 564-583.
- Zahler, A.M., Neugebauer, K.M., Lane, W.S., and Roth, M.B. (1993). Distinct functions of SR proteins in alternative pre-mRNA splicing. Science *260*, 219-222.
- Zerbe, L.K., Pino, I., Pio, R., Cosper, P.F., Dwyer-Nield, L.D., Meyer, A.M., Port, J.D., Montuenga, L.M., and Malkinson, A.M. (2004). Relative amounts of antagonistic splicing factors, hnRNP A1 and ASF/SF2, change during neoplastic lung growth: implications for pre-mRNA processing. Mol Carcinog *41*, 187-196.
- Zhang, Z., and Gilmour, D.S. (2006). Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Mol Cell *21*, 65-74.

VITA

Name: Lilia M. Turcios
Date of Birth: January 7, 1974
Place of Birth: Caracas, Venezuela

Education:

- M.S. Biological Sciences, Simon Bolivar University (2000), Caracas Venezuela
- B.S. Chemistry, Simon Bolivar University (1996), Caracas Venezuela

Professional positions held:

- Visitor Scholar, University of Kentucky, Department of Plant and Soil Sciences, 2004-2005.

- Research scientist, Venezuelan Institute for scientific Research (IVIC; Venezuela), 2001-2004

- Laboratory Technician , Environmental Management Unit, Universidad Simon Bolivar, (Venezuela), 2000-2001.

- Teaching Assistant, Department of Chemistry, Universidad "Simon Bolivar"1992-1995.

Publications:

- **Turcios, L**., Casart, Y., Florez, I., de Waard, J., Salazar, L. Characterization of IS6110 insertions in the dnaA-dnaN intergenic region of *Mycobacterium tuberculosis* clinical isolates. 2009. Clinical Microbiology and Infection. 15(2): 200-203.

- Casart, Y., **Turcios, L**., Florez, I., Jaspe, R., Guerrero, E., de Waard, J., Agular, D., Hernandez-Pando, R., Salazar, L. IS6110 in oriC affects the morphology and growth of Mycobacterium tuberculosis and attenuates virulence in mice. 2008. Tuberculosis. 88(6): 545-552.

- Ke J, Gururajan M, Kumar A, Simmons A, **Turcios L**, Chelvarajan RL, Cohen DM, Wiest DL, Monroe JG, Bondada S. The role of MAPKs in B cell receptor-induced down-regulation of Egr-1 in immature B lymphoma cells. 2006. J Biol Chem. 281(52): 39806-18.

- Salazar, L., Guerrero, E., Casart, Y., **Turcios, L**., Bartoli, F. Transcription analysis of the dnaA gene and oriC region of the chromosome of *Mycobacterium smegmatis* and *Mycobacterium bovis* BCG, and its regulation by the DnaA protein. Microbiology. 2003. 149:773-784.

Abstracts

Turcios, L., Morford, L., Spear, B., Peterson, M. (2011) Post-natal AFP and H19 repression by Zhx2 through RNA splicing inhibition. Experimental Biology, Anaheim-USA.

Turcios, L., Morford, L., Spear, B., Peterson, M. (2009) Post-natal AFP and H19 repression by Zhx2 through RNA splicing inhibition. RNA, Wisconsin-USA.

Turcios, L. and Peterson, M. Co-Transcriptional cleavage in the immunoglobin M (IgM) gene. Rustbelt RNA Meeting, Ohio-USA. 2008.

Turcios L., Casart Y., Flores I., Jaspe R., Guerrero E., de Waard J., Aguirre D., Hernández-Pando R. y Salazar L. (2007) IS*6110* insertion in the origin of chromosome replication of *Mycobacterium tuberculosis* affects bacilli morphology, growth and attenuates virulence in mice. 24°. Congreso Brasileño de Microbiología. XII Simposio Brasileño de Micobacterias y II Encuentro Latinoamericano de Micobacterias. Brasilia-Brasil.

Salazar , L., **Turcios L**., Casart Y., de Waard J., Gonzalez-y-Merchand J., Aguilar D., and Hérnandez-Pando R. (2005) Effect of the high rate of IS*6110* transposition in *oriC* on the virulence of *Mycobacterium tuberculosis.* Keystone Symposia: "Tuberculosis, Integrating Host and Pathogen Biology". Whistler-Canada.

Turcios L., Casart Y., de Waard J. y Salazar L. (2004) Characterization of the *dnaA-dnaN* intergenic region of *M. tuberculosis* from Venezuelan clinical isolates reveals recent IS*6110* transposition in a short chromosome region. XVII Congreso Latinoamericano de Microbiología y X Congreso Argentino de Microbiología. Asociación Latinoamericana de Microbiología y Asociación Argentina de Microbiología. Buenos Aires-Argentina.

Salazar, L., Guerrero, E., Casart, Y., **Turcios, L**., Bartolli, F. (2002) Análisis transcripcional del gen *dnaA* y la region *oriC* del cromosoma de *Mycobacterium smegmatis* y *Mycobacterium bovis,* y su regulación por la proteína DnaA. IV reunión del grupo de microbiología molecular de la sociedad española de microbiologia. Santander-Spain.

Gonzatti, M. I.; **Turcios, L**. M.; Fernández, I; González, L. E. y Aso, P. M. (1999) Proteasas de *Trypanosoma evansi*. Second Symposium on New World Trypanosomes and the others Hemoparasites. San Juan de los Morros-Venezuela.

Gonzatti, M., Salazar, L. y **Turcios, L**. (1999) Clonamiento, secuenciación y expresión del gen que codifica para la evansaína. IL Convención Anual de la AsoVAC, Maracay-Venezuela.

Turcios, L. y Gonzatti, M. I. (1998) Clonamiento, secuenciación y expresión del gen que codifica para la evansaína. I Simposium Nacional de Hemoparásitos y sus Vectores. Maracay-Venezuela.

Turcios, L. y Gonzatti, M. I. (1996) Parámetros cinéticos de la proteasa evansaína purificada. XLVI Convención Anual de la AsoVAC, Barquisimeto-Venezuela.

Gonzatti, M. I., Bremo, A., González, L. y **Turcios, L**. (1996) Evansain: A new member of the cysteine protease family. Keystone Symposia: Proteolytic Enzymes and Inhibitors in Biology and Medicine. Keystone-Colorado, USA.

Bremo, A., González, L., **Turcios, L**. y Gonzatti, M**.** (1995) Purificación y Caracterización de evansaína y su Inhibidor Cistatina. XLV Convención Anual de la AsoVAC, Caracas-Venezuela.