CHARACTERIZATION OF U2AF^{26}, A PARALOG OF
THE SPLICING FACTOR U2AF^{35}

APPROVED BY SUPERVISORY COMMITTEE

_____
Steve McKnight, Ph.D.
Chairman and Professor of Biochemistry

_____
Kristen Lynch, Ph.D.
Assistant Professor of Biochemistry

_____
Melanie Cobb, Ph.D.
Professor of Pharmacology

_____
Joachim Herz, Ph.D.
Professor of Molecular Genetics

_____
Yi Liu, Ph.D.
Assistant Professor of Physiology

DEDICATION

Dedicated to my parents Frances Stone, Jim Shepard, and to my wife, Dana Shepard, for all their encouragement during my graduate studies.

CHARACTERIZATION OF U2AF^{26}, A PARALOG OF THE SPLICING FACTOR U2AF^{35}

by

Jeremiah Brian Shepard

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

July, 2004

ACKNOWLEDGEMENTS

I would like to thank my mentors, Dr. Steve McKnight and Dr. Kristen Lynch, for their guidance and support during my graduate training. I also thank my co-workers from the McKnight/Lynch labs. From the McKnight lab, I thank Dr. Martin Reick, Dr. Jared Rutter, Dr. Jake Zheng, Dr. Noelle Williams, Leeju Wu, Dr. Carolyn Michnoff, Sandi Joe Estill, Carol Dudley, Claudia Erbel, Dr. Fred Robinson, and Brandon Probst. From the Lynch lab, I thank Caryn Rothrock, Paul Shieves, and Amy House.

I thank the Division of Cell and Molecular Biology and Integrative Biology program and I especially thank my committee members Dr. Melanie Cobb, Dr. Phil Perlman, Dr. Joachim Herz, and Dr. Yi Liu.

I thank my parents Frances Stone and Jim Shepard for their support and encouragement while I was growing up and during my graduate studies. I also thank my stepparents, Robin Shepard and Marion Stone for their support. I thank my grandfather, Charles Shepard, for his advice during my graduate training. I thank my father- and mother-in-law, Paul and Linda Chang for their encouragement. My sister and brother, Jennifer Shepard and Nick Shepard, have helped me by always being very caring and giving.

Thanks to Josh Bembenek, Joe Baur, Amelia Feulner, Jason Mendler, and Anthony Berget for their friendship and advice during my graduate studies. I want to especially thank to my wife, Dana Shepard, for her encouragement and advice. She has been an invaluable source of support during the time I have been in graduate school.

CHARACTERIZATION OF U2AF^{26}, A PARALOG OF
THE SPLICING FACTOR U2AF^{35}

Publication No._____

Jeremiah Brian Shepard

The University of Texas Southwestern Medical Center at Dallas, 2004

Supervising Professors: Kristen Lynch, Ph.D. and Steve McKnight, Ph.D.

ABSTRACT

The essential splicing factor U2 auxiliary factor (U2AF) mediates 3' splice site recognition during spliceosome assembly. The mammalian U2AF is composed of a large subunit, U2AF^{65}, and a small subunit, U2AF^{35}. U2AF^{65} recognizes the pyrimidine tract and U2AF^{35} binds to the AG dinucleotide, both of which are specific 3' splice site sequence motifs. In the present work U2AF^{26}, a paralog of the conventional U2AF^{35}, has been studied. U2AF^{26} shares 84% primary amino acid identity with U2AF^{35}, suggesting

functional homology. However, U2AF^{26} has two amino acid substitutions in ribonulceoprotein consensus sequence-2 (RNP-2) and significant differences within the RS domain, two regions thought to be important for the function of U2AF^{35}. The goal of this study was to characterize the functional differences between the two small subunits.

Western blot analysis revealed that U2AF^{26} protein expression varies relative to U2AF^{35} in different mouse tissues. Site-specific crosslinking analysis of sixteen permutations of the nucleotide composition upstream and downstream of the AG indicates that U2AF^{26} and U2AF^{35} bind to the UAGG motif with the highest affinity. Interestingly, U2AF^{26} binds the UAGU motif better than U2AF^{35}. This observation suggests that U2AF^{26} and U2AF^{35} have overlapping binding affinities, but that U2AF^{26} might be capable of recognizing a specific 3' splice site motif better than U2AF^{35}.

Initial evidence suggested that U2AF^{26} is regulated by circadian rhythm. Analysis of U2AF^{26} over a 24-hour period in the mouse forebrain indicates that expression of the full length transcript does not change significantly, but the alternative splicing of the U2AF^{26} transcript fluctuates during the day:night cycle. Examination of U2AF^{26} alternative splicing in other tissues revealed that this splicing event is temporally-regulated in the liver, but with a two-peaked pattern of splicing. Further analysis of other alternative splicing events in the liver indicates that the polypyrimidine tract binding (PTB) transcript is regulated in a similar manner. The two-peaked pattern of splicing in the liver suggests that the alternative splicing of U2AF^{26} and PTB is not regulated by circadian rhythm. However, this is the first time it has been observed that pre-mRNA splicing changes as a function of the day:night cycle.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Shyng, S., Ferrigni, T., **Shepard, J**., Nestorowicz, A., Glaser, B., Permutt, M., and Nichols, C. Functional analyses of novel mutations in the sulfonylurea receptor 1 associated with persistent hyperinsulinemic hypoglycemia of infancy. *Diabetes. 47(7):1145-51, 1998 Jul.*

Shepard, J., Reick, M., Olson, S., and Graveley, B. Characterization of $U2AF^{26}$, a splicing factor related to $U2AF^{35}$. *Molecular & Cellular Biology. 22(1):221-30, 2002 Jan.*

LIST OF FIRGURES

LIST OF ABBREVIATIONS

ASLV – Avian Sarcoleukosis Virus

bp – Base Pair(s)

BPS – Branch Point Sequence

BrdU – 5-Bromodeoxyuridine

cDNA – Complementary DNA

DNA – Deoxyribonucleic Acid

dsx – doublesex

*dsx*RE – *doublesex* Repeat Element

ESE – Exonic Splicing Enhancer

EST – Expressed Sequence Tag

GST – Glutathione S-Transferase

His – Histidine Protein Tag

hnRNP – Heterogeneous Nuclear Ribonucleoprotein Particle

kDa - Kilodalton

mRNA - Messenger RNA

MS2 – Bacteriophage RNA Binding Protein

Ni-NTA – Nickel NTA Agarose

NMD – Nonsense Mediated Decay

PCR –Polymerase Chain Reaction

pre-mRNA – Pre-messenger RNA

PTC – Premature Termination Codon

R – Purine Nucleotide

RNA - Ribonucleic Acid

RNAi – RNA interference

RT-PCR – Reverse Transcription PCR

RUST – Regulated Unproductive Splicing and Translation

RRM – RNA Recognition Motif

ΨRRM – Pseudo RRM

RNP cs – Ribonucleoprotein Consensus Sequence

RS – Arginine- and Serine-Rich Motif

SDS PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SELEX - Systematic Evolution of Ligands by Exponential Enrichment

Sf9 – *Spodoptera frugiperda* Cells

siRNA – Short Interfering RNAs

snRNA – Small Nuclear RNA

snRNP – Small Nuclear Ribonucleoprotein Particle

SR protein – Splicing Factor with N-terminal RRM and C-terminal RS Domain

Y – Pyrimidine Nucleotide

CHAPTER ONE
Introduction

The sequencing of several metazoan genomes revealed that differences in the complexity of metazoan organisms is not due to the number of genes contained within a genome. The increase in complexity of an organism must be due to a mechanism that increases the complexity of the proteins expressed from the respective genome (known as the proteome). There are several regulatory mechanisms that would allow for enhanced complexity of the proteome, but the most powerful is alternative splicing (Maniatis and Tasic, 2002). Alternative splicing arises from differential recognition of splice sites within a transcript. Nearly every example of alternative splicing is regulated during splice site selection and understanding the mechanism of the regulated of splice site selection will be critical for furthering our knowledge of alternative splicing.

Role of cis-acting elements in nuclear splicing. Nuclear splicing involves catalytic removal of intervening sequences (introns) from nascent pre-mRNA transcripts (Black and Grabowski, 2003). Removal of the intron requires specific interactions between sequence elements at the 5' and 3' boundaries of the intron and an ensemble of proteins. The 5' end of the intron, known as the 5' splice site, is defined by the AG/GURAG consensus sequence (R-purine nucleotide and / denotes the exon-intron boundary) (Fig. 1.1). The 3' end, termed the 3' splice site, is characterized by the YAG/R motif (Y-pyrimidine nucleotide). The pyrimidine tract, a stretch of 15-25 nucleotides enriched with pyrimidine residues, immediately precedes this motif. Further upstream of the acceptor

site, typically 18-40 nucleotides, is the branch point sequence (BPS). The BPS is characterized by a highly degenerate sequence that invariably contains an adenine residue and roughly follows the consensus YNCURAY. (Burge, 1999).

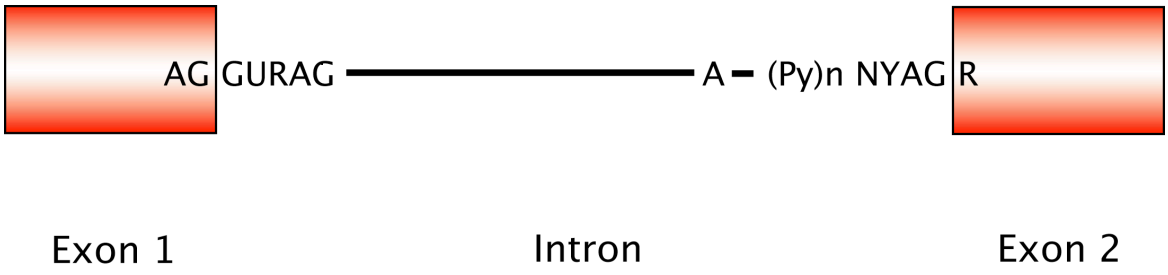


Figure 1.1: The exon-intron structure and the sequence motifs of the 5' and 3' splice sites. The upstream end of the intron contains the 5' splice site sequence motif AG/GURAG. The downstream end, or the 3' splice site, contains the branch point adenine, the pyrimidine tract, and the YAG/R motif.

Splicing of an intron is directed by a large, multicomponent RNA-protein complex known as the spliceosome. The catalytic removal of the intron occurs during two transesterification steps (Fig. 1.2). First, the 2'-hydroxyl group of the branch point adenine residue attacks the 3'-5' phosphodiester bond of the guanosine nucleotide at the 5' end of the intron. At the end of this step, the 5' end of the intron is cleaved from the upstream exon and covalently linked to the adenine. The second step is initiated when the free 3'-hydroxyl group from the excised exon attacks the phosphodiester bond at the 3' end of the intron. This releases the intron as a free lariat and produces an RNA with the two ligated exons (Black and Grabowski, 2003).

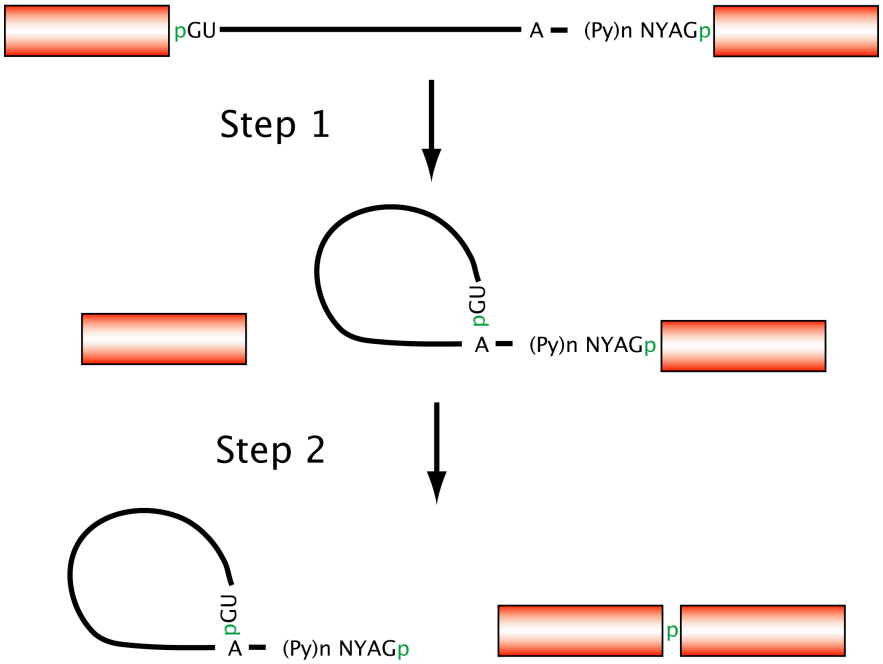


Figure 1.2: The two steps of splicing. During the first step the 2' hydroxyl of the branch adenine attacks the 3'-5'phosphodiester bond at the upstream exon-intron boundary, which causes the 5' end of the intron to be covalently linked to the branch point adenine. The second step begins when the free 3'-hydroxyl group from the excised exon attacks the phosphodiester bond of the 3' end of the intron. This produces a free lariat along with the ligated exons.

Spliceosome assembly. The formation of the spliceosome requires five small nuclear ribonucleoprotein particles (snRNPs) and several hundred non-snRNP proteins. Recent estimates suggest that roughly 300 proteins are involved in this process. (Zhou et al., 2002), (Rappsilber et al., 2002). In the earliest step of the spliceosome, the 5' sequence motif is recognized by the U1 snRNP, and this interaction is mediated by direct base pairing between the 5' splice site and the complementary sequence found within the U1 snRNA (Fig. 1.3). The 3' splice site is recognized by a complex containing the U2 auxiliary factor (U2AF) and splicing factor 1 (SF1). U2AF is an essential splicing factor composed of two subunits, $U2AF^{65}$ and $U2AF^{35}$, and together they recognize the

pyrimidine tract and YAG/R sequence respectively. The SF1 protein has been demonstrated to recognize the branch point sequence. Together, the interaction of the U1 snRNP with the 5' splice site and U2AF/SF1 complex association with the 3' splice site of the intron, define the first step in spliceosome assembly, the early (E) complex (Fig. 1.3, E complex) (Das et al., 2000).

The next step in spliceosome formation is the A complex. At this stage the SF1 protein is released and then the U2AF heterodimer recruits the U2 snRNP to the branch point sequence. The U2 snRNA anneals to the branch point, however this interaction is imperfect, causing the branch point adenine to bulge from the short helix created from the interaction (Fig 1.3, A complex). Next, this complex is joined by the U4/U6.U5 tri-snRNP to form the B complex and during the tri-snRNP association, the U2AF heterodimer is released from the 3' splice site (Fig 1.3, B complex). Upon extensive conformational rearrangement of the bound snRNPs and release of the U1 and U4 snRNP, the spliceosome enters the C complex and initiates the catalysis necessary for the removal of the intron.

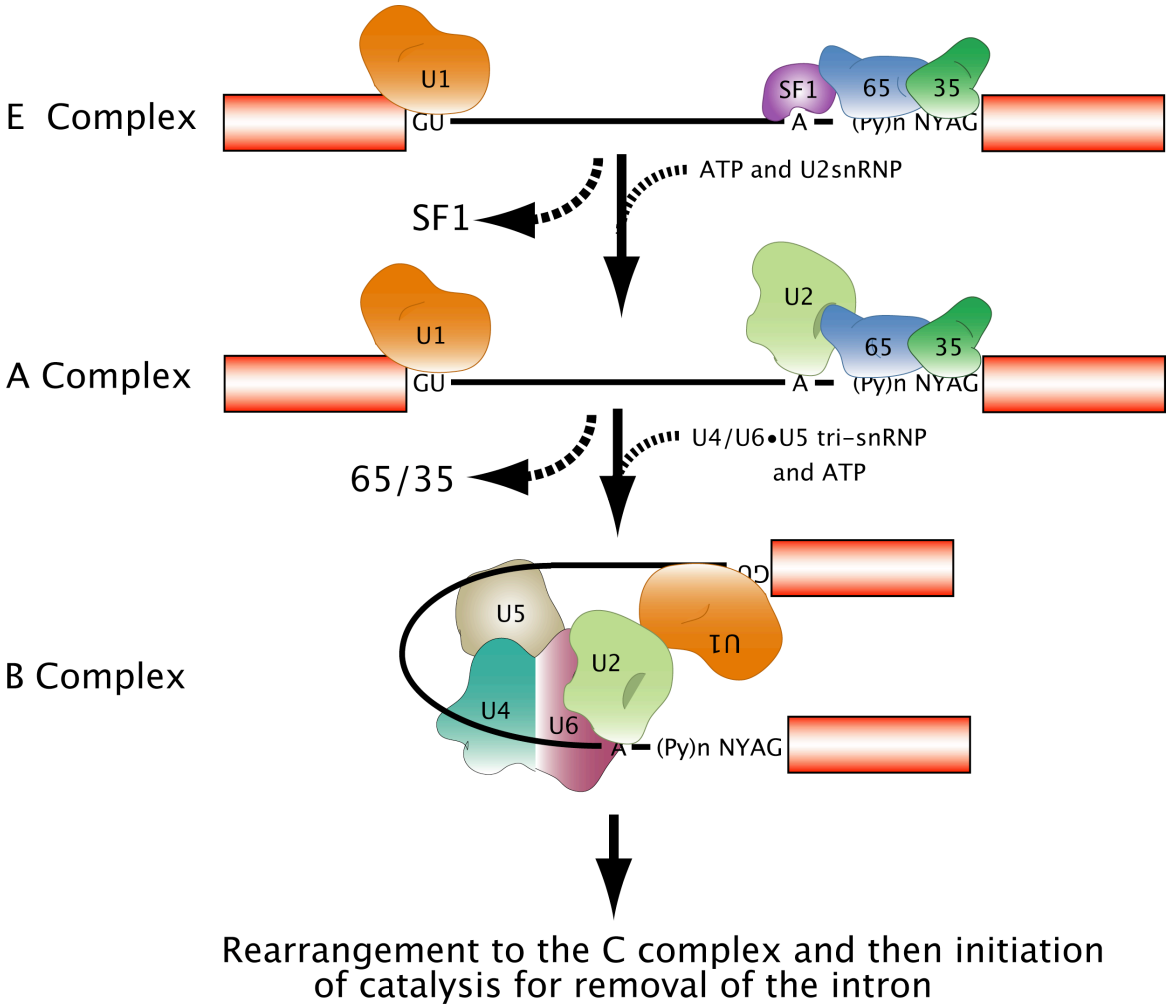


Figure 1.3: Different steps of spliceosome assembly. During the E complex, the U1 snRNP recognizes the 5' splice site and the U2AF/SF1 complex recognizes the 3' splice site. Next, SF1 is released from the branch point and U2AF^{65} recruits the U2 snRNP to the branch point sequence. This stage of spliceosome formation is called the A complex. This complex becomes the B complex with the loss of U2AF and the addition of the U4/U6.U5 tri-snRNP. Upon extensive rearrangement of the tri-snRNP and loss of the U1 and U4 snRNP, the complex becomes the C complex and is poised to initiate the catalysis necessary for the removal of the intron.

SR proteins. A class of splicing factors involved in spliceosome assembly is the arginine- and serine- rich (RS) domain proteins (SR proteins). These proteins are involved in every aspect of constitutive splicing and play a role in several examples of alternative splicing. SR proteins are found in all metazoan species studied and ten of

these proteins have been identified in humans. These proteins are not found in *Saccharomyces cerevisiae*, which is surprising given their critical role in constitutive splicing in metazoans. Since *S. cerevisiae* have very short introns, the absence of these proteins might be indicative of a different splicing mechanism in this organism.

SR proteins share a similar domain structure in that they have one or two amino-terminal RNA recognition motifs (RRM) and a carboxyl-terminal RS domain. These two domains are modular and can function independently of one another. Interestingly, the RRMs of each SR protein recognizes a specific sequence(s), and the consensus binding site of several SR proteins have been characterized. The sequences recognized by these proteins are typically 6-10 nucleotides in length and independent of secondary structure (Tacke and Manley, 1999).

The RS domain mediates protein-protein interaction between SR proteins and other splicing factors. The RS domains can also be exchanged between different SR proteins (Chandler et al., 1997; Wang et al., 1998). A study done by Graveley and Maniatis demonstrated that the activity of these domains differ from one another quantitatively (Graveley et al., 1998). In this experiment, the RS domains of six different SR proteins were tested by fusing each domain to the bacteriophage RNA binding protein MS2. The MS2 binding site was introduced downstream of a 3' splice site of a model template and each RS domain was then scored for the ability to stimulate splicing of the template. From this analysis, it was determined that the activity of the RS domain is directly related to the number of RS dipeptides present in the domain and that the RS domains differ quantitatively.

In vivo analyses of SR proteins suggests that some of these splicing factors may not have an equivalent role in metazoan splicing. For example, targeted disruption of the SRp20 gene in mice prevented embryos developing past the blastocyst stage (Jumaa et al., 1999). Mice homozygous for the SC35 null allele also have an embryonic lethal phenotype (Wang et al., 2001). Deletion of the *D. melanogaster* gene B52, a homolog of the human SRp55 gene, leads to a developmental defect as well, resulting in first- and second-instar lethality (Ring and Lis, 1994). Targeted reduction of SF2/ASF by RNA interference (RNAi) in *Caenorhabditis elegans* leads to an embryonic lethal phenotype. However, targeted reduction of six other SR proteins in *C. elegans* did not lead to any observable phenotype. Different combinations of targeted knock-down of those six SR proteins led to different phenotypes. This indicates that SR proteins are essential for certain development stages of metazoan organisms, but there may also be functional redundancy of certain SR proteins (Longman et al., 2000).

SR function during splice site selection. SR proteins have a diverse set of functions throughout the formation and processing of the spliceosome. One role for SR proteins includes 5' and 3' splice site recognition. In general, vertebrate genes are composed of several small exons separated by very large introns. The average human exon is about 150 nucleotides (Maniatis and Tasic, 2002), whereas introns average 3500 nucleotides (Deutsch and Long, 1999) and can be up to 500,000 nucleotides long (Rowen et al., 2002). This spacing suggests that 5' and 3' splice site recognition and pairing across introns is difficult to achieve. The exon-definition model proposes the 5' and 3' splice

sites are recognized across exons (Fig. 1.4) (Berget, 1995). According to this model, proteins that recognize 5' and 3' splice sites form a protein bridge, linked via protein-protein interactions, across the exon (Fig. 1.4) (Berget, 1995). This interaction is mediated by SR proteins bound to exonic splicing enhancer (ESE) within the exon (Shi et al., 1997; Watakabe et al., 1993), and it is thought that the majority of the constitutively spliced exons follow this model (Reed, 1996).

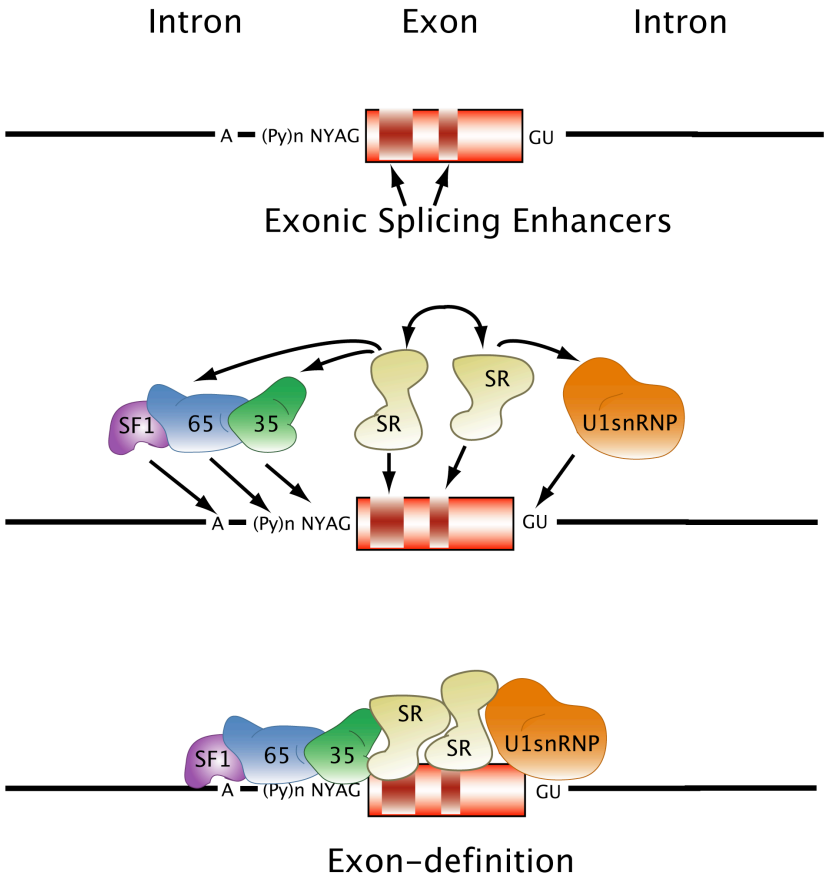


Figure 1.4: Diagram of the exon-definition model. The U1snRNP recognizes the 5' splice site and the U2AF/SF1 protein complex binds to the 3' splice site. This occurs in conjunction with the binding of SR proteins to the exonic splicing enhancers. The interaction between the splicing factors across the exon creates a synergy between the proteins, thereby strengthening their interaction with the RNA.

SR proteins also assist the binding of splicing factors to regulated splice sites. Regulated splice sites are typically sub-optimal and are not efficiently recognized by essential splicing factors. For a weak splice site, binding of SR protein(s) to the ESE is crucial for stimulating recognition of the splice site by the necessary splicing factors. The 3' splice site of the *dsx* pre-mRNA is an example of a weak 3' splice site that requires SR proteins for recognition. There are six copies of the *dsx* repeat element (*dsx*RE) downstream of this splice site. As mentioned previously, the heterodimeric complex of RBP1, Tra, and Tra2 recognize this element and stimulates 3' splice site recognition (Lynch and Maniatis, 1996). These SR proteins play an essential role in the recognition of this regulated 3' splice site.

SR proteins can play a role in recognition of regulated 5' splice sites. This has been observed for the regulation of the *fruitless* pre-mRNA. Upstream of the regulated 5' splice site, there are three copies of the *dsx*RE enhancer element. Again, this element binds the complex of RBP1, Tra, and Tra2 and this complex stimulates recognition of the 5' splice site for spliceosome assembly on the *fruitless* pre-mRNA (Ryner et al., 1996). Together these examples describe how SR proteins can play a role in recognition of constitutive and regulated splice sites.

Phosphorylation of the RS domain affects the function and localization of SR proteins. The phosphorylation state of this domain affects the ability of SR proteins to interact with other SR proteins (Wang et al., 1998; Xiao and Manley, 1998). Also, the phosphorylation state affects the sub-cellular localization of SR proteins. The hypo-phosphorylated proteins have been shown to be localized to nuclear speckles and then

recruited to active sites of transcription following phosphorylation of the RS domain (Spector et al., 1991).

SR protein kinases phosphorylate the serine residue of the RS dipeptide repeats within the RS domain. The kinases that phosphorylate RS domains *in vitro* include SR protein kinase 1 (SRPK1) (Gui et al., 1994), SRPK2 (Wang et al., 1998), the Clk/Sty family (Colwill et al., 1996), and DNA topoisomerase I (Rossi et al., 1996). *In vivo* analysis of SR protein kinases demonstrated a role for phosphorylation in the regulation of specific alternative splicing events. The *Drosophila melanogaster* homolog of Clk/Sty, *Doa*, has been shown to phosphorylate the *Drosophila* SR protein RBP1 and the splicing factors Tra and Tra2. As mentioned earlier, these three proteins influence the splicing of an 3' splice site of the *doublesex* (*dsx*) pre-mRNA and also regulate recognition of a 5' splice site of another pre-mRNA called *fruitless*. Interestingly, mutations of *Doa* affect the splicing of the *doublesex* 3' splice site pre-mRNA, but not the splicing of the alternative 5' splice site of the *fruitless* pre-mRNA. This suggests that phosphorylation of the RS domain of splicing factors is essential for specific splicing events, but is not required for other splicing events (Du et al., 1998).

SR protein functions outside of splice site recognition. SR proteins have roles outside of splice site recognition. SR proteins are thought to participate in pairing of 5' and 3' splice sites across an intron. It has been shown that the SR protein SC35 and SF2/ASF can simultaneously interact with the U1 and U2 snRNP across an intron (Fu and Maniatis, 1992; Wu and Maniatis, 1993). This role for SR proteins in spliceosome formation has

not been proven, but these observations suggest that they could be involved in the juxtaposition of splice sites across an intron.

Another role for SR proteins might be facilitating the interaction between the U4/6.U5 tri-snRNP and the proteins bound to the RNA during the B complex (Roscigno and Garcia-Blanco, 1995), (Tarn and Steitz, 1995). There are two non-snRNP proteins associated with the tri-snRNP, U4/6.U5 tri-snRNP 27K and U5-100K, that have RS domains and these proteins may help mediate interactions between the SR proteins and the tri-snRNP. These examples suggest that SR proteins could participate in other important steps outside of splice site recognition during the formation of the spliceosome.

There are other splicing factors, different from SR proteins, that contain an RS domain. These proteins, SR-related proteins (SRrps), have varied and distinct roles in pre-mRNA metabolism. The SRrps family includes splicing regulators Tra and Tra2, snRNP factors U1-70K, U5-100K, U4/U6.U5-27K, and hLuc7p, splicing coactivators SRm160 and SRm300, RNA helicases hPrp16 and HRH1, protein kinases Clk/Sty 1, 2, and 3, and the U2 Auxiliary Factor (U2AF) subunits $U2AF^{65}$ and $U2AF^{35}$ (reviewed in (Fu, 1995; Graveley, 2000)). This diverse group of proteins play a role in constitutive splicing and/or alternative splicing.

Identification and characterization of the large U2AF subunit. U2AF is an essential heterodimeric splicing factor required for 3' splice site recognition during spliceosome formation. The U2AF heterodimer was identified biochemically by its essential activity of assisting U2 snRNP binding to the branch point of the 3' splice site (Ruskin et al.,

1988). The U2AF heterodimer stably associates with the 3' splice site, in an ATP- and temperature-independent manner (Zamore and Green, 1989), and recruits the U2 snRNP to the branch point (Fig. 1.3, E complex). The requirement for U2AF for U2 snRNP binding is not surprising given that the metazoan branch point sequence is highly degenerate.

Further biochemical analysis led to the cloning and characterization of the large and small subunits of the U2AF heterodimer. The human $U2AF^{65}$ large subunit, $U2AF^{65}$, was cloned and characterized by Green and colleagues (Zamore et al., 1992). They determined that $U2AF^{65}$ is a functional splicing factor and it alone can support *in vitro* splicing of a model template. Orthologs of $U2AF^{65}$ have been observed in *Mus musculus* (Sailer et al., 1992), *C. elegans* (Zorio et al., 1997), *D. melanogaster*, (Kanaar et al., 1993), and *Schizosaccharomyces pombe* (Potashkin et al., 1993). Deletion of the large subunit in *S. pombe* also results in lethality (Potashkin et al., 1993). A deletion mutation of the large subunit in *D. melanogaster* is fully penetrant recessive lethal, and can be complemented by a genomic transgene containing the large subunit gene (Kanaar et al., 1993). This deletion can be complemented by the transgene containing human $U2AF^{65}$, suggesting that there is functional conservation between the two subunits.

Initial characterization of $U2AF^{65}$ revealed that it is present in the A complex of spliceosome formation (Zamore and Green, 1991). Binding analysis demonstrated that recombinant $U2AF^{65}$ binds to the pyrimidine tract of the 3' splice site with high affinity (Zamore et al., 1992). Analysis of the optimal binding site using systematic evolution of ligands by exponential enrichment (SELEX) identified the consensus site of $U2AF^{65}$ as

UUUUUUu/cCCc/uUUUUUUUUcc (Singh et al., 1995), providing more evidence that U2AF^{65} preferred pyrimidine rich regions.

Green and colleagues identified three RRM motifs and one N-terminal RS domain within the U2AF^{65} protein (Zamore et al., 1992). Deletion of the RRMs individually from the protein reduced the binding affinity for pyrimidine rich regions one hundred to five thousand times less than the full length protein, indicating that all three RRMs are required for efficient binding. Removal of the RS domain alone had no affect on the binding affinity of the recombinant protein for the model substrate tested. This experiment suggested that the RS domain was not required for substrate binding, yet each RRM is required for binding.

In vitro splicing assays using nuclear extract depleted of endogenous U2AF could not be efficiently complemented with recombinant U2AF^{65} lacking the RS domain. This observation suggested that the RS domain is required for the splicing activity of U2AF^{65}, but not necessarily for binding (Zamore et al., 1992). The RS domain was later implicated in assisting the U2 snRNP binding to the branch point sequence. It was proposed that the basic amino acids in the RS domain could assist in the direct base pairing between the U2 snRNP and branch point sequence (Valcarcel et al., 1996). Also, further analysis of the interaction between the U2 snRNP and U2AF^{65} demonstrated that the third RRM of the U2AF^{65} could interact with a protein, SAP 155, which was associated with the U2 snRNP (Gozani et al., 1998). Therefore, U2AF^{65} can potentially recruit in U2 snRNP through protein-protein interactions and then help stabilize in the interaction the U2 snRNA and the BPS.

Identification and analysis of the small subunit of U2AF. The human U2AF small subunit, $U2AF^{35}$, was cloned and characterized by Zhang *et al.* (Zhang et al., 1992). $U2AF^{35}$ has an N-terminal pseudo-RRM (ΨRRM) and a C-terminal RS domain interrupted by twelve consecutive glycine residues. Further analysis indicated that $U2AF^{35}$ stably associates with $U2AF^{65}$ in a 1:1 stoichiometry (Zamore and Green, 1989) and the region responsible for this interaction localized to the ΨRRM. There are two conserved zinc fingers on either side of the ΨRRM with the $\mathbf{C}X_8\mathbf{C}X_5\mathbf{C}X_3\mathbf{H}$ motif (the bold residues coordinate the Zn^{2+} metal ion) (Amann et al., 2003). Mutational analysis of the two zinc fingers of the small subunit in *S. pombe* revealed that each motif contributes to RNA binding independent of one another and each motif is essential for viability (Webb and Wise, 2004).

The crystal structure of the interaction of between the $U2AF^{65}$ and $U2AF^{35}$ was recently determined (Kielkopf et al., 2001). This analysis revealed a novel "tongue in groove" protein-protein interface responsible for the interaction between the two splicing factors. Also, the ΨRRM of $U2AF^{35}$ was carefully analyzed. This binding motif is termed a pseudo RRM because the length of the α-helix A deviates from canonical RRMs. This α-helix is 10 residues in length for the canonical RRMs of other RNA binding proteins, but it is 30 residues long in $U2AF^{35}$. Furthermore, all $U2AF^{35}$ orthologs have an α-helix A that deviates from the strict length of the helix of RRMs, suggesting that there is a functional significance to this unique characteristic of $U2AF^{35}$. Analysis of the amino acid composition of this structure reveals that the surface of the helix is coated

with negatively charged residues. This region of the protein might interact with the basic regions of RS domains found on SR proteins (Kielkopf et al., 2001).

The functional significance of the small subunit is further illustrated by the phylogenetic conservation of this subunit between different organisms. Orthologs of the small subunit have been identified in humans (Zhang et al., 1992), *C. elegans* (Zorio and Blumenthal, 1999b), *D. melanogaster* (Rudner et al., 1996), and *S. pombe* (Wentz-Hunter and Potashkin, 1996). Genetic analysis in *S. pombe* (Wentz-Hunter and Potashkin, 1996) and RNAi-mediated reduction analysis in *C. elegans* (Zorio and Blumenthal, 1999b) of the small subunit demonstrate a lethal phenotype, indicating that this splicing factor is required for viability.

Rudner and colleagues demonstrated that a mutation that disrupts the expression of the *Drosophila* small subunit causes a recessive lethal phenotype that can be rescued with a transgene carrying the small subunit gene (Rudner et al., 1996). A follow-up investigation revealed that the same mutation could be complemented with transgenes missing all or part of the RS domain (Rudner et al., 1998b) without a significant decrease in viability. Also, analysis of the rescued progeny, with the different deletions of the RS domain, indicated that there were no changes in constitutive or regulated splicing (Rudner et al., 1998b). This study provides strong evidence suggesting that the RS domain of $U2AF^{35}$ is not necessary for the regulation of alternative splicing events. Given the conservation of the small subunit and the requirement for viability, these observations indicate that the protein performs an essential function, however, the function of this protein remained elusive.

U2AF^{35} can assist U2AF^{65} with binding to pre-mRNAs. Biochemical analysis of the Drosophila U2AF heterodimer demonstrated that the small subunit significantly contributes to the high-affinity binding of the heterodimer (Rudner et al., 1998a). First, Rudner *et al*. demonstrated that small subunit alone had minimal RNA binding activity and that it could bind RNA with the 3'splice site sequence motifs. Also, binding experiments using RNA with a pyrimidine tract and the AG dinucleotide from the 3' splice site of a model template indicated that the small subunit increases the binding potential of the large subunit by twenty-fold. If the pyrimidine tract of the 3' splice site was mutated, by changing several pyrimidine nucleotides to purine residues, the dependence of large subunit on the small subunit for binding increased. This binding experiment was repeated using recombinant human U2AF^{65} and U2AF^{35}, and the human large subunit also required the human small subunit for efficient binding. This was the first time it was suggested that the small subunit could assist the large subunit with binding to a 3' splice site sequence through association with the RNA (Rudner et al., 1998a).

The basis of the U2AF^{35}-mediated assistance in substrate binding was conclusively demonstrated by studies showing that the small subunit specifically recognized the AG dinucleotide of the 3'splice site (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999a). This elucidated the function of the small subunit in 3' splice site recognition and explained how it could assist in U2AF^{65} binding. Further functional analysis of the heterodimer demonstrated that U2AF^{35} and U2AF^{65} could enhance the splicing efficiency of a substrate with a weak pyrimidine tract better than U2AF^{65} alone.

This type of intron was termed AG-dependent because of the need for the small subunit binding to the AG for efficient splicing (Wu et al., 1999).

Another study demonstrated that U2AF^{35}, when complexed with U2AF^{65}, could enhance the splicing efficiency of the AG-dependent substrate, IgM. Surprisingly the efficiency of the splicing was affected by the nucleotide immediately downstream of the AG dinucleotide (Guth et al., 2001). In the wild type IgM template, the AG is followed by a guanine. If the guanine at this position is replaced with a cytosine, then the splicing efficiency of the template was reduced. Further analysis revealed that U2AF^{35} binding was reduced for this template, suggesting that the reduction in splicing efficiency is due to a reduction in the U2AF^{35} biding affinity. This experiment demonstrated that U2AF^{35} can distinguish between sequence compositions around the AG dinucleotide (Guth et al., 2001).

Multiple forms of the small subunit in mammals. There have been four variants of the conventional U2AF^{35} described in mammals (Fig. 1.5). There could potentially be five functionally different heterodimers if each small subunit has a unique function. The difference in function could be conferred by two means, which specific sequences each unique small subunit recognizes around the AG dinucleotide and differences in protein-protein interactions.

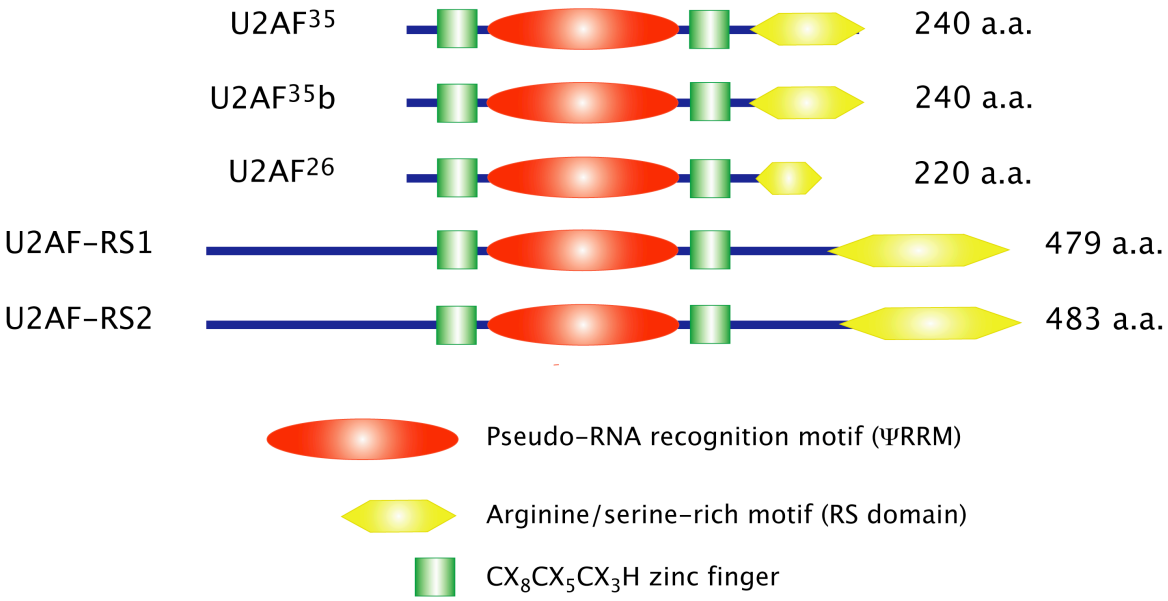


Figure 1.5: Alignment of U2AF^{35} and the four small subunit variants. All five proteins have the domains thought to be important for the function of U2AF^{35}. They all have the ΨRRM motif, the two zinc fingers on either side of the ΨRRM, and an C-terminal RS domain.

Recently it was reported that the *U2AF^{35}* gene has a mutually exclusive exon for exon 3, an exon observed in both mice and humans (Pacheco et al., 2004). When this exon is included in the transcript, it encodes a protein that is similar to U2AF^{35}, called U2AF^{35}b. Within this exon there are two motifs important for the function of the small subunit. First there is the RNP-2, which is predicted to partially mediate the interaction between the ΨRRM and RNA (Kielkopf et al., 2001). There are two amino acid substitutions within this hexamer motif in the alternative exon relative to the other exon. One significant substitution is the leucine amino acid at position 47 compared with the alanine residue found in the wild-type exon. This amino acid is predicted to directly contact the RNA (Kielkopf et al., 2001), therefore this change could affect the binding specificity of U2AF^{35}b.

The other motif different in the alternative exon is the α-helix A. The exon is predicted to encode only 4 amino acids of the helix (Pacheco et al., 2004), but two of these for amino acids are different. There is a serine and histidine at positions 65 and 66 respectively in the alternative exon and there is a leucine and argnine at the same positions in the wild-type exon. These two amino acids are on the surface of the α-helix, and the amino acids of the alternative exon could be phosphorylated by a specific kinase, thereby altering the surface charge of the helix and potentially affecting the function of the U2AF^{35}b.

Previously, other genes related to *U2AF^{35}* have been described, called *U2AF1-RS1* and *U2AF1-RS2* (Tronchere et al., 1997). U2AF1-RS2 can dimerize with U2AF^{65} and interact with SR proteins. U2AF1-RS1 has not been well characterized, but it shares 94% amino acid identity with U2AF-RS2, so it may function similar to U2AF-RS2.

A paralog of *U2AF^{35}*, *U2AF^{26}*, has been described, and in this study we describe our efforts to understand the differences between U2AF^{26} and U2AF^{35}. We sought to determine if there are functional differences between the two small subunits by comparing their function with *in vitro* splicing assays. Also, we examined the binding specificity of U2AF^{26} and U2AF^{35}. Finally we examined the regulation of the alternative splicing of the U2AF^{26} transcript in an effort to understand the *in vivo* function of the splicing factor.

CHAPTER TWO
Characterization of the U2AF^{26}, a Splicing Factor Related to U2AF^{35}

INTRODUCTION

Introns are removed from nascent transcripts by the process of pre-mRNA splicing. This reaction is catalyzed by the spliceosome, a large, multicomponent RNA-protein machine, which assembles on each intron targeted for removal (Burge, 1999). Spliceosome assembly is directed by conserved sequences within introns and at intron-exon boundaries. In mammals, these signals consist of a single sequence at the 5' splice site (AG/GURAG) and a set of three sequences at the 3' splice site (Burge, 1999). The latter consists of the branch point, a pyrimidine tract of ~20 nucleotides, and the sequence YAG/R that defines the end of the 3' splice site. The first step in spliceosome assembly, the formation of the E complex, involves the binding of U1 snRNP to the 5' splice site, SF1/mBBP to the branchpoint, and U2 snRNP Auxiliary Factor (U2AF) to the pyrimidine tract and YAG/R (Moore, 2000; Reed, 2000). In addition, U2 snRNP is associated with the spliceosome at this point, but not stably bound to the pre-mRNA (Das et al., 2000). Subsequently, U2 snRNP replaces SF1 at the branchpoint to form the A complex (Reed, 2000). Understanding the details of these early events in spliceosome assembly is important because alternative splicing is frequently regulated during these steps (Smith and Valcarcel, 2000).

U2AF is a heterodimeric complex conserved in all metazoan species. In mammals, this complex consists of 65-kDa (U2AF^{65}) and 35-kDa (U2AF^{35}) subunits (Zamore and

Green, 1989). U2AF^{65} contains an N-terminal arginine/serine (RS)-rich domain and three RNA recognition motifs (RRM) (Zamore et al., 1992). U2AF^{65} is an essential splicing factor *in vitro* and is required for viability in *Drosophila* (Kanaar et al., 1993), *C. elegans* (MacMorris et al., 1999) and *S. pombe* (Potashkin et al., 1993). U2AF^{65} interacts directly with the pyrimidine tract and is involved in stabilizing the interaction of U2 snRNP with the branchpoint (Gozani et al., 1998; Valcarcel et al., 1996). This activity requires the RS domain, which is thought to assist in the formation of the U2 snRNP:pre-mRNA duplex (Valcarcel et al., 1996), and the third RRM, which interacts with SAP 155, a component of U2 snRNP (Gozani et al., 1998).

The role of U2AF^{35} in pre-mRNA splicing has been more controversial than that of U2AF^{65}. U2AF^{35} contains two zinc fingers and a non-canonical RRM (Birney et al., 1993) in the N-terminus of the protein and an RS domain and a glycine tract at the C-terminus (Zhang et al., 1992). The precise role of U2AF^{35} in pre-mRNA splicing has been unclear because some studies have shown that U2AF^{35} is dispensable for splicing *in vitro* (Kan and Green, 1999; Zamore and Green, 1991), while other studies indicate that U2AF^{35} is an essential splicing factor *in vitro* (Guth et al., 1999; Zuo and Maniatis, 1996). Nonetheless, the small U2AF subunit is required for viability in *Drosophila* (Kanaar et al., 1993) and *C. elegans* (Zorio and Blumenthal, 1999b), as is its interaction with U2AF^{65} (Rudner et al., 1998c). One function of U2AF^{35} in pre-mRNA splicing was recently clarified when three groups demonstrated that U2AF^{35} associates with the AG at the 3' splice site (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999b).

In addition to constitutive splicing, U2AF^{35} appears to play an important role in

regulated splicing. The removal of many introns requires cis-acting sequences present in the downstream exon called exonic splicing enhancers (ESEs) (Blencowe, 2000; Graveley, 2000). Many ESEs are recognized by SR proteins – a conserved family of essential splicing factors (Blencowe, 2000; Graveley, 2000). Several studies have shown that enhancer-bound SR proteins function by recruiting U2AF^{65} to weak pyrimidine tracts (Bouck et al., 1998; Graveley et al., 2001; Wang et al., 1995; Zuo and Maniatis, 1996). It has been proposed that U2AF recruitment involves protein interactions between enhancer-bound SR proteins and the RS domain of U2AF^{35} (Wu and Maniatis, 1993; Zuo and Maniatis, 1996). However, other studies indicate that ESEs may function through a mechanism that requires neither U2AF recruitment (Kan and Green, 1999; Li and Blencowe, 1999), nor U2AF^{35} (Kan and Green, 1999). Interestingly, the small U2AF subunits from *C. elegans* (Zorio and Blumenthal, 1999b) and *S. pombe* (Wentz-Hunter and Potashkin, 1996) do not contain RS domains. Moreover, transgenic *Drosophila* expressing a version of dU2AF^{38} lacking its RS domain as the sole source of dU2AF^{38} are not only viable, but properly regulate the alternative splicing of the enhancer-dependent *doublesex* pre-mRNA (Rudner et al., 1998b). Thus, at least in non-vertebrate systems, the RS domain is not an essential feature of the small U2AF subunit.

Here we report the identification and characterization of a mammalian gene, *U2AF^{26}*, capable of encoding a protein product very similar to U2AF^{35}. The N-terminal portion of the U2AF^{26} polypeptide, which contains two zinc fingers, a non-canonical RRM, and a U2AF^{65} interaction domain, is 89% identical to U2AF^{35}. However, the C-terminal domain of U2AF^{26} is quite different from the C-terminal domain of U2AF^{35} –

most of the RS dipeptides, as well as the entire glycine tract present in U2AF^{35}, are absent in U2AF^{26}. We find that the relative levels of U2AF^{35} and U2AF^{26} mRNAs vary in different mouse tissues. U2AF^{26} is a nuclear protein that localizes to speckles and can physically associate with U2AF^{65}. Biochemical experiments demonstrate that U2AF^{26} functions as a pre-mRNA splicing factor and can functionally substitute for U2AF^{35} in constitutive splicing *in vitro*. Interestingly, U2AF^{26} can also replace U2AF^{35} in enhancer-dependent splicing, demonstrating that although a small U2AF subunit is necessary for efficient enhancer-dependent splicing, its RS domain is not. Finally, we show that U2AF^{26} enhances the binding of U2AF^{65} to weak 3' splice sites. These studies identify U2AF^{26} as a mammalian splicing factor that may participate in the regulation of alternative splicing.

RESULTS

Cloning of U2AF^{26} and its similarity to U2AF^{35}. A cDNA fragment encoding a protein similar to U2AF^{35} was identified in a screen for genes regulated by the transcription factor NPAS2 (Zhou et al., 1997). Using this cDNA fragment as a probe, two overlapping partial cDNAs were obtained from a mouse cDNA library and were used to assemble a full-length cDNA. This cDNA encodes a protein specifying a predicted molecular weight of 26 kDa that is 76% identical to mouse U2AF^{35}. We therefore designate this protein U2AF^{26}.

U2AF^{35} and U2AF^{26} share 89% primary amino acid sequence identity within the N-terminal 187 residues of the proteins (Fig. 2.1A). The N-terminal domain contains two

zinc fingers that are identical between the two proteins, as well as a non-canonical RNA recognition motif (RRM) (Birney et al., 1993; Kielkopf et al., 2001). The RNP-1 portion of the RRM is identical between U2AF^{35} and U2AF^{26} (Fig. 2.1A). However, the RNP-2 motif of U2AF^{26} contains two amino acid substitutions with respect to U2AF^{35} – an Ala to Val at position 47, and an Ile to Leu at position 51. The high degree of conservation of the non-canonical RRM suggests that U2AF^{26}, like U2AF^{35}, may have the ability to bind RNA (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999a). In addition, the majority of the residues in U2AF^{35} that directly interact with U2AF^{65} (Kielkopf et al., 2001; Zhang et al., 1992) are conserved in U2AF^{26}. For example, the crystal structure of the U2AF^{35}/U2AF^{65} heterodimer (Kielkopf et al., 2001) revealed that Glu80, Phe81, Glu84, Glu88, Arg133, and Phe135 of U2AF^{35} are in intimate contact with U2AF^{65}. Of these residues, only Glu80 is different in U2AF^{26} (changed to an Asn). Most importantly, Trp134, perhaps the most critical residue involved in the U2AF^{65}-U2AF^{35} interaction (Kielkopf et al., 2001), is conserved in U2AF^{26}. This suggests that U2AF^{26} most likely can physically interact with U2AF^{65}.

In contrast to the highly conserved N-terminus, the C-terminus of U2AF^{26} differs markedly from that of U2AF^{35}. The U2AF^{35} C-terminal domain contains an RS domain and a glycine tract (Zhang et al., 1992). U2AF^{35} orthologs have been cloned from *Drosophila* (Rudner et al., 1996), *C. elegans* (Zorio and Blumenthal, 1999b), and *S. pombe* (Wentz-Hunter and Potashkin, 1996). The glycine tract present in U2AF^{35} is conserved in both the *Drosophila* and *C. elegans* U2AF^{35} orthologs, dU2AF^{38} and CeU2AF^{35}, but not the *S. pombe* ortholog, SpU2AF^{23}. In addition, neither CeU2AF^{35} nor

SpU2AF[23] contain an RS domain, whereas dU2AF[38] does. The U2AF[35] RS domain contains 7 RS, 3 RE and 3 RD dipeptides which are all common features of RS domains contained in other proteins (Graveley, 2000). By contrast the C-terminal domain of U2AF[26] completely lacks a glycine tract, is fairly proline-rich, and contains only three RS, one RE, and no RD dipeptides. Moreover, none of the RS dipeptides in U2AF[26] are adjacent to one another. The RS domain of U2AF[35] has been shown to function as a nuclear localization signal (Gama-Carvalho et al., 2001) and is thought to participate in protein interactions with SR proteins (Wu and Maniatis, 1993). The sequence of the C-terminal domain of U2AF[26] suggests that both of these activities may be altered in U2AF[26].

Cloning of the *U2AF[26]* gene. To compare the genomic organization of the *U2AF[26]* gene with that of *U2AF[35]*, we cloned and sequenced the gene encoding U2AF[26] from a mouse genomic DNA library. The *U2AF[26]* gene is small, spanning only 2,059 bp (Fig. 2.1B). 5' RACE experiments suggest that transcription initiates approximately 50 bp upstream of the start codon. A poly(A) consensus signal resides 114 bp downstream of the stop codon. The gene contains 8 exons that when spliced together generate an mRNA of approximately 900 nucleotides. The exon-intron boundaries of the *U2AF[26]* gene are identical to those in the human *U2AF[35]* gene located on chromosome 21 (Lalioti et al., 1996).

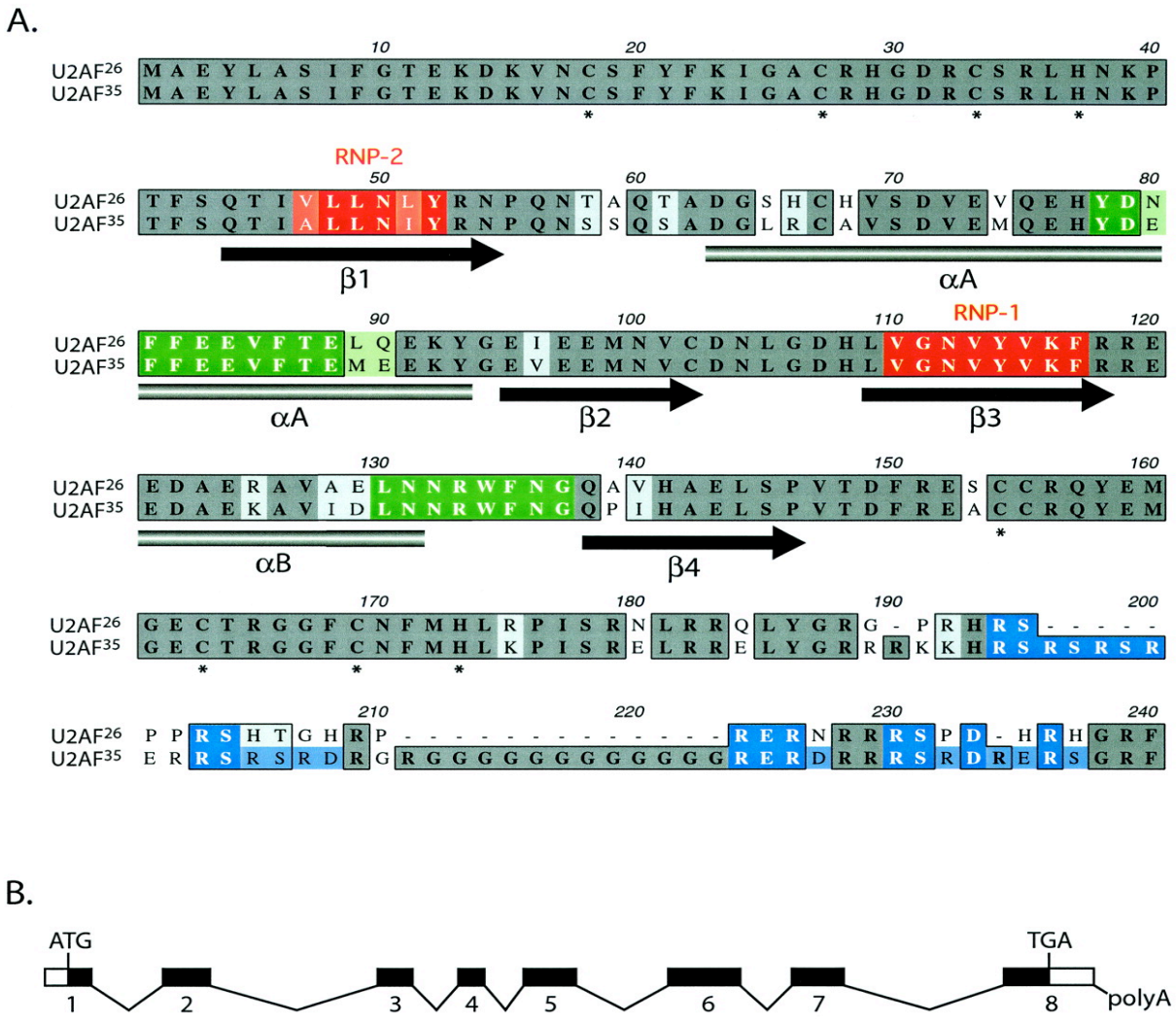


Figure 2.1: Sequence comparison of mouse U2AF^{26} and U2AF^{35} proteins and organization of the *U2AF^{26}* gene. (A) U2AF^{35} and U2AF^{26} share 89% sequence identity within the first 187 residues of the proteins. The dark grey boxes indicate identical amino acids and the light grey boxes indicate conserved amino acids. The location of the two zinc finger motifs are indicated with asterisks. The RNP1 and RNP2 motifs of the non-canonical RRM are highlighted in red. The amino acids involved in directly contacting U2AF^{65} are highlighted in green. In addition, RS, RE, or RD dipeptides in the C-terminus are highlighted in blue. (B) Organization of the mouse *U2AF^{26}* gene.

U2AF^{26} is differentially expressed. To examine the expression of U2AF^{26} in comparison to U2AF^{35}, we performed northern blots on poly(A)$^+$ RNA isolated from

different mouse tissues (Fig. 2.2A). Due to the fact that we could not correct for hybridization efficiency of the probes, we have expressed the results as a ratio of the $U2AF^{26}$ to $U2AF^{35}$ signal in each tissue, which was calculated after normalization to the cyclophilin control (see Materials and Methods). These experiments clearly demonstrate that the relative level of $U2AF^{26}$ and $U2AF^{35}$ mRNAs vary in different tissues (Fig. 2.2B). For example, the $U2AF^{26}$ to $U2AF^{35}$ ratio is highest in brain (~3) and lowest in liver (~0.5). Thus, $U2AF^{26}$ is differentially expressed in various mouse tissues.

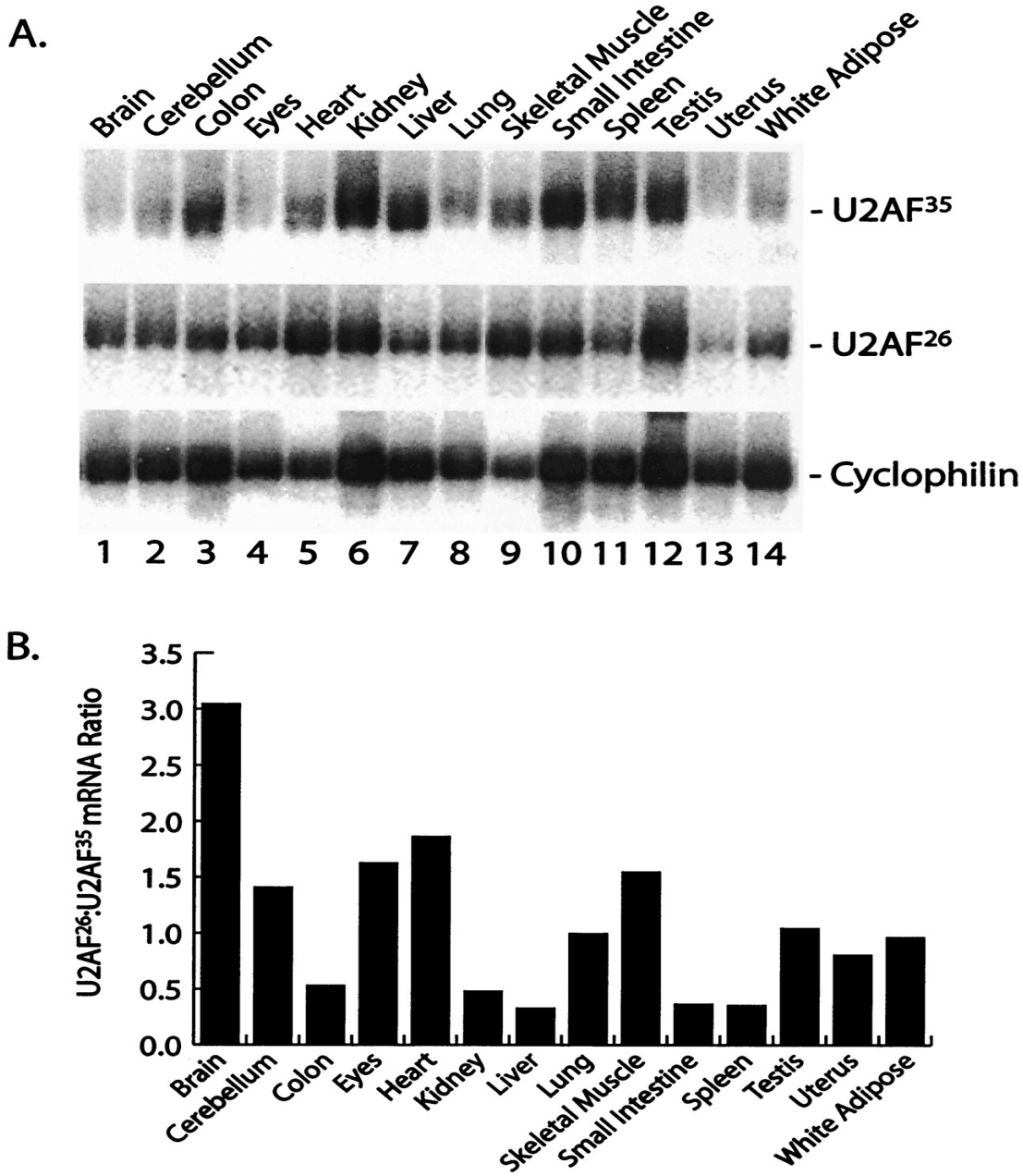


Figure 2.2: (A) Comparison of the expression patterns of U2AF^{26} and U2AF^{35} in mouse tissues. Northern blots containing poly(A)$^+$ RNA isolated from the indicated mouse tissues was hybridized with probes specific to the C-terminal domain of mouse U2AF^{35} (top), full-length mouse U2AF^{26} (middle), or a cyclophilin control (bottom). (B) The plot depicts the ratio of the U2AF^{26} to U2AF^{35} signals in each tissue that was calculated after normalization to the cyclophilin control as described in the Materials and Methods.

Subcellular localization of U2AF^{26}. To determine the subcellular localization of U2AF^{26}, HeLa cells were transfected with an expression vector encoding U2AF^{26} containing a C-terminal V5 epitope tag. As a control, HeLa cells were transfected with expression vectors encoding U2AF^{35}, or the SR protein SF2/ASF, each containing a C-terminal V5 epitope tag. Indirect immunofluorescence revealed that, as expected, U2AF^{35} (Gama-Carvalho et al., 2001) and SF2/ASF (Caceres et al., 1997) were exclusively localized to the nucleus (Fig. 2.3, bottom two rows). These two proteins were diffusely localized throughout the entire nucleus, but were excluded from the nucleolus, and were concentrated in a number of subnuclear speckles, a feature common to many splicing factors (Misteli and Spector, 1998). Likewise, U2AF^{26} was observed to be exclusively nuclear localized and associated with nuclear speckles (Fig. 2.3, top two rows). These results demonstrate that U2AF^{26}, like U2AF^{35}, is a nuclear protein concentrated in speckles, and thus is properly localized to function as a pre-mRNA splicing factor.

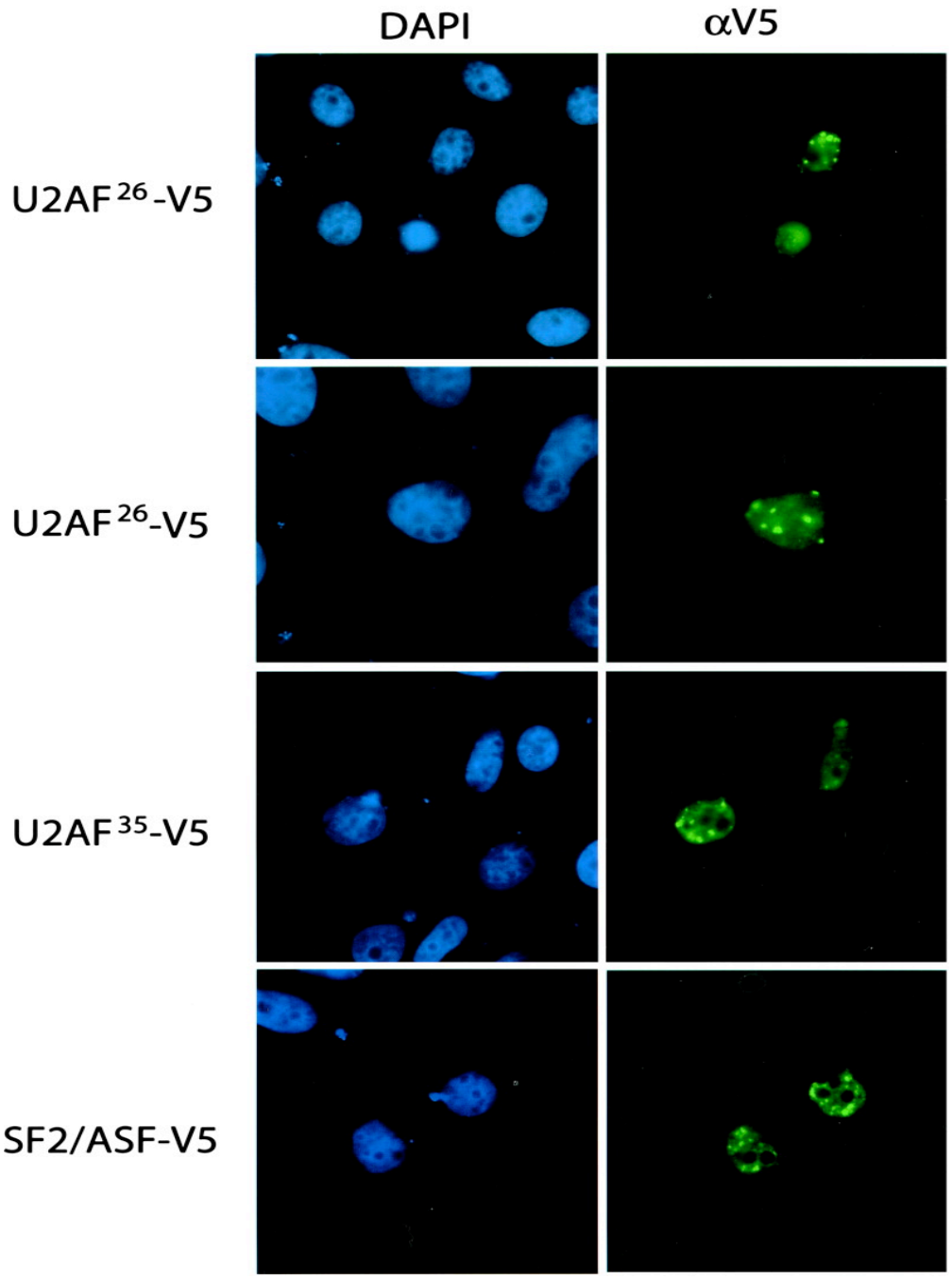


Figure 2.3: Immunolocalization of U2AF^{26}. HeLa cells were transfected with expression vectors encoding C-terminal V5-epitope tagged U2AF^{26} (top two rows), U2AF^{35} (third row), or SF2/ASF (bottom row). The cells were fixed and probed with an anti-V5 primary antibody and a FITC-tagged secondary antibody (right column). In addition, the cells were stained with DAPI to visualize the nucleus (left column). Two separate fields of U2AF^{26} are shown (first and second row). Speckles cannot be observed in the lower U2AF^{26} positive cell in the top row because it is out of the plane of focus.

Expression and purification of U2AF^{65}/U2AF^{26}. To examine whether U2AF^{26} might function as a pre-mRNA splicing factor, we first needed to express and purify the protein. Attempts to purify recombinant U2AF^{26} alone from *E. coli* or Sf9 cells yielded only insoluble protein. We therefore co-expressed His-tagged U2AF^{26} and GST-tagged U2AF^{65} in Sf9 cells using baculovirus expression vectors. In parallel, the conventional GST-U2AF^{65}/His-U2AF^{35} heterodimer, as well as GST-U2AF^{65} alone, were expressed and purified. The U2AF heterodimers were purified by a two-step procedure. First, the lysate was passed over a Ni-NTA column to purify His-tagged U2AF^{26} or U2AF^{35}. Secondly, the eluate from the Ni-NTA column was passed over a GST-agarose column to purify the GST-U2AF^{65}. The eluate from the GST-agarose column was analyzed by SDS-PAGE and stained with Coomassie brilliant blue. As expected, U2AF^{65} and U2AF^{35} co-purified by this method in a 1:1 stoichiometry (Fig. 2.4, lane 2). We found that U2AF^{65} and U2AF^{26} also co-purified over these two columns in a stoichiometry of approximately 1:1 (Fig. 2.4, lane 3). Gel filtration analysis of the purified recombinant proteins indicated that U2AF^{26} and U2AF^{65} are present in a complex (data not shown). These results demonstrate that U2AF^{26} can physically interact with U2AF^{65} suggesting that U2AF^{26} may have the ability to modify the activity of U2AF^{65} in pre-mRNA splicing.

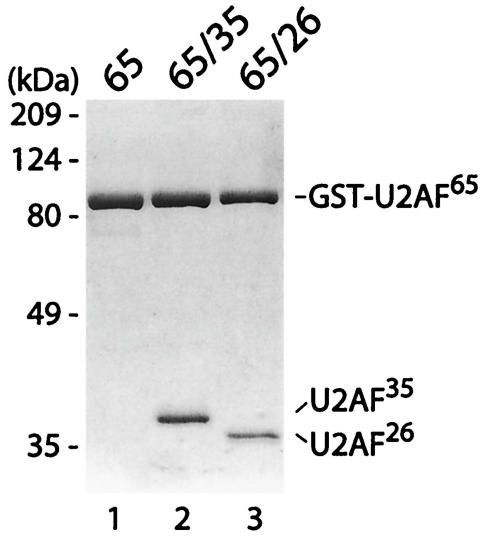


Figure 2.4: Expression and purification of U2AF complexes. Approximately 2 μg of recombinant U2AF^{65} (lane 1), U2AF^{35}/U2AF^{65} (lane 2), or U2AF^{26}/U2AF^{65} (lane 3) purified from baculovirus infected Sf9 cells were resolved on a 10% SDS polyacrylamide gel and stained with coomassie brilliant blue.

U2AF^{26} is a pre-mRNA splicing factor that can functionally replace U2AF^{35} *in vitro*. Given the similarities between U2AF^{26} and U2AF^{35}, we were interested in testing whether U2AF^{26} could function as a pre-mRNA splicing factor and, if so, comparing its activity to that of U2AF^{35}. We therefore compared the ability of U2AF^{65}, U2AF^{26}/U2AF^{65}, or U2AF^{35}/U2AF^{65} to restore splicing to HeLa cell nuclear extracts depleted of endogenous U2AF^{65} and U2AF^{35} (ΔNE). Previous work had shown that U2AF^{35} enhances the ability of U2AF^{65} to restore splicing in ΔNE (Graveley, 2000; Graveley et al., 2001; Misteli and Spector, 1998; Zuo and Maniatis, 1996). We first tested whether U2AF^{26} could function in constitutive splicing. As shown in Fig. 2.5A, U2AF^{65} alone weakly restores the splicing of β-globin pre-mRNA in ΔNE (lane 3). However, both U2AF^{35}/U2AF^{65} and U2AF^{26}/U2AF^{65} were approximately 2-fold more effective in restoring splicing of this substrate than U2AF^{65} alone (lanes 4 and 5).

Similar results were observed for the AdML pre-mRNA (data not shown). These observations indicate that U2AF^{26} can function as a pre-mRNA splicing factor and that it can functionally substitute for U2AF^{35} in constitutive splicing. We next tested whether U2AF^{26} could function in enhancer-dependent splicing. For this purpose we used the *dsx*-ASLV pre-mRNA (Staknis and Reed, 1994). This RNA contains the enhancer-dependent *Drosophila doublesex* intron and the ESE from avian sarcomaleukosis virus (ASLV) (Katz and Skalka, 1990) positioned 100 nucleotides downstream of the 3' splice site. The ASLV ESE is recognized by SR proteins present in the HeLa cell nuclear extract that, in turn, activate splicing by recruiting U2AF to the *dsx* 3' splice site (Zuo and Maniatis, 1996). U2AF^{65} alone weakly restored splicing of this pre-mRNA in ΔNE (lane 3). By contrast, both U2AF^{35}/U2AF^{65} and U2AF^{26}/U2AF^{65} were 5-6 fold more effective in restoring splicing of the *dsx*-ASLV pre-mRNA in ΔNE than U2AF^{65} alone (lanes 4 and 5). Thus, U2AF^{26} can functionally substitute for U2AF^{35} in both constitutive splicing and enhancer-dependent splicing. These results demonstrate that an RS domain in the small U2AF subunit is not required for enhancer-dependent splicing.

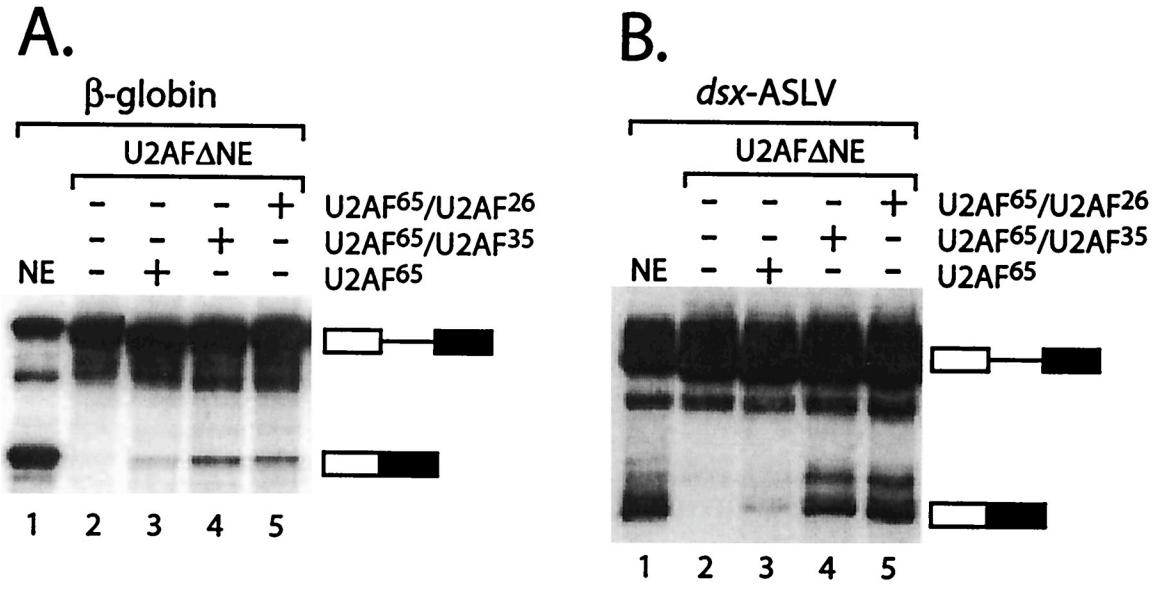


Figure 2.5: U2AF^{26} functions as a pre-mRNA splicing factor. A constitutively spliced β-globin pre-mRNA (A) or the enhancer-dependent *dsx*-ASLV pre-mRNA (B) were spliced in HeLa cell nuclear extract (lane 1) or U2AF-depleted extract either alone (lane 2), or with 25 nM U2AF^{65} (lane 3), U2AF^{35}/U2AF^{65} (lane 4), or U2AF^{26}/U2AF^{65} (lane 5).

U2AF^{26} enhances the binding of U2AF^{65} to weak 3' splice sites. Our data demonstrate that U2AF^{26} can functionally substitute for U2AF^{35} in both constitutive and enhancer-dependent splicing. We next wanted to determine the basis for the activity of U2AF^{26}. Given that U2AF^{26} appeared to be functionally similar to U2AF^{35}, we reasoned that the activity was most likely due to the sequences within the conserved N-terminus rather than the divergent C-terminus. Based upon the known function of U2AF^{35} (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999a), we thought U2AF^{26} might act by stabilizing the binding of U2AF^{65} on weak 3' splice sites and that this activity might involve an interaction between U2AF^{26} and the AG at the 3' splice site. To test this, we performed binding assays using the purified U2AF preparations and RNAs containing various 3' splice site sequences. As expected, the U2AF^{35}/U2AF^{65} heterodimer bound

much more efficiently to a 100 nucleotide RNA containing the weak *dsx* 3' splice site (K_D = 65 nM) than did U2AF^{65} alone (K_D = 2,000 nM) (Fig. 2.6 A). Surprisingly, the affinity of the U2AF^{26}/U2AF^{65} heterodimer for the *dsx* 3' splice site (K_D = 60 nM) was nearly identical to that of the U2AF^{35}/U2AF^{65} heterodimer (Fig. 2.6 A). Thus, both U2AF^{35} and U2AF^{26} enhance the affinity of U2AF for the *dsx* 3' splice site by greater than 30-fold. However, the impact of U2AF^{35} or U2AF^{26} on U2AF binding was substantially diminished on a 100 nucleotide RNA containing the pyrimidine tract from the β-globin pre-mRNA (Fig. 2.6 B) or an RNA containing the consensus U2AF^{65} binding site derived from SELEX (data not shown). In the case of the β-globin pyrimidine tract, the difference in affinity of the U2AF heterodimers (U2AF^{35}/U2AF^{65}, 120 nM; U2AF^{26}/U2AF^{65}, 100 nM) compared to U2AF^{65} alone (270 nM) was only about two-fold. This suggests that U2AF^{26}, similar to U2AF^{35}, functions to enhance the binding of U2AF^{65} to weak 3' splice sites.

To determine whether the ability of U2AF^{26} to enhance U2AF^{65} binding involves an interaction between U2AF^{26} and the 3' splice site AG, we generated two 100 nucleotide RNA substrates. Both substrates contain a variant of the β-globin pre-mRNA in which the pyrimidine tract was weakened by introducing two purines. However one substrate contained the 3' splice site AG dinucleotide and the other lacked this sequence. Both U2AF^{35} and U2AF^{26} enhanced the affinity of U2AF^{65} for the substrate containing the AG by 15-20 fold (U2AF^{65}, 750 nM; U2AF^{35}/U2AF^{65}, 50 nM; U2AF^{26}/U2AF^{65}, 35 nM) (Fig. 2.6 C). However, when the AG was deleted, the effect of U2AF^{35} and U2AF^{26} on U2AF binding is reduced to only 4-fold (U2AF^{65}, 60 nM; U2AF^{35}/U2AF^{65}, 15

nM;U2AF^{26}/U2AF^{65}, 16 nM) (Fig. 2.6 D). Interestingly, each of the protein preparations had a higher affinity for the RNA lacking the AG than the RNA containing the AG (Fig. 2.6 C and D). We believe that this may be due to differences in the RNA structure that present the pyrimidine tract of the RNA lacking the AG in a conformation more accessible to binding. However, the relative affinities of the U2AF heterodimers compared to U2AF^{65} alone are much higher for the RNA containing the AG (15-20 fold) than for the RNA lacking the AG (4-fold). These results show that U2AF^{26}, like U2AF^{35}, stabilizes the binding of U2AF^{65} to weak pyrimidine tracts. Given the similarity in the sequence of the RRM's of U2AF^{35} and U2AF^{26}, the well documented interaction between U2AF^{35} and the 3' splice site AG (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999a), and the nearly identical behavior of the two proteins in our assays, we believe that U2AF^{26}, like U2AF^{35}, can directly interact with the AG dinucleotide at the 3' splice site.

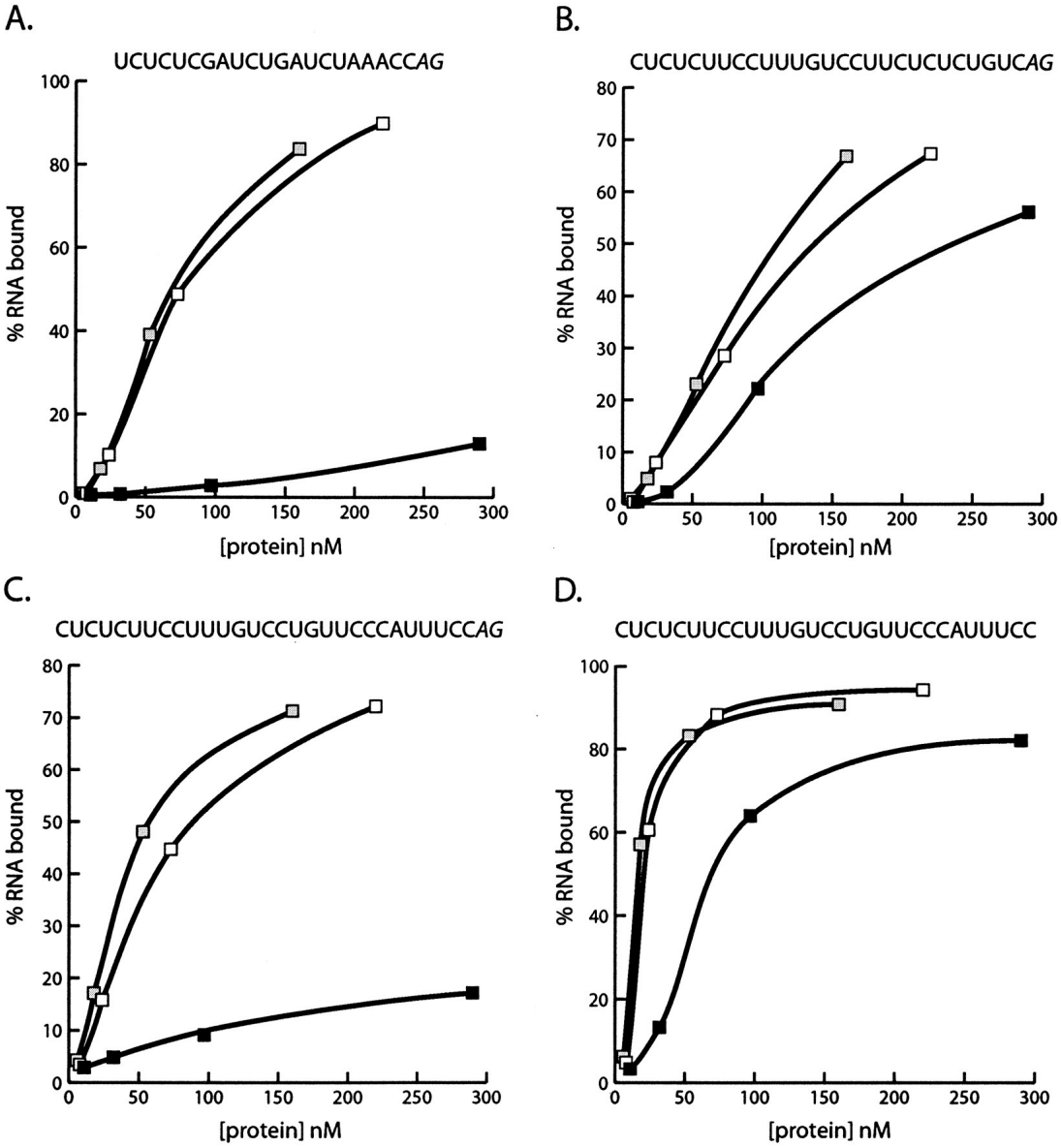


Figure 2.6: U2AF^{26} enhances the binding of U2AF^{65} to weak 3' splice sites. The binding of the various U2AF preparations was compared on RNAs containing the weak *dsx* 3' splice site (A) or the strong β-globin 3' splice site (B). In addition, U2AF binding was measured on RNAs containing a variant of the β-globin 3' splice site that either contained (C) or lacked (D) the 3' splice site AG dinucleotide. The binding reactions were resolved on non-denaturing polyacrylamide gels and quantitated with a Molecular Dynamics phosphorimager. The protein concentrations used in the binding experiments were 10 nM, 32nM, 97 nM, and 290 nM for U2AF^{65} (black squares), 8 nM, 24 nM, 73 nM, and 220 nM for U2AF^{35}/U2AF^{65} (white squares), and 6 nM, 18 nM, 53 nM, and 160 nM for U2AF^{26}/U2AF^{65} (gray squares). All of the RNAs used in these experiments were approximately 100 nucleotides. Only the sequence of the 3' splice site of each RNA is shown above the graph.

DISCUSSION

Here we report the identification and characterization of a protein designated U2AF^{26}, that is highly related to the splicing factor U2AF^{35}. Although the function of U2AF^{35} had been elusive for many years, it has recently been shown to function by interacting with the 3' splice site AG and thus stabilizing the binding of U2AF^{65} to weak pyrimidine tracts (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999a). Our results demonstrate that U2AF^{26} is a nuclear protein that interacts with U2AF^{65} and enhances its binding to weak pyrimidine tracts. We further show that this activity likely involves an interaction between U2AF^{26} and the AG dinucleotide at the 3' splice site. Thus, U2AF^{26} functions as a splicing factor much in the same way as U2AF^{35}.

Orthologs of U2AF^{26} do not appear to exist in lower eukaryotes, *Drosophila*, *C. elegans*, plants, or vertebrates such as *Xenopus* or zebrafish (data not shown). By contrast, ESTs more similar to U2AF^{26} than U2AF^{35} are also found in rat (AW142126, AW142127, AA848227), pig (BF193007, BF440828), and cow (BE589680). In addition, humans contain a *U2AF^{26}* gene located on chromosome 19 (NT_011296)**,** and several human ESTs in GenBank correspond to this gene. A comparison of the mouse and human *U2AF^{26}* revealed that the exon-intron boundaries are located in the same positions as in the human *U2AF^{35}* gene, although the introns are much smaller in the *U2AF^{26}* gene. In addition, the exon sequences of the human and mouse *U2AF^{26}* genes are 90% identical at the nucleotide level and the majority of the differences are neutral, third position changes. Thus, the gene encoding *U2AF^{26}* appears to have recently arisen in mammals by duplication of the *U2AF^{35}* gene. Given the high degree of

similarity between the human and mouse *U2AF*[26] genes, and the fact that human and rodents diverged approximately 96 million years ago (Nei et al., 2001), it is likely that U2AF[26] has a unique function.

Although proteins related to U2AF[35] have been described previously, U2AF[26] is the only one that has been shown to function in a manner similar to U2AF[35]. U2AF1-RS1 and U2AF1-RS2/Urp are 94% identical to one another and contain stretches that are approximately 50% identical to portions of U2AF[35] (Tronchere et al., 1997). Like U2AF[26], Urp interacts with U2AF[65] in a manner likely to be mutually exclusive with U2AF[35] (Tronchere et al., 1997). However, Urp contains an RS domain, interacts with SR proteins, and is functionally distinct from U2AF[35] because U2AF[35] cannot complement Urp-depleted extracts (Tronchere et al., 1997). In addition to U2AF1-RS1 and U2AF1-RS2/Urp, four proteins related to U2AF[35] were recently identified in the draft of the human genome sequence (Tupler et al., 2001). These observations, together with our data, suggest that mammalian cells contain multiple U2AF complexes.

In addition to its role in constitutive splicing, U2AF[35] has been shown in some studies to be required for efficient enhancer-dependent splicing (Graveley, 2000; Graveley et al., 2001; Zhu and Krainer, 2000; Zuo and Maniatis, 1996). One model proposes that SR proteins bound to exonic splicing enhancers act to recruit U2AF to the upstream 3' splice site (Bouck et al., 1998; Graveley et al., 2001; Wang et al., 1995; Zhu and Krainer, 2000; Zuo and Maniatis, 1996). This activity is thought to require interactions between SR proteins and the RS domain of U2AF[35] (Wu et al., 1999; Zuo and Maniatis, 1996). However, studies in *Drosophila* have shown that although the small U2AF subunit,

dU2AF^{38}, and its interaction with the large U2AF subunit are essential for viability (Rudner et al., 1996; Rudner et al., 1998c), yet the integrity of its RS domain is not (Rudner et al., 1998b). Moreover, the alternative splicing of the enhancer-dependent *dsx* pre-mRNA is not affected in flies lacking the dU2AF^{38} RS domain (Rudner et al., 1998b). Additionally, both the *C.elegans* and *S. pombe* small U2AF subunits naturally lack RS domains (Wentz-Hunter and Potashkin, 1996; Zorio and Blumenthal, 1999b). Together these observations suggest that enhancer-dependent splicing does not require interactions between enhancer-bound SR proteins and the RS domain of the small U2AF subunit. Our results support this model because U2AF^{26}, which lacks an RS domain, can functionally substitute for U2AF^{35} in enhancer-dependent splicing. It remains to be determined whether U2AF^{26} can directly interact with SR proteins. Although U2AF^{26} appears functionally identical to U2AF^{35} in our assays, it is our anticipation that U2AF^{26} will have a unique function *in vivo* that we have yet to uncover. The observation that U2AF^{35} and U2AF^{26} are differentially expressed raises the possibility that U2AF^{26} may function in tissue-specific alternative splicing. Interestingly, with respect to U2AF^{35}, U2AF^{26} is expressed most strongly in the brain where the occurrence of alternative splicing is unusually high (Grabowski and Black, 2001). Thus, alternative splicing may be affected by the regulated expression of U2AF^{26}.

What could be the functional differences between these two proteins? One possibility is that the U2AF heterodimers containing either U2AF^{35} or U2AF^{26} could recognize distinct, yet overlapping sets of 3' splice sites. In support of this model, it is interesting to note that in contrast to the RNP-1 sequences, the RNP-2 motif is slightly

different between the two proteins. In particular, the first residue of RNP-2 is an Ala in U2AF^{35}, and a Val in U2AF^{26}. Based on modeling with other RRM-RNA crystal structures, it was proposed that Ala47 of U2AF^{35} may directly contact RNA (Kielkopf et al., 2001). Thus, although both U2AF^{26} and U2AF^{35} appear to bind RNA, and recognize the AG dinucleotide at the 3' splice site, we believe that the binding specificities of the two heterodimers may be slightly different. Thus it is possible that the two heterodimers could recognize both common and distinct sets of 3' splice sites. In this scenario, differences in the expression levels of U2AF^{35} and U2AF^{26} could influence alternative splicing because the two U2AF isoforms might recognize and activate different 3' splice sites.

Alternatively, it is possible that the U2AF^{26}/U2AF^{65} and U2AF^{35}/U2AF^{65} heterodimers could interact differently with splicing regulatory factors. The crystal structure of the core U2AF^{35}/U2AF^{65} heterodimer revealed that helix A in the U2AF^{35} RRM is unusually long and amphipathic (Kielkopf et al., 2001). In particular, glutamate residues 73,76, 80, 83, 84, 88, 90, and 91, Asp79, and Thr87 (which would be negatively charged if phosphorylated) of U2AF^{35}, all reside on one side of helix A. Kielkopf *et al.*, point out that the highly negatively charged face of this helix may be an interaction surface for other splicing factors – possibly the basic RS domains contained in SR proteins (Kielkopf et al., 2001). Thus, changes in the residues contained in this alpha helix might alter the interactions between U2AF and other splicing factors. Interestingly, in U2AF^{26}, Glu80 is changed to Asn, and Glu90 is changed to Gln. These uncharged amino acids both reside on the negatively charged surface of helix A and should change

its electrostatic surface, possibly affecting the types of protein interactions that U2AF^{26} can participate in. Thus, U2AF may have distinct, yet overlapping responses to different splicing factors depending on whether U2AF^{35} of U2AF^{26} is associated with U2AF^{65}. Given that the levels of U2AF^{26} and U2AF^{35} vary with respect to one another in different tissues, we imagine that the relative concentrations of the different U2AF heterodimer isoforms may play an important role in determining the splicing patterns of a number of pre-mRNAs.

MATERIALS AND METHODS

Cloning of U2AF^{26}. We identified a cDNA fragment corresponding to the N-terminus of U2AF^{26} in a screen for genes regulated by the transcription factor NPAS2 (Zhou et al., 1997). This fragment was used to screen an adult mouse brain Uni-ZAP XR cDNA library (Stratagene). Two overlapping partial cDNA clones were used to assemble a full-length U2AF^{26} cDNA. The assembled cDNA sequence was confirmed by RT-PCR experiments from mouse brain RNA as well as multiple ESTs in GenBank. A mouse U2AF^{35} cDNA was cloned from total brain RNA using RT-PCR. Using the mouse U2AF^{26} cDNA as a probe, we screened a mouse 129/SvEV Tac f BR genomic library (Strategene) for the U2AF^{26} gene. Using overlapping genomic clones, we were able to assemble a contig containing the complete U2AF^{26} gene.

Northern blotting. Total RNA was isolated from dissected mouse brain, cerebellum, colon, eyes, heart, kidney, liver, lung, skeletal muscle, small intestine, spleen, testis, uterus and white adipose tissue using RNA Stat-60 (Tel-Test). Poly(A)$^+$ RNA was

isolated by passing the RNA samples twice through Oligo (dT) cellulose columns (Gibco). 5 μg of RNA was run on a 1.2% formaldehyde gel, transferred to a nytran supercharge membrane, and hybridized with ^{32}P-labeled DNA probes. The U2AF^{26} transcripts were detected with a full-length cDNA probe, while the U2AF^{35} transcripts were detected with a probe corresponding to the last 476 bp of the mouse cDNA. The blots were also probed with a cyclophilin probe as an internal control for RNA loading. The blots were developed and quantitated with a Fuji BAS1500 phosphoimager. After normalization to the cyclophilin control, the U2AF^{35} and U2AF^{26} signals in the lung tissue, which was the tissue expressing the median value of each mRNA, were set to a value of 1.0. The U2AF^{35} and U2AF^{26} signals in the remainder of the tissues were normalized to the signal in lung, and these values were used to calculate the ratio of U2AF^{26} to U2AF^{35}.

Indirect immunofluorescence. HeLa cells were transfected with 1 μg of pcDNA-mU2AF^{26}/V5, which encoded U2AF^{26} containing a C-terminal V5 epitope tag, using Fugene6 (Boehringer Mannheim). In addition, cells were transfected with control plasmids expressing either human U2AF^{35} or human SF2/ASF containing C-terminal V5 epitope tags. 24 hours after transfection, the cells were fixed in formaldehyde and probed with anti-V5 antibody (Invitrogen). The cells were then washed with 1XPBS and probed with an anti-mouse IgG-FITC secondary (Jackson ImmunoResearch Labs). The cells were washed, stained with DAPI, visualized with a Zeiss Axioplan 2 microscope, and the images acquired with a SenSys cooled CCD camera using OpenLab software (Improvision).

Expression and purification of recombinant proteins. His_6-mU2AF^{26} and His_6-

hU2AF^{35} were cloned into the baculovirus expression vector, pFastBac-1 (Gibco). The N-terminal His_6 tag was added by PCR. To construct a virus encoding the GSThU2AF^{65}, human U2AF^{65} was first cloned into pGEXT-4T (Amersham-Pharmacia), amplified by PCR to isolate the fragment encoding GST-U2AF^{65}, and subsequently cloned into pFastBac-1. Baculovirses encoding these proteins were produced as described by the manufacturer (Gibco). The U2AF^{26}/U2AF^{65} and U2AF^{35}/U2AF^{65} heterodimers were produced by co-infecting the Sf9 cells with the respective viruses, whereas U2AF^{65} was produced by infecting Sf9 cells with a single virus. 48 hours after infection, the Sf9 cells were centrifuged at 500Xg for 10 min and the pellet resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 0.1% Triton X-100, 5 mM β-ME, 1 mM PMSF, and 10 μg/mL leupeptin). The suspension was sonicated for 1 min and centrifuged at 18,000 RPM in a JA-20 rotor for 25 minutes. The supernatant was then incubated for 25 min at 4°C with Ni-NTA agarose beads (Qiagen) equilibrated with lysis buffer. After incubation, the beads were loaded onto a column, washed with lysis buffer, and eluted with lysis buffer containing 300 mM imidazole. The eluate was then incubated with equilibrated Glutathione Sepharose 4B (Amersham Pharmacia) for 25 minutes and then loaded onto a column. The beads were washed with lysis buffer and eluted with lysis buffer containing 50 mM glutathione. U2AF^{65} was purified on Glutathione Sepharose only. The purified proteins were dialyzed overnight in Buffer D (100 mM KCl, 20% glycerol, 20 mM HEPES pH 7.9, 0.2 mM EDTA, 0.1 mM DTT).

Pre-mRNA model substrates. Plasmids encoding the β-globin and *dsx*-ASLV pre-mRNAs were described previously (Reed, 1989; Staknis and Reed, 1994; Zhou et al.,

1997). Templates encoding the RNAs used for the gel-shift experiments were generated by PCR from the *dsx* or β-globin templates. ^{32}P-labeled RNAs were synthesized with either SP6 or T7 RNA polymerase.

***In vitro* splicing assays.** HeLa cell nuclear extract were depleted of U2AF^{65} and U2AF^{35} by oligo-dT chromatography as described previously (Graveley et al., 2001). The depleted extract used in these experiments is the same one used in the experiments reported in (Graveley et al., 2001). As shown in (Graveley et al., 2001), Western blot analysis suggests that >95% of U2AF^{65} and U2AF^{35} were depleted from this extract. Splicing assays were carried out using 30% U2AF depleted extract as previously described (Graveley et al., 2001).

U2AF binding assays. The recombinant U2AF preparations were incubated with 5 fmol of ^{32}P-labeled RNAs in reactions containing 800 ng of BSA, 250 ng of tRNA, 1.3% polyvinylalcohol, 1 mM DTT, 4 U of rRNasin (Promega), in a 10 μl volume for 15 min at 30°C. The reactions were transferred to ice, heparin was added to 0.05 mg/ml, and resolved by electrophoresis on 4% 80:1 polyacrylamide gels at 13 V/cm for three hours at 4°C. The gels were quantitated with a Molecular Dynamics Storm Phosphorimager and the percent of RNA bound at each protein concentration calculated. K_D values were calculated for U2AF^{65} by fitting the data to a single binding isotherm. K_D values for the U2AF heterodimers were calculated by fitting the data to a cooperative binding isotherm where $Y=(K_1[P]+K_1K_2K_C[P]_2])/(1+(K_1+K_2)[P]+K_1K_2K_C[P]_2)$, where K_1 and K_2 are the estimated equilibrium association constants for U2AF^{65} and U2AF^{35} or U2AF^{26}, Y is the fraction of RNA bound, P is the concentration of protein used, and K_C is the cooperativity

factor (Hertel and Maniatis, 1998; Witherell et al., 1990).

CHAPTER THREE
Analysis of the Functional Differences Between U2AF^{26} and U2AF^{35}

INTRODUCTION

Splice site selection is the first step in spliceosome assembly. During this step, trans-acting factors bind to specific sequences at the 5' and 3' splice sites, thereby defining which 5' and 3' splice sites will be utilized by the spliceosome. These factors include the U1snRNP and the U2 auxiliary factor (U2AF), and they recognize the 5' and 3' splice site respectively.

The U2AF heterodimer is composed of a large and small subunit. The large subunit preferentially binds to the pyrimidine tract (Singh et al., 1995; Zamore and Green, 1991) and the small subunit binds to the AG dinucleotide of the 3' splice site (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999a). Previously, it has been reported that U2AF^{35} can enhance U2AF^{65} binding to a 3' splice site nearly twenty-fold (Rudner et al., 1998a). Furthermore, the small subunit is necessary for efficient *in vitro* splicing of substrates that have weak or short pyrimidine tracts (Guth et al., 2001; Wu et al., 1999).

With the use of systematic evolution of ligands by exponential enrichment (SELEX), the optimal binding of U2AF^{35} was determined to be UAG/GU (/ denotes intron-exon boundary) (Wu et al., 1999). This provided more evidence that the small subunit recognized the AG dinucleotide of the 3' splice site and suggested that U2AF^{35} prefers a specific sequence upstream and downstream of the AG. Interestingly, this

sequence is not the most common sequence found around the AG dinucleotide. The consensus sequence for this region of the 3' splice site is CAG/GN (Shapiro and Senapathy, 1987; Stamm et al., 1994). This SELEX result suggests that $U2AF^{35}$ might preferentially recognize a splice site different from the 3' splice site consensus motif. One group tested whether mutating the position immediately downstream of the AG of an AG-dependent intron would affect the splicing efficiency of the intron (Guth et al., 2001). Surprisingly, changing the base from a guanine to cytosine (AG/G to AG/C) significantly reduced the splicing efficiency of the substrate. Further analysis indicated that the $U2AF^{35}$ binding was reduced, suggesting that the change in splicing was caused by a reduction in the $U2AF^{35}$ binding affinity. This is an incomplete analysis of possible permutations of the 3' splice site, but it does suggest that $U2AF^{35}$ can distinguish between different sequences around the AG dinucleotide.

Only one form of the large subunit has been described in mammals (Zamore et al., 1992) but there have been multiple variations of the small subunit. The different forms of $U2AF^{35}$ are products of alternative splicing of the conventional $U2AF^{35}$ transcript (Pacheco et al., 2004) or independent genes (Shepard et al., 2002; Tronchere et al., 1997). The presence of different small subunits raises the possibility that the various small subunits, when associated with $U2AF^{65}$, might confer a unique function for the respective heterodimers.

The unique U2AF heterodimers might lead to preferential recognition of different 3' splice sites. Tissue-specific differences in expression of the different small subunits could lead to tissue-specific recognition of certain 3' splice sites. This would be

consistent with the premise that different ratios of splicing factors, such as SR proteins and hnRNPs, leads to a "cellular code", where cell-specific splicing events are dictated by the expression pattern of the relevant splicing factors (Smith and Valcarcel, 2000). Examining the functional differences between the different small subunits would be essential to understand if there are functional differences between the potential heterodimers.

In this study we have analyzed the mouse U2AF^{26}, a paralog of the mouse U2AF^{35}, to determine if there are functional differences between the two small subunits. We hereby report that U2AF^{26} is differentially expressed in mouse tissues relative to U2AF^{35}. Also, U2AF^{26} and U2AF^{35} have overlapping binding specificity, but U2AF^{26} recognized a specific 3' splice site motif better than U2AF^{35}. These results suggest that the two small subunits are functionally similar, but there could be subtle differences that could lead to a functional distinction between the two splicing factors.

RESULTS

Protein expression of U2AF^{26} and U2AF^{35} in different tissues and cell lines. We first sought to determine if there are differences in expression of U2AF^{26} relative to U2AF^{35} in different mouse tissues. Using a polyclonal antibody raised against a peptide corresponding to the C-terminus of U2AF^{26}, we analyzed the protein expression of U2AF^{26} in whole tissue protein samples from brain, kidney, liver, lung, heart, spleen, and testis (Fig. 3.1A). The antibody detected a polypeptide in whole cell lysate from 293 cells in which U2AF^{26} was transiently overexpressed. Also, the antibody cross-reacted with the

U2AF^{35} polypeptide and this was confirmed by analyzing the same protein samples with a polyclonal antibody for U2AF^{35} (Fig. 3.1A). From this experiment, we concluded that U2AF^{26} expression varies significantly relative to U2AF^{35} expression.

Western blot analysis of eighteen different mouse- and rat-specific cell lines revealed that U2AF^{35} expression is much higher than U2AF^{26} in all cell lines tested (Fig. 3.1B). Additionally, expression of U2AF^{35} was much higher in the spleen and testis, tissues that have high levels of cell proliferation. Robust expression levels of U2AF^{35} in tissues that have high a proliferative rate and cell lines, which divide very rapidly relative to the parent tissue, suggests that U2AF^{35} is needed for rapid cell division. U2AF^{35} might be involved in the regulation of splicing events important for the progression of the cell cycle. Interestingly, U2AF^{26} expression was highest relative to U2AF^{35} in tissues that are terminally differentiated, and U2AF^{26} could potentially regulate recognition of 3' splice sites of substrates required for a cell to enter a differentiated, homeostatic state.

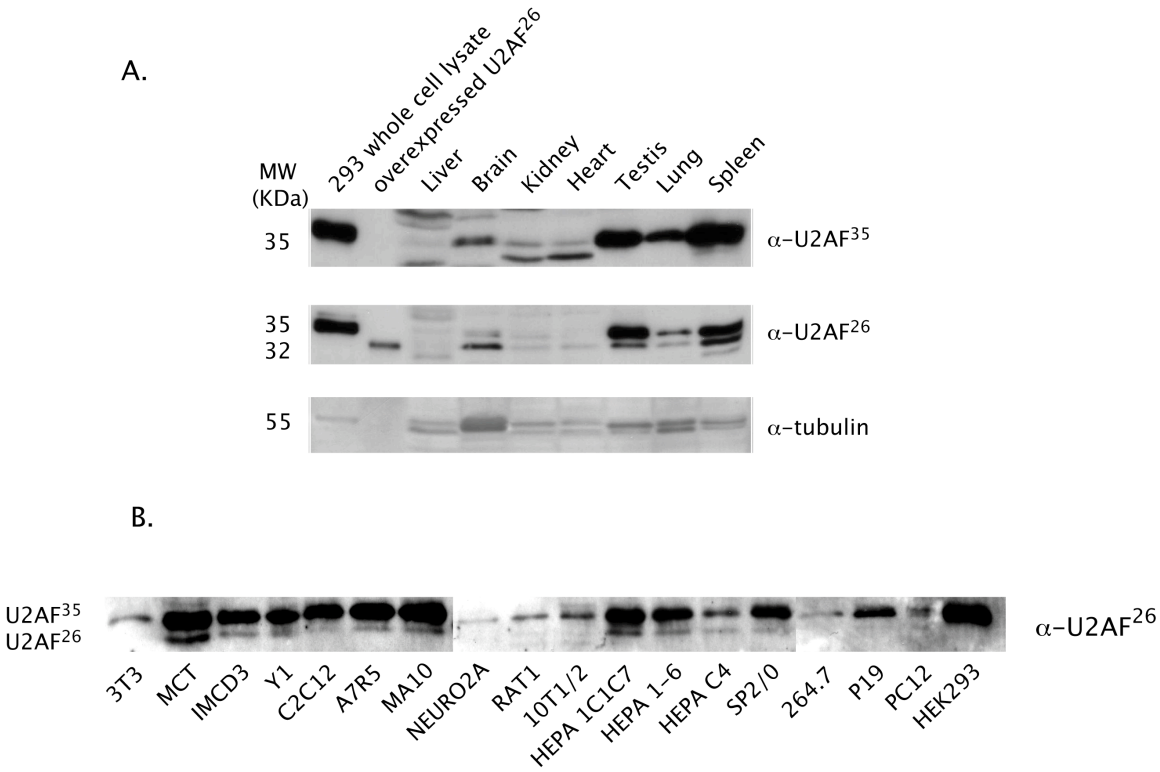


Figure 3.1: (A) Western blot analysis of U2AF^{35} and U2AF^{26} in different mouse tissues. The U2AF^{26} antibody recognizes a polypeptide that migrates at 32 kDa. This band corresponds with a polypeptide observed in a protein prep of 293 cells in which U2AF^{26} was over expressed. The U2AF^{26} antibody appears to cross-react with the U2AF^{35} protein as well. This was confirmed with a western blot analysis of the same samples using the polyclonal U2AF^{35} specific antibody. (B) Western blot analysis of U2AF^{26} and U2AF^{35} in different mouse- and rat-specific cell lines.

Creating cell lines that stably or conditionally overexpress U2AF^{26}. In an effort to examine the functional difference between the two small subunits, we set out to overexpress U2AF^{26} and then test for changes of known alternative splicing events. If there were changes in an alternative splicing event when U2AF^{26} is overexpressed, we could infer that those differences were directly or indirectly due to the increased expression of U2AF^{26}.

First we created stable cell lines using three different cell types. Western blot analysis indicated that we could produce clones that overexpressed U2AF^{26} with the given cell types (Fig. 3.2A). However, we could not overexpress U2AF^{26} above the expression levels of U2AF^{35}. Determination of the specific activity of the U2AF^{26} antibody indicated that the antibody preferentially recognized U2AF^{26} five times better than U2AF^{35}. Using this standard, we estimated that we could not stably overexpress U2AF^{26} beyond 20% of the U2AF^{35} expression.

Despite this limitation, we examined eighteen characterized alternative splicing events. We examined these alternative splicing events in Neuro2A clones that expressed U2AF^{26} at varying levels (Appendix A). We did not see any significant changes in the processing of the alternative splicing events. This could be because expression levels of U2AF^{26} were not high enough to induce changes in the splicing events examined, or this could be an indication that these splicing events are not influenced by U2AF^{26}.

Since there was a maximal level of expression we could achieve in the stable cell lines, we reasoned that stable overexpression of U2AF^{26} might be deleterious to mitotic growth. We decided to create cell lines that conditionally overexpress U2AF^{26} (Fig. 3.2B). However, upon induction of different clones that conditionally overexpress U2AF^{26}, we observed the same limit of overexpression. Since we could not conditionally overexpress U2AF^{26} beyond the levels we could with stable cell lines, we concluded that the limit to the expression was not due to any long-term effect of overexpressing U2AF^{26}, but to a short-term deleterious effect or some other phenomenon.

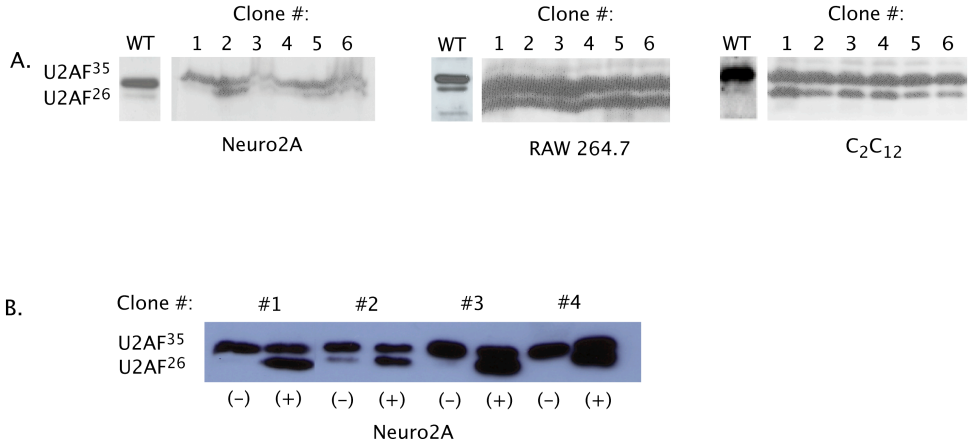


Figure 3.2: (A) Western blot analysis of clones that stably overexpress U2AF^{26} in Neuro2A, RAW 264.7, and C_2C_{12} cell lines. The antibody was the same U2AF^{26} antibody used in Fig. 1. The clone number is listed above the lane and the wild type cell lysate is to the left of the western blot of each set of U2AF^{26} stable clones. (B) Western blot analysis of four different Neuro2A clones that conditionally overexpress U2AF^{26}. (-) is with no treatment and (+) is with ponasterone treatment.

RNA interference of U2AF^{35} in a Neuro2A clone that conditionally overexpresses U2AF^{26}. In order to shift the ratio of the two small subunits in favor of U2AF^{26}, we performed RNA interference of the U2AF^{35} transcript in the Neuro2A clone #1 from figure 2. We hypothesized that if we reduced the expression of U2AF^{35} and then induced overexpression of U2AF^{26}, we would obtain cells that had a high expression of U2AF^{26} relative to U2AF^{35}.

We successfully performed this technique and we were able to achieve expression of U2AF^{26} roughly equal to the expression of U2AF^{35}, as estimated by western blotting (Fig. 3.3A). Once we had cells with this ratio, we purified total RNA from the cells and then analyzed the same eighteen alternative splicing events previously examined (Appendix A). Of the eighteen splicing events analyzed, only one regulated alternative splicing event was affected by this RNAi experiment (data not shown). The alternative

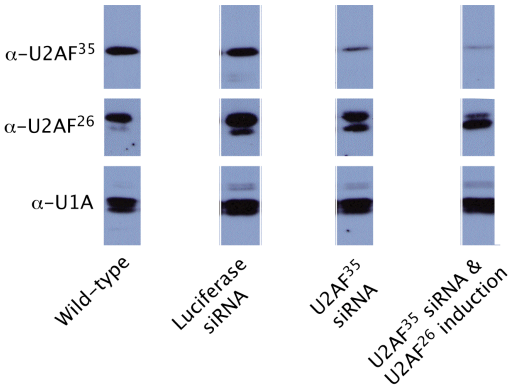
splicing event affected was the regulated inclusion of a cassette exon, exon EN, of the clathrin light chain B (CLC B) subunit. Surprisingly, we observed an increase of exon EN inclusion when we reduced U2AF^{35} expression and this splicing event was not effected by the induction of U2AF^{26} overexpression. We speculated that the increase of inclusion of this exon could be due to the increase in the ratio of the endogenous U2AF^{26} to U2AF^{35}. However, when we knocked down expression of U2AF^{26} and U2AF^{35} concurrently, we did not see any changes in the inclusion of the exon relative to the U2AF^{35} knock down alone (data not shown). This suggests that U2AF^{26} is not involved in the regulation of this splicing event.

To analyze the phenotypic changes of the cells after knock-down of the small subunits, we analyzed the cells for proliferation and cell cycle changes. We knocked-down U2AF^{35} alone, U2AF^{26} alone, and U2AF^{35} and U2AF^{26} concurrently and performed western blot analysis of this RNAi experiment (Fig. 3.3B). Using ^{3}H-thymidine, we measured the proliferation rate of the transfected cells. Knock-down of U2AF^{35} alone had a significant effect on the proliferation rate and knock-down of U2AF^{26} alone had minimal effect on the growth rate. Knock-down of U2AF^{35} and U2AF^{26} concurrently did not reduce the proliferation rate significantly lower than observed for the U2AF^{35} RNAi.

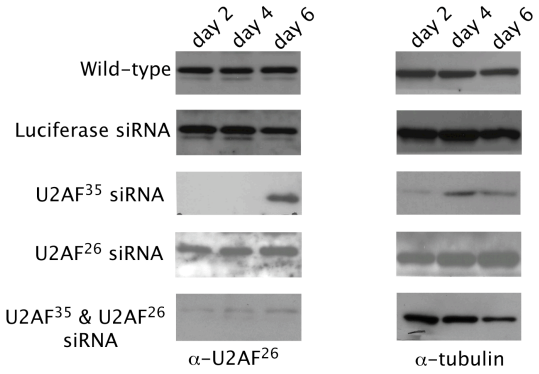
Flow cytometry analysis of 5-bromodeoxyuridine (BrdU) labeled of cells from Fig. 3.3B indicated that the reduction in U2AF^{35} expression leads to a severe cell cycle phenotype. The cells appear to be arrested in the G2/M stage and there is an increase in the number of apoptotic cells. Reduction of the U2AF^{26} expression has no observable effect on the cell cycle. Furthermore, the phenotype observed with knock-down of

U2AF^{35} and U2AF^{26} was not significantly different than with knock-down of U2AF^{35} alone. This suggests that U2AF^{35} has a role in the cell cycle, potentially in the exit from the G2/M stage. We were not able to determine the *in vivo* function of U2AF^{26} using this experimental technique.

A.

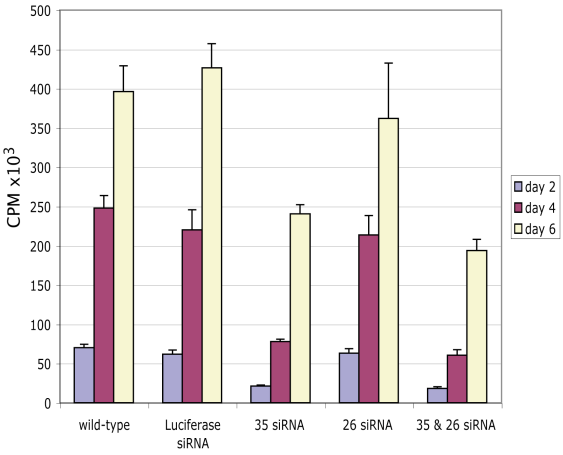


B.



C.

Quantitation of ^{3}H Thymidine Proliferation Assay:



D.

Flow Cytometry Analysis of BrdU Labeling:

Wild-type	day 2	day 4	day 6
Apoptotic	1.6	2.2	5.2
G0/G1	20.5	36.2	52.5
S	67.6	53.9	31.5
G2/M	6.5	5.0	8.7
Luciferase siRNA			
Apoptotic	2.1	1.8	4.7
G0/G1	16.7	27.6	46.9
S	67.8	62.9	39.6
G2/M	6.0	4.5	5.5
U2AF35 siRNA			
Apoptotic	3.5	10.0	4.0
G0/G1	18.1	26.7	29.2
S	54.9	49.0	58.9
G2/M	15.1	9.8	5.0
U2AF26 siRNA			
Apoptotic	2.6	3.0	4.0
G0/G1	25.7	29.9	50.7
S	56.7	59.2	38.5
G2/M	9.6	5.0	5.0
U2AF35 & U2AF26 siRNA			
Apoptotic	2.6	12.7	5.0
G0/G1	34.4	29.6	29.7
S	36.2	44.3	57.6
G2/M	22.8	9.5	5.4

Figure 3.3: (A) Western blot analysis of RNAi mediated knock-down of $U2AF^{35}$. The Luciferase siRNA served as the control for the specificity of the siRNA transfection. U1A is a splicing factor associated with the U1snRNP and the U1A western blot analysis served as the loading control for each lane. (B) Western blot of $U2AF^{26}$ and tubulin at day 2, day 4, and day 6 after transfection of the corresponding siRNA duplex. (C) Quantitation of cellular proliferation, using a ^{3}H-Thymidine incorporation assay, for the cells analyzed in Fig. 3.3B. (D) Flow cytometry analysis of the BrdU labeling of the same set of cells to assess changes in the cell cycle.

Examination of the optimal binding site the two small subunits using SELEX.

SELEX has been successfully implemented to identify the optimal binding site of several different splicing factors, including recombinant SR proteins [reviewed in (Graveley, 2000)], recombinant $U2AF^{65}$ (Singh et al., 1995), and for purified HeLa $U2AF^{35}$/$U2AF^{65}$ heterodimer (Wu et al., 1999). Using the purified heterodimer, the binding site for $U2AF^{35}$ was determined to be UAG/GU (Wu et al., 1999).

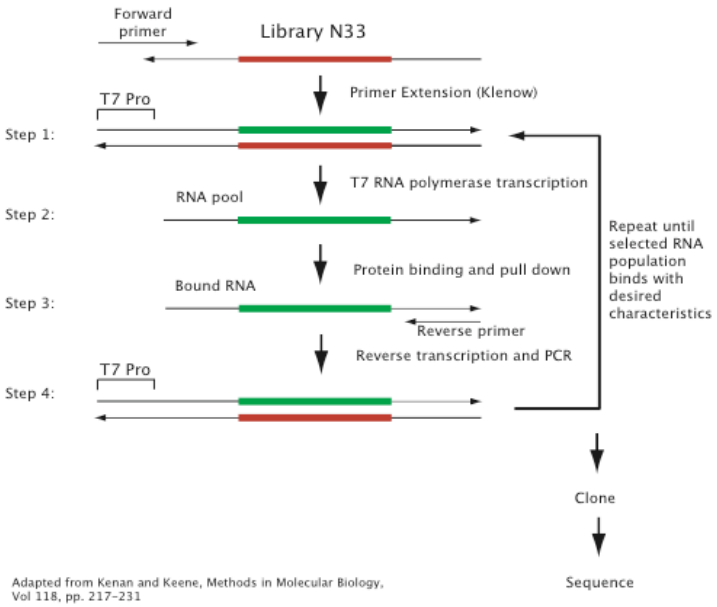
$U2AF^{26}$ has two amino acid substitution of in the RNP-2 motif, a region thought to directly contact RNA (Kielkopf et al., 2001). We thought this could confer a difference in the binding specificity for $U2AF^{26}$ relative to $U2AF^{35}$, so examined the binding affinity of recombinant $U2AF^{35}$/$U2AF^{65}$ and $U2AF^{26}$/$U2AF^{65}$. Using the two recombinant heterodimers, we performed a SELEX protocol (Kenan and Keene, 1999) under the same conditions used by Wu *et al*. (Wu et al., 1999) (Fig. 3.4A). After four rounds of selection, we examined the pools selected by $U2AF^{26}$/$U2AF^{65}$ using gel shift analysis. We observed that the percentage of the pool that bound to the $U2AF^{26}$ heterodimer increased significantly after each round of selection. Over 90% of the fourth pool bound well to the recombinant heterodimer (Fig. 3.4B). This analysis indicated that the enrichment had

worked properly and each progressive pool was increasingly enriched with RNAs with binding sites for $U2AF^{26}/U2AF^{65}$.

We performed a fifth round of selection and then sequenced forty unique clones of the $U2AF^{35}/U2AF^{65}$ and $U2AF^{26}/U2AF^{65}$ pools. Using this assay, we identified the binding site for $U2AF^{35}$ as CAG/GU (Fig. 3.4C), which is very similar to the binding site for $U2AF^{35}$ of the purified HeLa heterodimer (Wu et al., 1999). The only difference between the two binding sites is the first position before the AG dinucleotide (-3 position). Interestingly, the second most common base at this position for the different clones is uracil, which is the base observed at this position of the consensus sequence for the HeLa $U2AF^{35}$. This suggests that the SELEX protocol worked and that the recombinant protein binding affinity accurately reflects the binding preference of the purified HeLa $U2AF^{35}$.

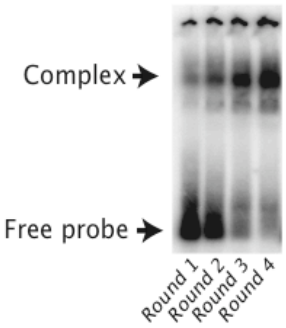
Analysis of the binding site for $U2AF^{26}$ indicated that there was a difference in the optimal binding site of $U2AF^{26}$ and $U2AF^{35}$. The binding site identified for $U2AF^{26}$ is CAG/CU (Fig. 4D). The difference between the two binding sites is the position immediately downstream of the AG dinucleotide, also known as the +1 position. For the $U2AF^{26}$ binding site, the base at this position is cytosine, which is different than the guanine found at the same position of the $U2AF^{35}$ binding site. This suggests that the differences between the two proteins binding preference is at the +1 position and this specificity might indicate a preference of $U2AF^{26}/U2AF^{65}$ for a distinct subset of splice sites.

A.



Adapted from Kenan and Keene, Methods in Molecular Biology, Vol 118, pp. 217-231

B.



c.


35.1 UAAAACUUCACUCUUUCCCCUUUUUCUUUUUAGGUU
35.2 UUAACACAGUUGCCUUGAUUUUUUCCCAUACAGGUU
35.3 UUUUCUCCCCUUACGACUCCGAGUCCGGUUAAC
35.4 UUAUACUAUUGAAACGCUUUUUUCCCCGUCUAGGUU
35.5 UGUGCAGCGCCUUUUAAUCAAUACCUAGAAAUG
35.6 CAAAAACUUUUCCUUCCAGGUUUCCUCCCUAAG
35.7 UUUCCCUCAACAGCACAGAGCACACAAAACACG
35.8 UUAUACUAUUGAAACGCUUUUUUCCCCGUCUAGGUU
35.9 UAACCCACGCUUCUUCUAAUUUUUCCCAAACAGUUG
35.10 UUUUUUCCAUUUAUUUUCCCCGACCAGUUAUCAAC
35.11 UUUCCCAUUGUAGCUUAUCAGGCACUCCCUCUAGGUU
35.12 UCUUCCAGACCAGAUUUCUCUCCGCCGCGCAAGUUG
35.13 UUUUCCAUUCCCUUACACCAGCUCGGCCAGAAA
35.14 CCUUUCUUCCCCAAUAUUUUUUUUUCCCCGAACAGGUU
35.15 CUUAAUUCCUUUCCUUUCCCCUCACAUUCAGGUA
35.16 UUCUUAUCCUCUCCUCUUUCUCAUUUCUAGGUCAU
35.17 UCUCUCCUAAACUUCUCUUUCCUCCUCCUCUAGGUU
35.18 CCUUUCUCCCUCCUUCCCUCCCUAAGUCUCCCU
35.19 UUUUUUCCAUUUAUUUUCCCCGACCAGUUAUCAAC
35.20 UUCAUCCUUUCUUCUUUUCUAUUUUCCCUUACAGUUG
35.21 CUCUCCUUUUCUUUCCUCAAACACAGCGCGGAC
35.22 ACCCCCCCGUUUUUCUUUCCUCCCAAUAAGCAG
35.23 UAAGCAAUGUAUACCCUUUCUUUUCUCUUACAGGUU
35.24 UUACCCCCUUUUAUUUUUUACAGGAUUUCCAAA
35.25 UCUUUCCCUUUUCACUAGACAACAGCGCCAAUC
35.26 CAUCUCCUCUUCCUUUCCCUCUCACAUUCAGGUA
35.27 UUUUUAAUACAGGUAACGCACUUAUUCAGGCAA
35.28 AAGCUUUUUCCUCCCUCCAGUAACCCCGCACCU
35.29 UCUCUCCCAAUAACACCGCACUUGACCCGACGAGUUG
35.30 AUUCCUUUAUUCCUUCCCUUACUUUCUACUACAGUUG
35.31 UUCUUUUCCCUUCUUGCUUCCUCUUUUAGCCAU
35.32 UCCUUUCUUCCCUAUCCCUAUUAGCACUCAUGG
35.33 GUUCCCAACUUUUUUCUCUCCCUUUUCUCCGUAGUUG
35.34 UUUCCCCUUCACGCCGCCCUGUCUGUACGCUAGCGU
35.35 UUCUCCAUUCCCUCUUCUACAUUUCUUUUUUAGGUU
35.36 UCACAACCUUUCCCCCCACCAUGACCCCGCCUCACAGUUG
35.37 CUUUUCAUUCUUCUUACUAGCUCUAUUUCCGU
35.38 AUCCUCCCUCCUAAUAUUUCUCUCUCCACACUCUAGAGCUAUG
35.39 CUUUUUUUUCAAGUUACGAAGCGCUAGCACCGCGU
35.40 UUUUUCCUUCCACAGGUUUAGACUGACCCAUGA

D.

```
26.1                    UUCUCUCCUUUCCCUCUCCUAGACAUCCGAUGC
26.2          UUUUCUCCCCAAAUAUUUCCCCCCAUGUAGUCC
26.3          UCCGAUUUUCCCUUAUUUUUUCCCCGUUAGACA
26.4     GGUCCCGAAUCUGAUCCCACUCAUCCCCUACCAGUGU
26.5                   UUUUCCCUCCCUAUGCCGUCAGACAAGACACCC
26.6             ACACUCCAACCUUUUCCUUCCCUUUUAGCAUCA
26.7                       UUUUUCUUUUUCUCCCUUAGCAGUCUAACAUUCC
26.8              UUAUACCUAUCAAUUUUUUUUCCCAAGCCCAGC
26.9       CUUUCUCUUUCCCCUCUUUCAAUAAUUUUGCAGGUU
26.10                         UCUUUUCCCUUCCUCAGAAUAACGCCUGCCAU
26.11    AUAACUUUUUCCUUCCUUUUCUUUCUUACCCAAGUUG
26.12     UUUUCUUCCAAUUUUUUUCUUGUGGUGUUGCAGUUG
26.13           UCUUUCCUCCUCUUUCCGCCACGAUCAGCUGAU
26.14    UUUCUCCCUAUUCUUUUUCCCCUCUGACGCGUAGUUG
26.15  UAUCUCUCCACAGACUUCUAUGUCCACCAACCCAGCCUU
26.16                  UCAUUCUCCCAAUCACGCCCAAGAUACCCUGCCCCG
26.17                          UAUUUUCCCUUAACAGUCUGCACUACCACCAAA
26.18                    UUUUCCAUUAAACCCAUCUAGAUAGAACCCCCC
26.19                           UUUCCUCCUUCUUAGAUACGCACAUACACUAUU
26.20                    UUUCCUCCUCCUUUUUUUUAGCUAGCCCACUUC
26.21             UUUUCUUCCUUCCAUUUAACCGUGAAGGCACAG
26.22    UUUCCCUAACUCAUUCUUCUUUUUCUCCCAACAGUUG
26.23     CCUACUAAUCCCACCAUCUUUUUCCCCCGACAGGUU
26.24  AUGCUUUAAUUUCUUCCCUUUUUCUCCCUCACAAGUUG
26.25                 GUAUUUUCUUCCUUUUAUUCACAGCUCAGCACU
26.26                 UUAGGUCCCUGUGCCCUCCUACAGCUACGGAAU
26.27                 UUCACUCCUUACUUUUUUCCGUAGCUUCGAUCUG
26.28    AAAAAUCAUUCACUCGACAAAAUUCCUCUAAAAGGUU
26.29                       UUUUCCUAACACUAACCAAGCACCUCCUGUCCU
26.30                            UUUUCUUCCUUUUAGACUGAUGCAGUAGUUAUU
26.31                  CAUUCCUCCUACAUCAUUACCAGCACAACCCCG
26.32        UUUAUCCUUUCCUUCCCCUUUUACGACCAAGCUG
26.33    UUUUUUUCCCUUUUUCUUAUCGUCUACUGAUCAGGGU
26.34        UUUUUCCUUCUCGUCCACCUUACUUUUAUAGCAG
26.35        UUCUUUCCCACCUUUUCUCCACCACAAACAGCCG
26.36                       UUUCUUUCCUUAUGUCACAGACAGCACCUUCAAC
26.37 UUUUUUGCUUUUUCCCAUUACUUAUUACCCUUAUAGUUG
26.38      UUUCUUCCUUUUCUAUCCCCCAUUCUAACGAGUGU
26.39    UAUUUUUCUUUAUCGUAUUUUGCUACUCCCAUAGAGCU
26.40                       UUUUCUCUCCUUACUAACAGAACACCGAUAACCG
```

E.

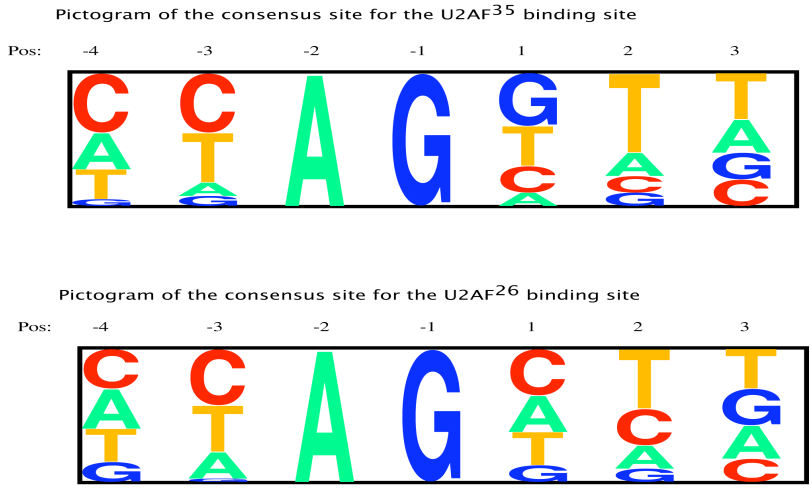


Figure 3.4: (A) Diagram of SELEX protocol adapted from Kenan and Keene (Kenan and Keene, 1999). (B) Gel shift analysis of the pools from the first four rounds of SELEX. After the fourth round, over 90% of the pool contained RNAs with binding sites preferred by the U2AF^{26}/U2AF^{65} heterodimer. (C) Alignment of forty unique clones from the SELEX analysis of U2AF^{35}/U2AF^{65}. The nucleotides highlighted in blue correspond to sequence of the template downstream of the randomized 33-mer. (D) Alignment of forty unique clones from the SELEX analysis of U2AF^{26}/U2AF^{35}. Again, the nucleotides highlighted in blue correspond to sequence of the template downstream of the randomized 33-mer. (E) Pictograms of U2AF^{35} and U2AF^{26} consensus binding sites from the SELEX experiment.

Site-specific crosslinking analysis of different permutations of the small subunit binding site. Next we sought to analyze different variations of the small subunit binding site to determine if there were differences in U2AF^{26} and U2AF^{35} binding affinities. This would allow us to directly test the differences between the consensus motifs for U2AF^{26} and U2AF^{35} identified using SELEX. Given that the position two bases downstream from the AG, or +2 position, is uracil in both motifs identified in the SELEX experiment, we decided to test all possible permutations at the –3 position and the +1 position of the binding site. There are sixteen possible variations of the binding site when varying these two positions (Fig. 3.5A).

We site-specifically labeled the sixteen variants and tested each one with both $U2AF^{26}$/$U2AF^{65}$ and $U2AF^{35}$/$U2AF^{65}$. We performed a binding experiment used by Wu *et al*. to analyze the binding affinity of $U2AF^{35}$ (Wu et al., 1999) (Fig. 3.5B). From this analysis, we determined the binding site recognized best by both $U2AF^{26}$ and $U2AF^{35}$ was UAG/GU (Fig. 3.5C). This is interesting because it is identical to the SELEX-derived optimal binding site previously described for $U2AF^{35}$ (Wu et al., 1999) and similar to the consensus site we observed for $U2AF^{35}$ for our SELEX analysis. However, $U2AF^{26}$ appeared to recognize the UAG/UU motif better than $U2AF^{35}$. This observation is interesting because it suggests that the difference between the two protein binding site preference may be at the +1 position and that $U2AF^{26}$ may recognize a uracil at this position better than $U2AF^{35}$.

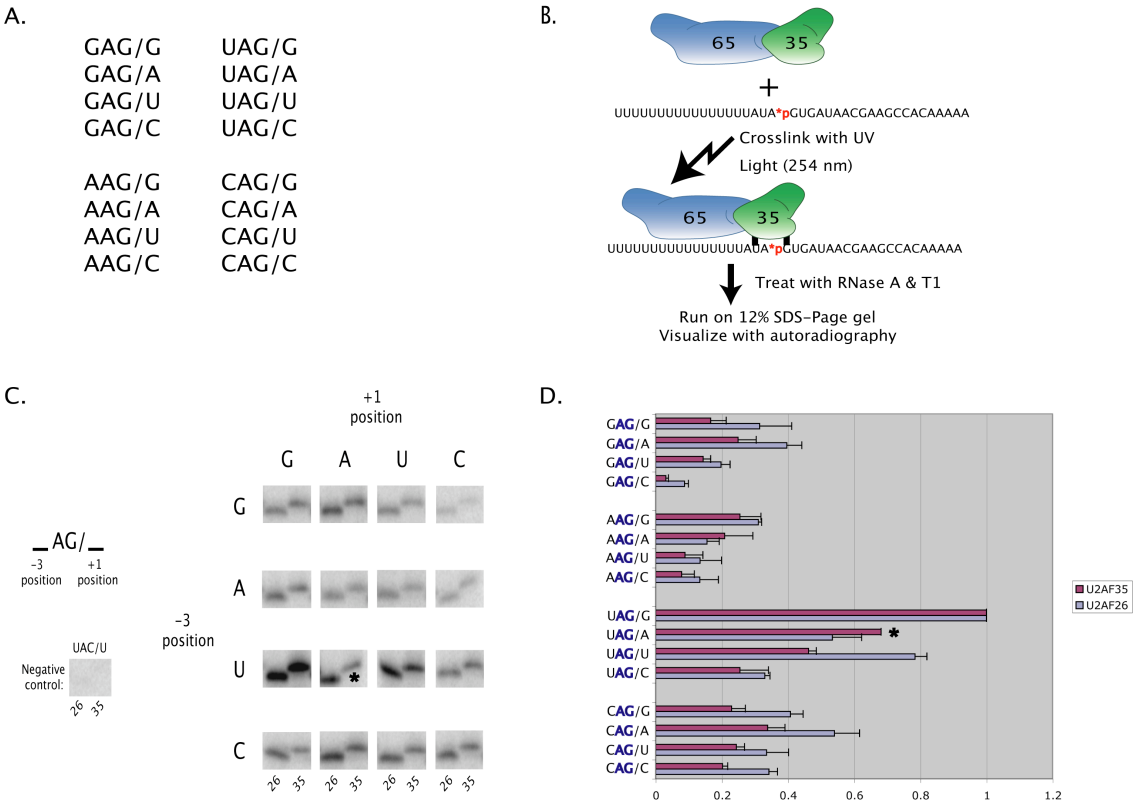


Figure 3.5: (A) List of the sixteen permutations of the 3' splice sites tested. (B) Diagram of crosslinking of site-specific labeled RNA. RNAs labeled with ^{32}P between the A and G nucleotides are incubated with recombinant heterodimer and then crosslinked with UV-light to bind the heterodimer to the RNA. The crosslinked complex is treated with RNase A and T1 and resolved on a 12% SDS-PAGE gel. (C) Results from crosslinking of sixteen different site-specific labeled RNAs. The RNA with a UAC/U small subunit binding site was used as a negative control to demonstrate the specificity of $U2AF^{26}$ and $U2AF^{35}$ for the AG dinucleotide. * This result is an outlier relative to the two other results. (D) Quantitation of the previous experiment. All values for the $U2AF^{26}$ and $U2AF^{35}$ heterodimer were normalized to the respective value for the binding to the UAG/G motif. * Only two results were averaged and the value for the one outlier was discarded.

DISCUSSION

The protein expression of the $U2AF^{26}$ relative to $U2AF^{35}$ varies markedly between different mouse tissues. Interestingly, $U2AF^{35}$ is expressed at the highest levels in tissues

(spleen and testis) that have a high rate of cell proliferation and U2AF^{26} is highest in the tissues (brain and lung) that are terminally differentiated. Also, the observation that expression of U2AF^{35} is highly expressed in the majority of cell lines tested is consistent with the correlation of high U2AF^{35} expression and a high rate of cell proliferation. Given the highly cooperative nature of spliceosome assembly, small changes in the expression of a single splicing factor could affect the ability of processing of regulated splice sites (Smith and Valcarcel, 2000). Therefore, differences in the expression of U2AF^{26} and U2AF^{35} could be representative of significant differences of the splicing of regulated introns in a tissue specific manner.

Attempts to overexpress U2AF^{26} in excess of U2AF^{35} were not successful. Using siRNAs to reduce U2AF^{35} expression in the clonal cell line that could conditionally overexpress U2AF^{26} indicated that we could make the U2AF^{26}:U2AF^{35} ratio roughly equal. When we analyzed known alternative splicing events under these conditions, we did not observe changes that could be correlated with U2AF^{26} up-regulation. Even when we knocked-down the expression of U2AF^{26} alone, we could not detect changes in the splicing events examined.

Phenotypic analysis of cells when U2AF^{26} is knocked-down indicates that there is no significant change in proliferation or cell cycle defects. However, when U2AF^{35} alone is reduced, there is a significant effect on both the proliferation rate and observable defects in the cell cycle of the targeted cells. The same phenotype for the U2AF^{35} was observed even with induction of overexpression or knock-down of U2AF^{26}. This suggests that U2AF^{26} does not have redundant function *in vivo*, since the phenotype observed when we

reduced U2AF^{35} and U2AF^{26} expression at the same time was nearly identical to the U2AF^{35} knock-down phenotype. However, the lack of phenotypic changes when U2AF^{26} expression was knocked down could be due to the low expression of U2AF^{26} relative to U2AF^{35} in wild-type cells.

SELEX experiments suggest that U2AF^{26} and U2AF^{35} have different binding affinities. The only difference between the consensus motifs identified for the splicing factors was the position immediately downstream from the AG (+1 position). U2AF^{26} prefers a cytosine at this position and a U2AF^{35} prefers a guanine at this position. This is significant because it suggests that U2AF^{26} may prefer 3' splice sites that are slightly different than U2AF^{35}.

When we tested binding affinities of the two small subunits directly with site-specific crosslinking, we found that U2AF^{26} and U2AF^{35} appear to have overlapping binding affinities. However, for one motif, UAG/UU, U2AF^{26} bound better than U2AF^{35}. It would be interesting to test observations from the SELEX and site-specific crosslinking experiments with *in vitro* splicing. It would be straight-forward to mutate the 3' splice site of an AG-dependent model substrate and determine if U2AF^{26} could preferentially recognize the specific splice sites identified in the SELEX and site-specific experiments. Previously, it was demonstrated that changing the +1 position of the IgM 3' splice site affected the ability of U2AF^{35} to bind that splice site and the splicing efficiency was reduced as well (Guth et al., 2001). This is an incomplete analysis of potential 3' splice sites, but it suggests that U2AF^{35} can distinguish between various sequences around AG dinucleotide.

One significant observation made in the site-specific experiments was that both U2AF^{26} and U2AF^{35} bound to the UAG/GU motif best. This suggests that within AG-dependent class of introns, different permutations of the sequence around the AG could directly affect the ability of U2AF^{26} and U2AF^{35} to recognize that site. Also, these observations suggest that the –3 and +1 position equally affect the small subunit binding to the 3' splice site.

This study suggests that recognition of the UAG/GU 3' splice sites that have weak pyrimidine tracts would be highly dependent on the level of expression of U2AF^{26} or U2AF^{35} in a given tissue. Potentially the UAG/UU motif could be affected by the level of U2AF^{26} expression in different tissues as well. From the phenotypic analysis of the knock down of U2AF^{35}, it is possible that certain cell cycle-dependent transcripts have such small subunit-dependent introns and the phenotype observed when we knocked-down U2AF^{35} could have been caused by the inefficient splicing of those introns.

MATERIALS AND METHODS

Western blot analysis of tissues and cell lines. Whole tissue protein preps were made by homogenizing the respective tissue in RIPA buffer. The homogenate was spun at 55,000 RPM in a TLA 100.3 rotor for one hour to produce S100 extracts. The extracts were standardized using the Bio-Rad Protein Assay. 40 μg of total protein was resolved on a 12% SDS-polyacrylamide gel and then transferred to Hybond-P hydrophobic polyvinylidene difluoride (PVDF) membrane. Once the protein was transferred to the

membrane, the membrane was blocked with 5% blotting grade blocker non-fat dry milk (Bio-Rad) for 1 hour at room temperature. The primary antibody used for $U2AF^{26}$ was the polyclonal antibody #3254 that was affinity purified with peptide 2237. The purified antibody was used at 1:100 in 1% milk and the membrane was probed for 1 hour at room temperature. The polyclonal-$U2AF^{35}$ was used at 1:3000 and the monoclonal-tubulin antibody was used at 1:1000. The membrane was rinsed 2X with TBS-T and then washed 3X TBS-T, 10 minutes for each wash. The blot was developed with an ECL Western Blotting Analysis System (Amersham Biosciences).

The same procedure was repeated for the cell lines, except that 2.5 μg of total cell protein was analyzed. Also, protein concentrations were first determined with Bio-Rad Protein Assay and then standardized by Coomassie-brilliant blue staining of 2.5 μg of each cell line on a SDS polyacrylamide gel.

Stable cell line production. Neuro2A, Raw 264.7 and C_2C_{12} cell lines were transfected with 80 μL of Lipofectamine (Invitrogen) and 4 μg of linearized pcDNA3-$U2AF^{26}$ in 100mm dishes. 24 hours after the start of the transfection, cells were split 1:10 and selection began 48 hours post-transfection with Geneticin (Invitrogen). The Neuro2A, Raw 264.7, and C_2C_{12} cell lines were selected in 400μg/mL of Geneticin for 2-3 weeks and then individual colonies were picked. Each individual colony was tested by the same western blot analysis described earlier.

Conditional cell line production. The Neuro2A cell line was co-transfected with Ecdysone-Inducible Expression System (Invitrogen) to conditionally overexpress $U2AF^{26}$. A 100 mm dish of Neuro2A cells was transfected with 5μg of linearized VgRXR and 5μg of linearized pIND-$U2AF^{26}$ plasmids with 80 μL of Lipofectamine. 24 hours post-transfection, the cells were split 1:10. After 48 hours, the cells were treated with full medium that contained 100μg/mL of Zeocin (Invitrogen) and 200 μg/mL of Geneticin. 2-3 weeks later, individual clones were picked and analyzed for induction of the $U2AF^{26}$ protein expression by western blot. Cells were induced with the ecdysone-analog ponasterone at 5 μM in complete medium 24 hours before analysis.

Semi-quantitative RT-PCR analysis. The RT-PCR technique involved analysis of total RNA purified from cell culture using RNA-STAT (Tel-Test). 500ng of RNA was reverse transcribed using transcript specific primers with the MMLV RT enzyme (Invitrogen). The cDNA was analyzed with PCR using a 5' end labeled primer set designed to the given alternative splicing event. The PCR reactions were mixed with equal volumes of formamide loading buffer and resolved on 5% denaturing polyacrylamide gels. The gel was fixed, dried down, and then analyzed by the typhoon phosphorimager (Molecular Dynamics). All gels were quantitated with ImageQuant (Molecular Dynamics).

RNA interference. One of the Neuro2A clones that conditionally overexpressed $U2AF^{26}$, clone #21, was transfected with siRNAs to knock-down expression of $U2AF^{26}$ and/or $U2AF^{35}$. The siRNAs were synthesized at the RNA Oligonucleotide Synthesis Core

facility at the UTSW North Campus. The siRNAs were annealed and stored as 20 μM stocks. The siRNAs were transfected using 2 μL (6 well plate) or 11 μL (100 mm dish) of Oligofectamine (Invitrogen) using OPTI-MEM. Complete medium with 3X the normal serum levels was added to the transfection medium 4 hours after transfection. 24 hours latter, the old medium was removed and replaced with fresh complete medium. Knock-down was confirmed by western blot analysis.

Proliferation assay. After transfection of siRNAs, cells were split and 40,000 cells were put into flat bottom 96 well plate for 4 hours. Next, 1 μL of [methyl-^{3}H] Thymidine (specific activity 5Ci/mmol, Amersham Biosciences) was added to each well and incubated at 37˚C for 16 hrs. The plate was removed and put at –20˚C until all time points were be collected. Each plate was harvested by a 24-well harvester and the labeled DNA was collected on glass microfiber Whatman paper (cat. 1827-887). Each filter paper was put into a plastic vial with 2 mLs of scintillation fluid and counted with a scintillation counter.

BrdU labeling and flow cytometry. After transfection of the siRNAs in 6-well plates, the cells were labeled with 10 μM of 5-bromodeoxyuridine for 45 minutes. After this incubation, the cells were trypsinized, permeabilized, stained with the anti-BrdU FITC antibody, and then treated with 7-amino-actinomycin D (7-AAD). The labeled cells were analyzed using flow cytometry (Becton Dickson FACscan) using CellQuest.

SELEX. A 33-mer random-nucleotide library was inserted in the template suggested by Kenan and Keene (Kenan and Keene, 1999). The 33-mer random-nucleotide library was synthesized by Integrated DNA Technology and to ensure accurate representation of each nucleotide in the library, the random library was synthesized with a hand-mixed cocktail. The production of the RNA library was done under the conditions recommended by Kenan and Keene (Kenan and Keene, 1999). The binding conditions were the same that were used by Wu *et al*. (Wu et al., 1999). The RT, PCR, and cloning techniques were the same as described by Kenan and Keene (Kenan and Keene, 1999).

Site-specific crosslinking analysis. The 3' halves of the RNAs were 5' end labeled with ^{32}P using T4 Polynucleotide Kinase (New England Biolabs). The labeled RNAs were phenol/chloroform precipitated and then ligated to the 5' end or specific RNAs as described previously (Moore and Sharp, 1992). The ligated RNAs were gel purified using a 10% denaturing polyacrylimide gel. The RNAs were eluted O/N at room temp in 0.5M sodium acetate and then phenol/chloroform precipitated. Each RNA was diluted to 30K CPM/μL and 1μL of each RNA was used for the binding reaction under the following conditions: 7.5 ng/μL of protein, 20 ng/μL of BSA, 1.3 % polyvinylalcohol, 1 mM DTT, 0.4 U of RNasin/μL, 3 ng/μL of tRNA, 5 mM $MgCL_2$, and 65 mM KCL. The binding reaction was incubated at 30˚C for 15 minutes and then transferred to 96 well round bottom plate. Next the reactions were treated with UV light (254nm) for 20 minutes on ice. The reaction was returned to the original 1.7 mL eppi tube and treated with RNase A & T1 (20U of each RNase/binding reaction) for 20 minutes at 37˚C. Then 12 μL of 2X

SDS sample buffer was added to each reaction and the binding reaction was resolved on a 12% SDS polyacrylamide gel. The gel was developed by the typhoon phosphoimager and quantitated with ImageQuant.

CHAPTER FOUR
Temporal- and Tissue-Specific Alternative Splicing of U2AF[26]

INTRODUCTION

An understanding of the pathway controlling mammalian circadian rhythm has dramatically increased during the last several years. The underlying molecular network of mammalian circadian rhythm has been elucidated by several discoveries, beginning with the identification of the positive circadian transcription factor, *Clock* (Antoch et al., 1997; King et al., 1997). Mammalian circadian rhythm is driven by the circadian oscillator, which is composed of positive and negative elements. The positive elements induce expression of the negative elements, which inhibit the activity of the positive elements. This negative transcription/translation feedback loop drives molecular circadian rhythm in mammals (Dunlap, 1999).

The rhythm generated by the circadian oscillator leads to rhythmic expression of specific genes over 24 hours. The rhythmic expression of these genes is thought to control the physiological changes associated with circadian rhythm. Analysis of the rhythmic genes in different mouse organs, such as the heart and liver (Panda et al., 2002; Storch et al., 2002), has been completed, and examining their function in the circadian change in physiology is the next step in understanding mammalian circadian rhythm.

A transcription factor enriched in neuronal tissues (NPAS2) has high sequence identity with Clock (Zhou et al., 1997). Functional analysis of NPAS2 through examination of putative target genes and NPAS2-deficient mice revealed that NPAS2 is a

functional analog of Clock (Reick et al., 2001). The identification of a transcription factor that can participate in a circadian oscillator as a positive element indicates that there may be another independent circadian loop in mammals. Studying the function of the NPAS2 target genes will help further our understanding of the function of this transcription factor and will prove to be critical for discovering the functional difference of NPAS2 and Clock.

The circadian rhythm in peripheral tissues (every tissue excluding the suprachiasmatic nucleus) can be entrained to a new rhythm with restricted feeding (Damiola et al., 2000; Stokkan et al., 2001). This suggests that circadian rhythm in the peripheral tissues might be driven by the metabolic state of the tissue. It has been demonstrated that the DNA binding activity of Clock and NPAS2 is influenced by the redox cofactors, NAD(H) and NADP(H), and this could be a molecular explanation for the change in circadian rhythm with restricted feeding.

In this study we examine the expression of a putative target gene of NPAS2, $U2AF^{26}$. We demonstrate that the full length transcript of $U2AF^{26}$ does not fluctuate in a circadian manner in the mouse forebrain. Interestingly, alternative splicing of the $U2AF^{26}$ mRNA changes over 24 hours and the alternative splicing of this transcript is advanced 12-hours in the forebrain with restricted feeding. Examination of the processing of the $U2AF^{26}$ transcript in the liver tissue indicates that the alternative splicing of this transcript fluctuates over 24 hours. Analysis of nine other alternative splicing events in the mouse liver indicates that the alternative splicing of the PTB transcript changes over 24 hours. However, the pattern of splicing of the two transcripts has a two-peaked pattern, different

from the one-peaked pattern that is the hallmark of circadian rhythm. Potentially, the alternative splicing of the U2AF^{26} and PTB transcripts is not linked to NPAS2 or circadian rhythm.

RESULTS

Examination of the alternative splicing of U2AF^{26} over 24-hours. A cDNA fragment of U2AF^{26} was identified in a representational difference analysis of cells that conditionally overexpress the NPAS2 transcription factor. This fragment was used to identify the full length U2AF^{26} cDNA (Shepard et al., 2002). Since U2AF^{26} shares high identity at the primary amino acid level with the splicing factor U2AF^{35}, we hypothesized that U2AF^{26} may regulate recognition of specific 3' splice sites in a circadian manner. If U2AF^{26} is linked to the NPAS2 circadian oscillator, then expression of the U2AF^{26} transcript would oscillate over a 24-hour period. To examine this possibility, we performed a RT-PCR/southern blot analysis of the full length U2AF^{26} transcript in poly-A$^+$ RNA isolated from mouse brain over 24-hours. Surprisingly, expression of the full length U2AF^{26} transcript changed very little over this period, however, the alternative splicing of the transcript changed dramatically (Fig. 4.1A). The isoform that changed in abundance skips exon 6 and 7 (Δ67) (Fig. 4.1C). We speculated that the alternative splicing of this transcript is a regulated splicing event and this regulation is linked to circadian rhythm.

It has been demonstrated that reverse feeding [feeding restricted to the light phase, opposite of when mice eat the majority of their food (Damiola et al., 2000)] of mice

advances the circadian rhythm 12 hours in the peripheral tissues (all tissues other than suprachiasmatic nucleus) (Damiola et al., 2000; Stokkan et al., 2001). We repeated this experiment and analyzed the alternative splicing of the U2AF^{26} transcript. The splicing of the transcript was also advanced 12-hours, suggesting that there was a link between circadian rhythm and the alternative splicing of U2AF^{26} (Fig. 4.1B).

To study this splicing event more closely, we performed semi-quantitative RT-PCR of the U2AF^{26} transcript. We performed this analysis on total RNA isolated from the somatosensory cortex of the mouse brain over 24 hours, when mice were fed *ad libidum*. The pattern of alternative splicing was similar to the one observed in the forebrain, but there was a second peak at the 2 A.M. time point (Fig. 4.1D). This is different from the pattern of splicing in Figure 4.1A. Either this is indicative of a significant difference in the regulation of splicing of the U2AF^{26} transcript in the cortex or this is a phenomenon specific to the technique of isolating and purifying the RNA for this experiment.

To determine if other alternative splicing events exhibit a temporally-regulated pattern of splicing, we examined seventeen more alternative splicing events in the same samples (Appendix A). No other splicing events changed over a 24-hour period. Notably, U2AF^{35} did not exhibit the same type of temporal alternative splicing and this splicing phenomenon does not apply for all U2AF small subunits in this region of the brain.

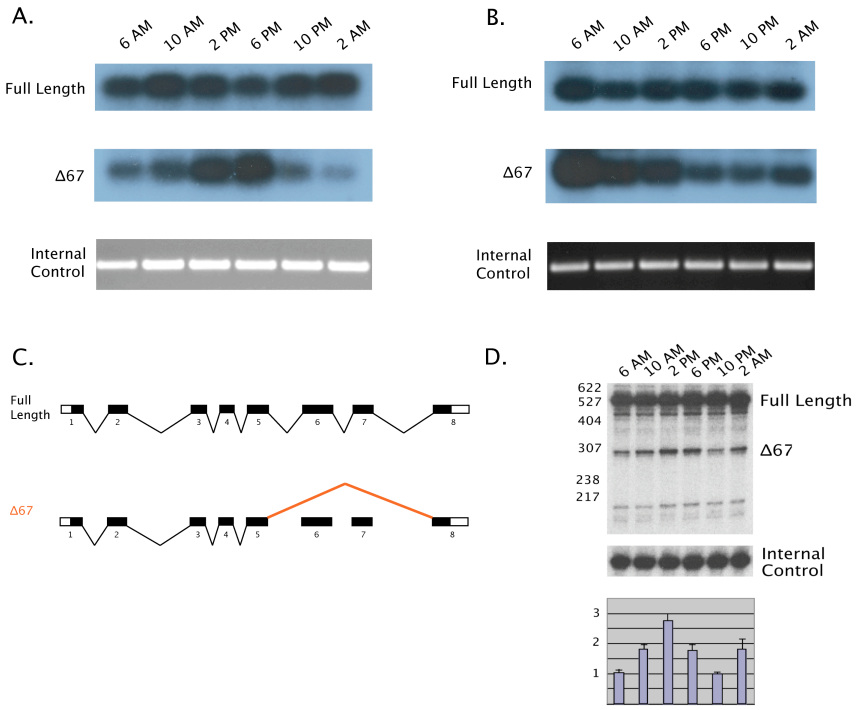


Figure 4.1: (A) RT-PCR/Southern blot analysis of the U2AF^{26} transcript. Mouse brain poly A^{+} RNA isolated every 4 hours over a 24-hour period was analyzed in this experiment. (B) The same RT-PCR/southern blot protocol was used as in (A). The mice were fed only during the daytime (6AM-6PM) and fasted at night for two weeks to advance their circadian rhythm. After the two weeks, the brains were isolated from the mice every 4 hours over a 24-hour period. (C) Diagram of the alternative isoform Δ67 relative to the full length isoform. (D) Semi-quantitative RT-PCR of U2AF^{26} transcript in the somatosensory cortex. Quantitation of the analysis is below the internal control.

Analysis of alternative splicing events in the mouse liver over 24 hours. Given the link between metabolism and circadian rhythm, we decided to test for temporally-regulated alternative splicing events in the mouse liver. We wanted to determine if the alternative splicing of U2AF^{26} and other regulated splicing events splice similarly to the pattern seen for U2AF^{26} in the brain. Semi-quantitative analysis of the U2AF^{26} transcript indicated that this transcript undergoes regulated alternative splicing over a 24-hour period in the liver (Fig 4.2A). However, the pattern of splicing and the isoform of the transcript

that fluctuated over this period is different than in the brain. The variant in the liver is missing part of exon 5, all of exon 6 and 7, and part of exon 8 (alt 567, Fig. 4.2B). There is an abundance of this alternative isoform during the night phase (6 P.M. to 6 A.M) and also at 10 A.M. The times the alt 567 isoform of U2AF^{26} is enriched is much different than times of enrichment of Δ67 in figure 4.1A and 4.1D. These observations suggest that the alternative splicing of the U2AF^{26} transcript in the liver is regulated in a different manner than in the brain.

We examined nine more alternative splicing events in the liver to determine if they were regulated in a temporal manner (Appendix A), and only one of the splicing events was temporally regulated. The polypyrimidine tract binding protein (PTB) transcript undergoes regulated alternative splicing over a 24-hour period (Fig 4.2C). This alternative splicing event involves skipping of exon 11 (trPTB) (Fig 4.2D), and the abundance of this isoform fluctuates up to ten-fold over 24 hours (Fig. 4.2C). Interestingly, the two-peaked pattern of the PTB transcript splicing is similar to the pattern for U2AF^{26}, which suggests that these two splicing events are coordinately regulated.

The skipping of exon 11 leads to a frame shift, thereby introducing several downstream in-frame premature termination codons (PTC) into the coding sequence. These in-frame stop codons lead to nonsense mediated decay (NMD) of the trPTB transcript. It was demonstrated that PTB is involved in the regulation of this splicing event (Wollerton et al., 2004). This suggests that PTB regulates the abundance of its own transcript in a temporal manner.

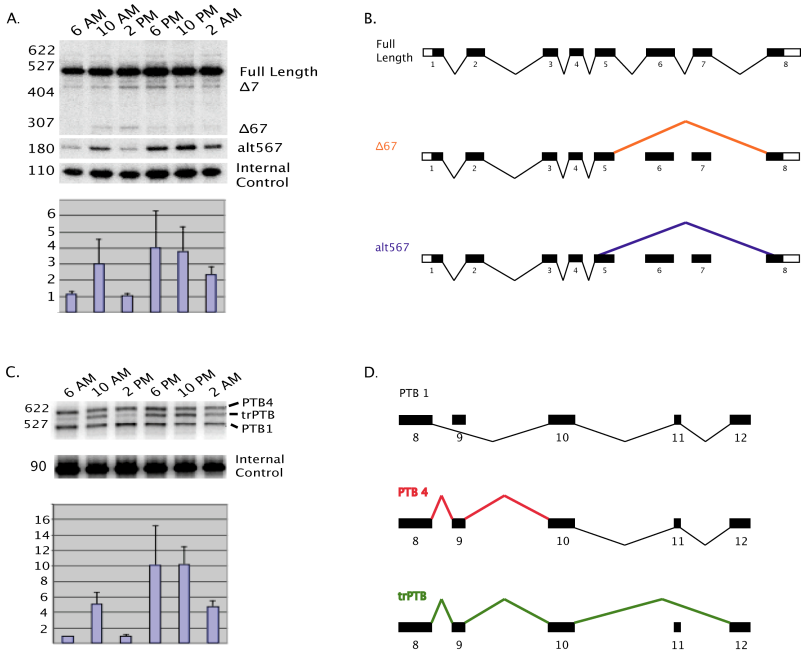


Figure 4.2: (A) Analysis of the alternative splicing of the U2AF^{26} transcript in the liver. Using semi-quantitative RT-PCR, we observed a novel isoform, alt 567, that changes in abundance over the day:night cycle. Quantitation of this isoform indicates that it fluctuates 4-fold over 24 hours. (B) Diagram of the alternative isoforms of the U2AF^{26} transcript. (C) Upon further analysis of known alternative splicing events, we observed a PTB isoform that fluctuates 10-fold in a correspondingly to the U2AF^{26} transcript. (D) Diagram of the region of pre-mRNA that undergoes alternative splicing to produce the different isoforms of PTB observed in (C).

Tissue-specific alternative splicing of U2AF^{26} and PTB. We hypothesized that U2AF^{26} and PTB are coordinately regulated. To test this hypothesis, we examined total RNA from two additional tissue sample and six different regions of the brain for changes in the processing of the U2AF^{26} and PTB transcripts. For the tissues tested, we only observed significant amounts of the alt 567 isoform of U2AF^{26} and trPTB isoform of PTB in the liver. We could not detect significant amounts of alt 567 in any region of the brain and there are only trace amounts of the trPTB isoform in tissues other than the liver. This is

not a comprehensive analysis, but this suggests that these two alternative splicing events are coordinately regulated.

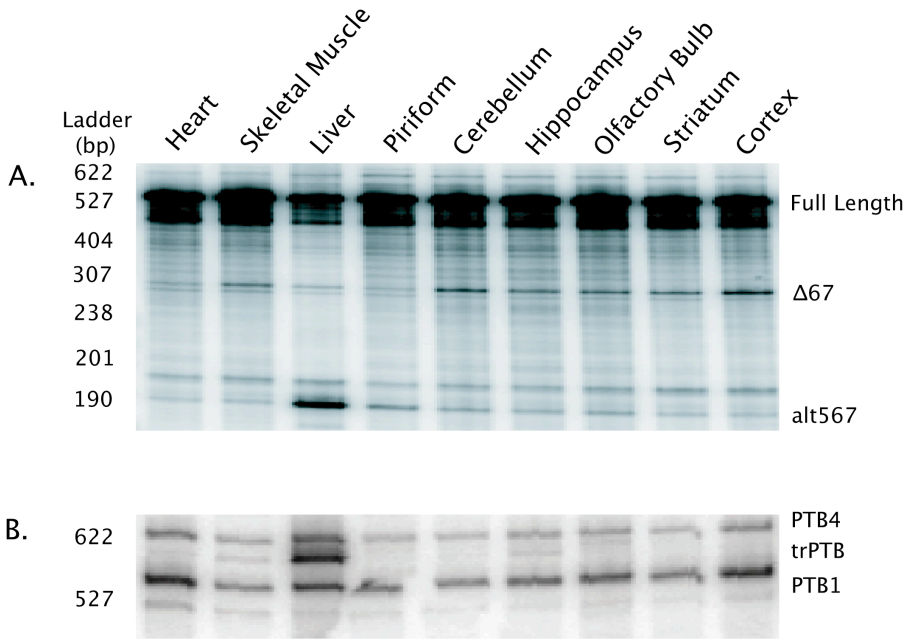


Figure 4.3: (A) Semi-quantitative RT-PCR of U2AF^{26} in different mouse tissues and regions of the brain. (B) Semi-quantitative RT-PCR of PTB in different mouse tissues and regions of the brain.

DISCUSSION

We examined U2AF^{26} expression over 24 hours because initial evidence suggested that it might be regulated by the positive circadian factor NPAS2. Expression of the full length transcript did not change significantly over 24 hours in the mouse forebrain, but the alternative splicing of the transcript did fluctuate over this period. Furthermore, regulation of this splicing event is affected in the mouse forebrain by restricted feeding.

Examination of this alternative splicing event in the somatosensory cortex revealed that this splicing event is temporally-regulated, but with a two-peaked pattern.

We examined seventeen more alternative splicing events in the mouse somatosensory cortex and none of the splicing events have a splicing pattern similar to U2AF^{26}. Next, we analyzed ten alternative splicing events in the liver, including U2AF^{26}. The splicing of the U2AF^{26} transcript was temporally-regulated in the liver. The timing of the splicing pattern was different than observed for the forebrain, but it had a two-peaked pattern, similar to the one observed in the somatosensory cortex. Also, the alternative splicing of PTB is temporally regulated and it exhibited the same splicing pattern as U2AF^{26} in the liver. Examination of two other tissues and five different regions of the brain indicate that the isoforms identified in the liver for U2AF^{26} and PTB are the most abundant in the liver. These observations suggest that these two splicing events may be coordinately regulated in the liver.

The two-peaked pattern of the U2AF^{26} transcript in the somatosensory cortex and the U2AF^{26} and PTB transcripts in the liver suggests that these splicing events are not influenced by circadian rhythm. However, a paper has been published that describes a two-peaked pattern of expression of the fibrinogen subunits in the liver (Sakao et al., 2003). Analysis of the expression of these subunits in the Clock mutant over 24 hours revealed that the expression is no longer rhythmic in the liver, suggesting that there is a link between expression of these genes and circadian rhythm. This studies provides an interesting observation, but without further study into the U2AF^{26} and PTB splicing phenomenon, there is only a correlation between the alternative splicing and the study.

Also, the two-peaked pattern of alternative splicing we observed in the somatosensory cortex and liver might be a by-product of the experimental method used for collecting the different tissues for analysis. We advanced one-half of the mice for two weeks to prepare the nighttime-points for harvesting during daytime hours. This might cause the two-peaked pattern observed for the processing of U2AF^{26} and PTB. Advancing the circadian rhythm of the mice by 12-hours requires the altering of the feeding schedule of the mice. This might cause the circadian controlled genes to undergo a two-peaked expression until the rhythm can be reach a steady-state. In another study, restricted feeding of mice altered the one-peaked expression of the circadian controlled gene, coumarin 7-hydroxylase (Cyp2a5) in the liver. After restricted feeding, the expression exhibited a two-peaked pattern. This suggests that restricted feeding can cause a two-peaked pattern of expression in liver. Since we collected the somatosensory cortex and liver samples at the time points from the same mice (independent of the forebrain samples), this observation may explain the two-peaked pattern of splicing. However, with the observations we have made about the two-peaked temporally-regulated splicing in the brain and liver, we cannot conclude that NPAS2 or circadian rhythm is linked to the alternative splicing of the U2AF^{26} and PTB transcripts.

The alternative splicing event observed for PTB, trPTB, has been described as product of the negative feedback loop between PTB and its transcript (Wollerton et al., 2004). The trPTB transcript is lacks exon 11, causing a frame-shift that introduces downstream in-frame PTCs. This causes the transcript to be targeted for degradation via nonsense mediated decay (NMD). This is the first well-defined example of regulated

unproductive splicing and translation (RUST), where changes in alternative splicing of a transcript lead to a decrease in translation competent transcripts and ultimately down regulation of expression of the protein.

35% of the EST-suggested human splice variants analyzed by Lewis *et al.* (n=8820) are predicted to contain in-frame PTCs, which would target the splice variants for destruction by NMD (Lewis et al., 2003). This suggests that RUST is a pervasive mechanism for gene regulation. What is interesting about our observations is that this type of gene regulation is temporally-regulated in the liver. Also, PTB is thought to be involved in the regulated alternative splicing of α-tropomyosin, β-tropomyosin, α-actinin, c-src, γ-2 $GABA_A$, clathrin light chain B, NMDA, FGFR-1 and -2, fibronectin, CASP-2, tau, and CT/CGRP reviewed in (Wagner and Garcia-Blanco, 2001). If PTB-mediated alternative splicing of the PTB-transcript fluctuates over 24 hours in the liver, then other alternative splicing events affected by PTB may fluctuate with the same pattern. Furthermore, alternative splicing events regulated by PTB, and potentially other splicing factors, could be temporally-regulated by RUST as well.

MATERIALS AND METHODS

Semi-quantitative RT-PCR analysis. The RT-PCR technique involved analysis of total RNA purified from cell culture using RNA-STAT (Tel-Test). 500ng of RNA was reverse transcribed using transcript specific primers with the MMLV reverse transcriptase enzyme (Invitrogen). The cDNA was analyzed with PCR using a 5' end labeled primer set

designed to the given alternative splicing event. The PCR reactions were mixed with equal volumes of formamide loading buffer and then resolved on 5% denaturing polyacrylamide gels. The gel was fixed, dried down, and then analyzed by the typhoon phosphorimager (Molecular Dynamics). All gels were quantitated with ImageQuant (Molecular Dynamics).

RT-PCR/Southern blot. cDNA for the mouse poly-A^+ RNA samples for the corresponding time points was produced using the Superscript II kit (Invitrogen) with a 15-mer oligo dT primer. Next PCR using primers specific to the cDNA was carried out at 18 cycles for the full length and 22 cycles for the Δ67 isoform. The PCR products were resolved on a 2% agarose gel and then transferred overnight to a nitrocellulose membrane. The splice variant specific oligo was 5' end labeled with ^{32}P. The membrane was pre-hybed overnight at 45˚C and then the oligo was added to the hybridization buffer at 1M CPM/ml of hybridization buffer. The blot was hybed at 45˚C for 8 hours. Then the blot was washed 2X with 2X SSC, 0.1%SDS for 20 min at 25˚C and then 2X with 0.2X SSC, 0.1% SDS at 45˚C for 15 min. The blot was developed and analyzed with a Fuji BAS1500 phosphorimager.

CHAPTER FIVE
Conclusion and Future Directions

The U2 auxiliary factor is an essential splicing factor that binds the 3' splice site and recruits the U2 snRNP to the branch point sequence. U2AF is a heterodimer composed of a large and small subunit. To date, one large subunit and five small subunits have been described. The large subunit, $U2AF^{65}$, binds the pyrimidine tract of the 3' splice site (Singh et al., 1995; Zamore et al., 1992) and small subunit, $U2AF^{35}$, recognizes the AG dinucleotide of the 3' splice site (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999a). Since there are five different small subunits, there could be five distinct heterodimers that preferentially recognize different 3' splice sites. $U2AF^{65}$ is promiscuous in nature, because it recognizes pyrimidine rich regions independent of sequence, therefore the small subunit might be responsible for the specificity of the U2AF heterodimer. Determining the difference in specificity between the small subunits will be critical to understanding the functional significance of the small subunits.

Characterization of $U2AF^{26}$. We decided to study the function of the mouse $U2AF^{35}$ paralog, $U2AF^{26}$. $U2AF^{26}$ can associate with $U2AF^{65}$ to form a heterodimer functionally similar to the $U2AF^{35}$/$U2AF^{65}$ heterodimer. Also, $U2AF^{26}$ can aid $U2AF^{65}$ binding to weak pyrimidine tracts, in a manner similar to $U2AF^{35}$. These observations confirm that $U2AF^{26}$ is a functional splicing factor.

Several observations about U2AF^{26} suggest that there may be *in vivo* functional difference between the two subunits. One is that there are two amino acid substitutions within RNP-2, a domain of U2AF^{26} predicted to directly contact RNA (Kielkopf et al., 2001). Another is that the α-helix A of U2AF^{26} and U2AF^{35} are quite different. The crystal structure of U2AF^{35} revealed that the surface of this long amphipathic helix is coated negatively charged amino acids (Kielkopf et al., 2001) and this region could potentially interact with the basic RS domain found of SR proteins. On the surface of helix A of U2AF^{26}, there are two amino acid substitutions, Gln80 is changed to Asn and Glu90 is changed to a Gln (relative to U2AF^{35}), which reduce the net negative charge.

Additionally, there are three other amino acid substitutions within the U2AF^{26} helix A that could be functionally significant. L65S, R66H, and A68H (substitutions relative to U2AF^{35} at the same position) are on or near the surface of the U2AF^{26} helix and they could carry a negative charge if phosphorylated. Also, since these amino acids might be phosphorylated by a specific kinase(s), this might be a point of regulation for the ability of U2AF^{26} to interact with SR proteins. If this helix if important for the interaction with other SR proteins, then U2AF^{26} could potentially interact with SR proteins in a distinct manner.

The C-terminal region of U2AF^{26} and U2AF^{35} are divergent. U2AF^{26} lacks several of the RS, RE, and RD dipeptides found within the RS domains of SR proteins (Graveley, 2000). Also, all metazoan U2AF^{35} orthologs have a glycine rich region within the C-terminus, which is typically a long stretch of consecutive glycines. Mammalian U2AF^{35}

orthologs all have twelve consecutive glycine residues that interrupt the RS domain. U2AF^{26} completely lacks the twelve consecutive glycines. The function of the glycine tract is unknown, but given the strong conservation it might be important for the function of U2AF^{35}. These C-terminal differences might confer a unique *in vivo* function for U2AF^{26}.

We wanted to determine if there were any differences in protein expression of U2AF^{26} and U2AF^{35}. Western blot analysis revealed varying levels in expression of the two subunits between different mouse tissues. Notably, there is higher expression of U2AF^{35} in tissues with high levels of cellular proliferation and the highest levels of expression of U2AF^{26} in terminally differentiated tissues.

Functional analysis of U2AF^{26} and U2AF^{35}. RNA interference of U2AF^{35} led to cell cycle defects and reduction of the proliferation rate of a cell line. This suggests U2AF^{35} plays a role in cell division and this is consistent with the observations made with the western blot. Knock-down of U2AF^{26} failed to induce a significant phenotype of the targeted cells. Also, RT-PCR of eighteen known alternative splicing events were not affected by the loss of U2AF^{26}. From these experiments we could not infer any *in vivo* function for U2AF^{26}.

We decided to examine the binding affinity of U2AF^{26} and U2AF^{35} with two *in vitro* binding experiments. First we used SELEX to determine the optimal binding site of U2AF^{26}. Previously the consensus sequence for purified HeLa U2AF^{35} was determined to be UAG/GU (Wu et al., 1999) and we used the recombinant U2AF^{35}/U2AF^{65} heterodimer

as a control for our SELEX experiment. We identified the consensus binding site for recombinant U2AF^{35} to be CAG/GU, which demonstrated that our experimental results are consistent with the experiment using purified HeLa U2AF^{35}. For U2AF^{26}, we identified the consensus binding motif to be CAG/CU. Interestingly, of the forty independent clones from the U2AF^{26} selection, the least common base immediately downstream of the AG (+1 position) is guanine. This could be significant because statistical analysis of a pool of mammalian 3' splice sites revealed that a pyrimidine is present at this position in only 24% of the 3' splice sites (n=1432) (Shapiro and Senapathy, 1987).

Next, we performed site-specific crosslinking to directly test the binding affinity of the two subunits. We tested all sixteen possible permutations of a 3' splice site when co-varying the sequence at the –3 and +1 position. We tested each one with U2AF^{26} and U2AF^{35} individually, and then scored for the relative binding affinity of each splice site sequence. UAG/GU was the motif that both subunits bound to with the highest affinity. This is very close to the consensus sequence we identified for U2AF^{35} using SELEX. Although U2AF^{26} bound best to the same motif as U2AF^{35}, U2AF^{26} bound to the UAG/UU motif better than U2AF^{35}. This suggests that U2AF^{26} may have a similar binding affinity as U2AF^{35}, but it might be able to recognize specific sequences better than U2AF^{35}. This could lead to better recognition of certain 3' splice sites with the UAG/UU motif in tissues with high U2AF^{26} expression.

It has been demonstrated that U2AF^{35} specifically recognizes the AG dinucleotide of the 3' splice site (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal,

1999b). These studies demonstrated that U2AF^{35} bound to RNA and required the AG dinucleotide for this interaction. The binding affinity of U2AF^{35} has previously been examined for only two different permutations of the 3' splice site. It was observed that U2AF^{35} binding was less efficient when the CAG/GU motif was changed to CAG/CU (Guth et al., 2001). This suggests the U2AF^{35} can distinguish between different sequences at the 3' splice site and it prefers a guanine at the +1 position. Our analysis of the sixteen permutations of the 3' splice site is consistent with this experiment, but we conclude the –3 position and the +1 are equally important to the binding of U2AF^{35}. Also, this is the first time it has been demonstrated that U2AF^{35} prefers a specific site to all other possible 3' splice sites.

Significance of the specific binding affinities of the two small subunits. U2AF^{65} has a low affinity for pyrimidine tracts that are short and/or are interrupted by several purines. For introns with weak pyrimidine tracts, the U2AF^{35} recognition of the AG is required for efficient U2AF binding to these sites and these introns are termed AG-dependent introns (Guth et al., 2001; Wu et al., 1999). The site-specific crosslinking results we present suggest that not all AG-dependent introns are the equal. Since U2AF^{26} and U2AF^{35} recognize the UAG/GU motif best, we propose that there are small subunit-dependent introns within the class of AG-dependent introns.

Statistical analysis of a pool of mammalian 3' splice sites indicates that uracil is present at the –3 position for 22% of all 3' splice sites examined and there is a guanine at the +1 position for 50% of the 3' splice sites (Shapiro and Senapathy, 1987). Also, uracil

is present in about 25% of all 3' splice sites at the +2 position (Stamm et al., 1994). If these three positions (the –3, +1, and +2 positions) co-vary independently, the UAG/GU motif would be rare. Additionally, it is would be even more unusual to have a UAG/GU motif combined with a weak pyrimidine tract. Given our site-specific crosslinking results, introns that have 3' splice sites with a weak pyrimidine tract and the UAG/GU would be predicted to be highly dependent on $U2AF^{26}$ and/or $U2AF^{35}$ for efficient recognition.

One interesting observation from the site-specific crosslinking experiments is that $U2AF^{26}$ can recognize the UAG/UU motif better than $U2AF^{35}$. Potentially $U2AF^{26}$ can recognize 3' splice sites with this motif and a weak pyrimidine tract better than $U2AF^{35}$. This suggests that there could be small subunit-dependent introns as well as $U2AF^{26}$-dependent introns. Given that there are two other $U2AF^{35}$ homologs (Tronchere et al., 1997) and one other isoform of $U2AF^{35}$, $U2AF^{35}$b (Pacheco et al., 2004), there could be other AG-dependent introns that preferentially recognized by the other forms of the small subunit.

If the small subunits preferentially recognize specific 3' splice sites, then the expression level of the small subunits could directly affect the splicing ability of introns within transcripts that contain those splice sites. The small subunit could be a significant factor in the "cellular code" and affect the steady-state of certain transcripts. Interestingly, one transcript that has UAG/GU motif in conjunction with a weak pyrimidine tract is the $U2AF^{65}$ pre-mRNA. There are ten introns in this pre-mRNA and only one has this site. This suggests that $U2AF^{26}$ or $U2AF^{35}$ could feedback on the transcript of $U2AF^{65}$ and increase the efficiency of splicing of the intron. Western blot analysis of $U2AF^{65}$ indicates

that the expression of the large subunit is highest in tissues that have the highest expression of U2AF^{26} and/or U2AF^{35} (data not shown), which is consistent with our prediction.

This is the first time it has been demonstrated that the nucleotide upstream and downstream of the AG dinucleotide can affect the small subunit binding to a 3' splice site. This suggests that certain motifs are "stronger" than others. It is possible to predict if a 5' splice site or pyrimidine tract will be recognized by U1snRNP and U2AF^{65} respectively by how closely they follow the consensus motif identified for those sites. However, an optimal sequence around the AG dinucleotide has not been described. Now, it may be possible to predict the ability of motifs to be recognized by the small subunit and this would be extremely important for 3' splice sites with weak pyrimidine tract. These studies advance our understanding of U2AF^{26} and U2AF^{35} recognition of the 3' splice site.

Temporal- and tissue-specific splicing. RT-PCR analysis indicates that the alternative splicing of U2AF^{26} is temporally regulated in the mouse forebrain. This splicing event is influenced by restricted feeding, suggesting that metabolism can affect the processing of this splicing event. Semi-quantitative RT-PCR analysis of U2AF^{26} in the somatosensory cortex indicates that U2AF^{26} is regulated in a temporal manner in this region of the brain. Interestingly, the pattern exhibited a two-peaked pattern of alternative splicing in the somatosensory cortex, different from the one-peaked pattern observed in the forebrain.

No other splicing events tested in the somatosensory cortex appeared to be regulated over 24 hours. We examined the splicing of the U2AF^{26} transcript in the liver and observed that the alternative splicing in this tissue was regulated over 24 hours.

However, the timing of the splicing pattern was different than the pattern in the somatosensory cortex, but it did exhibit a two-peaked pattern. We then examined nine more alternative splicing events in the liver and noticed that one, PTB, had a similar pattern as U2AF^{26} in the liver. Analysis of the U2AF^{26} and PTB transcripts in two other tissues and five more regions of the brain indicates that the alternative isoforms of U2AF^{26} and PTB are specific to the liver. This suggested that these two alternative splicing events are coordinately regulated. More thorough analysis of other tissues is needed to make a stronger conclusion about the specificity.

Interestingly, the isoform of the PTB transcript, trPTB, that fluctuates over 24 hours, skips exon 11. This splice variant of PTB introduces several in-frame premature termination codons (PTC), which leads to the degradation of the transcript via nonsense mediated decay (NMD) (Wollerton et al., 2004). The regulation of this alternative splicing event involves the PTB protein. PTB binds to the pyrimidine tract upstream of the 3' splice site of exon 11 and inhibits the recognition of this splice site by the U2AF heterodimer (Wollerton et al., 2004). This negative feedback loop controls expression of the transcript through alternative splicing and NMD. This is the first characterized example of regulated unproductive splicing and translation (RUST). Since we have confirmed that the alternative splicing event of PTB that produces the exon 11-skipped transcript is temporally-regulated, this observation suggests that control of gene expression via RUST can change over 24 hours in the liver.

Not only is this first time anyone has demonstrated that alternative splicing is temporally-regulated, this also the first time that this type of gene expression in the liver

has been described. Since 35% of EST-suggested human splice variants contain in frame PTCs (Lewis et al., 2003), other genes could be regulated in this manner. RUST is an unappreciated mechanism of gene expression and this observation suggests that this type of gene mechanism may be more dynamic than previously known. This highlights another powerful affect alternative splicing has over the proteome.

Significance of temporal changes of splicing of PTB transcript over 24 hours. The temporal changes in the splicing of the PTB transcript and the U2AF^{26} transcript in the liver mirror one another. The fact that the changes in the splicing occur at the same time suggests that the two events are co-regulated. It has been described that the PTB protein feeds-back on its own transcript to produce the trPTB transcript. This suggests that PTB may be responsible for the changes in the splicing of the U2AF^{26} transcript over 24 hours.

The binding site for PTB has been described to be UCUU (within a pyrimidine rich context) (Perez et al., 1997). The U2AF^{26} transcript that fluctuates over 24 hours in the liver is alt567, which is missing most of exon 5, all of exon 6 and 7, and part of exon 8. There is one UCUU motif in the pyrimidine tract of exon 6 and two UCUU motifs in the pyrimidine tracts of exon 7 and exon 8. This suggests that PTB might be responsible for the alt567 transcript in the liver. More analysis will be necessary to confirm the relationship between PTB and the temporal changes of the U2AF^{26} transcript.

FUTURE DIRECTIONS

The site-specific crosslinking results suggest that the sequence around the AG impacts the small subunit recognition of the 3' splice site. Functional analysis of the site-specific crosslinking results will need to be completed to conclude if the binding differences between motifs is significant. There are two AG-dependent model substrates that have been characterized (Guth et al., 1999; Wu et al., 1999). Mutating the nucleotides upstream and/or downstream of the AG of these templates could affect the splicing efficiency due to reduced small subunit binding. This would provide a functional assay to test these motifs and provide another proven experiment to confirm these findings.

It would be interesting to determine if there are differences in protein-protein interactions between $U2AF^{26}$ and $U2AF^{35}$. Potentially $U2AF^{26}$ could interact with a set of SR proteins in a manner different from $U2AF^{35}$ and the differences between the two proteins in helix A suggest that this could be a possibility. A two-hybrid analysis could identify specific SR protein(s) or that could interact with $U2AF^{26}$. This could provide insight into the *in vivo* function of $U2AF^{26}$.

It would be interesting to determine if the +2 position can influence the binding of the small subunits. The SELEX results indicate that a uracil is preferred at this position for $U2AF^{26}$ and $U2AF^{35}$. The –3 and +1 position appear to equally influence the binding of the small subunits, so it would be interesting to determine if the +2 position is equally important. Creating other permutations of the UAG/GU motif at the +2 position and then testing binding with site-specific crosslinking would indicate if this position affects

binding affinity. Since the consensus binding site for U2AF^{26} and U2AF^{35} determined by SELEX has 5 nucleotides, analysis of the +2 position would allow for complete examination of the sequence positions recognized by the small subunits around the AG dinucleotide.

There are three other variants of the small subunit. It would be interesting to determine if these small subunits have unique binding affinities using SELEX and site-specific crosslinking. This would indicate if the other subunits do have any functional significance and if they regulate distinct 3' splice sites. If the other small subunits do recognize a specific motif, this could be followed up with site specific crosslinking with *in vitro* splicing assays.

The fluctuation in the abundance of the PTB isoform missing exon 11 suggests that PTB protein expression changes over 24 hours. There are fourteen alternative splicing events influenced by PTB (Wagner and Garcia-Blanco, 2001). It would be interesting to examine these alternative splicing events over 24 hours in the liver to determine if these alternative splicing events fluctuate in a manner similar to the PTB transcript. This is a direct method to examine more alternative splicing events that could fluctuate over 24 hours in the liver.

APPENDIX A

Alternative splicing events examined in the stable cell lines, conditional cell lines, and the mouse somatosensory cortex:

1. U2AF^{26}
2. Polypyrimidine Tract Binding protein
3. Clathrin Light Chain B
4. cAMP Responsive Element Binding Protein
5. U2AF^{35}
6. cAMP Responsive Element Modulator 5' end
7. cAMP Responsive Element Modulator 3' end
8. Neural Cell Adhesion Molecule
9. SLO gene-STREX exon
10. N-Methyl-D-Aspartate Receptor 1
11. c-Src
12. Agrin
13. Neurexin #1 splice site #1
14. Neurexin #2 splice site #1
15. Dopamine D3 Receptor
16. Dopamine D2 Receptor
17. Neurotrophin 3
18. Tau

Alternative splicing events examined in the liver:

1. U2AF^{26}
2. Polypyrimidine Tract Binding protein
3. Neuronal PAS domain protein
4. Clock
5. U2AF^{35}
6. cAMP Responsive Element Modulator 5' end
7. cAMP Responsive Element Modulator 3' end
8. cAMP Responsive Element Binding
9. Tau
10. Neuronal Polypyrimidine Tract Binding protein

BIBLIOGRAPHY

Amann, B. T., Worthington, M. T., and Berg, J. M. (2003). A Cys3His zinc-binding domain from Nup475/tristetraprolin: a novel fold with a disklike structure. Biochemistry *42*, 217-221.

Antoch, M. P., Song, E. J., Chang, A. M., Vitaterna, M. H., Zhao, Y., Wilsbacher, L. D., Sangoram, A. M., King, D. P., Pinto, L. H., and Takahashi, J. S. (1997). Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. Cell *89*, 655-667.

Berget, S. M. (1995). Exon recognition in vertebrate splicing. J Biol Chem *270*, 2411-2414.

Birney, E., Kumar, S., and Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res *21*, 5803-5816.

Black, D. L., and Grabowski, P. J. (2003). Alternative pre-mRNA splicing and neuronal function. Prog Mol Subcell Biol *31*, 187-216.

Blencowe, B. J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci *25*, 106-110.

Bouck, J., Fu, X. D., Skalka, A. M., and Katz, R. A. (1998). Role of the constitutive splicing factors $U2AF^{65}$ and SAP49 in suboptimal RNA splicing of novel retroviral mutants. J Biol Chem *273*, 15169-15176.

Burge, C. B., T. Tuschl, and P.A. Shard (1999). Splicing of precursors to mRNAs by the spliceosomes. In The RNA world, T. R. C. R.F. Gestland, and J.F. Atkins, ed. (Cold Springs Harbor, Cold Springs Harbor Press), pp. p. 525-560.

Caceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L., and Krainer, A. R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J Cell Biol *138*, 225-238.

Chandler, S. D., Mayeda, A., Yeakley, J. M., Krainer, A. R., and Fu, X. D. (1997). RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. Proc Natl Acad Sci U S A *94*, 3596-3601.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., and Duncan, P. I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. Embo J *15*, 265-275.

Damiola, F., Le Minh, N., Preitner, N., Kornmann, B., Fleury-Olela, F., and Schibler, U. (2000). Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. Genes Dev *14*, 2950-2961.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol Cell *5*, 779-787.

Deutsch, M., and Long, M. (1999). Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res *27*, 3219-3228.

Du, C., McGuffin, M. E., Dauwalder, B., Rabinow, L., and Mattox, W. (1998). Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in Drosophila. Mol Cell *2*, 741-750.

Dunlap, J. C. (1999). Molecular bases for circadian clocks. Cell *96*, 271-290.

Fu, X. D. (1995). The superfamily of arginine/serine-rich splicing factors. Rna *1*, 663-680.

Fu, X. D., and Maniatis, T. (1992). The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proc Natl Acad Sci U S A *89*, 1725-1729.

Gama-Carvalho, M., Carvalho, M. P., Kehlenbach, A., Valcarcel, J., and Carmo-Fonseca, M. (2001). Nucleocytoplasmic shuttling of heterodimeric splicing factor U2AF. J Biol Chem *276*, 13104-13112.

Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Mol Cell Biol *18*, 4752-4760.

Grabowski, P. J., and Black, D. L. (2001). Alternative RNA splicing in the nervous system. Prog Neurobiol *65*, 289-308.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. Rna *6*, 1197-1211.

Graveley, B. R., Hertel, K. J., and Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. Embo J *17*, 6747-6756.

Graveley, B. R., Hertel, K. J., and Maniatis, T. (2001). The role of $U2AF^{35}$ and $U2AF^{65}$ in enhancer-dependent splicing. Rna *7*, 806-818.

Gui, J. F., Lane, W. S., and Fu, X. D. (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature *369*, 678-682.

Guth, S., Martinez, C., Gaur, R. K., and Valcarcel, J. (1999). Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). Mol Cell Biol *19*, 8263-8271.

Guth, S., Tange, T. O., Kellenberger, E., and Valcarcel, J. (2001). Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. Mol Cell Biol *21*, 7673-7681.

Hertel, K. J., and Maniatis, T. (1998). The function of multisite splicing enhancers. Mol Cell *1*, 449-455.

Jumaa, H., Wei, G., and Nielsen, P. J. (1999). Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. Curr Biol *9*, 899-902.

Kan, J. L., and Green, M. R. (1999). Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. Genes Dev *13*, 462-471.

Kanaar, R., Roche, S. E., Beall, E. L., Green, M. R., and Rio, D. C. (1993). The conserved pre-mRNA splicing factor U2AF from Drosophila: requirement for viability. Science *262*, 569-573.

Katz, R. A., and Skalka, A. M. (1990). Control of retroviral RNA splicing through maintenance of suboptimal processing signals. Mol Cell Biol *10*, 696-704.

Kenan, D. J., and Keene, J. D. (1999). In vitro selection of aptamers from RNA libraries. Methods Mol Biol *118*, 217-231.

Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core $U2AF^{35}$/$U2AF^{65}$ heterodimer. Cell *106*, 595-605.

King, D. P., Zhao, Y., Sangoram, A. M., Wilsbacher, L. D., Tanaka, M., Antoch, M. P., Steeves, T. D., Vitaterna, M. H., Kornhauser, J. M., Lowrey, P. L., *et al*. (1997). Positional cloning of the mouse circadian clock gene. Cell *89*, 641-653.

Lalioti, M. D., Gos, A., Green, M. R., Rossier, C., Morris, M. A., and Antonarakis, S. E. (1996). The gene for human U2 snRNP auxiliary factor small 35-kDa subunit (U2AF1) maps to the progressive myoclonus epilepsy (EPM1) critical region on chromosome 21q22.3. Genomics *33*, 298-300.

Lewis, B. P., Green, R. E., and Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A *100*, 189-192.

Li, Y., and Blencowe, B. J. (1999). Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. J Biol Chem *274*, 35074-35079.

Longman, D., Johnstone, I. L., and Caceres, J. F. (2000). Functional characterization of SR and SR-related genes in Caenorhabditis elegans. Embo J *19*, 1625-1637.

Lynch, K. W., and Maniatis, T. (1996). Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. Genes Dev *10*, 2089-2101.

MacMorris, M. A., Zorio, D. A., and Blumenthal, T. (1999). An exon that prevents transport of a mature mRNA. Proc Natl Acad Sci U S A *96*, 3813-3818.

Maniatis, T., and Tasic, B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature *418*, 236-243.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF^{35} and the 3' splice site AG. Nature *402*, 838-841.

Misteli, T., and Spector, D. L. (1998). The cellular organization of gene expression. Curr Opin Cell Biol *10*, 323-331.

Moore, M. J. (2000). Intron recognition comes of AGe. Nat Struct Biol *7*, 14-16.

Moore, M. J., and Sharp, P. A. (1992). Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. Science *256*, 992-997.

Nei, M., Xu, P., and Glazko, G. (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc Natl Acad Sci U S A *98*, 2497-2502.

Pacheco, T. R., Gomes, A. Q., Barbosa-Morais, N. L., Benes, V., Ansorge, W., Wollerton, M., Smith, C. W., Valcarcel, J., and Carmo-Fonseca, M. (2004). Diversity of vertebrate splicing factor U2AF^{35}: Identification of alternatively spliced U2AF1 mRNAs. J Biol Chem.

Panda, S., Antoch, M. P., Miller, B. H., Su, A. I., Schook, A. B., Straume, M., Schultz, P. G., Kay, S. A., Takahashi, J. S., and Hogenesch, J. B. (2002). Coordinated transcription of key pathways in the mouse by the circadian clock. Cell *109*, 307-320.

Perez, I., Lin, C. H., McAfee, J. G., and Patton, J. G. (1997). Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. Rna *3*, 764-778.

Potashkin, J., Naik, K., and Wentz-Hunter, K. (1993). U2AF homolog required for splicing in vivo. Science *262*, 573-575.

Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245.

Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. Genes Dev *3*, 2113-2123.

Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. Curr Opin Genet Dev *6*, 215-220.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Curr Opin Cell Biol *12*, 340-345.

Reick, M., Garcia, J. A., Dudley, C., and McKnight, S. L. (2001). NPAS2: an analog of clock operative in the mammalian forebrain. Science *293*, 506-509.

Ring, H. Z., and Lis, J. T. (1994). The SR protein B52/SRp55 is essential for Drosophila development. Mol Cell Biol *14*, 7499-7506.

Roscigno, R. F., and Garcia-Blanco, M. A. (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. Rna *1*, 692-706.

Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C., and Tazi, J. (1996). Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. Nature *381*, 80-82.

Rowen, L., Young, J., Birditt, B., Kaur, A., Madan, A., Philipps, D. L., Qin, S., Minx, P., Wilson, R. K., Hood, L., and Graveley, B. R. (2002). Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. Genomics *79*, 587-597.

Rudner, D. Z., Breger, K. S., Kanaar, R., Adams, M. D., and Rio, D. C. (1998a). RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high-affinity binding. Mol Cell Biol *18*, 4004-4011.

Rudner, D. Z., Breger, K. S., and Rio, D. C. (1998b). Molecular genetic analysis of the heterodimeric splicing factor U2AF: the RS domain on either the large or small Drosophila subunit is dispensable in vivo. Genes Dev *12*, 1010-1021.

Rudner, D. Z., Kanaar, R., Breger, K. S., and Rio, D. C. (1996). Mutations in the small subunit of the Drosophila U2AF splicing factor cause lethality and developmental defects. Proc Natl Acad Sci U S A *93*, 10333-10337.

Rudner, D. Z., Kanaar, R., Breger, K. S., and Rio, D. C. (1998c). Interaction between subunits of heterodimeric splicing factor U2AF is essential in vivo. Mol Cell Biol *18*, 1765-1773.

Ruskin, B., Zamore, P. D., and Green, M. R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell *52*, 207-219.

Ryner, L. C., Goodwin, S. F., Castrillon, D. H., Anand, A., Villella, A., Baker, B. S., Hall, J. C., Taylor, B. J., and Wasserman, S. A. (1996). Control of male sexual behavior and sexual orientation in Drosophila by the fruitless gene. Cell *87*, 1079-1089.

Sailer, A., MacDonald, N. J., and Weissmann, C. (1992). Cloning and sequencing of the murine homologue of the human splicing factor U2AF^{65}. Nucleic Acids Res *20*, 2374.

Sakao, E., Ishihara, A., Horikawa, K., Akiyama, M., Arai, M., Kato, M., Seki, N., Fukunaga, K., Shimizu-Yabe, A., Iwase, K.*, et al*. (2003). Two-peaked synchronization in

day/night expression rhythms of the fibrinogen gene cluster in the mouse liver. J Biol Chem *278*, 30450-30457.

Shapiro, M. B., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res *15*, 7155-7174.

Shepard, J., Reick, M., Olson, S., and Graveley, B. R. (2002). Characterization of U2AF(26), a splicing factor related to U2AF(35). Mol Cell Biol *22*, 221-230.

Shi, H., Hoffman, B. E., and Lis, J. T. (1997). A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. Mol Cell Biol *17*, 2649-2657.

Singh, R., Valcarcel, J., and Green, M. R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science *268*, 1173-1176.

Smith, C. W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci *25*, 381-388.

Spector, D. L., Fu, X. D., and Maniatis, T. (1991). Associations between distinct pre-mRNA splicing components and the cell nucleus. Embo J *10*, 3467-3481.

Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol Cell Biol *14*, 7670-7682.

Stamm, S., Zhang, M. Q., Marr, T. G., and Helfman, D. M. (1994). A sequence compilation and comparison of exons that are alternatively spliced in neurons. Nucleic Acids Res *22*, 1515-1526.

Stokkan, K. A., Yamazaki, S., Tei, H., Sakaki, Y., and Menaker, M. (2001). Entrainment of the circadian clock in the liver by feeding. Science *291*, 490-493.

Storch, K. F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F. C., Wong, W. H., and Weitz, C. J. (2002). Extensive and divergent circadian gene expression in liver and heart. Nature *417*, 78-83.

Tacke, R., and Manley, J. L. (1999). Determinants of SR protein specificity. Curr Opin Cell Biol *11*, 358-362.

Tarn, W. Y., and Steitz, J. A. (1995). Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. Proc Natl Acad Sci U S A *92*, 2504-2508.

Tronchere, H., Wang, J., and Fu, X. D. (1997). A protein related to splicing factor U2AF^{35} that interacts with U2AF^{65} and SR proteins in splicing of pre-mRNA. Nature *388*, 397-400.

Tupler, R., Perini, G., and Green, M. R. (2001). Expressing the human genome. Nature *409*, 832-833.

Valcarcel, J., Gaur, R. K., Singh, R., and Green, M. R. (1996). Interaction of U2AF^{65} RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. Science *273*, 1706-1709.

Wagner, E. J., and Garcia-Blanco, M. A. (2001). Polypyrimidine tract binding protein antagonizes exon definition. Mol Cell Biol *21*, 3281-3288.

Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C., and Fu, X. D. (1998). SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. J Cell Biol *140*, 737-750.

Wang, H. Y., Xu, X., Ding, J. H., Bermingham, J. R., Jr., and Fu, X. D. (2001). SC35 plays a role in T cell development and alternative splicing of CD45. Mol Cell *7*, 331-342.

Wang, Z., Hoffmann, H. M., and Grabowski, P. J. (1995). Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. Rna *1*, 21-35.

Watakabe, A., Tanaka, K., and Shimura, Y. (1993). The role of exon sequences in splice site selection. Genes Dev *7*, 407-418.

Webb, C. J., and Wise, J. A. (2004). The splicing factor U2AF small subunit is functionally conserved between fission yeast and humans. Mol Cell Biol *24*, 4229-4240.

Wentz-Hunter, K., and Potashkin, J. (1996). The small subunit of the splicing factor U2AF is conserved in fission yeast. Nucleic Acids Res *24*, 1849-1854.

Witherell, G. W., Wu, H. N., and Uhlenbeck, O. C. (1990). Cooperative binding of R17 coat protein to RNA. Biochemistry *29*, 11051-11057.

Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A., and Smith, C. W. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell *13*, 91-100.

Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF^{35}. Nature *402*, 832-835.

Xiao, S. H., and Manley, J. L. (1998). Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. Embo J *17*, 6359-6367.

Zamore, P. D., and Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci U S A *86*, 9243-9247.

Zamore, P. D., and Green, M. R. (1991). Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. Embo J *10*, 207-214.

Zamore, P. D., Patton, J. G., and Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zhang, M., Zamore, P. D., Carmo-Fonseca, M., Lamond, A. I., and Green, M. R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. Proc Natl Acad Sci U S A *89*, 8769-8773.

Zhou, Y. D., Barnard, M., Tian, H., Li, X., Ring, H. Z., Francke, U., Shelton, J., Richardson, J., Russell, D. W., and McKnight, S. L. (1997). Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. Proc Natl Acad Sci U S A *94*, 713-718.

Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182-185.

Zhu, J., and Krainer, A. R. (2000). Pre-mRNA splicing in the absence of an SR protein RS domain. Genes Dev *14*, 3166-3178.

Zorio, D. A., and Blumenthal, T. (1999a). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. Nature *402*, 835-838.

Zorio, D. A., and Blumenthal, T. (1999b). $U2AF^{35}$ is encoded by an essential gene clustered in an operon with RRM/cyclophilin in Caenorhabditis elegans. Rna *5*, 487-494.

Zorio, D. A., Lea, K., and Blumenthal, T. (1997). Cloning of Caenorhabditis $U2AF^{65}$: an alternatively spliced RNA containing a novel exon. Mol Cell Biol *17*, 946-953.

Zuo, P., and Maniatis, T. (1996). The splicing factor $U2AF^{35}$ mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev *10*, 1356-1368.

VITA

Jeremiah Brian Shepard was born on the 27th of February, 1975. He was the only child of Jim A. Shepard and Frances S. Stone. He grew up in Arlington, TX and then moved to San Antonio prior to his eighth grade year. He attended St. Mary's Hall from eighth grade until graduation in 1993. Fortunately, he gained two siblings, Jennifer Shepard and Nick Shepard during his high school years. He then traveled to St. Louis, MO to attend Washington University and he graduated with a Bachelor of Arts in Biology in 1997. In 1998 he left St. Louis to enter graduate school at UT Southwestern. He joined Steve McKnight's lab in 1999 and then Kristen Lynch's lab in 2001. He married Dana H. Chang on October 26th, 2002 in St. Louis, MO.

Permanent Address: 5712 Highgate Dr., Arlington, TX 76016