2016

RNA-Protein Interactions in the U12-Dependent Spliceosome

Jagjit Singh

RNA-PROTEIN INTERACTIONS IN THE U12-DEPENDENT SPLICEOSOME

JAGJIT SINGH

Bachelor of Science
Panjab University
May 2003

Master of Science
Punjab Agricultural University
May 2006

submitted in partial fulfillment of requirements for the degree of

DOCTOR OF PHILOSOPHY IN REGULATORY BIOLOGY

at the

CLEVELAND STATE UNIVERSITY

December 2016

We hereby approve this dissertation for

Jagjit Singh

Candidate for the Doctor of Philosophy in Regulatory Biology degree for the
Department of Biological, Geological and Environmental Sciences and
CLEVELAND STATE UNIVERSITY'S College of Graduate Studies by

Dr. Girish C. Shukla

Dissertation Chairperson and Major Advisor
GRHD/BGES, Cleveland State University Date: 12/12/2016

Dr. Barsanjit Mazumder

Dissertation Committee Member
GRHD/BGES, Cleveland State University Date: 12/12/2016

Dr. Sailen Barik

Dissertation Committee Member
GRHD/BGES, Cleveland State University Date: 12/12/2016

Dr. Aaron F. Severson

Dissertation Committee Member
GRHD/BGES, Cleveland State University Date: 12/12/2016

Dr. Sadashiva Karnik

Dissertation Committee Member
Dept. of Molecular Cardiology, LRI Date: 12/12/2016

Dr. Anton A. Komar

Internal Examiner
GRHD/BGES, Cleveland State University Date: 12/12/2016

Dr. Richard A. Padgett

External Examiner
Dept. of Cellular and Molecular Medicine, LRI Date: 12/12/2016

Student's Date of Defense: December 12, 2016

Dedication

Dedicated to Almighty,

My Family

&

My Friends

Acknowledgement

First and foremost, I would like to thank my advisor Dr. Girish C. Shukla who has been the guiding force during this work. This day would not have been possible without him. He has been a great mentor and has taught me everything about science, from day-to-day running of experiments to how to think about my own work, and biology in general, in a larger context. His friendly and understanding nature has made working with him very enjoyable. Not only scientific discussions, his door was always open for healthy discussions, even if it is related to something personal. Words are not enough to express my gratitude for his patience and endless support towards me.

I would also like to thank the Advisory Committee Members, Dr. Mazumder, Dr. Barik, Dr. Severson and Dr. Karnik. All of them really appreciated the quality of my work and inspired me to work harder. They provided me with their valuable feedbacks during the course of my Ph.D. research.

I am very thankful to Dr. Padgett (external reviewer) and Dr. Komar (internal reviewer) for agreeing to serve in my Ph.D. dissertation defense committee. Thank you Dr. Komar for your wonderful piece of advice in my protein related work and travel awards for attending the scientific meetings without any hesitations.

The members of the Shukla lab, both former (Kavleen, Jinani, Tupa, Kevin & Neha) and current (Savita, Alexis, Narmada, Asmita & Maitri), have been great colleagues and have provided many valuable inputs into the work. At the same time, they have been very good friends and working with them has been a pleasure

and full of fun. I owe my sincere thanks to Jey Sabith, from the bottom of my heart, for supporting me out of the way in every aspect during this journey of Ph.D.

I would like to thank all the administrative personnel especially Monica, Lisa and Carolee for their all-time help with any kind of official work. Also, the friends around the hallway are thanked for their support.

I would like to acknowledge the Cellular and Molecular Medicine Specialization program for two years funding support towards this work and Cleveland State University for providing the Doctoral Dissertation Research Expense Award Fellowship to finish my doctoral studies.

I am really blessed to have my older sister and brother-in-law. Thank you for always being there around the corner through my ups and downs in life. There can be no other older sister who worries about her little brother as she does for me. My nephew and my niece are my stress busters.

Last but not the least, my parents are the continuous source of inspiration that helped me chase my dream to complete my Ph.D. studies thousands of miles away from home which otherwise would have been remained only as a dream.

All the names might not have been included but none is forgotten.

This work is partially reprinted from

Jagjit Singh, Kavleen Sikand, Heike Conrad, Cindy L. Will, Anton A. Komar and Girish C. Shukla.

U6atac snRNA stem-loop interacts with U12 p65 RNA binding protein and is functionally interchangeable with the U12 apical stem-loop III. Scientific Reports. 2016; 6: Article number: 31393. Doi:10.1038/srep31393

RNA-PROTEIN INTERACTIONS IN THE U12-DEPENDENT SPLICEOSOME

JAGJIT SINGH

ABSTRACT

Nuclear precursor messenger RNA (Pre-mRNA) splicing is an important regulatory step in metazoan gene expression. More than 99% of nuclear pre-mRNA introns are U2-type that are spliced by U2-dependent spliceosome containing U1, U2, U4, U5 and U6 snRNAs. Only less than 1% of the introns are U12-type and spliced by U11, U12, U4atac, U5 and U6atac snRNAs. U12 and U6atac snRNAs play a central role in the splicing of U12-dependent introns. Our previous work demonstrated that the conserved 3′ stem-loop region of U6atac snRNA contains a U12-dependent spliceosome-specific targeting activity, however any potential molecular mechanism was unclear. We discovered that the distal 3′ stem-loop of U6atac has structural and sequence similarities with stem-loop III of U12 snRNA. These observations convinced us to investigate the structure-function requirement of the substructure of the U6atac 3′ stem-loop in U12-dependent *in vivo* splicing. Our results show that the C-terminal RNA recognition motif of p65, a U12 snRNA binding protein, also binds to the distal 3′ stem-loop of U6atac. Using *in vivo* genetic suppressor assay, we demonstrate that stem-loop III of U12 snRNA which binds to p65 protein can be functionally replaced by U6atac distal stem-loop and vice-versa. Furthermore, we tested the compatibility of the U6atac 3' end from phylogenetically distant species in a human U6atac suppressor background to establish the evolutionary relatedness of these

structures and *in vivo* functionality. In conclusion, we demonstrate that p65 C-terminal RNA recognition motif interacts with the U6atac distal 3′ stem-loop. Although the significance of p65 binding to U6atac snRNA is not clear, our study suggests that both the helix structure, as well as the sequence of U6atac distal 3′ stem-loop is important for snRNA-protein interactions and U12-dependent intron splicing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Splicing and its role in gene regulation

The comprehensive advancements in the transcriptomic studies revealed that 98% of a human genome is transcriptionally active and producing both the protein coding as well as the non-protein coding transcripts. In human genome, from the total of 26, 564 annotated genes are considered to contain approximately 233,785 exons and 207,344 introns. Almost all the eukaryotic genes are initially transcribed into precursor mRNAs (pre-mRNAs) which are interrupted with at least one intron. On an average, there are 8.8 exons and 7.8 introns per gene (Sakharkar M.K. et al., 2004). Since only a mature mRNA can escape the nuclear surveillance and degradation mechanism, these introns have to be removed, in order to make the genetic message continuous. Therefore, in order to make mature mRNA that could actively engage to the translational process, all pre-mRNAs must undergo different modifications such as modifications at the 5' end, 3' end, and the removal of introns. In eukaryotes, the 5' and 3' end of the primary transcript are protected against the exonucleases with the help of the 5' 7mG cap

(Gao M., 2000) and the 3' polyadenylation tail respectively (Guhaniyogi J. and Brewer G., 2001). All three of the modifications mentioned above play an important role in gene regulation by different possible mechanisms, such as nuclear export, degradation, and translational enhancement (Visa N., 1996; Gao M., 2000; Meijer H. A., 2007; Sonenberg N. and Gingras A.C. 1998; Guhaniyogi J. and Brewer G., 2001; Braddock M. et al., 1994; Matsumoto K. et al., 1998). Surprisingly, it was discovered that the introns and their act of removal, influences various other stages involved in RNA metabolism. These stages include the transcription of the gene, pre-mRNA editing and polyadenylation, nuclear export, translation, and mRNA decay. Hence, the process of intron removal contributes to significant differential expression profiles for the intron-containing and intron-less version of the same gene (Buchman A.R. and Berg P., 1988; Callis J. et al., 1987; Duncker B.P. et al., 1997).

In Drosophila and humans, the electron micrograph analysis of actively transcribing genes depicts that pre-mRNA splicing occurs with fairly high frequency on the nascent transcripts. This analysis also supports the notion that the splice site recognition event generally precedes the polyadenylation. This existing evidence indicates that the splicing event occurs co-transcriptionally. (Beyer A.L. and Osheim Y.N., 1988; Bauren G. and Wieslander L., 1994; Kiseleva E. et al., 1994; Zhang G. et al., 1994; Pandya-Jones A. and Black D.L., 2009; de la Mata M. et al., 2010).

Introns could be either spliced by a constitutive splicing mechanism where all the exon in a gene at the 5' end of the intron are ligated to the exon at the 3’

end of the intron. No exon skipping event takes place in the constitutive splicing mechanism. However, another mechanism with which an intron splices is the alternative splicing. Alternative splicing process involves the joining of different 5′ and 3′ splice sites, producing different spliced variants of the same gene. The proteins that are encoded by multiple mRNAs spliced variants, perform various diverse and sometimes even antagonistic functions. Up to 59% of the total human genes undergo alternative splicing and generate multiple mRNAs (Lander E. S. et al., 2001), and approximately 80% of alternatively spliced variants alters the encoded protein (Modrek B. and Lee C., 2002). The variability in mRNA, due to alternative splicing, results in the a) introduction of a termination codon in the mRNA b) insertion or removal amino acids or c) frame shift of the reading frame in the expressed protein. The removal or insertion of important regulatory elements by the alternative splicing further affects the gene expression by controlling translation, mRNA stability, or localization is regulated. Alternative splicing alters the coding capacity of genomes by bringing huge diversity to the proteome capacity of a given cell (Sultan M. et al., 2008; Wang E.T. et al., 2008), and thereby playing a very significant role in regulating the gene expression (Blencowe B.J., 2006).

Nuclear pre-mRNA splicing is a two-step process involving two transesterification reactions catalyzed by a dynamic, multi-megaDalton ribonucleoprotein (RNP) machinery known as spliceosome (Will C.L. and Luhrmann R., 2005). In almost all metazoans, two types of introns have been identified so far. These are U2-dependent (major class) introns comprising more

than 99% and U12-dependent (minor class) introns comprising less than 1% of the total introns. These two types of introns differ in their consensus signature sequences and require different splicing machineries with almost similar splicing mechanisms. For the splicing of two different types of introns, two specific spliceosomes with unique components have been characterized. The U2-dependent spliceosome, contains U2-type snRNAs and catalyzes the removal of U2-dependent introns (Moore M.J. et al., 1993; Burge C.B. et al., 1999) whereas, the rare class U12-dependent pre-mRNA introns are spliced by the U12-dependent spliceosome and contain U12-dependent snRNAs (Tarn W.Y. and Steitz J.A., 1997; Burge C.B. et al., 1999; Wu Q. and Krainer A.R., 1999; Patel A.A. and Steitz J.A., 2003).

It has been suggested that both U2 and U12 type introns co-evolved during the course of evolution, but U12-dependent introns were found to be missing in a number of organisms, such as budding yeast and nematodes. The U2-dependent introns as well as U2-dependent spliceosome are ubiquitously present in all eukaryotes whereas, the U12-dependent introns and their respective machinery is only present in a subset of eukaryotes (Burge C.B. et al., 1998). Both the U2- and U12-dependent spliceosomes coexist in eukaryotic cells, and the pre-mRNAs transcripts containing both types of introns serve as substrates for both the spliceosomes.

Primarily, the spliceosome helps in the recognition of the intron/exon boundaries and the catalysis of the reaction that removes the intron and joining of the exons. Spliceosomes are assembled on the pre-mRNA intron in a multistep

process during which the snRNPs, including five small non-protein coding nuclear RNAs, recognize and interact with the target intron cis-acting elements with the help of RNA-RNA base pair interactions. These events are also facilitated by additional splicing factors interacting stepwise with the pre-mRNA. Most mammalian pre-mRNAs contain multiple introns with remarkably variable length that range from few bases, to hundreds of thousands of kilobases. Whereas the exons have rather fixed average length of only ~120 nucleotides (nts) as compared to the introns. However, when an intron's length exceeds ~200-250 nts, the formation of the splicing complex initiate across the exon. This particular process is known as exon definition. It is evident that serine-arginine rich proteins (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNP) proteins bind to splicing regulatory elements near both the U2- and U12-dependent splice sites and help in the recruitment of the respective spliceosome (Lewandowska D. et al., 2004; Hastings M.L. and Krainer A.R., 2001; Wu Q. and Krainer A.R., 1998; McNally L.M., 2006). Being a very dynamic process, splicing complex involves various RNA:RNA, RNA:protein, and protein:protein interactions. The splicing process begins with the sequential recognition and binding of the 5' splice site and the branch site by their complementary snRNAs. After the complete assembly of spliceosome across an intron is achieved, two transesterification reactions result in the removal of the intron, resulting in the joining of the flanking exons. The detailed process of spliceosomal assembly and the mechanism are explained in section 1.2.

1.2 Two different types of introns and their splicing mechanisms

The 5' splice site (AG/GURAGU), the branch site (CURACU) and the 3' splice site (YAG/G) of U2-dependent introns contain a characteristic sequence signature. The adenosine (A) underlined in the branch site sequence has been shown to be involved in the nucleophilic reaction. In the U2-dependent introns, the branch site sequence and the 3' splice site sequence are roughly 40 nucleotides apart and is pyrimidine rich region hence known as polypyrimidine tract. This polypyrimidine region has been shown to be important for the recognition of upstream branch site sequence of the intron. RNA-RNA base pair interactions occurring between the splice sites of the introns and the snRNAs accounts to the specificity of the splicing reaction. In case of U2-dependent intron splicing mechanism, the 5' splice site and the branch site are recognized and base paired with U1 snRNA and U2 snRNA respectively (Will C.L. and Luhrmann R., 2005).

In contrast, the U12-dependent introns consensus splice site sequences signatures are different from U2- dependent introns. The 5' splice site (RTATCCTTT) as well branch site (UCCUUAACU, where the underlined A is the branch point adenosine) of U12-type introns are longer and more tightly constrained as compared to the U2-type introns. U12-dependent introns do not contain polypyrimidine tract. Therefore, the average distance between the U12-type branch site and the 3' splice site is significantly shorter (i.e., 11–13 nucleotides) than that of the U2-type introns (i.e., 18–40 nucleotides) and has been demonstrated to be an important determinant for the recognition of U12-dependent introns (Hall S.L. and Padgett R.A., 1994; Dietrich R.A. et al., 2001a; Levine A.

and Durbin R.A., 2001).

The U12-dependent introns were named as “atac” introns because they were first identified on the basis of their terminal AT–AC di-nucleotides at their 5' and 3’ ends respectively (Jackson I.J. 1991; Hall S.L. and Padgett R.A. 1994). Later on, with the detailed bioinformatics studies, it was realized that these terminal dinucleotides are not exclusive for all the annotated U12-dependent introns (Wu Q. and Krainer A.R., 1997). Rather than the terminal di-nucleotides signature, highly conserved 5′ splice site and branch point sequences (BPS) are considered to be the major determinants of the U12-type intron as compared to U2-type introns. Whereas the 3' splice site of U12-type intron is more variable (Jackson I.J., 1991) (Figure 1). Surprisingly, 70-80% of the total U12-type introns contain GT–AG terminal di-nucleotide and is actually the more common subtype (Burge C.B. et al., 1998). Therefore, a different nomenclature was adopted in which the introns are referred to as U2-dependent introns, and the less abundant introns as U12-dependent introns (Dietrich R. A. et al., 1997). Similarly, the splicing machineries which catalyze the removal of U2- or U12-dependent introns are named as U2- or U12-dependent spliceosomes respectively. Based on their relative abundance in the cell, these spliceosomes are also often referred to as the ‘major’ and ‘minor’ spliceosomes respectively (Dietrich R. A. et al., 1997; Sharp P.A. and Burge C.B., 1997).

U12-type introns coexist with neighbouring U2-type introns in a host gene. Typically, U12-type introns occur one per any given host gene. However, there are multiple genes annotated in the U12 Data-Base that are considered to have two

or more U12-dependent introns. Human *NHE-6* gene has been reported to contain three U12-type introns (Levine A. and Durbin, R.A, 2001).

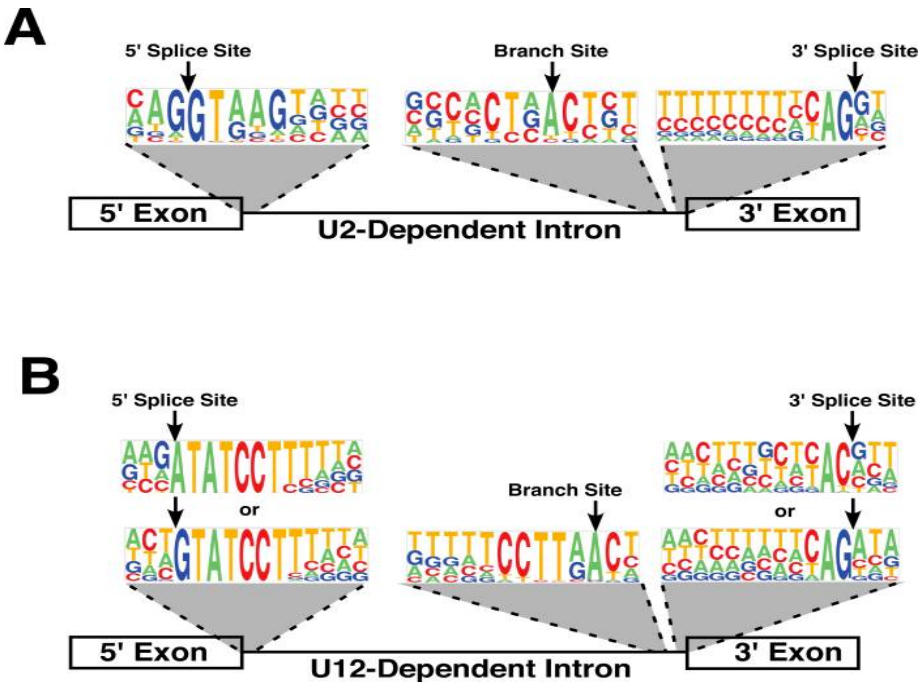


Figure 1: Consensus splice site sequences of (a) U2-dependent and (b) U12-dependent introns. The colored consensus sequences and the size of each letter represents the frequency of each base at the 5' and 3' splice sites respectively of majority of the introns. In case of U2-dependent introns, GT and AG are the terminal dinucleotides. Whereas for U12-dependent introns they could be either AT-AC or GT-AG (Padgett R.A., 2012).

1.2.1 Mechanism of U2-dependent intron splicing

The efficient and precise interactions of splicing factors with the consensus pre-mRNA signature sequences plays most important role in the assembly of the complete spliceosomal complex. The reactive groups of the pre-mRNA such as the 5' splice site, the branch point adenine, and the 3' splice site are spatially positioned in such a way to efficiently catalyze the removal of an intron by

sequential transesterification reactions. Several conformational rearrangements of the small nuclear ribonucleoproteins (snRNPs) within the ribonucleoprotein complex result in this proper spatial positioning of all the splice sites.

The assembly of U2-dependent spliceosome is achieved by the involvement of the U1, U2, U5, and U4/U6 snRNPs as well as numerous other splicing factors. The process of spliceosomal assembly on pre-mRNA and formation of active catalytic complex involves highly dynamic RNA-RNA base pair interactions between pre-mRNA:snRNA and snRNA:snRNA (Nilsen T.W., 1998). This complex, dynamic, and multistep process involves the formation of sequential complexes known as E, A, B, B*, and C (represented in Figure 3) (Reed R. and Palandjian L., 1997). The initial step involves the recognition and interaction of the U1 snRNA at the 5' splice and formation of the commitment complex or E complex. The proteins that are specifically associated with the U1 snRNA, such as U1-70K and U1-C, facilitate and stabilize the interaction between the 5' splice site sequence and the U1 snRNA (Will C.L. and Luhrmann R., 1997). In the following step, the U2 snRNA and its associated proteins interacts with the branch site, forming the A complex. The U2 snRNA associated proteins such as SF3a and SF3b interact with the pre-mRNA at or near the branch site sequence and enhance the stabilization of U2 snRNA interactions with the branch site (Reed R., 1996; Will C.L. and Luhrmann R., 1997). After the formation of A complex, the U4, U6 and U5 snRNAs interact with each other to form a U4.U6/U5 tri-snRNP complex. In this tri-snRNP complex, the U6 and U4 interact by RNA-RNA and U5 by RNA protein interactions. This U2-dependent spliceosome specific tri-snRNP complex

associates with the existing splicing complex, resulting in the formation of complex B. At this step, the snRNAs undergo major structural and spatial rearrangements which further lead to the formation of catalytically active complex B*. During this step, the RNA-RNA base-pair interactions between U4/U6 are disrupted and as a consequence of these structural rearrangements, the U1 snRNA interactions with the 5' splice site and U4 snRNA interactions with the U6 within the spliceosome are destabilized and, hence both the snRNAs dissociate from the spliceosome (Will C. L. and Luhrmann R., 2005).

After the release of U1 snRNA, U6 snRNA interacts with and occupies the 5' splice site. After the release of U4 snRNA, a region of U6 snRNA forming stem-I and stem-II rearranges into an intramolecular stem-loop (U6-ISL) structure, that is positioned very close to the catalytic site of U2-type splicing and is implicated in the catalysis of splicing. In addition, the U5 snRNA works as a tethering snRNA by interacting nucleotides of the exon both near the 5' splice site and the 3' splice site. Subsequently, the first catalytic step of splicing occurs and the C complex is generated. The active catalytic core complex involves only three snRNAs including U6, U2, and U5 for U2-dependent spliceosome. At this step, with the help of two sequential transesterification reactions (Figure 2) the intron is removed as lariat structure and the 5' and 3' exons are joined together. After the second catalytic step, the components of the spliceosome dissociate resulting in the release of the mRNA and the excised intron, as well as the snRNPs, which are recycled to take part in new rounds of splicing (Will C. L. and Luhrmann R., 2005; Chen H.C. and Cheng S.C., 2012). Despite the fact that spliceosomal proteins predominantly

comprise the major composition of the spliceosome, which also play important role in the activation of the spliceosome, but the catalysis of pre-mRNA intron splicing appears to be completely RNA-based where U2 and U6 snRNAs playing the pivotal role (Collins C. A. and Guthrie C., 2000).

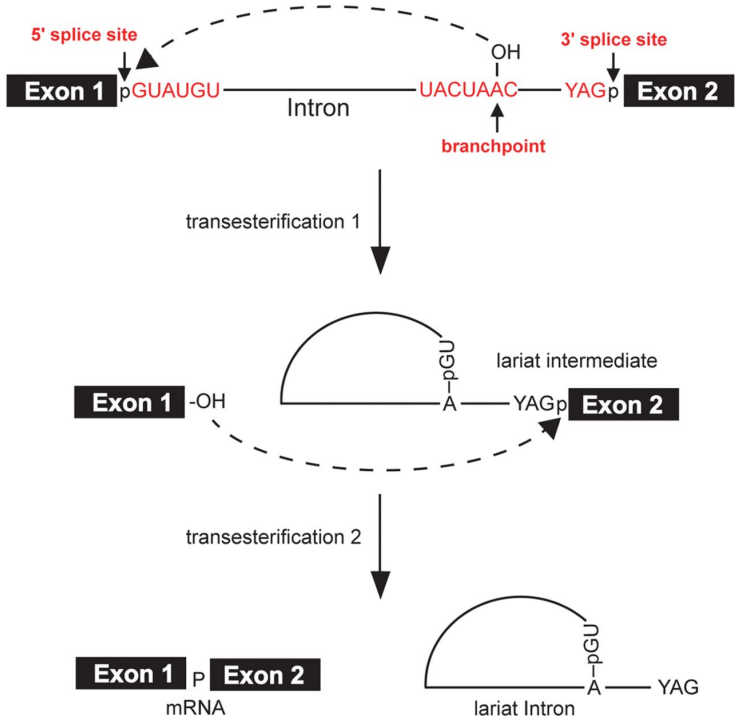


Figure 2: Two step trans-esterification reaction involved in pre-mRNA splicing. The conserved splice site sequences are shown in red. The splicing reaction takes place in two steps. The first step involves the cleavage of the 5' splice site resulting in the formation of lariat intron and Exon 2 with the help of 2'-5' phosphodiester linkage. The second step is the cleavage at the 3' splice site and the ligation of the two exons. (Chen H.C. and Cheng S.C., 2012).

1.2.2 Mechanism of U12-dependent intron splicing

The U12-dependent spliceosome includes U11, U12, U5, and U4atac/U6atac snRNPs as well as numerous non-snRNP proteins (Hall S.L. and Padgett R.A., 1996; Tarn W.Y. and Steitz J.A., 1996 a, b). These snRNAs are the functional analogs of the snRNAs catalyzing U2-dependent splicing (Hall S.L. and Padgett R.A., 1996; Tarn W.Y. and Steitz J.A., 1996a, b; Kolossova I. and Padgett R.A., 1997; Yu Y.T. and Steitz J.A., 1997; Incorvaia R. and Padgett R.A., 1998). Thus, among all the spliceosomal snRNAs, only U5 is common between both the spliceosomes. Assembly of the U12-dependent spliceosome is almost similar to that of the U2-dependent spliceosome, with a major difference occurring at the initial step (Figure 3). Prior to the recognition and their interactions with the pre-mRNA, the U11 and U12 snRNAs interact with each other and form a highly stable di-snRNP complex that binds in a cooperative manner to the 5' splice site and the branch site, respectively during the first step of the minor spliceosome assembly (Wassarman D.A. and Steitz J.A., 1992; Frilander M.J. and Steitz J.A., 1999). Therefore, in contrast to the major spliceosome, the earliest assembly step involves the formation of the A complex with missing commitment E complex. In the subsequent step, the U4atac.U6atac/U5 tri-snRNP complex interacts with the A complex and results in the formation of the B complex (Tarn W.Y. and Steitz J.A., 1996 a,b). Similar to the U2 dependent spliceosome, activation of the minor spliceosome also involves the destabilization of the U11 and U4atac snRNPs. The role of the U4 or U4atac snRNA is thought to be that of a chaperone by sequestering the U6 or U6atac snRNA via extensive base-pairing interactions prior

to catalytic activation of the spliceosome (Guthrie C. and Patterson B., 1988; Tarn W.Y. and Steitz J.A., 1996a). The remaining steps appear to be similar to those involved in the U2-dependent spliceosome such as the formation of a C-like complex at the time of the first transesterification reaction followed by dissociation of the minor spliceosomal components after the second transesterification. Recent studies have revealed that at the time of A complex formation, the nucleotides at the 5' end of U12 snRNA also transiently interact with exon nucleotides just upstream of the 5' splice site (Frilander M.J. and Meng X., 2005). These interactions are destabilized when the U4atac.U6atac/U5 tri-snRNP joins the spliceosome. Thus, during the early stages of minor spliceosome assembly, U12 can simultaneously interact with both the branch site as well as nucleotides of the pre-mRNA near the 5' splice site. Upon integration of the minor U4atac.U6atac/U5 tri-snRNP, the U4atac and U6atac base-pairing interaction is disrupted. The U6atac then interacts with the 5' splice site (displacing U11) and also with the 5' end of the U12 snRNA, whereas the U5 snRNA interacts with exon nucleotides near the 5' splice site and, at later stage, with exon nucleotides near the 3' splice site (Figure 3; Tarn W.Y. and Steitz J.A., 1996a, b; Yu Y.T. and Steitz J.A., 1997; Incorvaia R. and Padgett R.A., 1998; Frilander M.J. and Steitz J.A., 2001). The intermolecular structures formed by the pre-mRNA and the U12, U5, and U6atac snRNAs appear to be similar to those formed by their counterpart snRNAs involved in the U2-dependent spliceosome (Figure 3). Similar to the U6-intramolecular stem-loop structure, U6atac also forms an intramolecular stem loop structure with similar significance in the catalysis of intron removal. It has been shown that the

first step of the pre-mRNA intron splicing catalysis is mediated by the metal ion along with a nucleotide in the intramolecular stem-loop structure of U6 or U6atac snRNA. This secondary structure resembles the catalytically active structure of the domain 5 of self-splicing group II introns which self-splice without involving any protein (Gordon P. M. et al., 2000; Keating et al., 2010), thus indicating that spliceosome is an RNA-enzyme (a ribozyme) where the active site is composed of RNA molecules (Sashital et al., 2004).

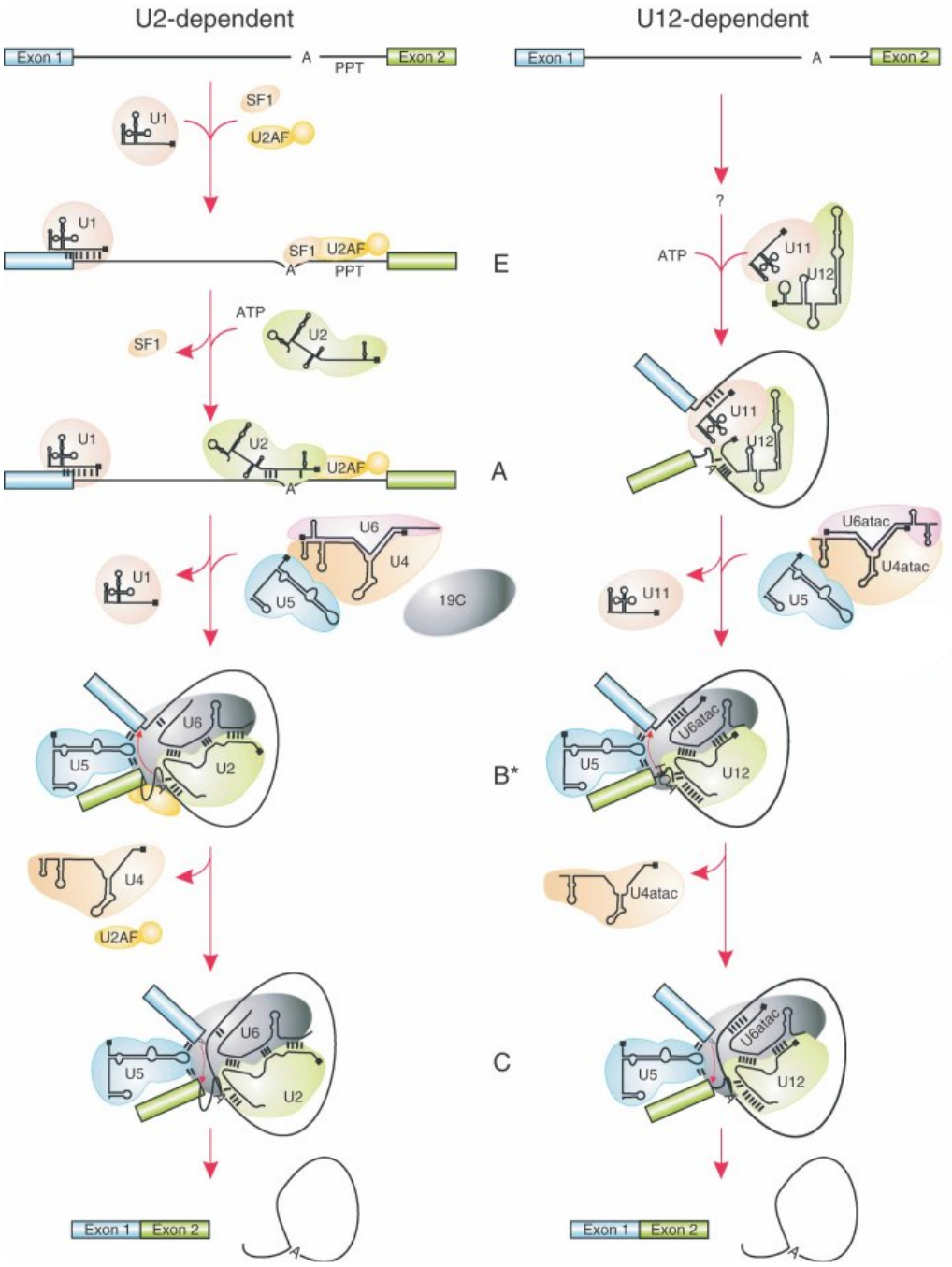


Figure 3: Spliceosome assembly (Left panel: U2-dependent spliceosome. Right panel: U12-dependent spliceosome). Schematic representation of the interaction of the spliceosomal snRNPs at various stages of the dynamic process. The different stages and complexes are depicted as E, A, B*, and C for both the U2- and U12-dependent spliceosomes (Turunen J.J. et al., 2012).

1.3 Predicted secondary structures of U12-dependent snRNAs

Thus far, we know that both types of spliceosomes involve the set of four unique and one common snRNA. Although the sequences of these analogous snRNAs involved in two different spliceosomes are quite divergent but to some extent they share a common overall secondary structures (Figure 4). Both the RNA as well as protein composition of U4atac.U6atac/U5 tri-snRNP of U12-dependent spliceosome appear to be very similar to that of U4.U6/U5 tri-snRNP complex involved in U2-dependent spliceosome (Schneider C. et al., 2002). Interestingly, it has also been demonstrated that both the U4 and the U4atac snRNAs from different spliceosome form equivalent secondary stem-loop structures which serve as binding platforms for the proteins that are essential for the formation of the tri-snRNP complex (Schneider C. et al., 2002; Nottrott S. et al., 2002). Furthermore, U1 and U2 snRNAs are available as distinct snRNPs and interact with the pre-mRNA in a sequential manner. Whereas, their counterparts (U11 and U12 snRNAs) are always bound to each other by specific RNA-protein interactions in the nucleus as a U11/U12 di-snRNP complex and interact with pre-mRNA in a cooperative manner (Wassarman K.M., 1992; Montzka K.A. and Steitz J.A., 1988). Shukla G. C. and Padgett R. A. (1999) have shown that the U12 and U6atac snRNAs of U12-dependent spliceosome share an overall structure as well as many of the sequences similarity to their highly conserved counterpart U2 and U6 snRNAs of U2-dependent spliceosome. In this report, the group identified the homologs of U6atac and U12 snRNAs in Arabidopsis thaliana. Higher percentage of conservation of sequence and structures for U6atac (65%) and for U12 (55%)

snRNA, suggest that these snRNAs in plants are significantly diverged from the human. The U6atac nucleotides of A. *thaliana* that base pair with the 5' splice site sequence of U12-dependent introns shows complete sequence conservation to that of Human U6atac (Shukla G. C. and Padgett R. A., 1999)).

These snRNAs contain a conserved single stranded sequence, normally flanked by two stem-loop structures, known as Sm-site with consensus of $PuAU_{4-6}$ (where Pu is purine). These Sm sites provide the binding platform for Sm-proteins (Will C. L. and Luhrmann R., 2001).

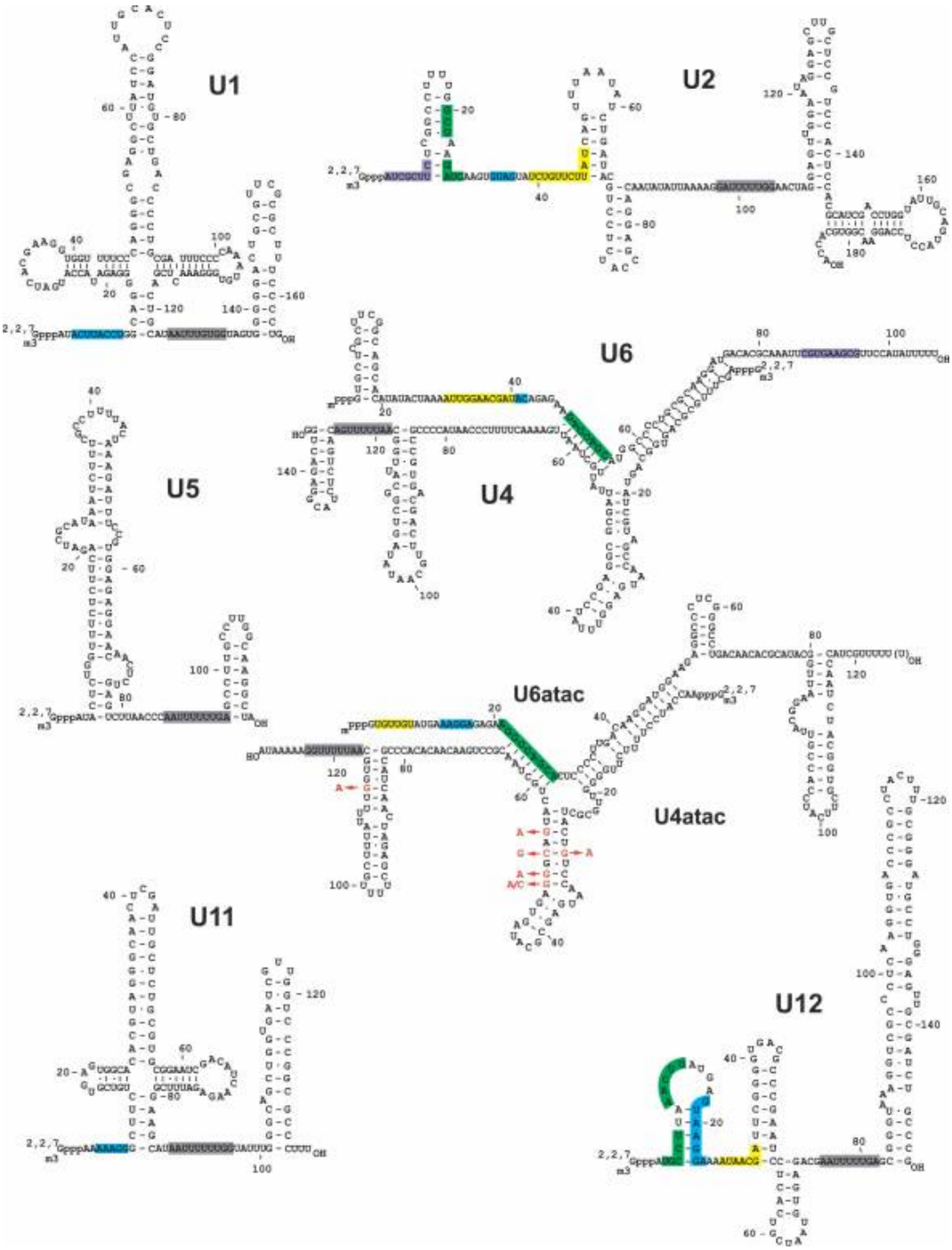


Figure 4: The predicted secondary structures of the human spliceosomal snRNAs. Different regions of the snRNAs have been highlighted in different colors. Same color coded regions have been shown to interact with each other during the process of splicing. The binding sites for Sm proteins are shaded in gray, and the sequences interacting with the 5′ss or BPS is shaded in cyan. U2/U6 or U12/U6atac interacting sequences are indicated by green (helix I), purple (helix II), and yellow shading (helix III). Nucleotides highlighted in red show the position of the microcephalic osteodysplastic primordial dwarfism type I (MOPD1) mutations in the U4atac snRNA (Turunen J.J. et al., 2012).

Similar to those U1, U2, and U4 snRNAs catalyzing U2-dependent intron splicing, their counterparts snRNAs such as U11, U12, and U4atac along with U5 snRNA possesses 2,7,7 tri-methylguanosine 5' cap structure. Whereas, U6atac (and U6) has monomethyl phosphate cap and lacks Sm-binding site (Tarn W.Y. and Steitz J.A. 1996a).

1.4 RNA-RNA base-pair interactions in the catalytic core complex of spliceosomes

Although the formation of the catalytic core appears to be somewhat more flexible in the U12-dependent as compared to the U2-dependent spliceosome, the structure and function of both the core complexes appear to be very similar (Frilander M.J. and Steitz J.A., 2001). Moreover, both spliceosomes involve the same sequential two-step transesterification reaction mechanism for the removal of an intron (Tarn W.Y. and Steitz J.A. 1996) (Figure 2). Similar RNA-RNA base pair interactions can be drawn among U12, U6atac, U5, and with the U12-dependent intron in the active catalytic core of the U12-dependent spliceosome to those happening in the components of the major spliceosome. The RNA-RNA base pair interactions between U2 and U6 snRNA which form the helix II structure is specific to the major spliceosomal complex. However, similar U12-U6atac intermolecular interactions between these two molecules has not yet been validated in the minor spliceosome (Figure 5). The predicted helix III formed by U12 and U6atac interactions is still not yet validated and it is not conserved in plants (Shukla G.C. and Padgett R.A., 1999), but reduced splicing activity in mammals was observed when mutations were introduced in this region of U12

snRNA that weakens the helix III structure (Sikand K. and Shukla G.C., 2011). Furthermore, it is suggested that these mutations could possibly be interfering with the RNA-RNA base pair interactions resulting in the formation of the predicted helix III in the U12-dependent catalytic core complex. In addition, the structural domain similarity of snRNAs between the two distinct spliceosomes suggest that the minor spliceosome is likely to involve RNA-based catalysis, similar to major spliceosome (Valadkhan S. et al., 2009).

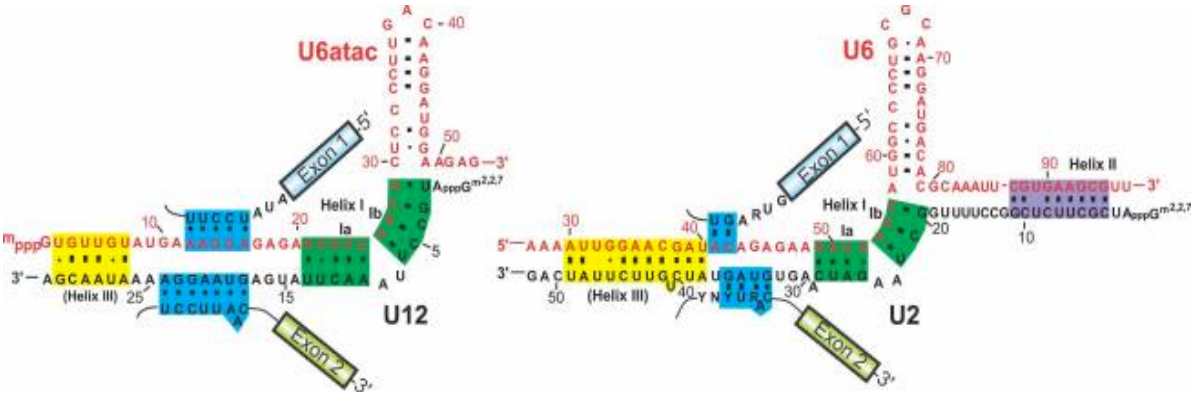


Figure 5: RNA–RNA base-pair interactions occurring in the active catalytic cores of U12- and U2-dependent spliceosomes. The boxes with the dark outline represent the 5' and 3' exons. U6 or U6atac interactions at the 5' splice site and U2 or U12 interactions at the branch site of the introns are shown in cyan. The helix I region formed by the RNA-RNA base pair interactions between U2/U6 and U12/U6atac have been highlighted in green. Purple and yellow shading represent the formation of helix II and helix III respectively. (Turunen J.J. et al., 2012).

1.5 U11/U12 di-snRNP complex, specific and shared proteins

The protein compositions of the major snRNPs and the spliceosome formed by these snRNPs are very well characterized. Due to the fact that the U12-dependent snRNPs are 100 times less abundant in the cell as compared to their

counterparts U2-dependent snRNAs (Yu Y.T. et al., 1999), our comprehensive knowledge about the protein composition involved in the minor spliceosome is still far from complete.

Characterization of the human U11/ U12 di-snRNP and U4atac/U6atac.U5 tri-snRNP complexes revealed most of the identified proteins present in the minor spliceosome, indicating that many snRNP-associated proteins are shared by both the spliceosomes. The sharing of the protein components by the two spliceosomes is consistent with the conserved catalytic mechanisms of both spliceosomes. To date, only seven (65K, 59K, 48K, 35K, 31K, 25K and 20K) (Figure 6) protein components, specific to U12-dependent spliceosome, have been identified. All the seven protein were found to be associated with the 18S U11/U12 di-snRNP complex (Will C.L. et al., 2004; Will C.L. et al., 1999). The lack of all the U1 associated specific proteins and some of the U2-associated proteins in U11/U12 di-snRNP complex makes it the most divergent component of the minor spliceosome so far (Will C. L. et al., 2001).

Four proteins (i.e., 59K, 48K, 35K, and 25K), out of seven U11/U12 di-snRNP specific associated proteins, were found in affinity-selected 12S snRNPs enriched in U11 (Will C.L. et al., 2004). Association of the 59K, 35K, and 25K proteins with the U11 snRNA was further confirmed by immunoprecipitation studies. Thus, the U11 snRNP is comprised of at least four novel proteins, three of which (59K, 48K, and 25K) bear no resemblance to the U1-70K, U1-A, and U1-C proteins found in the U1 snRNP (Will C.L. and Luhrmann R.1997).

Furthermore, because only one of the U11-associated proteins (i.e., 35K) contains a RNA recognition motif known to mediate RNA binding, it is likely that only one of the stem–loop structures of the U11 snRNA is bound to protein. These observations suggest that within the U11 snRNP, stem–loop III is bound by the 35K protein. There are evidences that suggest that the remaining U11-proteins associate in the U11/U12 do-snRNP complex via protein-protein interactions (Will C.L. et al., 2004).

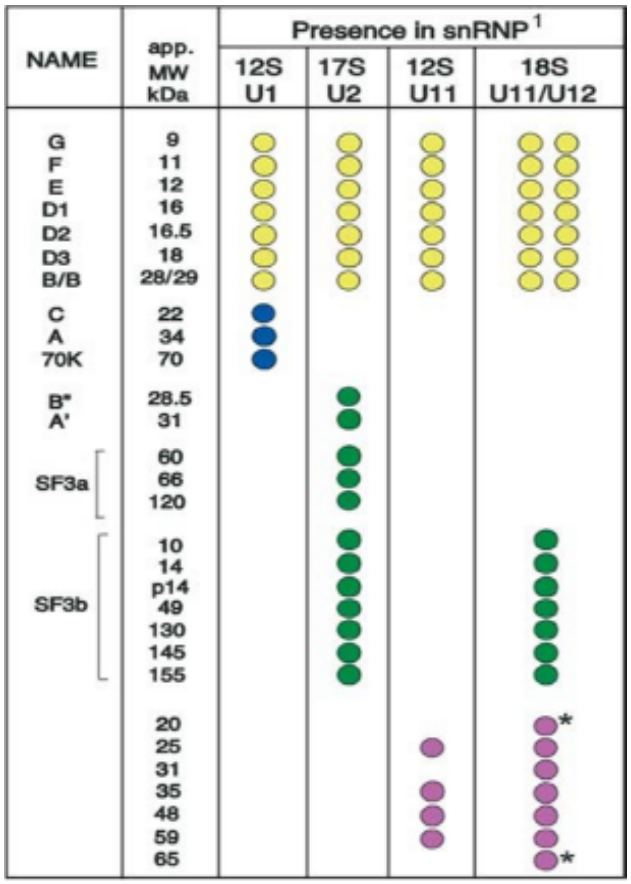


Figure 6: Protein composition of the human U2- and U12-dependent snRNPs. The presence of a given protein in a snRNP is indicated by a colored circle. Mass spectrometry revealed the presence of seven U12-dependent spliceosome specific proteins in purified human 18S U11/U12 snRNPs. These proteins are represented with purple colored circles. Four (25K, 35K, 48 and 59 K) out of seven are associated with 12S U11 snRNPs. (Will C.L. et al., 2004).

1.6 Phylogenetic conservation of U12-dependent spliceosome

The significant similarities related to the structure of different components and the mechanism involved by the two distinct spliceosomes suggest that both machineries are evolved from common ancestry. Putative U12-dependent introns can be recognized computationally on the basis of their conserved 5′ splice site and the branch site sequence signatures. Using different computational tools, U12-dependent RNAs (including introns as well as the snRNAs, catalyzing the intron splicing) were identified in a diverse range of phylogenetically distant species such as Fungi, *Acanthamoeba*/Mycetozoa and Streptophyta. Presence of U12-dependent introns and the splicing machinery in the lower eukaryotes strongly supports the notion that the U12-dependent spliceosome pre-existed the U2-dependent spliceosome. The two different theories, among many, which can explain the presence of U12-dependent introns as only a handful of total introns exhibited by the highly evolved human genome is either the complete loss of U12-type intron or the conversion of less efficient U12-type introns to more efficiently spliced U2-type introns over a period of time (Russell A.G. et al., 2006; Dietrich R.C. et al., 1997).

In contrast with U2-dependent introns, which are ubiquitously distributed among all the eukaryotes, the presence of U12-dependent introns has so far only been identified in vertebrates, insects, cnidarians, *Rhizopus oryzae*, *Phytophthora* and *Acanthamoeba castellanii* (Burge C.B. et al., 1998). Surprisingly, the genome of the budding yeast *Saccharomyces cerevisiae* and the nematode *Caenorhabditis elegans* do not contain any U12-dependent introns or

components of the spliceosome (Burge C.B. et al., 1998; Russell A.G. et al., 2006; Mewes H.W. et al., 1997). Both the computational as well as experimental analyses in depth identified a large number of spliceosomal RNAs from different organisms. However, there are certain phylogenetic groups in which spliceosomal RNAs were not identified at all (Dietrich R.C. et al., 1997; Griffiths-Jones S, 2007) (Figure 7).

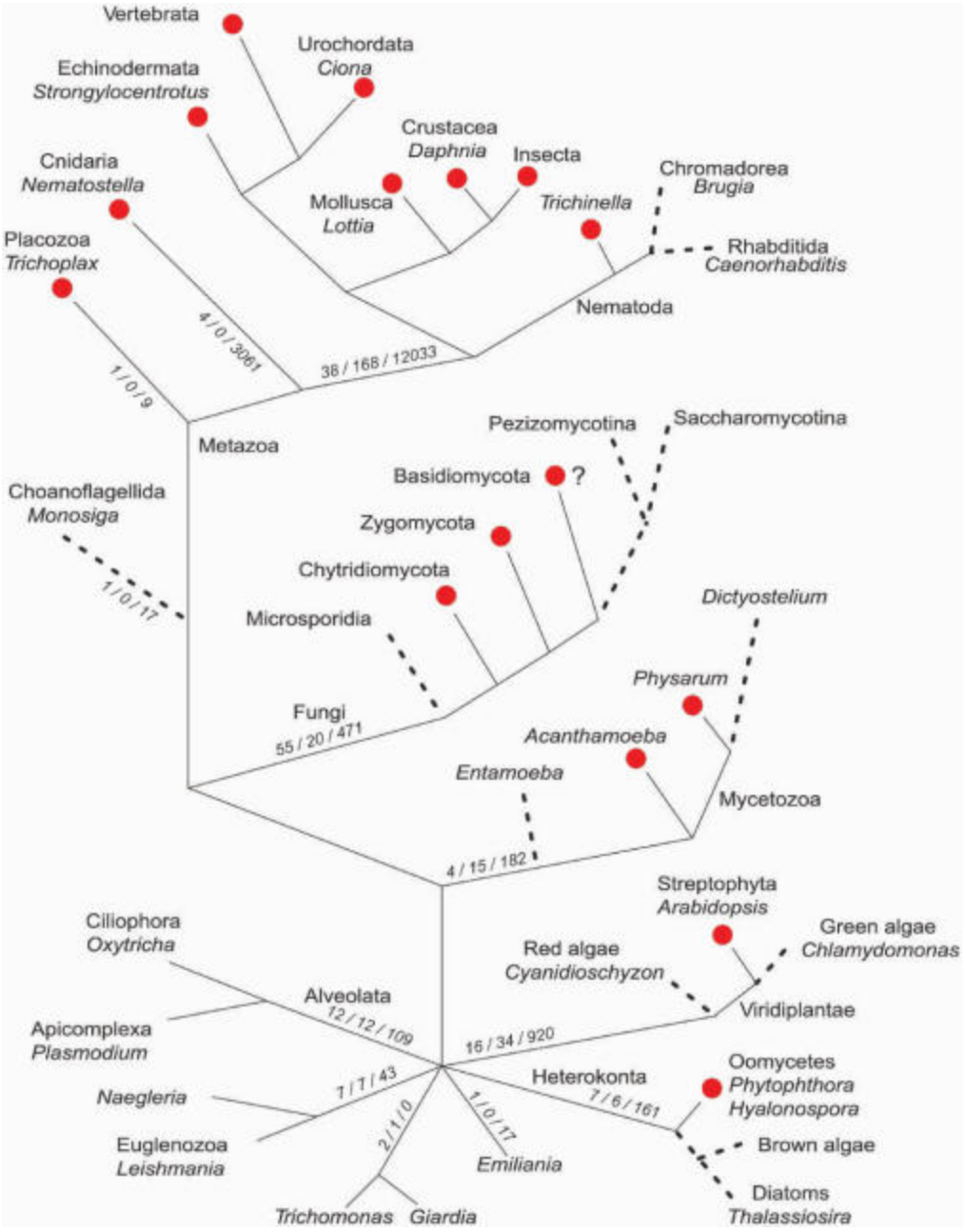


Figure 7: Schematic of a phylogenetic tree indicating where U12 introns were lost. Phylogenetic groups and their relationships have been shown together with example species. Species where one or more U12-dependent spliceosomal RNAs were found are highlighted in red circles as well as branches where the U12-type RNAs seem to have been lost are shown with dotted lines. Numbers at branches indicate 1) number of genomes analyzed, 2) number of query sequences used and 3) number of new sequences identified. (Lopez M.D. et al., 2008).

1.7 Slower kinetics of U12-dependent intron splicing regulates global gene expression

Both types of introns are removed cotranscriptionally in the nucleus (Singh J. and Padgett R.A., 2009) with almost similar splicing mechanism but with different kinetics. However, the rate of removal of U12-dependent intron is significantly slower. *In vitro* splicing experiments suggested a three to five-fold slower rate of U12-type intron splicing as compared to U2-type introns (Frilander M.J. and Steitz J.A.., 1999; Tarn W.Y. and Steitz J.A., 1996; Santoro B. et al., 1994). However, only two-fold higher accumulation of unspliced U12-dependent introns were detected *in vivo* in the steady-state transcript pools.

Further studies have shown that by converting U12-dependent intron to the U2-dependent intron in a protein coding transcript can increase the expression of the protein by almost six-folds (Santoro B. et al., 1994; Pessa H.K., 2006; Patel A.A. et al., 2002). The factors contributing to the slower rate of the minor spliceosome could be the relative lesser (100 fold less) abundance of the snRNPs as compared to the major snRNPs (Montzka K.A. and Steitz J.A., 1988; Tarn W.Y. and Steitz J.A., 1996). It is also hypothesized that high conservation of the splice sites might be resulting in rigid recognition of U12-dependent introns and therefore, linked to the slower kinetic rate of intron splicing. Regardless of the underlying mechanism, the slower rate of splicing of U12-dependent introns contribute to the rate-limiting mechanism for the expression of host genes (Patel A. A. et al., 2002). The transcripts with the unspliced U12-dependent intron are unable to escape the

nucleus and hence are targeted by nuclear surveillance mechanisms (Patel A. A. et al.,2002; Younis I. et al., 2013).

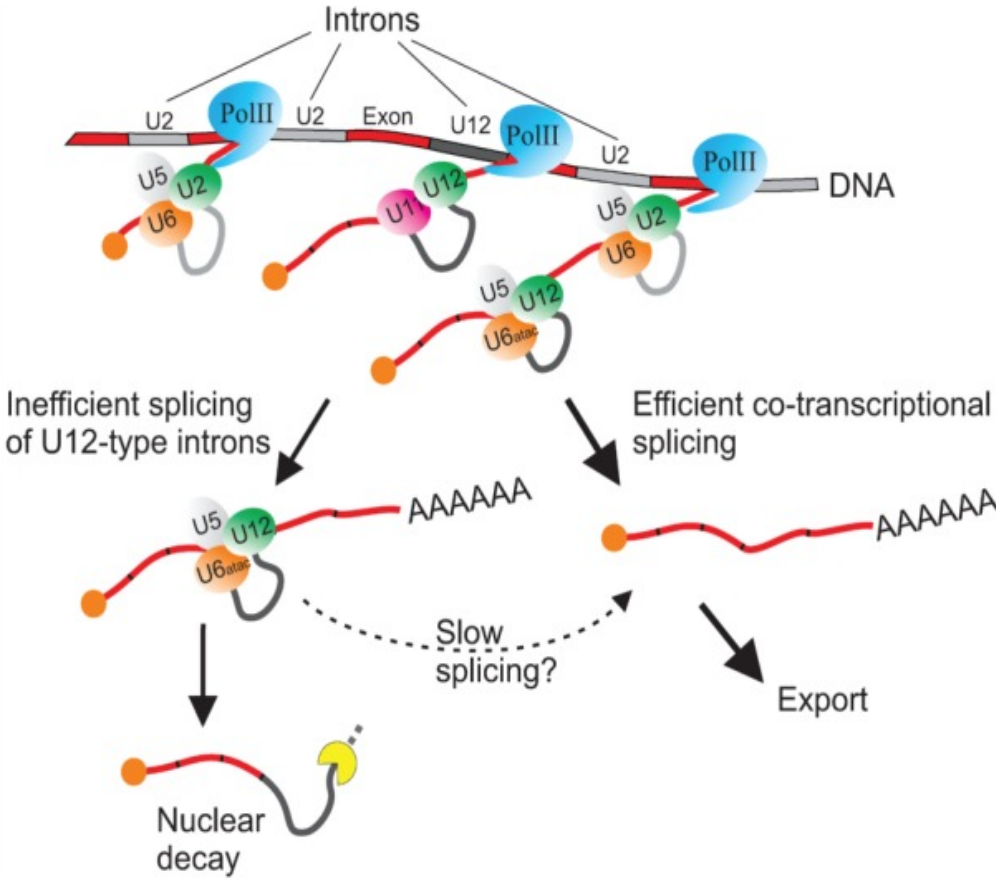


Figure 8: Schematic representation of rate-limiting regulation of gene expression by U12-type introns. Exons are indicated with a red color, U2-dependent introns are indicated in light gray, and U12-dependent introns are represented in dark gray. Similar to the U2-dependent introns, U12-dependent introns are also spliced co-transcriptionally. A majority, or all, U12-dependent introns are correctly recognized by the U11/U12 di-snRNP, and most likely can also assemble spliceosomal complexes. However, a subset of them fails to carry out the splicing reaction and is targeted by the nuclear quality control mechanisms. Alternatively, it is possible that this subset can be spliced more slowly post-transcriptionally (Niemela E. H. and Frilander M. J., 2014).

Stabilization of the transcripts hosting unspliced U12-dependent introns has been demonstrated upon knocking-down of the nuclear exosome. This observation clearly indicates that transcripts containing unspliced U12-dependent

introns are actively degraded by the nuclear quality control machinery (Figure 8). Together, these two observations strongly support for the rate-limiting control hypothesis which states that U12-dependent introns limit the expression levels of their host genes. Most likely the degradation activity is linked to nuclear retention whereby a delay in processing increases the probability of mRNA degradation, as suggested by the kinetic surveillance hypothesis of nascent pre-mRNAs (Burgess S.M. and Guthrie C., 1993; Doma M.K. and Parker R., 2007). Recently, it has been evident that the surveillance mechanisms are actively recruited to transcripts containing unspliced introns through interactions between the components of the stalled spliceosome and the nuclear quality control machinery (Nag A. and Steitz J.A., 2012).

The less abundant U6atac snRNP (approx. 2000 copies/cell) is strikingly unstable with a shorter half-life (t½<2 hr). The expression of U6atac snRNA depends on both RNA polymerases II and III, which can be rapidly enhanced by cell stress-activated kinase (p38MAPK), thus resulting in stabilization of U6atac. p38MAPK kinase mediated stabilization of U6atac further enhances downstream mRNA expression of genes containing minor introns which are otherwise suppressed by limiting U6atac snRNA. Interestingly, the minor intron-containing tumor suppressor PTEN (Phosphatase and Tensin Homolog) expression also depends upon p38MAPK-dependent U6atac mediated regulation. Therefore, minor class introns are considered as embedded molecular switches regulated by U6atac abundance (Younis I. et al., 2013).

1.8 Chimeric snRNAs can functionally activate *in vivo* U12-dependent intron splicing

The functional cores of U6 snRNA of U2-dependent and U6atac snRNA of U12-dependent spliceosomes are highly similar in sequence and are functionally interchangeable. Experimental evidences have demonstrated that a chimeric U6 snRNA generated by fusing the unique and highly conserved 3′ end of U6atac snRNA could efficiently activate the *in vivo* U12-dependent intron splicing (Figure 9). Furthermore, deletions of the 3' end of U6atac snRNA *in vivo*, as well as *in vitro* antisense experiments, suggested that the 3′ end of U6atac snRNA has U12-dependent spliceosome specific guiding activity for the U4atac.U6atac/U5 tri-snRNP complex, therefore directing the complex to the U12-dependent spliceosome (Dietrich R.C. et al., 2009).

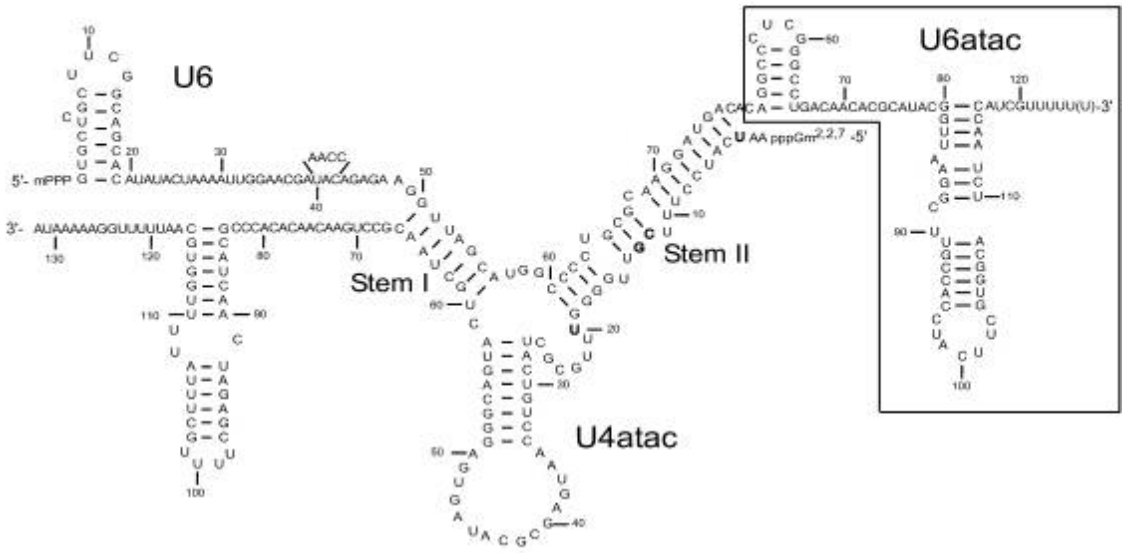


Figure 9: RNA-RNA base-pairing interactions of the chimeric U6/U6atac snRNA with the modified U4atac snRNA. The chimeric snRNA consists of U6 nucleotides 1–79 joined to U6atac nucleotides 52–125 (3′ stem–loop region shown within the box). Nucleotides at positions U40, A41,

and C42 were replaced by the sequence AACC, which are known to base pair with the mutant 5' splice site of U12-dependent intron F of P120 minigene. Mutations were introduced in the U4atac snRNA so that it can form Stem I and Stem II base pairing with chimeric U6/U6atac (Dietrich R.C. et al., 2009).

U4 and U6 snRNA form a base-paired di-snRNP complex that is essential for pre-mRNA splicing of the U2-dependent introns. The functionally analogous but highly diverged U4atac and U6atac snRNAs which are involved in splicing of U12-dependent introns also form a similar complex. It has been shown earlier that wild-type U4 snRNA interacts with the mutant U6atac snRNA (Shukla G. C. and Padgett R. A., 2001). Recent evidence shows that a mutant U4 snRNA designed to base pair with a mutant U6atac snRNA can activate U12-dependent splicing when co-expressed in an *in vivo* genetic suppression assay (Shukla G. C. and Padgett R. A., 2003). These potential interactions between U4 and U6atac snRNAs were also demonstrated by an *in vitro* crosslinking assay, suggesting that a U4/U6atac di-snRNP can be recruited to the U12-dependent spliceosome which could efficiently and precisely splice the U12-dependent intron. Moreover, it has been shown that the spliceosome type specificity is exhibited by U6 and U6atac snRNAs (Shukla G.C. and Padgett R.A., 2004).

Surprisingly, there is an evidence that the U6atac-ISL can be functionally replaced either by U6-ISL either from humans or from budding yeast (Shukla G.C. and Padgett R. A., 2001) (Figure 10). This demonstrates that an important substructure of the spliceosomal RNAs can be functionally substituted by the other between the two distinct splicing systems. Therefore, supporting both the significance of these similar elements, as well as the notion that these elements

are derived from parallel but distinct splicing systems in ancestral prokaryotic genomes that might have combined during the genesis of the eukaryotic lineage (Burge C.B. et al., 1998).

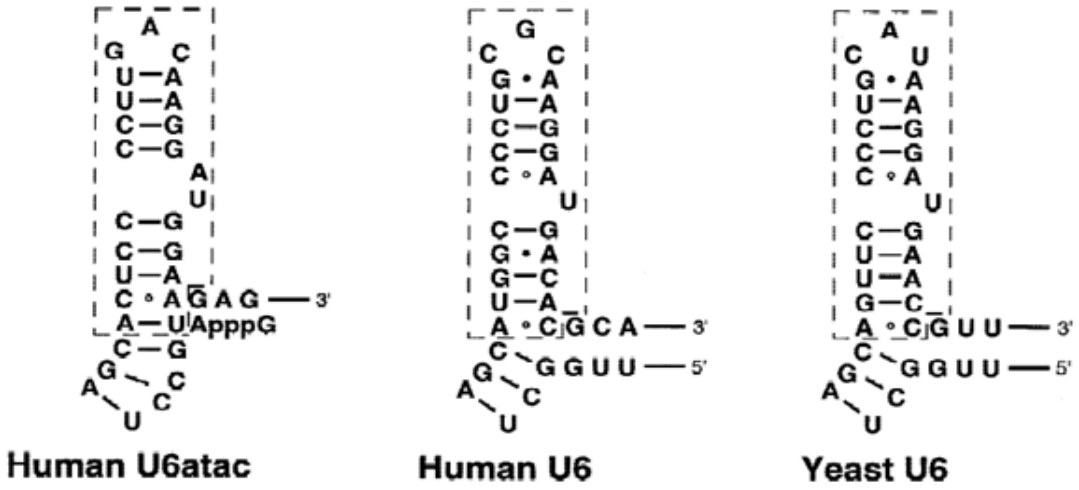


Figure10: Intramolecular stem-loop structures of various U6 and U6atac snRNAs. The left panel shows the sequence of Human U6atac forming intramolecular stem-loop (ISL) structure. The middle and the right panel represent the sequences of ISL from Human and yeast U6 snRNA. The Human U6 and yeast U6 ISL were substituted into the human U6atac snRNA (Shukla G.C. and Padgett R. A., 2001).

The first step of the catalytic reaction involved in the pre-mRNA intron splicing is mediated by the metal ion and a nucleotide in the intramolecular stem-loop structure of either U6 or U6atac snRNA. The secondary structure of ISL resembles the catalytically active structure of the domain 5 of self-splicing group II introns. Thus indicating that spliceosome is a ribozyme, where the active site is composed of RNA (Sashital et al., 2004).

U6atac and U12 snRNAs have been shown to interact with each other, as well as with the pre-mRNA by RNA-RNA base-pair interactions. The overall

secondary structures and many of the sequences are very similar to the highly conserved analogous regions of U6 and U2 snRNAs of U2-dependent spliceosome. The homologs of U6atac and U12 snRNAs in the plant *Arabidopsis thaliana* are significantly diverged in terms of overall sequence similarities from human. However, the sequences and structures which are involved in the splicing i.e. 5' splice site, branch site, intramolecular stem-loop structure and helix 1a and 1b regions formed by RNA-RNA base pair interactions between U12 and U6atac snRNAs are almost completely conserved. The intramolecular stem-loop structure formed by U6atac snRNA, after the U4atac snRNA is released, differs from the human sequence. This chimeric snRNA generated by transplanting this stem-loop into human U6atac was functionally inactive for *in vivo* U12-dependent intron splicing. However, the coexpression of a U4atac snRNA expression construct containing compensatory mutations that restored base pairing to the chimeric U6atac snRNA was able to restore the U12-dependent intron splicing (Shukla G.C. and Padgett R. A., 1999).

Interestingly, the interchangeability of D5, a catalytically active domain of the group II intron sequence and the intramolecular stem loop of U6atac formed in the catalytic core of the U12-dependent spliceosome, has been demonstrated (Shukla G.C. and Padgett R. A., 2002) (Figure 11). This suggests that a catalytic RNA structure lies at the heart of the spliceosome and shares an evolutionary history with group II introns. Taken together, these results clearly indicate that the catalysis of U12-dependent intron splicing proceeded by an identical RNA-based mechanism. Thus, mechanistic differences between U2- and U12-dependent

splicing appear to be limited to early intron recognition events, rather than to catalysis.

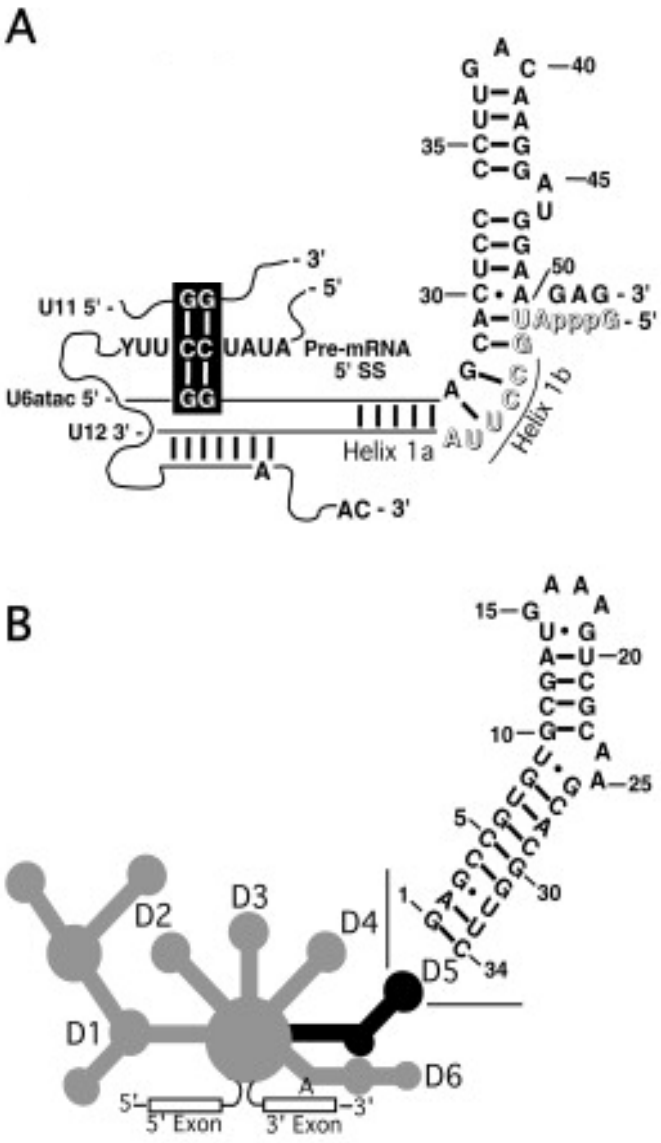


Figure11: Similar RNA structures in the U12-dependent Spliceosome and Group II Self-Splicing Introns. (A) RNA-RNA interactions in the catalytic core of U12-dependent spliceosome. The nucleotides which are mutated in the pre-mRNA and the U11 and U6atac snRNAs in order to perform the *in vivo* mutation suppression assay are highlighted in the dark box. The U11 and U6atac snRNA contains compensatory mutations to the ones introduced at the 5' splice site of the U12 dependent intron. The sequence of the U6atac-ISL has been shown. (B) Secondary structure of a group II intron showing the six domains. The sequence of the domain 5 is shown (Shukla G.C. and Padgett R. A., 2002).

1.9 P65 is highly conserved and an essential component of U11/U12 di-snRNP complex

Recently, recessive lethal point mutation in the RNA binding protein rnpc3 [RNA-binding region (RNP1, RRM) containing 3] gene has been associated with severe developmental arrest of digestive organs in zebrafish. The rnpc3 expresses the zebrafish ortholog of human RNPC3 which is also recognized as the U11/U12 di-snRNP 65-kDa protein, and is one of the unique protein components of the U12 spliceosome specific U11/U12 di-snRNP complex. This mutation in the rnpc3 gene is considered to interfere in the formation of the U11/U12 di-snRNP complex, therefore, resulting in impaired efficiency of U12-dependent intron splicing. Furthermore, it was suggested that the retention of the intron resulted in differential gene expression of multiple genes involved in various important cellular steps of mRNA processing such as transcription, splicing, and nuclear export are compromised in mutant zebrafish (Sebastian M. et al., 2014).

The functional role of a U12-dependent spliceosomal specific protein, U11/U12-65K in Arabidopsis thaliana, has been characterized. The miRNA-mediated transgenic knockdown generated plants showed severe growth and developmental defects. Experiments revealed compromised splicing of majority of predicted U12-dependent intron-containing genes in the U11/U12-65k mutant (knockdown). The U11/U12-65K is very conserved among different species (Figure 12) and its mutation also interferes with alternative splicing. The genes involved in cell wall biogenesis and function, plant development, and metabolic processes were the ones which were greatly affected in the mutant plant.

U11/U12-65K protein binds specifically to the highly conserved stem loop III of U12 snRNA (Benecke H. et al., 2005), which is necessary for branch site recognition. These results clearly demonstrate that U11/U12-65K is an essential component of the minor spliceosome, and plays a vital role in both U12- dependent intron splicing as well as in regulation of alternative splicing events of many introns, which are crucial for plant development (Jung H.J. and Kang H., 2014).

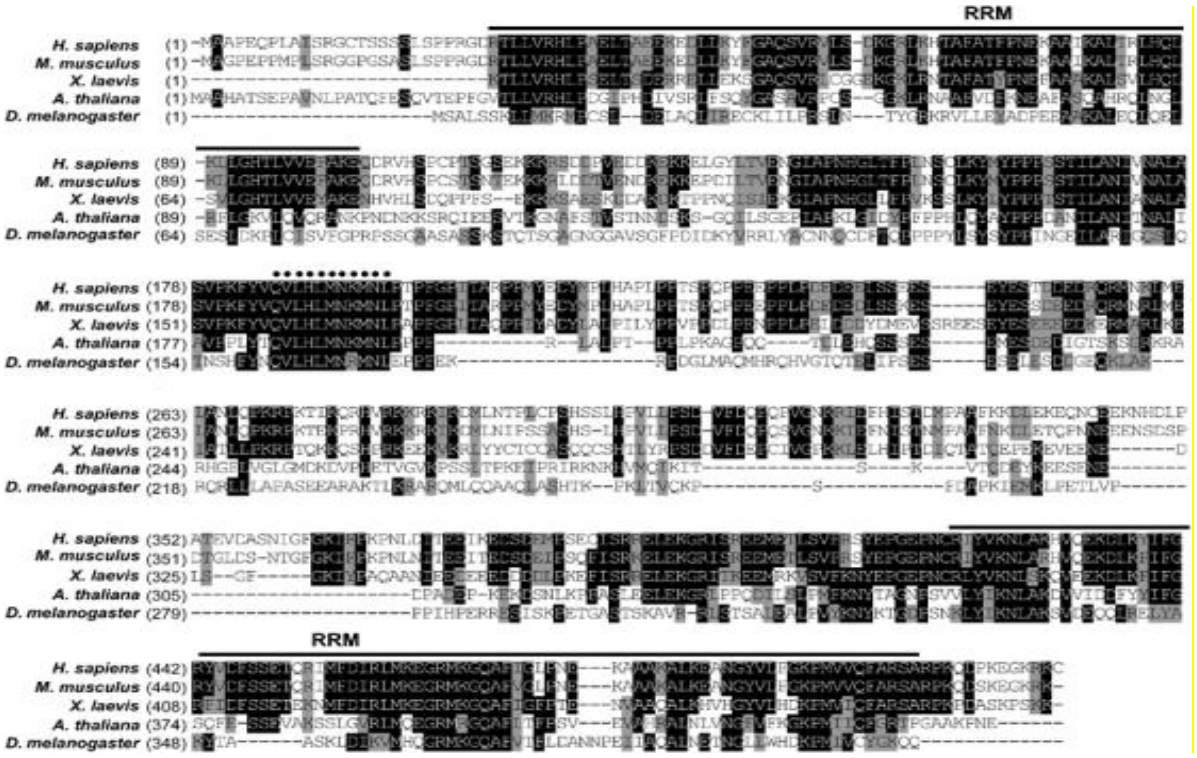


Figure 12: The U11/U12 di-snRNP specific 65K protein is evolutionarily conserved. Sequence alignment of 65K from *H. sapiens* and putative orthologs from *M. musculus*, *X. laevis*, *A. thaliana* and *D. melanogaster*. Residues identical in at least three proteins are highlighted in black and conserved residues are highlighted in gray. The RNA recognitions motifs are indicated by a bar on the top. The conserved sequence QVLHLMN(K/R) MNL is marked by dots (Benecke H. et al., 2005).

1.10 U12 Stem-loop III is indispensable for *in vivo* U12-dependent intron splicing

In both the spliceosomes U2 and U12 snRNAs interact with the branch site. U12 snRNA is a functional analog of U2 snRNA of the U2-dependent spliceosome and is essential for the splicing of U12-dependent introns. Recently, the secondary structures or regions of U12 snRNA which are outside of the branch site base pairing region are analyzed, and the functional requirement of these structures or elements in U12-dependent intron splicing has been studied in detail. The intricate predicted secondary structure of U12 snRNA contains several stem-loops separated by single-stranded regions (Figure 13). Data from the well characterized genetic suppression assay demonstrated that stem-loop IIa is an essential element and is required for *in vivo* U12-dependent intron splicing. Surprisingly, an evolutionarily conserved stem-loop IIb was found to be dispensable for *in vivo* U12-dependent intron splicing. Also, the stem-loop III, which has been shown to provide an interacting platform for C-terminal RNA recognition motif of p65 (Benecke H. et al., 2005) of the U11/U12 di-snRNP complex, is indispensable for *in vivo* U12-dependent intron splicing (Sikand K. and Shukla G.C., 2011) (Figure 13).

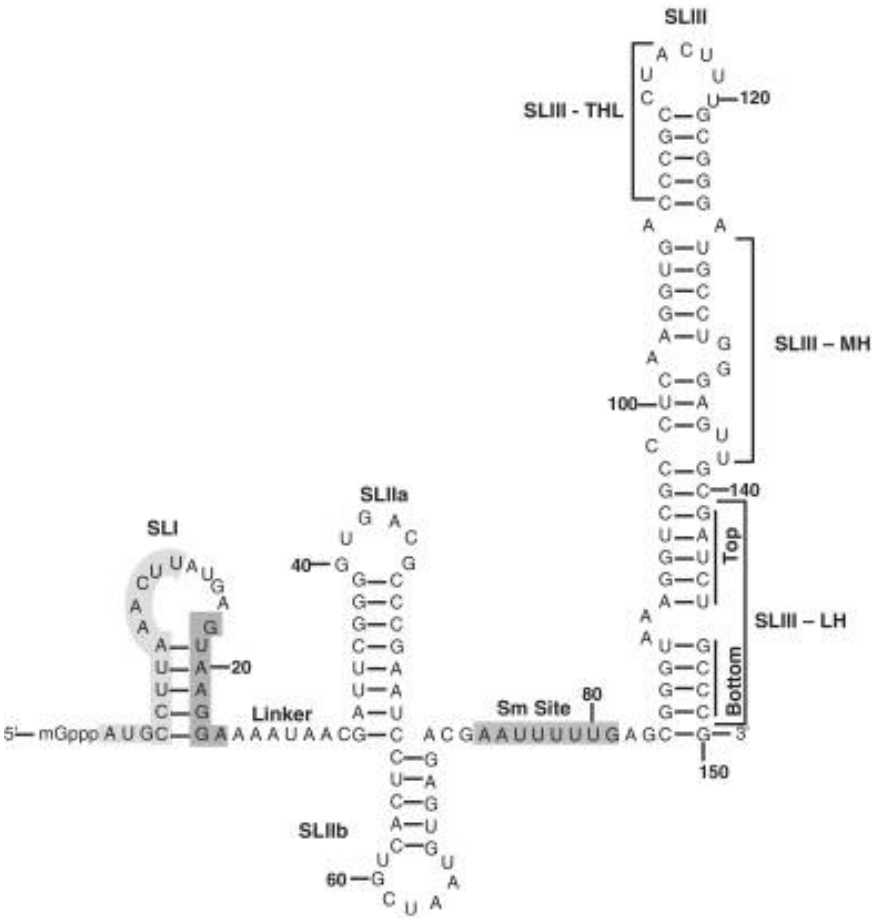


Figure 13: Predicted secondary structure of human U12 snRNA. The structure is based on the earlier predicted models. SLI, SLIIa, SLIIb, SLIII represent the four different stem-loop structures. SLIII-THL, MH and LH are the Terminal Helix and Loop, Middle Helix and the Lower Helix of stem loop III. Nucleotides 1–13 in SLI highlighted in light grey are predicted to interact with U6atac snRNA. Nucleotides 18–24, highlighted in dark grey SLI bind to branch site of a U12-dependent intron. The predicted Sm site is also indicated in grey shade between SL IIB and SL III (Sikand K. Shukla G.C., 2011).

1.11 U12 Stem-loop III interacts with C-terminal RRM of 65K protein

U11 and U12 snRNAs interact with each other to form a stable U11/U12 18S di-snRNP complex. This complex interacts with the pre-mRNA, in a cooperative manner, at the 5' splice site and the branch site, thereby acting as a molecular bridge between the 5' and 3' ends of the intron within the U12-dependent pre-spliceosome. By performing detailed *in vitro* RNA-protein interaction

experiments, it has been demonstrated that the U11/U12-65K protein interacts directly with the U12 snRNA stem loop III via its C-terminal RRM and with the U11-associated 59K protein via its N-terminal half (Benecke H. et al., 2005) (Figure 14). The 3' end of U12 snRNA forms an extended stem-loop with a highly conserved terminal seven-nucleotide loop. This stem-loop III of U12 snRNA also serves as the 65K binding site. The addition of an oligonucleotide designed to express the interacting site for 65K to an *in vitro* splicing reaction inhibited U12-dependent. However, the U2-dependent pre mRNA splicing was not affected at all. Taken together, these data suggest that U11/U12-65K and U11-59K contribute to di-snRNP formation and intron bridging in the minor pre-spliceosome (Benecke H. et al., 2005).

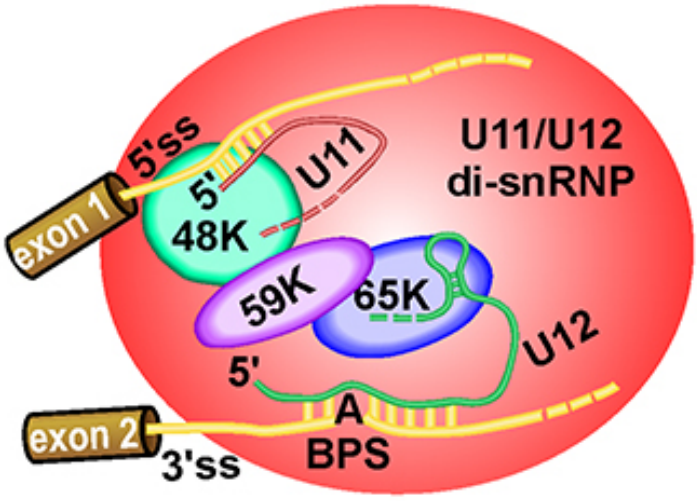


Figure 14: Schematic of the human U11/U12 di-snRNP complex. 65 K interacts with U12 snRNA via c-terminal RRM and acts as a bridging molecule with U11 snRNA through 59 K protein. Branch point sequence (BPS); "A" represents the branch point adenosine (Sebastian M. et al., 2014).

In this study we observed structural and sequence similarities between the distal 3' stem-loop of U6atac and stem-loop III of U12 snRNA. After successfully demonstrating the interchangeability of the substructures of the two snRNAs, we provided the evidence that the C-terminal RNA recognition motif of p65, a U12 snRNA binding protein, also interacts with the distal 3′ stem-loop of U6atac. By using well characterized genetic suppression assay we demonstrated the functional requirement of the distal 3' stem-loop of U6atac in U12-dependent *in vivo* splicing. Furthermore, we also tested the compatibility of the U6atac 3' end from phylogenetically distant species in a human U6atac suppressor background to establish the evolutionary relatedness of these structures and *in vivo* functionality. Although the significance of p65 interaction to U6atac snRNA in *in vivo* splicing is not clear, we have suggested that both the helix structure, as well as the sequence of U6atac distal 3' stem-loop is important for snRNA-protein interactions and U12-dependent intron splicing.

Chapter II

MATERIALS AND METHODS

2.1 Cell culture

Human Hela cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and antibiotics. CHO-K1 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 5% FBS, 2 mM L-glutamine, 1 mM L-proline, 10 mM 4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES) and antibiotics. All cell lines were maintained in a humidified 5% CO_2 atmosphere at 37 ^{0}C. Hela and CHO-K1 cells were obtained from ATCC (Manassas, VA).

2.2 Construction of snRNA expression plasmids

The U11 GG6/7CC, U6atac GG14/15CC and U12 GA23/24CU expression plasmids have been described previously (Hall S.L. and Padgett R.A., 1996; Kolossova I. and Padgett R. A., 1997; Incorvaia R. and Padgett R. A., 1998; Shukla G. C. et al.,

2002; Brock J. E. et al., 2008). Second site mutations were introduced in U6atac GG14/15CC and U12 GA23/24CU snRNAs to construct swapped SL snRNA plasmids and U6atac distal 3′ SL mutant plasmids. The swapped SL snRNA plasmids included two modified plasmids, namely, U6atac GG14/15CC w U12 SL and U12 GA23/24CU w U6atac SL. The distal 3' SL (nts. 91-109) of U6atac GG14/15CC plasmid was substituted with SLIII (nts. 109-125) of U12 snRNA to give rise to the U6atac GG14/15CC w U12 SL construct. Similarly, the U12 GA23/24CU w U6atac SL construct was made by replacing SLIII (nts. 109-125) of U12 GA23/24CU plasmid with distal 3′ SL (nts. 91-109) of U6atac snRNA. 5′ phosphorylated mutagenic oligonucleotides were used for site directed mutagenesis using the Change-IT mutagenesis kit (USB Corporation, Santa Clara, CA). The sequences of mutant snRNAs were confirmed by DNA sequencing.

2.3 P120 minigene construct

The construction of P120 minigene and the mutations introduced in the intron F are described previously by Kolossova I. and Padgett R. A. (1997). P120 minigene contains four exons (Exon 5 - 8) and three introns (intron E, F and G). The intron F is a U12-dependent intron flanked by U2-dependent introns (Figure 15). The information about the complete genomic sequence of P120 minigene is provided in the appendix.

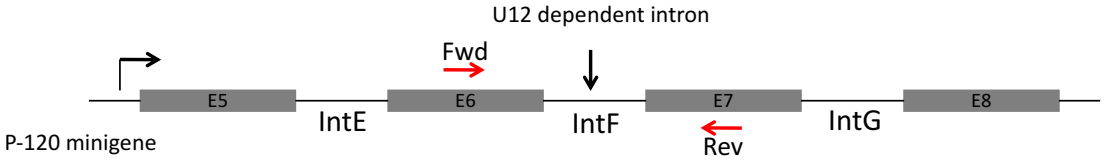


Figure 15: Schematic representation of U12-dependent intron containing P120 minigene. The above sequence comprising four exons (E5-7) and three introns (Int E, F and G) of P120 gene was amplified and cloned in mammalian expression vector pCB6. Intron F is the U12 dependent intron and the red arrows represent the forward and reverse gene specific primers used for NESTED PCR.

2.4 *In vivo* 5' splice site genetic suppression assay

The 5′ splice site, *in vivo* genetic suppression assays have been described previously (Shukla G.C. et al., 2002; Dietrich R.C. et al., 2009; Shukla G.C. and Padgett R.A., 2004). This assay uses a minigene derived from the nucleolar proliferating antigen gene P120 or NOL1. The P120 minigene plasmid contains four exons (5–8) and three introns (E, F and G). Introns E and G are U2-dependent introns whereas intron F is a U12-dependent intron. For the 5′ splice site suppression assay, the P120 minigene carrying a CC5/6GG mutation in the 5′ splice site of intron F was used. CHO cells were co-transfected with P120 CC5/6GG plasmid, U11 GG6/7CC suppressor snRNA and either U6atac GG14/15CC suppressor snRNA or U6atac GG14/15CC w U12 SL snRNA. For this experiment, 0.5 µg of P120 plasmid and 5 µg of each of the snRNA expression plasmids were used for transient transfection as described previously (Hall S.L. & Padgett R. A., 1996; Sikand K. and Shukla G.C., 2011). Where one or more snRNA plasmids were omitted, a corresponding amount of pUC19 plasmid DNA was substituted. Transfections with empty vector and wild type (WT) P120 plasmid

were carried out as controls in all experiments. Total RNA was isolated from cells 48 h after transfection and further processed as described by Sikand K. and Shukla G.C. (2011). Briefly, 500 ng of DNased RNA was reverse transcribed and PCR amplified using the Thermostable rTth Reverse Transcriptase RNA PCR kit (Applied Biosystems) and the following primers:

Forward primer: GGCCCGGGAAGCTGCTGCTGGGATC

Reverse primer: CTTCTAAGAACTCCACCAGCTCAGA

One microliter of the cDNA was then subjected to nested PCR under the following conditions: 94°C for 3 min, 25 cycles of 94°C for 1 min and 60°C for 1 min followed by a final extension at 68°C for 5 min. A reverse transcriptase minus control was performed to monitor DNA template contamination. PCR products were resolved on 3% agarose gel and were visualized using Ethidium Bromide (EtBr). The gel was scanned on a Typhoon 9410 variable mode imager (GE Healthcare, Little Chalfont, UK). Using ImageJ software, the intensity of bands was quantified. For each lane, the band intensity of each product (unspliced, U12 spliced, U12-cryptic spliced and U2 cryptic spliced) was expressed as the percentage of the total product.

Sequence of nested PCR primers used in 5′ splice site genetic suppression assays.

Forward primer: TTGTGCTGCCCCCTGCTGGGGAGATG

Reverse primer: TGAGCCCCAAAATCACGCAGAATTCC

2.5 *In vivo* branch site genetic suppression assay

The branch site of the U12-dependent intron F if P120 minigene contains the UC84/85AG mutation, which abolishes *in vivo* U12-dependent splicing. The U12 snRNA with compensatory mutation as GA23/24CU was sufficient to restore the wild type U12-dependent *in vivo* splicing (Hall S.L. and Padgett R.A., 1996). Second site mutations were introduced in the U12 snRNA, where the stem loop III was replaced with distal 3' stem loop of U6atac (nucleotide 91-109). U12 suppressor plasmid having GA23/24CU mutation and U12 GA23/24CU with U6atac stem loop were co-transfected with branch site P120 mutant into cultured CHO cells as described previously (Hall S.L. and Padgett R.A., 1996; Brock J.E. et al., 2008). For these experiments, 0.5 µg of P120 plasmid and 5 µg of each of second site mutation carrying U12 snRNA expression plasmids were added to CHO cells. Where one or more U12 snRNA plasmids were omitted, a corresponding amount of pUC19 plasmid DNA was substituted. Total RNA was isolated from cells 48 h after transfection and further processed as described by Sikand K. and Shukla G.C. (2011). Briefly, 500 ng of DNase-treated RNA was reverse transcribed and PCR amplified using the Thermostable rTth Reverse Transcriptase RNA PCR kit (Applied Biosystems) and the following primers:

Forward primer: GGCCCGGGAAGCTGCTGCTGGGATC

Reverse primer: CTTCTAAGAACTCCACCAGCTCAGA

One microliter of cDNA was then subjected to nested PCR under the following conditions: 94°C for 3 min, 25 cycles of 94°C for 1 min and 60°C for 1 min followed by a final extension at 68°C for 5 min. A reverse transcriptase minus

control was performed in all experiments to monitor DNA template contamination The nested PCR products were analyzed by agarose gel electrophoresis. The DNA bands were visualized using EtBr and scanned on Typhoon Image Scanner (GE Healthcare). Intensity of bands was quantified using the ImageJ software. Sequence of the primers used for nested PCR in branch site genetic suppression assays.

Sequence of nested PCR primers used in branch site genetic suppression assays.

Forward primer: TTGTGCTGCCCCCTGCTGGGGAGATG

Reverse primer: TCAGACAGAGGGAAGAGGTCCATGA

2.6 *In vivo* genetic suppression assay using binary splice site substrate

This assay also uses the same minigene derived from the nucleolar proliferating antigen gene P120 or NOL1. For the binary splice site suppression assay, intron F of the P120 minigene contained both the 5′ splice site CC5/6GG and the branch site UC84/85AG mutations. In this assay, CHO cells were co-transfected with P120 CC5/6GG + UC84/85AG plasmid, U11 GG6/7CC snRNA, U6atac GG14/15CC snRNA or U6atac GG14/15CC w U12 SL snRNA or each of U6atac GG14/15CC distal 3′ SL mutants or U6atac chimera snRNAs and U12 GA23/24CU snRNA or U12 GA23/24CU w U6atac SL snRNA. For all experiments, 0.5 µg of P120 plasmid and 5 µg of each of the snRNA expression plasmids were used for transient transfection as described previously. Where one or more snRNA plasmids were omitted, a corresponding amount of pUC19 plasmid DNA was

substituted. Transfections with empty vector and wild type (WT) P120 plasmid were carried out as controls in all experiments. Total RNA was isolated from cells 48 h after transfection, reverse-transcribed and PCR-amplified as described earlier (Sikand K. and Shukla G.C., 2011). A total of 500 ng of DNase-treated RNA was reverse transcribed and PCR amplified using the Thermostable rTth Reverse Transcriptase RNA PCR kit (Applied Biosystems) and the following primers:

Forward primer: GGCCCGGGAAGCTGCTGCTGGGATC

Reverse primer: CTTCTAAGAACTCCACCAGCTCAGA

One microliter of the cDNA was then subjected to nested PCR under the following conditions: 94°C for 3 min, 25 cycles of 94°C for 1 min and 60°C for 1 min followed by a final extension at 68°C for 5 min. A reverse transcriptase minus control was performed in all experiments to monitor DNA template contamination. PCR products were visualized by agarose gel electrophoresis using EtBr and scanned on a Typhoon 9410 variable mode imager (GE Healthcare, Little Chalfont, UK). The intensity of bands was quantified using ImageJ software. For each lane, the band intensity of each product (unspliced, U12 spliced, U12-cryptic spliced and U2 cryptic spliced) was expressed as the percentage of the total product. Each snRNA suppressor construct was transfected a minimum of three times in two stocks of cells. Independent transfections and analyses gave essentially similar results.

Sequence of nested PCR primers used in 5′ splice site genetic suppression assays.

Forward primer: TTGTGCTGCCCCCTGCTGGGGAGATG

Reverse primer: TGAGCCCCAAAATCACGCAGAATTCC

2.7 IPTG based induction of GST-tagged C-terminal RNA Recognition Motif and protein purification using affinity based method

GST-p65-C-RRM construct was received as a gift from Dr. Luhrmann's lab as a bacterial stock. The bacterial culture (E. coli, BL 21 cells) was streaked on the Luria-broth (LB) plates with ampicillin as a drug resistance. A mini culture from single colony was prepared by picking single colony in 5ml of LB media and ampicillin. The inoculated mini-culture was incubated in a shaker at 37 ^{0}C, 300 RPM for 12-16 hours. Next day the 40ml of LB media was inoculated with 0.1% of bacterial culture and let it grow under same conditions for 2-3 hours with constant monitoring of the optical density (O.D.) using spectrophotometer (from Fisher Scientific) at 260nm. Once the O.D. was reached in the range from 0.4-0.6 the bacterial culture was induced using 0.5 mM IPTG. The induced culture was again kept for shaking for another 4-5 hours. The culture without the addition of IPTG was used as a negative control for our protein induction experiment. After 4-5 hours of induction 3 ml of the culture was spun at 14, 000 RPM at room temperature for 30 seconds and rest of the volume was spun in a different culture tube under same spinning conditions. The bacterial cell pellet from 3 ml of induced and un-induced culture was lysed in 4X loading buffer. The cells re-suspended in the loading buffer were boiled for 10 mins and the lysed cells were spun at 10,000 × g for 30 seconds. The supernatant thus obtained was loaded on SDS-PAGE gel

with 4-12% gradient. Once the induction was confirmed The cell pellet obtained from the bigger culture was lysed with Bacterial Protein Extraction Reagent (B-PER) and further processed for protein purification. Fusion protein was purified by using the B-Per® GST Spin Purification Kit (Thermo Scientific/Pierce, Waltham, MA) according to the manufacturer's recommendation and the kit protocol. Briefly, the column(s) were equilibrated to working temperature and all purifications were performed at room temperature. Samples were prepared by mixing protein extract with Equilibration/Wash Buffer so that the total volume equals at least two resin-bed volumes. The bottom tab from the Pierce Glutathione Spin Column was removed by gently twisting. After snapping the bottom part of the column, it was assembled into a collection tube. The columns were centrifuged at 700 × g for 2 minutes to remove storage buffer. The column was equilibrated with two resin-bed volumes of Equilibration/Wash Buffer. The buffer was allowed to enter the resin bed. Another round of column centrifugation was done at 700 × g for 2 minutes to remove buffer. The protein lysate prepared in B-PER reagent was added to the column and allowed it to enter the resin bed. For maximal binding, the sample was incubated for 30-60 minutes at room temperature or 4 ^{0}C on a rocking platform. The column was centrifuged again at 700 × g for 2 minutes and the flow-through was collected in a new centrifuge tube. Resin washing was performed with two resin-bed volumes of Equilibration/Wash Buffer and centrifugation was done at 700 × g for 2 minutes. The wash step was repeated twice and GST-tagged fusion protein bound to the resin was eluted by adding one resin-bed volume of Elution

Buffer and centrifugation was done at 700 × g for 2 minutes. This step was repeated twice, collecting each fraction in a separate tube. The eluted protein was analyzed by resolving the eluate SDS-PAGE and staining with commassie brilliant blue.

2.8 5' end labeling of the RNA oligos and Electrophoretic Mobility Shift Assay

RNA oligonucleotides with the WT U12 SLIII (nts. 109-125) sequence, and the WT and mutant U6atac distal 3′ SL (nts. 91-109) sequences were obtained from IDT (Coralville, IA). The sequence of the oligonucleotides from 5' to 3' orientation is given below and the mutations introduced are highlighted in red.

WT U12	5′ – CCC GCC UAC UUU GCG GG – 3′
WT U6atac	5′ – UGC CAC CUA CUU CGU GGC A – 3′
U6atac U98G	5′ – UGC CAC CGA CUU CGU GGC A – 3′
U6atac A99C	5′ – UGC CAC CUC CUU CGU GGC A – 3′
U6atac C100G	5′ – UGC CAC CUA GUU CGU GGC A – 3′
U6atac Stem 1	5′ – UGC CAU CUA CUU CAU GGC A – 3′
U6atac Stem 2	5′ – UGC CAC CUA CUU CCA CCG U – 3′
U6atac Stem 3	5′ – ACG GUG CUA CUU CCA CCG U – 3′
U6atac Stem 4	5′ – UGC CAG CUA CUU CCU GGC A – 3′
U6atac Stem 5	5′ – ACG GUC CUA CUU CGA CCG U – 3′

U6atac Stem 6 5′ – GUG GCA CUA CUU CGU GGC A – 3′

U6atac Comp loop 5′ – UGC CAC GAU GAA GGU GGC A – 3′

These oligonucleotides were labeled at their 5′ ends using ATP γ-^{32} P (MP Biomedicals, Cleveland, Ohio) and T4 polynucleotide kinase (from New England Bio-labs) with the recommended recipe

RNA oligonucleotides	Up to 50pmol
10X T4PNK Reaction Buffer	5 µl
ATP	50 pmol of [γ-32P] ATP
T4 Polynucleotide kinase	2 µl (20 units)
Nuclease free water	up to 50 µl

The labelled reaction was then passed through the Quick Spin Columns pre-containing RNase free suspension of Sephadex G-25 in STE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl) for radiolabeled RNA purification purpose.

2 µl of the radiolabeled RNA eluate was resuspended in 5 ml of scintillation fluid and was subjected to the scintillation counter at Cleveland Stated University (Instrumentation facility). Two hundred thousand counts of ^{32}P-labeled oligonucleotides were incubated with or without GST-p65-C-RRM (0, 20, 40, 60 nmoles). A 20 µl reaction was prepared in the binding buffer containing 10 mM HEPES (pH 7.6), 5 mM $MgCl_2$, 100 mM KCl, 1 mM DDT and 5% glycerol. After 40 min of incubation at room temperature, the reaction was loaded on a 6% native polyacrylamide gel with 5% glycerol. The gel was run for 3.5 h at 150 Volts and

then exposed overnight to a storage phosphor screen. The exposed screen was read using a Typhoon 9410 variable mode imager (GE Healthcare, Little Chalfont, UK). The intensity of the bands was quantified using ImageJ software.

2.9 *In vivo* RNA protein pull down assay

Full length p65 (1603bp) open reading frame (ORF) was cloned in pcDNA 3.1 (-) mammalian expression plasmid with FLAG and 6 × His tag upstream (N-terminal) and downstream (C-terminal) of the ORF respectively. Two hundred thousand Hela cells/ well were seeded in 6-well plate. 2 µg of the expression plasmid was transfected in each well using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Cells were washed with ice cold 1X phosphate buffer saline (PBS) buffer prior to collection and resuspension in 200 µl ice cold lysis buffer [100 mM KCl, 5 mM $MgCl_2$, 10 mM HEPES (pH 7.0), 0.5% NP-40, 1 mM DDT, 100 units/ml RNase inhibitor] for overnight storage at -80 ^{0}C. 100 µl of Agarose A/G beads were centrifuged at 2000 × g for 2 min. The beads were washed with 300 µl of ice cold NT-2 buffer [50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM MgCl2, 0.05% IGEPAL (NP-40)] thrice. Finally, the beads were resuspended in 5 ml of NT-2 buffer along with 5 µg of anti-FLAG antibody in a 15 ml conical tube. The resuspended beads were mixed end to end overnight at 4 ^{0}C. Beads with no antibody were used as a negative control for this experiment. After 12-14 hours, the cells resuspended in lysis buffer were centrifuged at 15000 × g for 15 min at 4 ^{0}C. The supernatant was saved for downstream experiments and the pellet was discarded. Simultaneously, the beads were centrifuged at 2000 × g

for 2 min at 4 ^{0}C. The anti-FLAG antibody conjugated beads were washed with ice cold NT-2 buffer three times and resuspended in 800 µl of IP buffer (100 mM DDT, 0.5 M EDTA, 200 units of RNase inhibitor, add ice cold NT-2 buffer to make the volume up to 850µl). The cell lysate was added to this mixture and incubated at 4 ^{0}C for 4 h with end to end mixing. After this incubation, 100 µl of the resuspended beads were collected in a new tube for total RNA extraction. All samples were centrifuged at 2000 × g for 2 min at 4 ^{0}C. The beads were washed 5 times with ice cold NT-2 buffer. For the final wash, beads were transferred to a new tube and RNA was extracted using Trizol following the manufacturer's protocol. Total RNA was obtained after digesting DNA with DNase I (Promega). RNA was then reverse transcribed using an Improm-II cDNA synthesis Kit (Promega). This step was achieved by mixing the random hexamer and the oligo dT primers 9:1. The reaction where reverse transcriptase was omitted served as a control for RT specificity. The cDNA thus obtained was subjected to PCR amplification using sets of gene specific primers for U12, U6atac and U5 snRNAs. The schematic of the immuno-precipitation is represented in Figure 16.

The primers used for the PCR are listed below.

U12 forward primer sequence: GAGTAAGGAAAATAACGATTCGGGG

U12 reverse primer sequence: CAGGCATCCCGCAAAGTAGGC

U6atac forward primer sequence: GTATGAAAGGAGAGAAGGTTAGC

U6atac reverse primer sequence: GGTTAGATGCCACGAAGTAG

U5 forward primer sequence: CTCTGGTTTCTCTTCAGATCGC

U5 reverse primer sequence: GCCAAGGCAAGGCTCAAAAAATTG

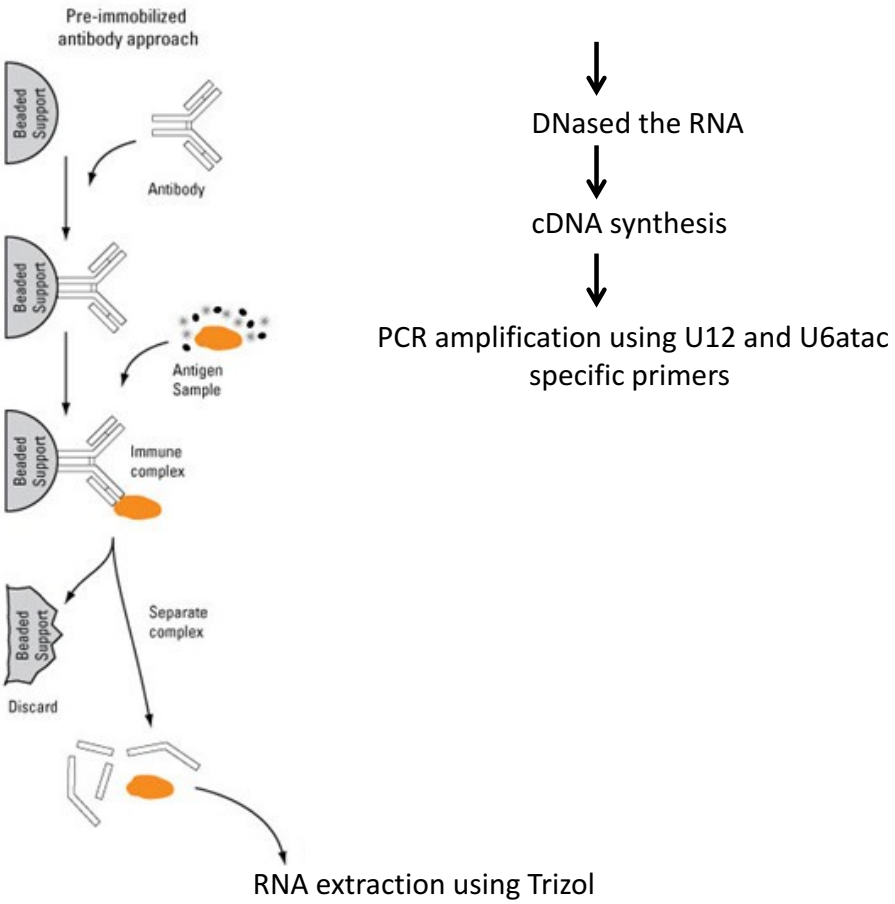


Figure 16: Schematic representation of immuno-precipitation of FLAG tagged protein. Anti-FLAG antibodies were immobilized on agarose A/G beads. The beads conjugated with antibody were incubated with total protein lysate. After washing the unbound protein, the beads/antibody/protein complex was harvested for total RNA. DNA contamination free RNA was obtained by treating the RNA with DNase. Using reverse transcription primers, the cDNA thus obtained was further used to amplify U12 and U6atac with gene specific primers.

2.10 Hybrid U6atac snRNA designing strategy

In order to test the evolutionary relatedness of the 3' ends of U6atac from phylogenetically distant species, the following hybrid U6atac sequences were designed with GG14/15CC background. The designed fragments were synthesized from the Bio-Basic, Amherst, New York. These sequences were designed with Not I and BamH I restriction sites at the 5' and the 3' end respectively. The sequences thus obtained were sub-cloned in the mammalian expression vector pcDNA 3.1(-). The clones were further confirmed via both the restriction digestion as well as sequencing method. Using these hybrid U6atac snRNAs genetic suppression assay was performed by using binary splice site mutant intron as a substrate. This assay required the co-expression of U11GG6/7CC, U12GA23/24CU and U6atac hybrid with GG14/15CC background mutation along with the substrate. The designed U6atac hybrid sequences are given below from 5' to 3' orientation.

U6atac Human

GTGTTGTATGAAACCAGAGAAGGTTAGCACTCCCCTTGACAAGGATGGAAGAGGCCCTC
GGGCCTGACAACACGCATACGGTTAAGGCATTGCCACCTACTTCGTGGCATCTAACCAT
CGTTTTT

U6atac Human/Arabidopsis

GTGTTGTATGAAACCAGAGAAGGTTAGCACTCCCCTTGACAAGGATGGAAGGACCTTCG
GGTCTTTGAACACATCCGGTTAAGGCTCTCCACATTCGTGTGGATCTAAACCCAATTTT
TT

U6atac Human/Phytophthora

GTGTTGTATGAAACCAGAGAAGGTTAGCACTCCCCTTGACAAGGATGGAAGTGCGCGTA
TGCGTATCCAACCACTGGATGGTTTAAGCTCTGTCCTCCTCCGTGAAGACATCTACCAG
TTTTTTT

U6atac Human/Drosophillo

GTGTTGTATGAAACCAGAGAAGGTTAGCACTCCCCTTGACAAGGATGGAAGCACATAAA
CGGTCGGCTAGGCACAGACAAAAGCCGTCCACAAATTTTTTT

U6atac Human/Trichinella

GTGTTGTATGAAACCAGAGAAGGTTAGCACTCCCCTTGACAAGGATGGAAGAGAATTGT
CAGACAAGCACAACAGCAAAGATCGCCTACCTTTGCTCCCCATTTTTT

2.11 Design and Expression of U11/U12 di-snRNP complex specific protein constructs

The cDNA sequences of Open Reading Frame (ORF) of seven U11/U12 di-snRNP specific proteins were obtained from the NCBI website. The NCBI or GeneBank Reference sequence information is provided in the table below. The codon optimization and the genes were synthesized from Gene-Art (Life Technologies Corporation, CA, US). These genes were designed with the with NotI and BamHI restriction sites at the 5' and 3' end respectively for the sub-cloning purposes in mammalian expression vector, pcDNA3.1(-). The cloned fragments were confirmed by both double digestion using the same restriction enzymes as well as by sequencing. For the purpose of *in vivo* immuno-precipitation experiments these cDNAs were tagged with FLAG tag and 6 × His tag at the N-terminal and C-terminal end respectively. The sequences for the following genes with the complete design strategy is given in the appendix I section.

Protein	NCBI Ref. Sequence	Gene Bank Ref. sequence
U11/U12 20K	NM_001003692.1	***
U11/U12 25K	NM_024571.3	***
U11/U12 31K	***	BK005200.1
U11/U12 35K	NM_180699.2	***
U11/U12 48K	***	BK005199.1
U11/U12 59K	***	BK005196.1
U11/U12 65K	***	BK005195.1

Table 1: U11/U12 di-snRNP specific proteins with their respective NCBI reference sequence number or Gene Bank reference sequence number.

Hela cells were seeded in 100mm plates a day before transfections. At the confluency of 70-80 %, cells were transfected with the 5 micrograms of U11/U12 specific protein expressing constructs (described in 2.10) using Lipofectamine 2000 (Invitrogen, Carlsbad, CA). 48 hours post transfections the cells were collected in centrifuge tubes and were harvested for protein by re-suspending in 100 ul of M-PER (Mammalian Protein Expression Reagent; Pierce) supplemented with protease inhibitor and kinase inhibitor cocktail. The tubes were incubated at room temperature for 40 mins with end to end mixing. After quantifying the protein concentration using Bradford reagent. 10 micrograms of the total protein was loaded and resolved on NuPAGE 4–12 % Bis-Tris gels (120 volt for 2 hours at room temperature) and electro-transferred to nitrocellulose membranes (80 volts

for 1.5 hours at 4 ^{0}C). The following antibodies were used: mouse monoclonal anti-FLAG antibody (1:1000, SIGMA) and horseradish peroxidase conjugated anti-mouse secondary antibody (1:10,000, GE Healthcare, Piscataway, NJ). Bands were detected using the Western blotting detection reagent (ROCHE) and LI-COR.

2.12 Designing, induction and purification of U11/U12 di-snRNP complex specific protein in bacterial cells

The open reading frame sequences of the seven U11/U12 di-snRNP specific proteins were obtained from the sources mentioned above in 2.11. For bacterial expression, the codon optimization for all the sequences was done by Dr. Anton A. Komar and the genes were synthesized from the Gene-Art (Life Technologies Corporation, CA, US). These genes were designed with Nde-I and BamH-I restriction sites on their 5' and 3' ends respectively for subsequent sub-cloning purposes in the bacterial expression pET3a vector. The cloned fragments were confirmed by both double digestion using the same restriction enzymes as wells by sequencing. For the purification purposes these fragments were tagged with FLAG tag and 6 His tag at the N-terminal and C-terminal end respectively. The sequences for the following genes with complete design strategy is given in the appendices section.

The pET3a vectors containing gene of interest were transformed in the BL-21 gold cells (Bacterial strain). A single bacterial colony was used to inoculate the 5ml miniculture using drug resistance. The culture was allowed to grow over night at 37 ^{0}C at 350 RPM. 0.1 % of the this miniculture was used to inoculate 40ml of

Luria broth media. The growth of the bacterial cells was monitored by using spectrophotometer at 600 nm. The plane LB media was used for blanking the machine. When the O.D. was reached in the range of 0.4-0.6 the bacterial culture was induced by using 0.5 mM IPTG. The induced culture was incubated for 4-5 hours in the shaker under the same conditions mentioned above. The culture with no IPTG induction was used as the negative control.

After 4-5 hours of induction the bacterial cells were harvested by centrifuging the culture at 5000 RPM for 5 minutes at room temperature. The cell pellet thus obtained was resuspended in 8 ml of the Guanidinium Lysis Buffer, pH 7.8 (6 M Guanidine HCl, 20 mM Sodium phosphate, pH7.8 and 500 mM Nacl). After proper resuspension the cells with the lysis buffer were kept on the rocker for 5-10 mins in order to insure thorough lysis. The cell lysates were sonicated with three 5 seconds pulse at high intensity followed by the centrifugation at 3000 × g for 15 mins to remove the cell debris. The supernatant was transferred to the fresh tube. Before the purification step 50 ul of the supernatant was resolved in denaturing SDS-PAGE gels to confirm the induction.

Since all the fusion proteins are expressing the 6 × His tag at the N-terminal, we used the His-Bind resin (EMD Millipore, USA). Before using it the resin was resuspended gently by end to end mixing in a tube. 2ml of the resin was used in a 10 ml Purification Column and allowed the resin to settle down completely by low speed centrifugation at 800 × g for 1 min. The supernatant was removed very carefully avoiding the loss of resin. His-bind resin was washed with 6ml of nuclease free water followed by above mentioned low speed centrifugation. The resin pellet

was washed twice with the denaturing binding buffer (8 M Urea, 20 mM sodium phosphate pH 7.8, 500 mM NaCl). After washing the resin pellet was resuspended in 7 ml of the bacterial cell lysate and was incubated for 2 hours with gentle agitation followed by low speed centrifugation. The pellet was subsequently washed twice with denaturing binding buffer and twice with the denaturing wash buffer (8 M Urea, 20 mM NaH_2PO_4 pH 6.0, 500 mM NaCl). Finally, the His bound fusion protein was eluted by re-suspending the pellet in 2 ml of denaturing elution buffer (8 M Urea, 20 mM NaH_2PO_4 pH 4.0, 500 mM NaCl). In order to ensure the purification, the eluate was resolved on the denaturing SDS-PAGE gel with 4-12 % gradient.

Chapter III

RESULTS

3.1 Structural and sequence similar elements of U6atac and U12 snRNAs

U6atac snRNA interacts with the 5' splice site after the U11 snRNA interactions with the 5' splice site is destabilized and U11 is released during B* complex formation. Simultaneously, U12 snRNA interacts with the branch site of U12-dependent intron. Both the U12-type specific snRNAs play a pivotal role in the catalyzing the splicing. The secondary structure present at the 3' end of these two snRNAs exhibit structural and sequence similarity. The alignment of these elements (nucleotides 109-125 of U12 snRNA and nucleotides 91-109 of U6atac) from two different snRNAs shows identical 7 nucleotides base in the loop region of both the stem loops (Figure 17).

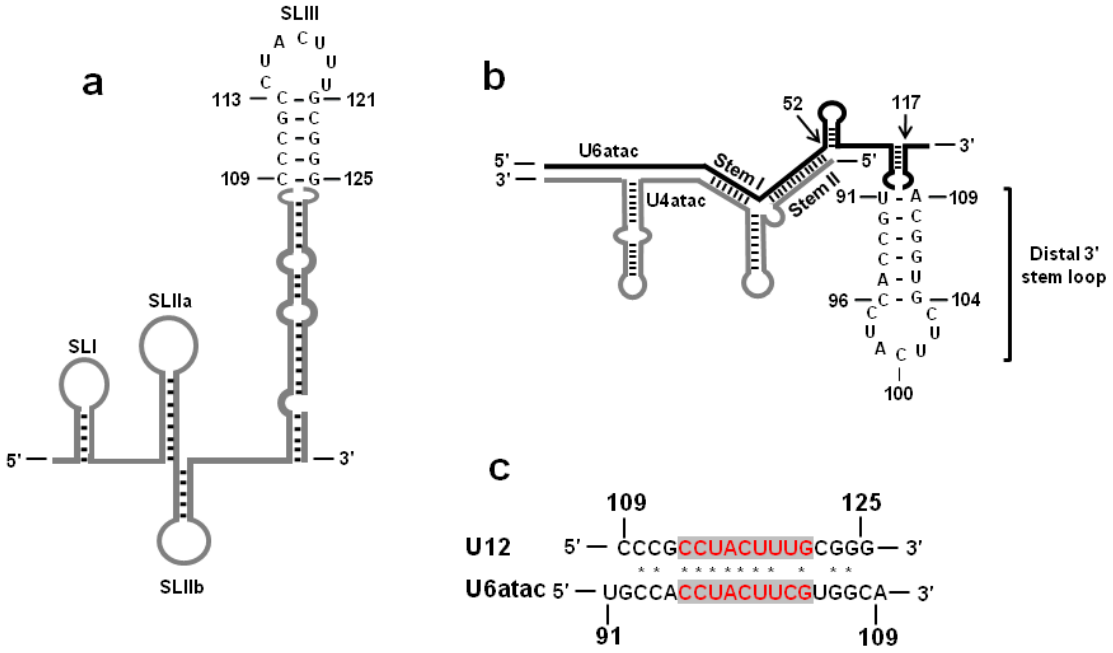


Figure 17. Structure and sequences of the human U12 and U6atac snRNAs. (a) Line diagram showing the secondary structure of human U12 snRNA based on the model of Benecke H. et al., 2005. The sequence and base pairing interactions of the SLIII are shown. (SL:stem-loop) (b) Schematic of the secondary structure of human U6atac (black) and U4atac (grey) snRNAs based on the model of Padgett and Shukla[19]. The sequence and base pairing interactions of the distal 3' SL of U6atac snRNA are shown. (c) Sequence comparison of the SLIII of U12 snRNA and distal 3' SL of U6atac snRNA. Asterisks denote identical nucleotides. Loop nucleotides are highlighted in grey.

3.2 U12 snRNA with U6atac distal 3' Stem-loop activates U12-dependent intron splicing

To test if the U6atac distal 3′ SL (nts. 91-109, Figure 17b) can functionally replace the p65 binding SL (nts. 109-125, Figure 17a) in U12 snRNA, we used the U12 branch site mutation suppressor assay. This assay relies on the base pairing of the P120 branch site UC84/85AG mutant with U12 snRNA containing a compensatory GA23/24CU mutation (Figure 18a). We modified the suppressor U12 snRNA containing a first site GA23/24CU mutation by replacing its p65 binding apical SLIII (nts. 109-125, Figure 17a) with U6atac distal 3′ SL (nts. 91-109, Figure 17b) as a second site mutation. P120 UC84/85AG alone was inactive in WT U12-dependent splicing; instead, the intron was spliced using a cryptic 3′ U12-dependent branch site (Figure 18b and c, lane 3) (Hall S.L. and Padgett R.A., 1996; Sikand K. and Shukla G.C., 2011). Co-transfection of P120 UC84/85AG and U12 GA23/24CU suppressed the downstream cryptic splicing and restored the splicing from the WT U12-dependent splice sites (Figure 18b and c, lane 4). U12 GA23/24CU with complementary loop nucleotide (nts. 114-120, Figure 17a) sequence show a reduced spliced phenotype as compared to WT with very little splicing from the cryptic splice site (Figure 18b and c, lane 5). U12 GA23/24CU with deleted SL (nts. 109-125, Figure 17a) was largely inactive for WT splicing (Figure 18b and c, lane 6) (Sikand K and Shukla G.C., 2011). Similar activation of WT splicing was observed when U12 GA23/24CU suppressor snRNA containing the U6atac distal 3′ SL (U12 GA23/24CU w U6atac SL) was co-transfected with P120 UC84/85AG (Figure 18b and c; compare lane 7 with lanes 2 and 4). These

results demonstrate that the U6atac distal 3′ SL, when expressed in the context of the U12 snRNA, can functionally replace the U12 SL that is bound by p65 during U12-dependent splicing.

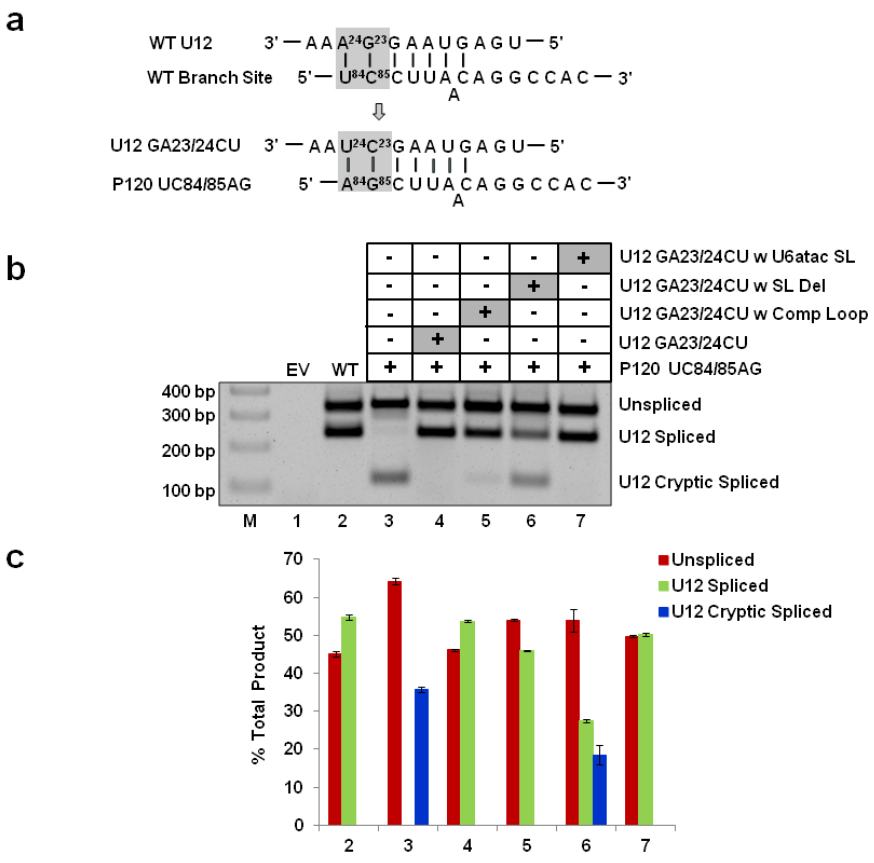


Figure 18: Effect of replacing the SLIII in U12 snRNA with the U6atac distal 3' SL on *in vivo* splicing. (a) Features of the branch site suppression assay. Wild type (WT) base pairing between human U12 snRNA and the branch site of the U12-dependent intron of P120 pre-mRNA is shown. The boxed nucleotides were mutated to their complementary nucleotides as shown. GA nucleotides at positions 23/24 in U12 snRNA were mutated to CU and the corresponding nucleotides UC at positions 84/85 in the branch site were mutated to AG. U12 GA23/24CU mutations are required to fully suppress the effect of the branch site UC84/85AG mutation. (b) Splicing phenotypes of P120 WT and the P120 UC84/85AG mutant coexpressed with the indicated U12 snRNA mutants. CHO cells were transfected with the indicated constructs and splicing phenotypes were assayed by RT–PCR. Lane M: 100 bp ladder. U12 GA23/24CU w U6atac SL denotes the U12 GA23/24CU snRNA construct containing the U6atac distal 3' SL in place of U12 SLIII. The positions of bands corresponding to unspliced RNA, RNA spliced at the normal U12-dependent splice sites (U12 spliced) and RNA spliced at the cryptic U12-dependent splice sites (U12 cryptic spliced) are indicated. The cryptic spliced product results from the activation of a U12-dependent cryptic splice site in the downstream exon. (c) Quantitative analysis of spliced/unspliced products. Numbers (x-

axis) correspond to the respective lanes of the gel shown in “b”. Error bars represent ±SE of three experiments.

3.3 U12 snRNA Stem-loop III can functionally replace distal 3' Stem-loop of U6atac snRNA

Next, we wanted to test if the U12 snRNA p65 binding SL (nts.109-125) is compatible for U6atac snRNA function *in vivo*. For this, we modified the suppressor U6atac snRNA containing a first site GG14/15CC mutation by replacing its distal 3′ SL (nts. 91-109, Figure 17b) with U12 SL (nts. 109-125, Figure 17a) as second site mutation. This U12 SL-containing U6atac mutant was transfected along with the P120 CC5/6GG and U11 GG6/7CC mutants (Figure 19). The P120 CC5/6GG mutant was defective for WT U12-dependent splicing. Instead, cryptic splice sites were activated and a smaller intron was spliced via the U2-dependent splicing pathway (Figure 19b and c, lane 3) (Hall S.L. and Padgett R.A., 1996). The transfection of the U11 suppressor alone did not suppress the splicing defect (Figure 19b and c, lane 4). However, the U6atac GG14/15CC suppressor alone suppressed the splice site defect and partially restored splicing at the WT splice sites (Figure 19b and c, lane 5), demonstrating that the assay is U6atac-dependent. Coexpression of the U11 and U6atac suppressors activated the P120 CC5/6GG U12-type splicing to nearly WT levels, as both snRNAs are required for enhanced splicing activity of the P120 mutant intron (Figure 19b and c, lane 6) (Incorvaia R. and Padgett R.A., 1998; Dietrich R.C. et al., 2001). When we cotransfected P120 CC5/6GG with U11 GG6/7CC and the U6atac GG14/15CC suppressor containing the U12 snRNA SL as a second site mutation (U6atac GG14/15CC w U12 SL), the U12-dependent splicing of the P120 5′ splice site

mutant was restored to almost WT levels (Figure 19b and c; compare lane 7 with lanes 2, 6). This demonstrates that the U12 snRNA SL can functionally replace the distal 3′ SL of U6atac during U12-dependent splicing when expressed in the context of the U6atac suppressor snRNA.

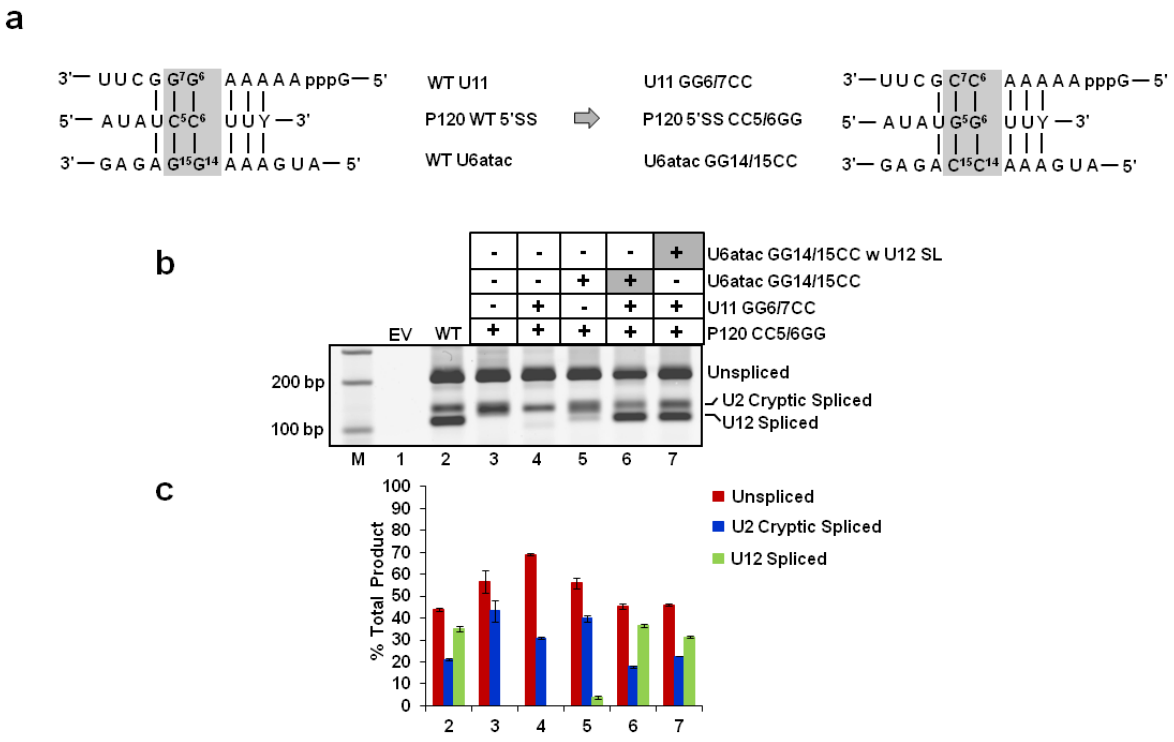


Figure 19: Effect of swapping the distal 3' SL in U6atac snRNA with the U12 SLIII on *in vivo* splicing. (a) Diagram of the 5′ splice site (SS) *in vivo* genetic suppression assay. The base pairing interactions between the WT (wild type) U11 snRNA and WT P120 pre-mRNA 5′ SS, and between the WT U6atac snRNA and pre-mRNA 5′SS are shown. The mutations introduced in U11 snRNA, U6atac snRNA and the P120 pre-mRNA's 5′SS are also indicated. The boxed nucleotides (shown on the left) were mutated to their complementary nucleotides as shown in the box on the right. Both the U11 GG6/7CC and U6atac GG14/15CC mutations are required to fully suppress the effect of the 5′SS CC5/6GG mutation. (b) Splicing phenotypes of the P120 WT and P120 CC5/6GG mutant co-expressed with the indicated U11 and U6atac snRNA mutants. Total RNA was extracted from CHO cells transfected with the indicated constructs and the *in vivo* splicing pattern was analyzed by RT-PCR. Lane M: 100 bp ladder. U6atac GG14/15CC w U12 SL denotes the U6atac GG14/15CC snRNA construct containing the SLIII of U12 snRNA in place of the U6atac distal 3' SL. The positions of bands corresponding to unspliced RNA, RNA spliced at the normal U12-dependent splice sites (U12 spliced) and RNA spliced at the cryptic U2-dependent splice sites (U2 cryptic) are indicated. The cryptic spliced product is a result of activation of U2-dependent cryptic splice sites in the U12-dependent intron of the P120 minigene. (c) Quantitative analysis of spliced/unspliced bands. Numbers (x-axis) correspond to the respective lanes of the gel shown in "b". Error bars represent ±SE of three experiments.

3.4 Both U6atac and U12 Stem-loop elements are functionally interchangeable during U12-dependent intron splicing as determined by an *in vivo* binary splice site suppression assay

To demonstrate that both U6atac and U12 SL elements are functional when swapped between the U6atac and U12 snRNAs, we developed a novel minor intron's splicing specific binary 5' splice site and branch site mutation suppression assay. In this assay, we combined previously characterized 5′ splice site CC5/6GG and branch site UC84/85AG mutations of the P120 intron (Figure 20a) (Hall S.L. and Padgett R.A., 1996; Kolossova I. and Padgett R.A., 1997 Sikand K. and Shukla G.C., 2011; Dietrich R.C. et al., 2001; Shukla G.C. and Padgett R.A., 2004). This binary splice site (P120 CC5/6GG + UC84/85AG) mutant was inactive in U12-dependent splicing and did not lead to cryptic splicing events that were observed in our previous 5′ splice site and branch site mutation suppressor assays (Figure 20b and c, lane 3). The splicing of the P120 binary splice site mutant was restored only in the presence of all three suppressors (U11 GG6/7CC + U6atac GG14/15CC + U12 GA23/24CU) (Figure 20b and c, lane 6, also compare with lanes 4 and 5; also see Figures 21c and 27b, lanes 3-6), suggesting that viable RNA-RNA base pairing interactions among all molecules are essential for the restoration of splicing. When we transfected the U6atac GG14/15CC suppressor snRNA containing the U12 SL (Figure 20b and c, lane 7) or the U12 GA23/24CU suppressor snRNA containing the U6atac SL (lane 8), or when U12 and U6atac suppressor snRNAs containing their respective second site SL mutations were

both cotransfected at the same time (lane 9), U12-dependent splicing was observed and restored to almost WT levels (Figure 20b and c, compare lanes 7, 8, 9 with lanes 2 and 6). In summary, the above data show that SLs of U12 and U6atac snRNAs that have similar loop sequences and stem lengths are functional in *in vivo* U12-dependent splicing in different structural contexts within the spliceosome.

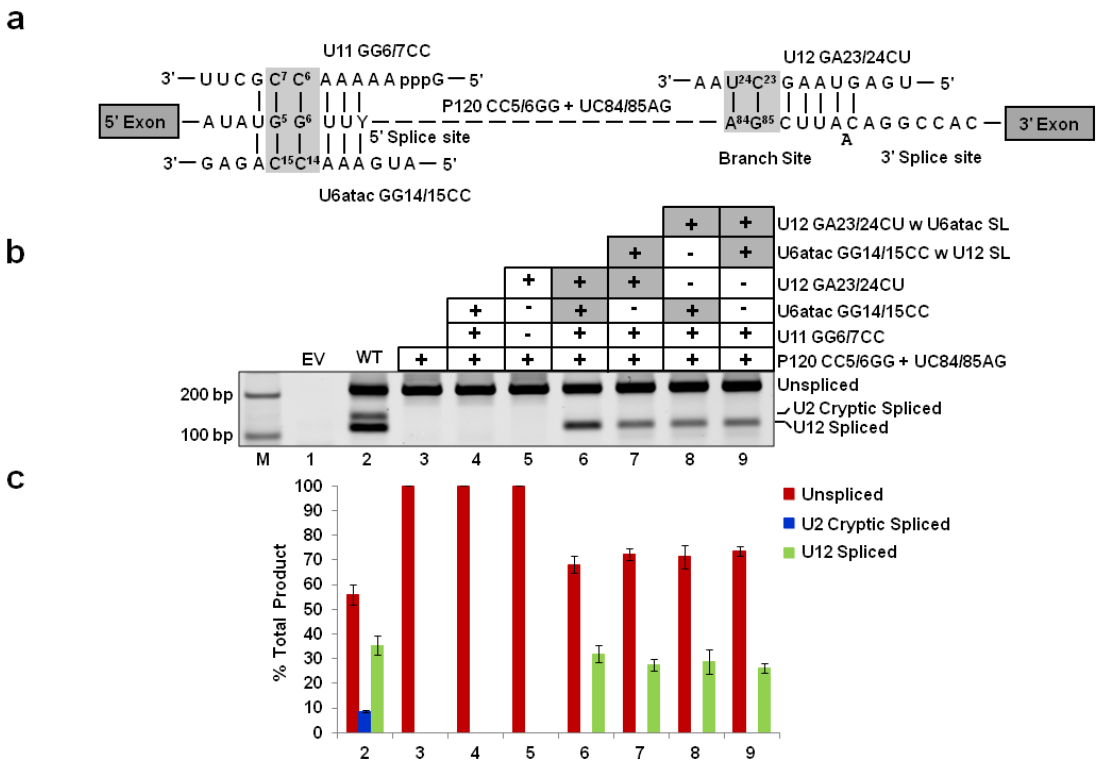


Figure 20: Combined effect of the exchange of SLs in both U6atac and U12 snRNAs on *in vivo* splicing. (a) Features of the binary splice site suppression assay. In this assay, both 5′ SS and branch site mutations shown in Fig. 2 and 3 were present in the P120 U12-dependent intron. The U12-dependent intron containing 5′ SS CC5/6GG and branch site UC84/85AG mutations is shown. The base pairing interactions between U11 GG6/7CC, U6atac GG14/15CC, U12 GA23/24CU snRNAs and the mutant 5′ SS and branch site of P120 intron are also shown. Boxed nucleotides denote the mutated nucleotides as shown in Fig. 2 and 3. (b) Splicing phenotypes of P120 WT and the P120 CC5/6GG + UC84/85AG mutant coexpressed with the indicated suppressor snRNA constructs. CHO cells were transiently transfected with the indicated constructs and total RNA was extracted. The splicing pattern of the U12-dependent P120 intron was analyzed by RT-PCR using primers designed to bind flanking exons. Lane M: 100 bp ladder. The positions

of bands corresponding to unspliced RNA, RNA spliced at the normal U12-dependent splice sites (U12 spliced) and RNA spliced at the cryptic U2-dependent splice sites (U2 cryptic) are indicated. (c) Quantitative analysis of spliced/unspliced products. Numbers on the x-axis correspond to the respective lanes of the gel shown in “b”. Error bars represent ±SE of three experiments.

3.5 Differential *in vivo* U12-dependent intron splicing activation by evolutionarily distant chimeric (hybrid) U6atac snRNAs

With respect to both U2- and U12-dependent snRNAs, it is clear that only the regions or domains that are functional “business ends” of the snRNAs are conserved among most species. However, in regions other than the “business ends” there are some less pronounced secondary structure similarities that are shared by the snRNAs among phylogenetically distant species including Arabidopsis, Phytophthora, Drosophila and Trichinella (Lopez D. et al., 2008). Figure 21a and b shows the alignment of U6atac sequences and M-fold predicted secondary structure of the 3′ ends of U6atac snRNAs from different species. Previously, we have shown the importance of the 3' end of human U6atac snRNA as a guide element for recruiting minor tri-snRNP to the U12-dependent spliceosome. We next investigated if the 3' ends of U6atac snRNAs from phylogenetically distant species could activate U12-dependent splicing in our binary splice site mutation suppressor assay. For this, we replaced the human U6atac snRNA 3′ stem-loop (nucleotides 50 to 125) with that from Arabidopsis, Phytophthora, Drosophila and Trichinella in a GG14/15CC background. All chimeric U6atac snRNAs were coexpressed with the U11 suppressor (GG6/7CC), U12 suppressor (GA23/24CU) and P120 binary splice site mutant (CC5/6GG + UC84/85AG) as a reporter plasmid. Figure 21c shows the splicing phenotypes obtained with our binary splice site mutation suppressor assay. The splicing of the

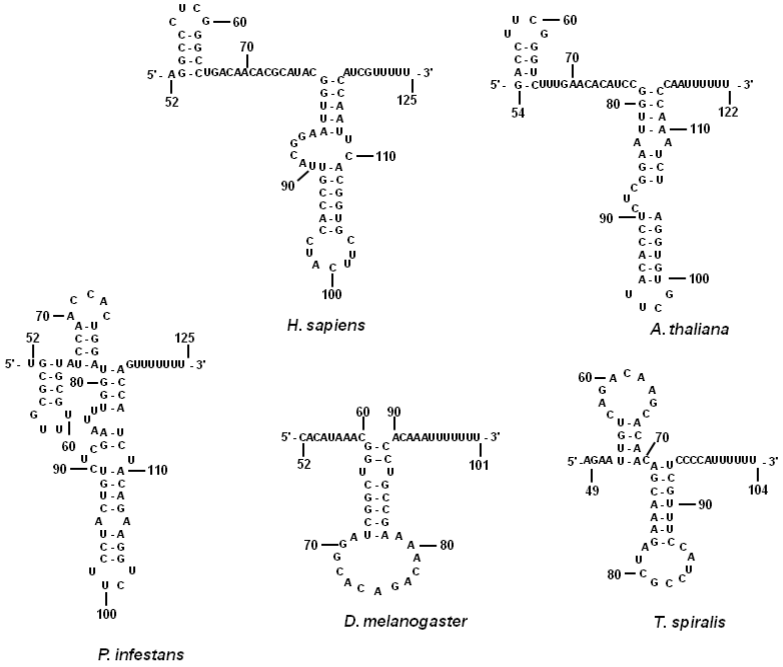
binary splice site mutant was restored to WT levels only in the presence of the three suppressor snRNAs (U11 GG6/7CC + U6atac GG14/15CC + U12 GA23/24CU; Figure 21c and d lanes 7, compare with lanes 2). Human U6atac GG14/15CC containing the 3' end of Arabidopsis U6atac snRNA only weakly supported U12-dependent splicing (Figure 21c and d, lane 8, compare with lane 7), whereas a moderate level of U12-dependent splicing was observed with the 3' end of U6atac snRNA from Phytophthora (Figure 21c and d, lane 9). On the other hand, the 3′ end of U6atac snRNA from Drosophila (lane 10) and Trichinella (lane 11) failed to restore the splicing of the binary mutant P120 minigene. This *in vivo* experiment shows that the 3' ends of the U6atac snRNAs from distant species are not interchangeable for U12-dependent splicing.

a

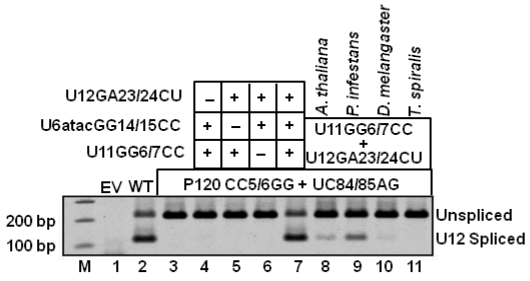
```
Majority          GTGTTGXTTGAAAGGAGAGAAGGTTAGCACTCCCCTTGAC-AAGGACGGAATXXX-CCTTGGXXTGTCCGACAAGCAXAC
                  ---------+---------+---------+---------+---------+---------+---------+---------+
                           10        20        30        40        50        60        70        80
                  ---------+---------+---------+---------+---------+---------+---------+---------+
H. sapiens        GTGTTGTATGAAAGGAGAGAAGGTTAGCACTCCCCTTGAC-AAGGATGGAAGAGGCCCTCGGGCCTGACAACACGCATAC    79
A. thaliana       GTGTTCGTAGAAAGGAGAGATGGTTGGCATCTCCTCTGAC-AGAGACGGGATTTGACCTTCGGGTCTTTGAACAC-ATCC    78
P. infestans      GTGTTCGTTGAGCCGAGAGAAGGTTAGCATCTCCCTCGAC-AAGGACGGGATTGCGCGTATGCGTATCCAACCACTGGAT    79
D. melanogaster   GTGTTGTTTGGAAGGAGAGCAAGTTAGCACTCCCCTAGACCAAGGATGGAACA----CATAAACGGTCGGCTAGGCACA-    75
T. spiralis       GTGCGGAATGAAAGGAGAGCTAGTTAGCACTCCCCTTCAC-AAGGACGGA----------GAATTGTCAGACAAGCACAA    69
                  ****** ********************** ********* **********     ***  *  * **  ** **** *
```

```
Majority          GGTTAAAGCTCTGCCACCXTTXXXGXXGXTCXCAACCX--TTTTTT
                  ---------+---------+---------+---------+------
                           90        100       110       120
                  ---------+---------+---------+---------+------
H. sapiens        GGTTAAGGCATTGCCACCTACTTCGTGGCATCTAACCATCGTTTTT    125
A. thaliana       GGTTAAGGCTCT--CCACATTCGTGTGGATCTAAACCCAATTTTTT    122
P. infestans      GGTTTAAGCTCTGTCCTCCTCCGTGAAGACATCTACCAGTTTTTTT    125
D. melanogaster   -GACAAAAGCCGTCCACAAAT-------------------TTTTTT    101
T. spiralis       CAGCAAAGATCGCCTACCTTT------GCTCCCCA-----TTTTTT    104
                  ************ ***** **   *  *      ***   ******
```

b



c



d

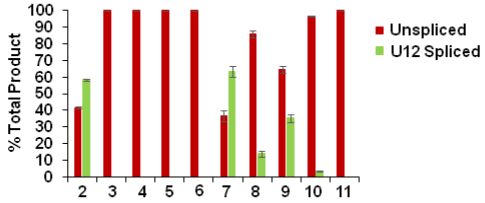


Figure 21: Activity of chimeric U6atac snRNAs in *in vivo* U12-dependent splicing. (a) Comparative sequence analysis of U6atac snRNAs from the indicated species. '*' denotes the conservation of a nucleotide in at least three species. Nucleotide sequences in green boxes represent stem I and stem II formed by RNA-RNA base pair interactions between U4atac:U6atac snRNAs. The region of U6atac snRNA denoted by a solid line below the consensus nucleotide sequences at the 5' end of U6atac snRNA is referred to as the "business end" of the molecule. (b) Predicted MFold secondary structures of the 3' end of U6atac snRNA from Homo sapiens, Arabidopsis thaliana, Phytophthora infestans, Drosophila melanogaster and Trichnella spiralis. (c) Splicing phenotypes of the binary splice site mutant (P120 CC5/6GG + UC84/85AG) coexpresssed with Human chimeric U6atac snRNA containing the 3' end of U6atac snRNA from the species shown in Panel a & b. Transfections, RT-PCR, cDNA amplification and gel electrophoresis was performed as described in the legend to Fig. 4. (d) Quantitation of unspliced and spliced products using Image J software. Numbers (x-axis) correspond to the respective lanes of the gel shown in "c".

3.6 The U11/U12 di-snRNP specific p65 protein interacts with U6atac snRNA

As shown by Benecke H. et al., 2005 and illustrated in Figure 17a, the human U12 snRNA (nucleotides 114-119; CUACUU), which form the loop region of SLIII, bind to the C-terminal RRM of the p65 protein. U6atac snRNA contains an identical sequence (nucleotides97-102; CUACUU) in the loop of its distal 3' SL (Figure 17a, b and c), suggesting that this U6atac SL has the potential to interact with the U11/U12 p65 protein. To test if U6atac interacts with the p65 protein, we performed Electrophoretic Mobility Shift Assays (EMSA) using GST-fused purified, full length p65 protein (data not shown), as well as GST/p65 fusion protein containing only the C-terminal RRM domain, and the WT U6atac distal 3′ SL (nts. 91-109) or WT U12 SLIII (nts. 109-125). C-terminal RRM of p65 was induced by using 0.5 mM IPTG. After inducing for 4-5 hours the bacterial cells were lysed and GST fused p65 protein was purified by using GST affinity based purification columns. In order to confirm, a fraction of purification protein was resolved on a denaturing acrylaminde/bis-acrylaminde gel (4-12%) (Figure 22).

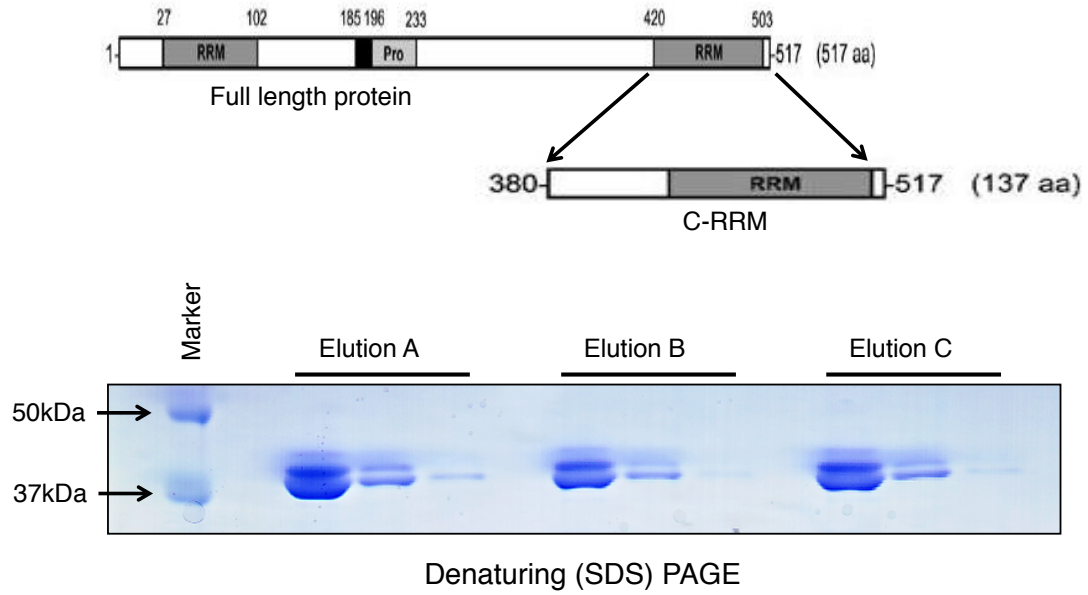


Figure 22: GST fused C-terminal RRM domain purification. Upper panel represents the schematic of both full length as well as truncated version (containing only the c-terminal RRM) of U11/U12 di-snRNP specific 65K protein. Lower panel shows the purified protein using GST spin columns. Elution A, B and C shows the induction and purification for three different clones. Different lanes under each sample shows the sequential elution.

The purified protein thus obtained was incubated in increasing concentration (20, 40 and 60 nmoles) with 200,000 counts of 5' end radiolabeled U12 (nucleotides 109-125) and U6atac (nucleotides 91-109) stem-loop RNA oligos. After resolving the RNA-protein complex on native PAGE gel with 5% glycerol, the p65 C-terminal RRM domain was found to interacting with WT U6atac distal 3' stem loop in a dose dependent pattern (Figure 23b). Consistent with the previous studies (Benecke H. et al., 2005) the U12 Stem loop III also followed the same interaction pattern when incubated with C-terminal RRM of the p65 protein (Figure 23a). Lane 1 in both the panels a & b of Figure 23 shows only the free radiolabeled U12 stem loop III and U6atac distal 3' stem loop oligo nucleotides at the bottom as this lane was not incubated with any purified protein. Lanes 2-4 (both

a & b) show the dose dependent shift of labeled RNA oligos for both U12 as well as U6atac. Lane 5 (panel a) shows the competition when radiolabeled U12 stem loop III was incubated 5-fold of un-labelled U6atac distal 3' stem loop along with the purified protein. Lane 5 (panel b) competition when radiolabeled U6atac distal 3' stem loop was incubated 5-fold of un-labelled U12 stem loop III along with the purified protein. This observation clearly suggests that this interesting RNA-protein interaction happening between U6atac and p65 may also occur in the minor spliceosome *in vivo*.

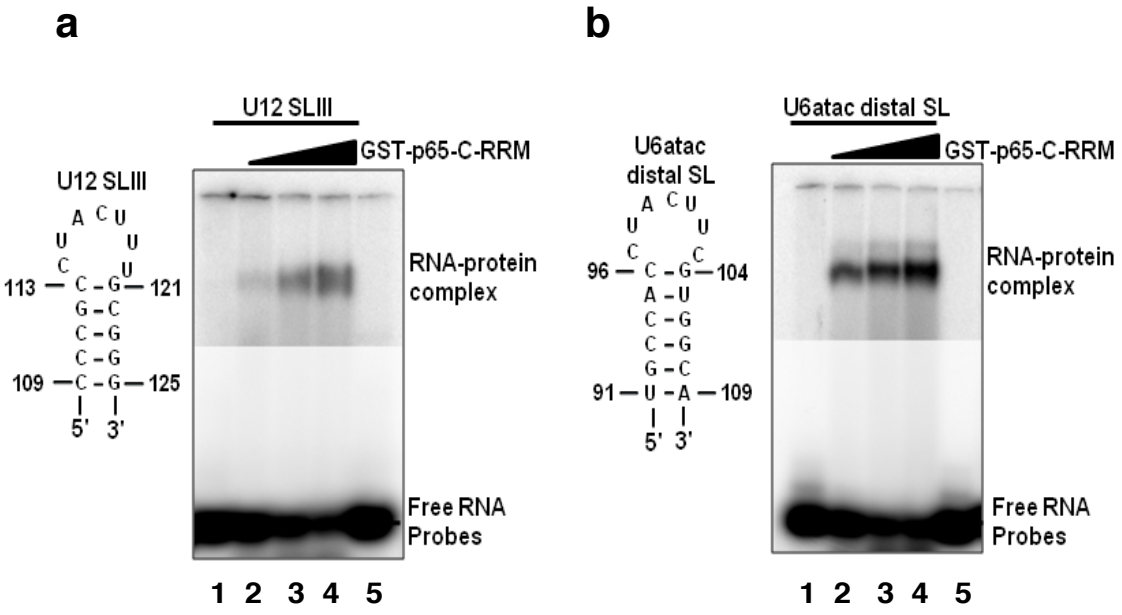


Figure 23: The C-terminal RRM of p65 binds to the distal 3' SL of U6atac snRNA. (a) EMSA of U12 SLIII with GST-p65-C-RRM. The sequence of the WT U12 SLIII RNA oligonucleotide is shown. ^{32}P-labeled oligonucleotide was incubated with increasing concentrations of GST-p65-C-RRM (0, 20, 40, 60 nmoles). RNA–protein complexes were separated on a 6% native polyacrylamide gel. (b) EMSA of the U6atac distal 3' SL with GST-p65-C-RRM. The sequence of the WT U6atac distal 3' SL RNA oligonucleotide is shown. ^{32}P-labeled oligonucleotide was incubated with increasing concentrations of GST-p65-C-RRM (0, 20, 40, 60 nmoles). RNA–protein complexes were separated on a native 6% polyacrylamide gel. On the right side are labeled the position of the RNA-protein complex band and unbound RNA at the bottom of the gel. The upper and lower parts of the gel represent different exposures.

As far as U12-dependent spliceosome specific proteins are concerned, a set of seven (65K, 59K, 48K, 35K, 31K, 25K and 20K) proteins were identified in U11/U12 di-snRNP 18S complex (Will C.L. et al., 2004; Will C.L. et al., 1999). Four proteins (i.e., 59K, 48K, 35K, and 25K), of this complex were found to be associated with U11 snRNA (Will C.L. et al., 2004; Will C. L. et al., 2001).

Among above mentioned proteins, Benecke *et al*. (2005) have demonstrated that p65 of 18S U11/U12 di-snRNP complex interacts to the stem-loop III of U12 snRNA. Structural and sequence similarities of the stem-loop III of U12 snRNA with distal 3' stem-loop of U6atac snRNA gives rise to the notion that p65 might also potentially interact with U6atac. To test if p65 also interacts with the distal 3' stem-loop of U6atac snRNA *in vivo*, we generated recombinant p65 protein with FLAG-tag at the N-terminal and 6 × His tag at the C-terminal end (the sequence and designing strategy is provided in the Appendix 2). Just in case, if p65 interactions with U6atac snRNA is dependent on other U11/U12 di-snRNP specific proteins, we designed the fusion proteins for all the seven proteins of U11/U12 di-snRNP complex [20K (170 aa), 25K (123 aa), 31K (217 aa), 35K (246 aa), 48K (342aa), 59K (485 aa) and 65K (517aa)] (Will C.L. et al., 2004). Before performing the immuno-precipitation experiment, we first tested the expression of these recombinant proteins in HeLa cells. Full length recombinant expression plasmids were transiently transfected into HeLa cells. The expression of the FLAG tagged expressing protein was determined by western blot analysis using anti-FLAG antibody as shown in Figure 24. Further, for our *in vivo* RNA-protein interactions experiments, we only overexpressed 65K recombinant protein

expressing both the N-terminal as well as the C-terminal RRM in Hela cells (Figure 24, lane 7).

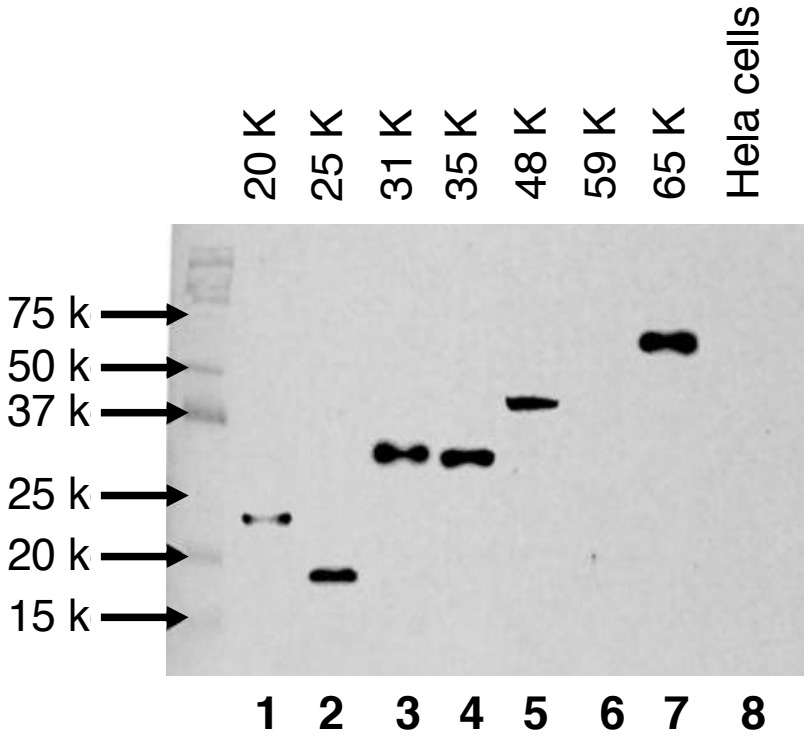


Figure 24: Ectopic expression of U11/U12 di-snRNP specific proteins in mammalian cells. Open reading frames of seven U11/U12 di-snRNP specific proteins, with FLAG tag and 6 His tag at the N- and C-terminal respectively, were cloned in mammalian expression vector pcDNA 3.1 (-) using Not I and BamH I restriction sites. These vectors were transfected in Hela cells and the after transfer of proteins, the membrane was probed with anti-FLAG antibody. Lane 1 shows the protein marker (Bio-Rad) and the following lanes (lane 2-8) show the expression of seven different proteins in Hela cells.

Protein pull-down experiments were performed using anti-FLAG antibody conjugated to agarose A/G beads. To test the specificity of immune-precipitation of FLAG- p65- 6 x His using agarose A/G beads conjugated with the anti-FLAG antibody, western blot analysis was performed. The agarose A/G beads conjugated with the anti-FLAG antibody showed a clear band of fusion protein

when probed with anti-FLAG antibody (Figure 25, lane 2). A similar size protein was detected in the total protein lysate in which p65 fusion protein was over expressed ectopically (Figure 25, lane 4). For the specificity purposes, the agarose A/G beads without any conjugation with anti-FLAG antibody served as negative control for the immune precipitation experiment (Figure 25, lane 3).

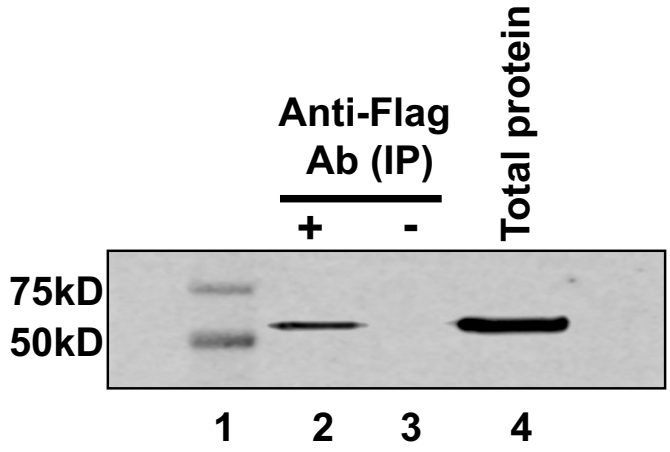


Figure 25: Immuno-precipitation of FLAG tagged p65. Anti-FLAG antibody conjugated agarose A/G beads were incubated with total protein lysate of cells over-expressing FLAG-p65-6xHis protein. The protein conjugated beads were denatured, resolved and probed with anti-FLAG antibody. Lane 1 shows the protein marker. Lane 2 and 4 shows the FLAG-p65-6xHis protein after immune-precipitation and from the total cell protein lysate. Lane 3 shows the pull down where beads were not conjugated with anti-FLAG antibody.

After washing the unbound protein with NT-2 buffer, total RNA was extracted from the bound fraction and reverse-transcribed followed by PCR amplification for U6atac, U12 and U5 snRNAs using gene specific primer sets. Interestingly, not only the positive control U12 snRNA (Figure 26a, lane 3) was pulled down together with p65, but also the U6atac snRNA was identified in the immunoprecipitation of p65 (Figure 26c, lane 3). However, we failed to identify U5

snRNA in the same pull down (Figure 26e, lane 3) indicating the specificity of p65 protein interactions with the two important snRNAs involved in catalyzing the removal of U12-dependent intron. We also performed a similar experiment with the N-terminal half of the p65 protein, but no U12 or U6atac snRNA was pulled down together with p65, suggesting that only the C-terminal RRM binds to both snRNAs (data not shown).

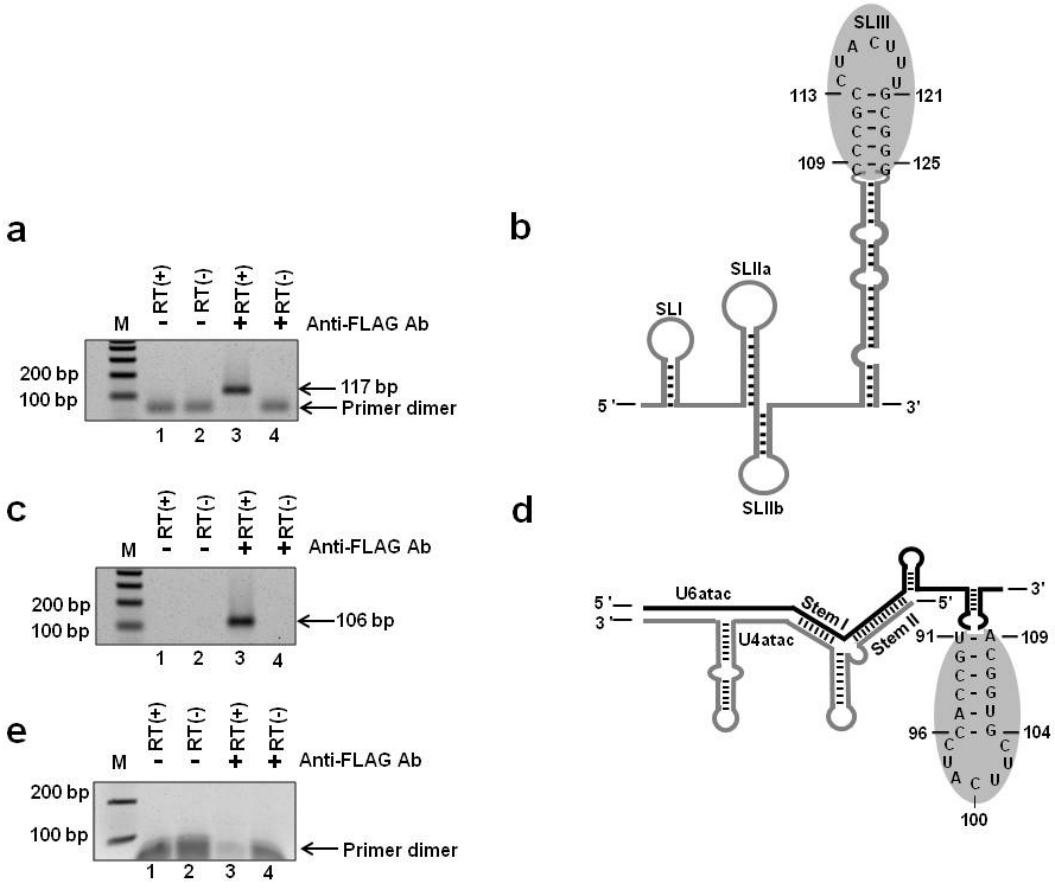


Figure 26: p65 interacts with endogenous U6atac snRNA. RT-PCR amplicons of U12 snRNA (117 bp) (Figure 26a), U6atac snRNA (106 bp) (Figure 26c) and U5 snRNA (111 bp) (Figure 26e). Panels b and d show the structure and p65 binding domains (illustrated by gray circles) of the U12 and U6atac snRNAs, respectively. FLAG-tagged full length recombinant p65 was transfected into Hela cells. 48 h post transfection, p65 protein was pulled down using anti-FLAG antibody conjugated with agarose A/G beads and snRNA was analyzed by RT-PCR (lane 3). Agarose beads without anti-FLAG antibody (lanes 1 and 2) and +/- reverse transcriptase (RT), or beads conjugated with antibody but -RT (lane 4) serve as negative controls.

3.7 Sequence and secondary structure requirement of U6atac distal 3′ Stem-loop for *in vivo* U12-dependent intron splicing

Earlier studies of interaction of the U12 snRNA with the C-terminal RRM of the p65 protein showed that the U12 loop region, as well as specific ribonucleotides in U12 SLIII, are required for p65 binding (Benecke H. et al., 2005). In light of the ability of p65 to bind the U6atac distal 3′ SL, we investigated, in detail, the requirement of the latter for *in vivo* U12-dependent splicing, including the requirement of nucleotides in the U6atac distal 3′ SL, whose equivalent nucleotides in U12 SLIII are important for p65 binding (Benecke H. et al., 2005). For this, we used the binary splice site mutation suppression assay described above. We created a series of U6atac second site mutants in the loop and the double-stranded helix region of the distal 3' SL and cotransfected them together with the U12 suppressor (GA23/24CU), U11 suppressor (GG6/7CC) and P120 double splice site mutant (CC5/6GG + UC84/85AG) reporter plasmid. The predicted secondary structure of the U6atac distal 3′ SL, positions of mutations and splicing phenotypes are shown in Figure 27a, b and c. Splicing of the P120 binary splice site mutant was restored in the presence of all three suppressors (U11 GG6/7CC + U6atac GG14/15CC + U12 GA23/24CU; Figure 27b and c, lane 7). The loop mutants U98G, A99C and C100G supported *in vivo* splicing nearly to the positive control levels (Figure 27a – loops; Figure 27b and c, compare lane 7 positive control with lanes 8, 9 and 10). Interestingly, the identity of the corresponding nucleotides (U115, A116, C117) in SLIII of human U12 snRNA is crucial for binding the C-terminal RRM of the p65 protein (p65-C-RRM) (Benecke H. et al., 2005). Stem 1

mutant, 96C-104G/96U-104A (Figure 27a) was functional in *in vivo* splicing (Figure 27b and c, lane 11). Similarly, the stem 4 mutant 96C-104G/96G-104C (Figure 27a) was also active in U12-dependent splicing (Figure 27b and c, lane 14); significantly the equivalent base pair (113C-121G) in human U12 SLIII was shown to be important for the interaction of p65-C-RRM (Benecke H. et al., 2005). In the stem 2 mutant, nts. 104-109 were mutated to their complementary sequence to disrupt the formation of the putative stem structure (Figure 27a). Similarly, in the stem 6 mutant, nts. 91-96 were mutated to disrupt the stem (Figure 27a), whereas in the stem 3 mutant the putative helix structure was restored, albeit in a complementary orientation (Figure 27a). The stem 2 and stem 6 mutants were largely inactive in U12-dependent splicing (Figure 27b and c, lane 12 and 16), and the restoration of the stem (stem 3 mutant) only partially restored U12-dependent splicing (Figure 27b and c, lane 13). Taken together, these results suggest that a U6atac 3' helix structure and its sequence are important for U12-dependent splicing.

Because the corresponding base pair (113C-121G) in the human U12 SLIII is important for p65 binding, we next investigated if the restoration of wild type base pairing to the stem 3 mutant only at positions 96 and 104 (i.e. stem 5 mutant) would be sufficient to restore splicing activity. Comparison of lanes 13 (stem 3 mutant) and 15 (stem 5 mutant) in Figure 27b and c show that splicing was similarly inhibited with both the mutants. Taken together, the splicing phenotypes obtained with the stem 1, 4 and 5 mutants demonstrate that the identity of nucleotides

forming the loop closing base pair in the U6atac distal 3′ SL is not important for *in vivo* U12-dependent splicing.

To further test the requirement of the U6atac distal 3′ SL in U12-dependent splicing, we completely deleted the SL (nucleotides 91-109; SL Del mutant, Figure 27a) or mutated the loop nucleotides 97-103 to their complementary sequence [Comp Loop mutant, Figure 27a and described in our earlier study (Dietrich R.C. et al., 2009)]. Both the SL Del and Comp Loop mutants were active in U12-dependent splicing (Figure 27b and c, lanes 17 and 18). These results are consistent with our previous study, in which the SL Del mutant and the Comp Loop mutant were active in U12-dependent splicing in two different U6atac suppressor snRNA backgrounds (i.e., U6atac GG14/15CC and a U6/U6atac hybrid background) (Dietrich R.C. et al., 2009). Moreover, both the comp loop and SL Del second site mutants in U12 GA23/24CU branch-site background appears to have reduced or no splicing, respectively as compared to WT (Figure 18b and c, lanes 5 and 6 this study which is consistent with Sikand K. and Shukla G.C., 2011 shown in the Figure 27 lanes 4 and 6). These observations indicate that the U6atac distal 3′ SL is dispensable for *in vivo* U12- dependent splicing. However, when the SL is present, the sequence and structure of its stem appears to be important for efficient and productive U12-dependent splicing.

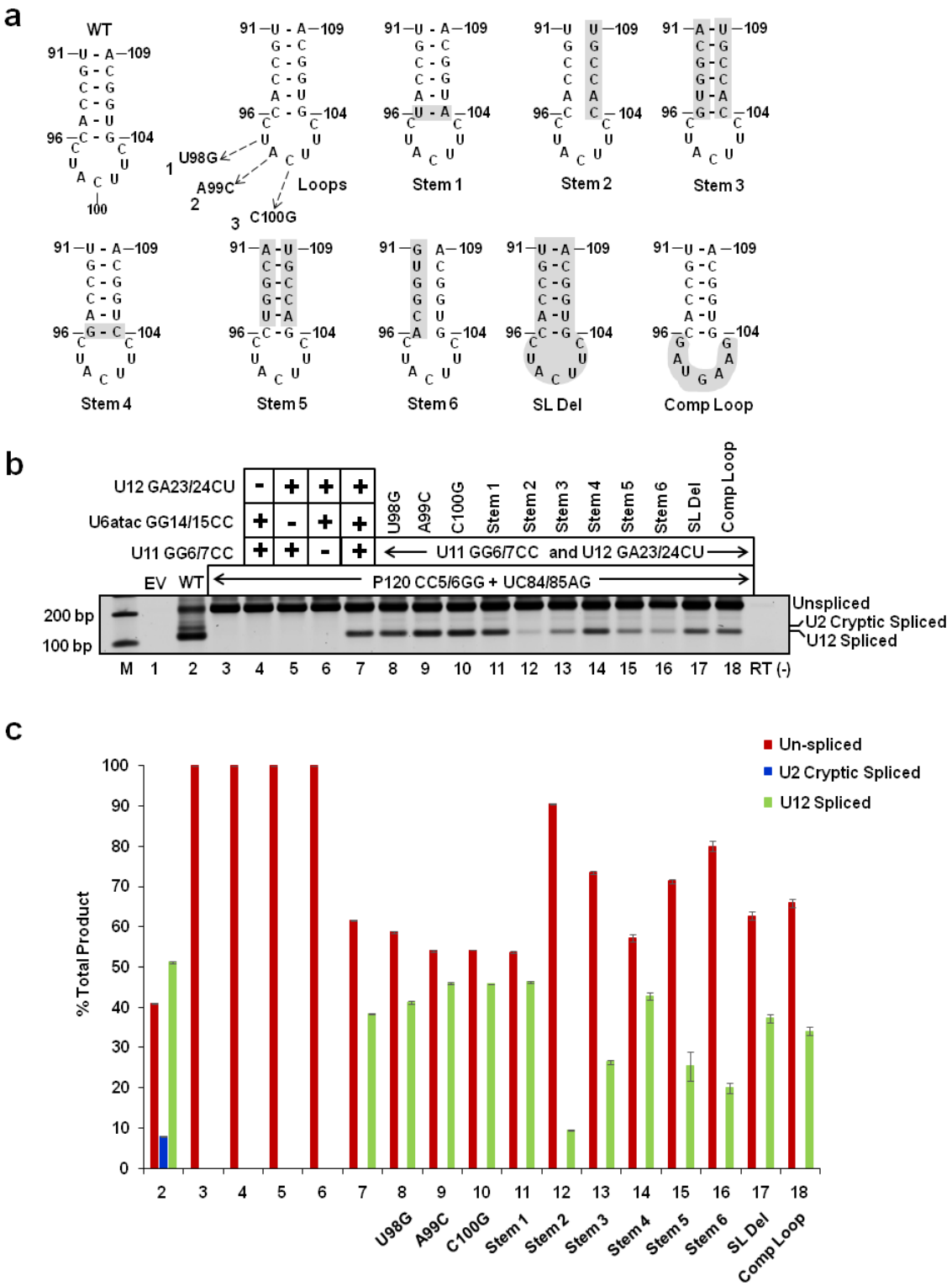

a
WT
U98G
A99C
C100G
Loops
Stem 1
Stem 2
Stem 3
Stem 4
Stem 5
Stem 6
SL Del
Comp Loop
b
U12 GA23/24CU
U6atac GG14/15CC
U11 GG6/7CC
U11 GG6/7CC and U12 GA23/24CU
EV
WT
P120 CC5/6GG + UC84/85AG
200 bp
100 bp
Unspliced
U2 Cryptic Spliced
U12 Spliced
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 RT (-)
c
Un-spliced
U2 Cryptic Spliced
U12 Spliced
% Total Product
0 10 20 30 40 50 60 70 80 90 100
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
U98G
A99C
C100G
Stem 1
Stem 2
Stem 3
Stem 4
Stem 5
Stem 6
SL Del
Comp Loop

Figure 27: Effect of mutations in the distal 3' SL of U6atac snRNA on *in vivo* U12-dependent splicing. (a) Sequence of the WT U6atac distal 3' SL and mutations made in the SL are illustrated. These mutations were made in the U6atac snRNA carrying the GG14/15CC first site mutation. (b) Splicing phenotypes of the P120 WT and the P120 CC5/6GG + UC84/85AG mutant coexpressed with the indicated suppressor snRNA constructs. CHO cells were transiently transfected with the indicated constructs and total RNA was extracted. The splicing pattern of the U12-dependent P120 intron was analyzed by RT-PCR using primers designed to bind flanking exons. Lane M: 100 bp ladder, lane E: Empty vector; RT(-), without reverse transcriptase. The positions of bands corresponding to unspliced RNA, RNA spliced at the normal U12-dependent splice sites (U12 spliced) and RNA spliced at the cryptic U2-dependent splice sites (U2 cryptic) are indicated. (c) Quantitative analysis of the U12 unspliced/spliced bands. Lanes (x-axis) correspond to the respective lanes of the gel shown in (b). Error bars represent ±SE of three experiments.

3.8 Differential binding of p65 protein with U6atac distal 3′ Stem-loop mutants

Next, we determined the structure/sequence of the U6atac distal 3′ SL that is required for interaction with p65. To this end, we analyzed the interaction of the p65 protein with the U6atac SL mutants tested for *in vivo* splicing activity. We used synthetic 5′ end, ^{32}P-labeled RNAs spanning nts. 91-109 of each U6atac SL mutant described in Figure 28a and performed EMSA with p65-C-RRM-GST fusion protein. As shown in Figure 28a and b, the binding of the U98G and A99C loop mutants to p65 was compromised (Figure 28a, lane 6 and 8, compare with lane 3). p65 binding was completely abolished with the C100G mutant (Figure 28a, lane 10), indicating a critical requirement of loop nucleotide C100 for binding. In contrast, all three equivalent loop mutations in the human U12 SLIII (U115G, A116C and C117G) abolished *in vitro* binding of p65 (Benecke H. et al., 2005). As all three loop mutants were fully active in U12-dependent splicing (Figure 28b and c, lanes 8-10), this indicates that p65 binding to the U6atac distal 3′ SL is not required for splicing. Stem 1 mutant that carried a 96C-104G to 96U-104A mutation and stem 4 mutant that carried 96C-104G to 96G-104C changes were able to bind

the p65 protein (Figure 28a and b, lane 12 and 18). These mutants were also active for *in vivo* splicing (Figure 28b and c, lanes 11 and 14). Similar mutations in the equivalent base pair in the U12 SL showed reduced binding to p65 (Benecke H. et al., 2005). The stem 2 and stem 6 mutants, which disrupt stem formation, were unable to bind p65 (Figure 28a and b, lane 14 and 22). The stem 3 and stem 5 mutants that were designed to maintain the structure of the U6atac distal 3′ SL but change the sequence of the stem, showed a moderate reduction in p65 binding as compared to the WT (Figure 28a and b, lane 16 and 20), indicating that the helix sequence may be important for *in vivo* binding of p65 to U6atac. The U12 SLIII equivalent of the stem 3 mutant did not bind p65 c-terminal RRM domain (Benecke H. et al., 2005). The inability of a WT loop closing base pair (Stem 5) to completely restore p65 binding, plus the nearly wildtype binding of RNAs with mutated loop closing base pairs U-A (Stem 1) and G-C (Stem 4), suggests that the identity of loop closing base pair is not important for p65 interaction with the human U6atac distal 3′ SL. As expected, the Comp Loop mutant, in which all loop nucleotides of the U6atac distal 3′ SL were mutated to their complementary sequences, was not bound by p65 (Figure 28a and b, lane 24). Interestingly, this mutant was active in U12-dependent splicing *in vivo* (Figure 28b and c, lane 18) (Dietrich R.C. et al., 2009). Taken together, these data indicate that the SL structure, as well as the identity of loop nucleotides are important for the interaction of p65 with the distal SL of U6atac snRNA, but that this interaction is not essential for U12-dependent splicing.

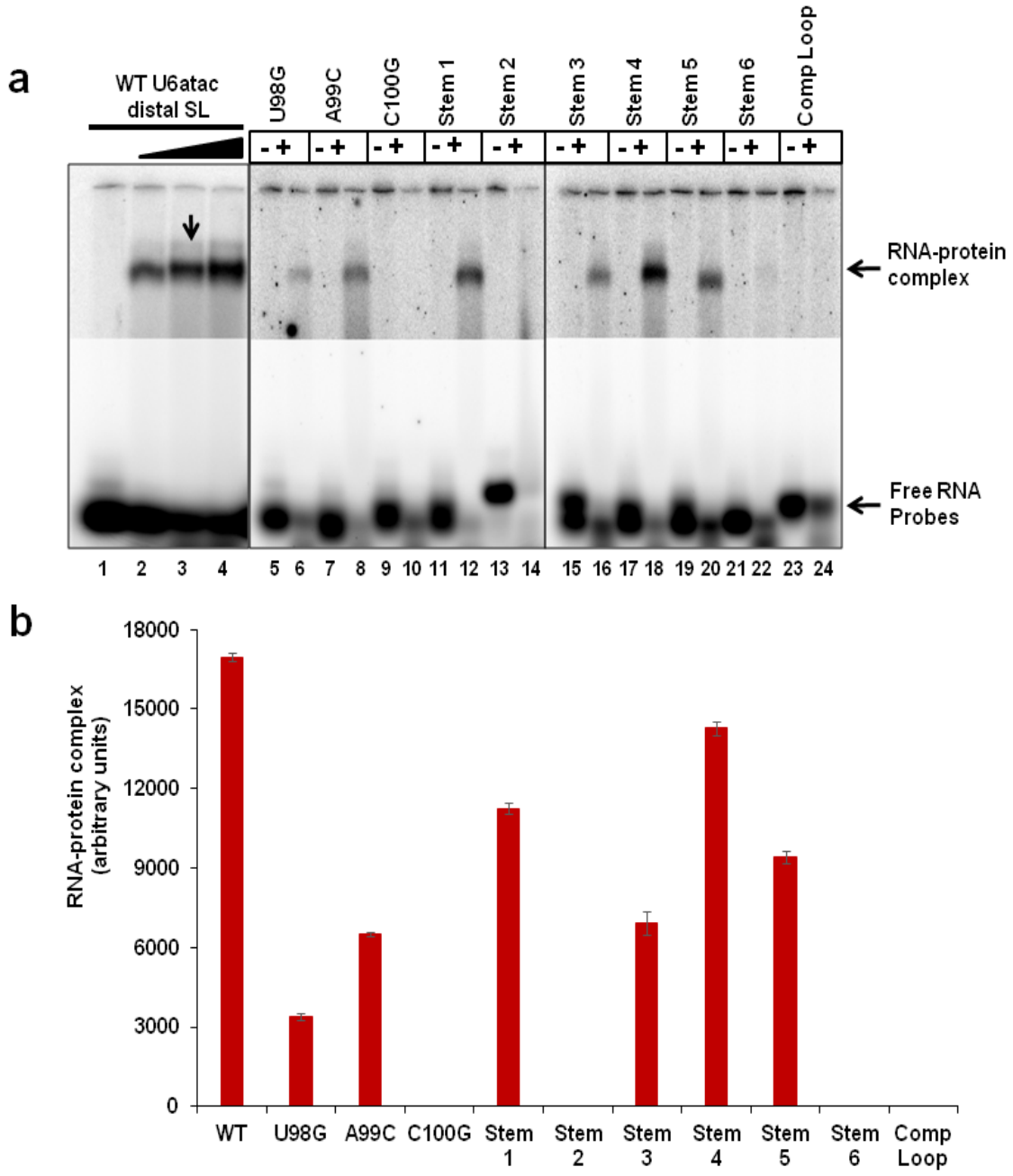


Figure 28. Effect of U6atac distal 3' SL mutations on p65-C-RRM binding. (a) EMSA of WT and mutant U6atac distal 3' SL (nts. 91-109) oligonucleotides with GST-p65-C-RRM. The location of the respective mutations in the U6atac SL RNA oligonucleotides are illustrated in Fig. 8a. WT U6atac distal 3' SL ^{32}P-labeled oligonucleotides were incubated with 0 (lane 1), 20, (lane 2) 40 (lane 3) and 60 (lane 4) nmoles of GST-p65-C-RRM. ^{32}P-labeled oligonucleotides were incubated with (+) or without (-) GST-p65-C-RRM (40 nmoles). RNA–protein complexes were resolved on a 6% native polyacrylamide gel. Arrows on the right denotes the position of the RNA-protein complex band and free RNA. The upper and lower parts of the gel represent different exposures. (b) Quantitation of the RNA-protein complex formed with 40 nmoles of GST-p65-C-RRM in the gel shown in “a”. The RNA-protein complex for the WT corresponds to lane 3 in Figure 28a.

For convenience purpose Table 2 summarizes the *in vivo* U12-dependent intron splicing results from section 3.7 (Figure 27) and *in vitro* RNA-protein interactions represented in section 3.9 (Figure 28).

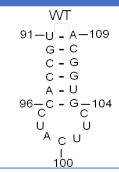
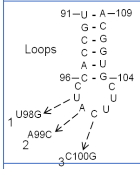
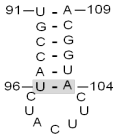
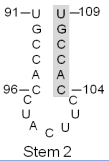
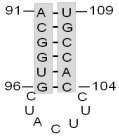
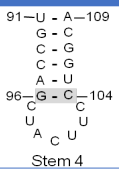
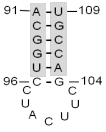
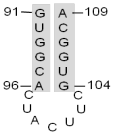
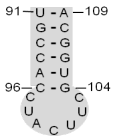
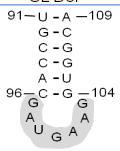
U6atac 3' SL	*In-vivo* Splicing Phenotype	*In-vitro* p65 binding
 WT	+ + +	✓
 Loops	+ + +	C100G
 Stem1	+ + +	✓
 Stem 2	- - -	✖
 Stem 3	+ - -	✓
 Stem 4	+ + +	✓
 Stem 5	+ - -	✓
 Stem 6	- - -	✖
 SL Del	+ + -	
 Comp Loop	+ + +	✖

Table 2: Summarizing the effect of mutations introduced in distal 3' stem loop of U6atac snRNA on *in vivo* U12-dependent splicing and *in vitro* RNA-protein interactions. The +++ indicate the wild type or restoration of splicing to the wild type levels. The ++ indicate the weak level of restoration as compared to the wild type. The + represents very weak restoration of splicing phenotype. The check mark indicates the *in vitro* RNA-protein interactions whereas, the cross sign shows the defect in *in vitro* interactions.

Chapter IV

DISCUSSIONS

The efficient and precise removal of non-coding intervening introns, from the precursor mRNAs, is an essential step in eukaryotic gene expression. Almost all metazoans contain two different types of introns classified on the bases of their relative occurrence in the genome. To date, approximately 800 introns have been annotated as U12-dependent introns (U12 DataBase), which comprise only less than 0.5 % of the total introns in human genome or any given genome. The recognition of the splice sites and the catalytic reactions that result in splicing take place within one of the most complex mega-dalton molecular machineries in the cell called the spliceosome (Wahl M.C. et al., 2009; Will C.L. and Luhrmann R., 2006). As these U12-dependent introns differ in their splice site sequences from the U2-dependent introns, their processing requires similar but distinct spliceosomal machinery. Each spliceosome is comprised of four distinct snRNAs, one common snRNA and a large number of associated proteins that along with snRNAs form small nuclear ribonucleoprotein particles called snRNPs. In addition, large number of additional

non-snRNP proteins are also associated with the spliceosome. Proteomic studies of *in vitro* assembled spliceosomes have suggested the involvement of 150–300 individual polypeptides approximately (Wahl M.C. et al., 2009; Jurica M.S. and Moore M.J., 2002). The U2- and U12-dependent spliceosomes differ mainly in their snRNAs and snRNPs. U12-type introns are spliced less efficiently than the U2-dependent introns, and is believed add another layer of gene regulation by limiting the expression of the host genes.

Recent advancements in the functional assays identified mutations (30G>A, 51G>A, 55G>A, and 111G>A) in the gene encoding U4atac snRNA and linked these mutations with severe developmental disorder, MOPD I. Since U4atac snRNA plays an essential role in catalyzing U12-dependent intron splicing, these mutations cause defective U12-dependent splicing. Experimental evidences have demonstrated defective endogenous U12-dependent intron splicing, whereas the U2-dependent intron splicing was found to be normal in MOPD I patient fibroblast cells, suggesting the role of U12-dependent splicing in the development and human disease. This clearly indicates the significance of U-12 dependent intron splicing in regulating multiple important cellular processes other than directly related to the expression and function of the host genes (He H. et al., 2011).

All the snRNAs involved in the catalysis of either U2- or U12-dependent introns form complex secondary structures. The role of snRNAs in splicing is defined by their inter- and intra-molecular RNA-RNA interactions (Turunen J.J. et al., 2013; Staley J.P. & Guthrie C., 1998). In addition, snRNP proteins facilitate the recognition of introns and the assembly of the catalytic scaffold of the spliceosome

(Will C.L. and Luhrmann R., 2011; Will C.L. and Luhrmann R., 2005; Yan C. et al., 2015; Hang J. et al., 2015). Mass spectrometry analyses of the major snRNPs including various spliceosomal complexes generated a comprehensive list of U2-dependent spliceosomal proteins (Hartmuth K. et al., 2002; Jurica M.S. et al., 2002; Makarov E.M. et al., 2002, Rappsilber J. et al., 2002; Zhou Z. et al., 2002). However, due to the relatively low abundance of the U11, U12, and U4atac/U6atac snRNPs, the knowledge of the protein composition of the minor spliceosome is far from complete (Montzka K.A. and Steitz J.A., 1988).

Several studies have suggested that many spliceosomal proteins are shared by both the splicing machineries. Further, immunoprecipitation studies demonstrate that most of the proteins specifically associated with the tri-snRNP complexes involved in both the U2- as well as U12-dependent spliceosome are common (Luo H.R. et al, 1999; Schneider C. et al., 2002), which was also supported by *in vitro* binding studies (Nottrott S. et al., 2002). These evidences strongly support the idea for a homologous origin (i.e., common ancestry) of both the spliceosomes (Burge C.B. et al., 1998). SR (Serene-Arginine) protein family are considered as the accessory proteins facilitating the splicing mechanism, in addition to the snRNPs. Thus the members of the SR protein family which are required for the splicing of both the type introns are also shared by both spliceosomes (Hastings M.L. and Krainer A.R., 2001).

Characterization of human 18S U11/U12 and 12S U11 snRNPs revealed the set of associated proteins. A set of seven novel U11/U12 proteins were identified which were not associated with major spliceosomes. The subset of seven

novel proteins found associates with the 12S U11 monoparticle are considered to contribute in the recognition of U12-type 5′ splice site of the intron. This demonstrates that the proteins, specifically involved in the formation of pre-spliceosome, are unique to the U12- dependent spliceosome only (Will C.L. et al., 2004).

The most extensively studied and highly abundant snRNAs of the U2-dependent spliceosome, namely, U1, U2, U4, U5 and U6, perform various tasks during splicing, including the recognition of authentic splice sites. Similarly, their counterparts in the U12-dependent spliceosome (U11, U12, U4atac, U5 and U6atac) also function in intron recognition, as well as the establishment of the catalytic core of the U12-dependent spliceosome. The assembly of both spliceosomes involves the formation of a similar dynamic RNA network. The minor class spliceosomal snRNAs present an interesting model to study evolutionarily conserved inter- and intramolecular RNA-RNA interactions important for splicing.

In the U12-dependent spliceosome, various intramolecular RNA-RNA interactions, although evolutionarily conserved, appear to be functionally dispensable for *in vivo* U12-dependent splicing. U12 snRNA is predicted to form an intricate secondary structure containing several stem–loops (SL I, SL IIa, SLIIb and SL III) and single-stranded regions. By performing the robust genetic suppression assays, recent evidence has shown that the highly conserved SLIIb structure of human U12 snRNA is redundant for U12-dependent *in vivo* splicing (Sikand K. and Shukla G.C., 2011). Shukla G.C and Padgett R.A. (2001) have also

demonstrated the interchangeability of similar sequence and structural elements between the snRNAs involved in U2- and U12-dependent spliceosome.

U6, U2 and U5 snRNAs form the active catalytic core of U2 dependent spliceosome. During the spliceosome assembly process, the U6 snRNA forms an intramolecular stem-loop (ISL). This ISL element is formed after the U4atac snRNA interactions with the U6 are destabilized and is released. This structure has been shown to be important for the *in vivo* U2 dependent intron splicing and is proposed to be present near the catalytic center of the spliceosome. U6atac snRNA of U12-dependent spliceosome is the functional analog of U6 snRNA, contains a similar stem-loop whose structure but not sequence is conserved between humans and plants. It has been shown that the chimeric U6/U6atac snRNAs in which the ISL of U6atac was replaced either by the human U6 or budding yeast U6-ISL were functional to activate the *in vivo* U12 dependent intron splicing (Shukla G.C. and Padgett R.A., 2001).

These results further suggested that wild-type U4 snRNA might be able to interact productively with the chimeric U6/U6atac snRNA where the ISL of U6atac was replaced with that of U6 from either human or budding yeast. In order to show the functionality of U4 snRNA in the minor class mutant U4 snRNA was designed to base pair with a mutant U6atac snRNA and was tested for activating *in vivo* U12-dependent intron splicing. Crosslinking assay also supported this genetic interaction happening *in vitro*. These results indicate that a U4/U6atac di-snRNP can base pair and precisely splice a U12-dependent intron (Shukla G.C. and Padgett R.A., 2003).

It has been shown that both spliceosomal and self-splicing group II introns require the function of similar small, metal binding RNA stem-loop elements located in U6 or U6atac snRNAs of the spliceosome or domain 5 (D5) of group II introns. Further, in an interesting study Shukla G.C. and Padgett R.A., (2002) reported that two different D5 elements can functionally replace the U6atac snRNA intramolecular stem-loop structure in an *in vivo* splicing assay. These data strongly argue for the interchangeability of components and structures between the two spliceosomes, and more importantly, for the dispensable nature of many of the substructures of snRNAs, which are outside of the "business end" of the molecules.

In the work described here, we show that the human U12 snRNA (nucleotides 109-125) and U6atac (Nucleotides 91-109) snRNA specific stem loops are functionally interchangeable in U12-dependent *in vivo* splicing. Our data suggest that a) the U6atac distal 3′ stem loop (nts. 91-109) plays an ambiguous role in U12-dependent splicing; b) the U6atac distal 3′ stem loop can bind *in vitro* as well as *in vivo* to the C-terminal RRM domain of p65, a protein of the U11/U12 di-snRNP complex; and c) the U6atac distal 3′ stem loop can functionally replace the U12 SLIII in the context of the U12 snRNA. In addition, we also show the practicality of the binary splice site mutation suppressor assay for the *in vivo* study of U12-dependent splicing.

We began with the simple observation that human U12 and U6atac snRNAs have similar sequence and structure of stem loops. Using three different genetic suppression assay of U12-dependent *in vivo* splicing including at the 5' splice site,

branch site and the binary splice site (where 5' splice site and the branch site were simultaneously mutated) we have demonstrated that SLIII (nts. 109-125) of U12 snRNA and distal 3′ SL (nts. 91-109) of U6atac snRNA can functionally replace each other (Figure 18-20). These experiments demonstrate that the structural and sequence similarity of the two stem loops is sufficient to allow them to be interchangeable between U12 and U6atac snRNA. These results also support the preexisting evidences showing the flexibility of snRNA elements in U12-dependent spliceosome.

As mentioned earlier, the U12 snRNA is predicted to fold into an intricate secondary structure containing multiple stem loops and single stranded sequence for Sm binding proteins. Stem loop IIb is dispensable, stem loop III (terminal stem loop) of U12 snRNA is indispensable for *in vivo* U12-dependent intron splicing (Sikand K. and Shukla G.C., 2011). The stem loop III of U12 snRNA also serves as a binding site for the p65 protein, one of the seven U11/12 di-snRNP specific complex proteins (Benecke H. et al., 2005). It has been shown that the p65 protein performs dual binding activity by interacting directly with U12 snRNA stem-loop III via its C-terminal RRM and the U11-associated p59 protein via its N-terminal half (Will C.L. and Luhrmann R., 1997; Benecke H. et al., 2005).

Clear evidence for an interaction between p65 and the U6atac snRNA was provided by *in vitro* EMSA experiments. Our data show that the distal 3′ SL of U6atac snRNA can bind to the C-terminal RRM of p65 protein both *in vitro* and *in vivo* (Figure 23 and 26). The structure of the distal 3′ SL and the identity of loop nucleotides appears to play an important role in the binding of p65 (Figure 28).

Comparison of our p65-U6atac distal 3′ SL interaction data with the previously reported p65-U12 SLIII interaction data suggests that the determinants for p65 interaction appear to be similar. The loops of human U12 SLIII and U6atac distal 3′ SL have an identical nucleotide sequence, except for the 3′ most loop nucleotide which is U in human U12 and C in U6atac (Figure 17). The identity of loop nucleotides U115, A116 and C117 is critical for U12 SLIII interaction with p65-C-RRM (Benecke H. et al., 2005). However, our data show that of the three equivalent loop nucleotides U98, A99 and C100 in the U6atac distal 3′ SL, only the identity of C100 is crucial for p65 binding (Figure 28). The loop closing base pair in both the U6atac distal 3′ SL and U12 SLIII is identical: 96C-104G in U6atac and 113C-121G in U12 (Figure 17). However, whereas the identity of the loop closing base pair is an important determinant for p65 binding to the U12 SLIII (Benecke H. et al., 2005), these nucleotides do not appear to be important for the p65-U6atac distal 3′ SL interaction (Figure 28, Stem 1, 4 and 5 mutants). Similarly, altering the stem sequence of the U6atac distal 3′ SL had only a moderate effect on p65 binding (Figure 28, stem 3 and 5 mutants). The disruption of SL structure abolished p65 binding to both U12 (Benecke H. et al., 2005) and U6atac SLs (Figure 28, stem 2 and 6 mutants), showing the requirement of a SL structure for p65 binding. The basis for these differences is not clear and we cannot currently rule out that these differences could be attributed to altered sample handling or experimental conditions used in this report and that of Benecke H. et al (2005). As the U6atac stem is longer and thus more stable due to the presence of an additional base, it may play a more important role in p65 binding. Likewise, the minor difference in

the loop sequence between U12 SLIII and the U6atac distal 3′ SL could also potentially contribute to the observed differences, although the 3′ most loop nucleotide is also C (as opposed to U) in the U12 snRNA of some organisms (Benecke H. et al., 2005). To investigate the role of p65-U6atac snRNA interaction in U12-dependent splicing, we tested the activity of U6atac distal 3′ SL mutants in a binary splice site suppression assay. The C100G and complementary sequence (Comp Loop) loop mutants of U6atac distal 3′ SL, which were unable to bind the p65 protein (Figure 28), exhibited nearly WT activity in U12-dependent *in vivo* splicing (Figure 27b and c), thus indicating that the p65-U6atac distal 3′ SL interaction is not important at least for *in vivo* U12-dependent splicing. The ability of p65 to bind the other U6atac distal 3′ SL mutants (Figure 28) correlated with the *in vivo* splicing activity of these mutants (Figure 27b). It is possible that disruption of the p65-U6atac interaction can be compensated by other structures in the U12-dependent spliceosome, hence, resulting in little or no effect on splicing activity. The structural similarity between the U6atac distal 3′ SL and U12 SLIII, and the functional interchangeability between the SLs as demonstrated by domain swapping experiments, seem to support this possibility.

Published reports have shown that the RNA structures and interacting proteins serve to facilitate the formation of the functional RNA-protein scaffold necessary for biological reactions. Such RNA scaffolds appear to provide a dynamic option for multiple stoichiometric protein assemblies (Delebecque C.J. et

al., 2012). Nevertheless, in the absence of stable RNA-protein interactions, the key chemical reactions could still be carried out, although with compromised efficiency.

Recent findings show that many RNAs undergo various structural conformations through transitory dynamic stages to assist the formation of higher order RNP complexes to accomplish complicated processing events such as RNA splicing and translation. In the yeast telomere RNA scaffold, the telomerase RNP was still functional despite the presence of significant stiffening of its RNA component (Lebo K.J. and Zappulla D.C, 2012). Massive RNA scaffolds that are constructed by a large number of RNA structures of dispensable nature, perhaps provide an important evolutionary advantage to the chemical or enzymatic reaction to be accomplished by the whole complex, in this case, the spliceosome. The self-catalytic group II introns occur in a variety of sizes and are conserved in evolutionarily divergent organisms (Candales M.A. et al., 2012). However, most of the intronic RNA is dispensable for at least *in vitro* splicing of the intron. Many reports have shown that mutations that are lethal for *in vitro* splicing appear to be functional in *in vivo* splicing (Lambowitz A.M. and Zimmerly S., 2004; Michel F. et al., 2009; Simon D.M. et al., 2009; Keating K.S. et al., 2010). The proposal that the snRNAs are evolutionary descendants of group II intron domains, is thus consistent with the dispensable nature of many substructures of snRNAs. These precedents are suggestive of structural and spatial support for many RNA secondary structures in the assembly and function of RNP machinery.

CHAPTER V

CONCLUSIONS

While a lot is known about the U2-dependent spliceosome, the structure-function analysis of U12-dependent spliceosome is still under investigation. Intensive research has been conducted to explore the significant role of different secondary structures formed by the RNA-RNA base pair interactions outside the active catalytic core of the minor spliceosome. However, a lot of additional research studies related to the function of RNA-RNA and RNA-protein interactions in minor spliceosome is still on the way. There is a big gap in the understanding the role of numerous transient yet, important interactions and also the mechanism of how conserved but still dispensable elements contribute in regulating the U12-dependent intron splicing. Here in this study, we have investigated the functional role of a substructure (distal 3' stem-loop) present in the 3' end of U6atac snRNA in mammals. The entire 3' end has two stem loops and a single stranded region and has been demonstrated to act as a guide element for the tri-snRNP complex specifically to the minor class intron.

Results obtained from the comprehensive studies that were conducted during this work are novel and demonstrate the functional significance of the distal 3' stem loop of U6atac snRNA in U12-dependent spliceosome. Our studies conclude that a sub structure in the conserved 3' end of the U6atac snRNA has both structural as well as sequence similarity with the terminal element of stem loop III in U12 snRNA. Both U6atac and U12 snRNAs play a vital role in the catalysis of U12-depepdnent intron splicing. Alignment of these two elements shows a seven nucleotide base identity in the loop region and the conservation of these nucleotides in U12 snRNA signifies its importance in U6atac snRNA also. Domain swapping experiments conducted between U12 and U6atac clearly indicates that U6atac modified with terminal stem loop III of U12 snRNA and vice versa can efficiently activate the U12-dependent intron splicing.

Further, in-vitro studies have revealed the p65 C-terminal RRM interactions with terminal U12 stem loop III, which is indispensable for *in vivo* U12-dependent splicing. All the above observation convinced to hypothesize that p65 can apparently interact with U6atac and our electrophoretic mobility shift assay supported our hypothesis. *In vitro* analysis with various mutations introduced in both the helix as well as the loop region of the distal stem loop of U6atac revealed that the determinants which affect the p65 interaction with U6atac are different from that of the U12 snRNAs. Briefly, the structure and sequence of the helix region is important for *in vitro* RNA-protein interactions. Immuno-precipitation assay with total cell lysate confirmed the *in vivo* p65 interactions with U6atac snRNA.

In addition, using robust and stringent binary splice site genetic suppression

assay, which demonstrate various RNA-RNA base pair interactions in the spliceosome, we concluded that single nucleotide mutations introduced in the loop region of the distal 3' stem loop of U6atac snRNA does not affect the *in vivo* splicing activity of U12-dependent intron. Mutations that disrupted the RNA-RNA base pair interactions in the helix region of distal 3' stem loop resulted in defective *in vivo* splicing. Restoration of the interactions by complementing the helix sequence could restore the splicing back but not up to the wild type levels. These results also support the *in vitro* p65-U6atac interactions. To our surprise, the distal 3' stem loop mutant of U6atac was able to activate the *in vivo* U12-dependent splicing. This observation makes it hard to draw direct significance of p65 interactions with U6atac snRNA for U12-dependent *in vivo* splicing. Finally, we have shown that the 3' ends of U6atac from phylogenetically distant species are functionally active for *in vivo* splicing.

There are evidences which show that subtle mutations affect the *in vivo* splicing whereas, full deletion of the RNA element still continue to support the *in vivo* U12 dependent intron splicing (Dietrich R.C. et al., 2009; also consistent with our findings). However, the cause with detailed mechanism has not yet been discovered. Recent report from Younis I. et al. (2013) has demonstrated the supporting mechanism for the slower rate of U12-dependent intron splicing as compared to U2-dependent splicing. In addition to the lower copy number (approximately 2000/ cell), shorter half-life ($t_{1/2}$<2 hr) indicates that U6atac is highly unstable as compared to the other U12-depenednt snRNAs which further suggest that level of U6atac indeed is a rate limiting factor U12-type intron splicing.

p38MAPK mediated modulation of the levels of U6atac snRNA reflected the corresponding change in the mRNA expression of the minor class intron containing host genes (Younis I. et al., 20013).

Our work has potentially demonstrated the regulatory role of distal 3' stem-loop of U6atac snRNA in U12-dependent intron splicing. The fact that the minor class intron splicing can happen in distal 3' stem-loop deleted mutant of U6atac snRNA, supports the notion that this RNA element might be involved in fine-tuning the U12-dependent intron splicing. Different mutations introduced in the distal 3' stem-loop of U6atac leading to differential splicing phenotypes could be possibly due to the variability in half-life of the snRNA. The inability of p65 protein interactions with those mutants where the stem-loop structure was completely disrupted (Stem 2 and 6, Figure 27) appears to regulate splicing by modulating the stability of U6atac snRNA (Figure 29). More experimental evidences are needed to demonstrate the role of p65 interactions in U12-dependent splicing. Based on our findings from these observations, exploring the effect of similar mutations on the stability of U6atac snRNA could be one of the potential future directions of this project.

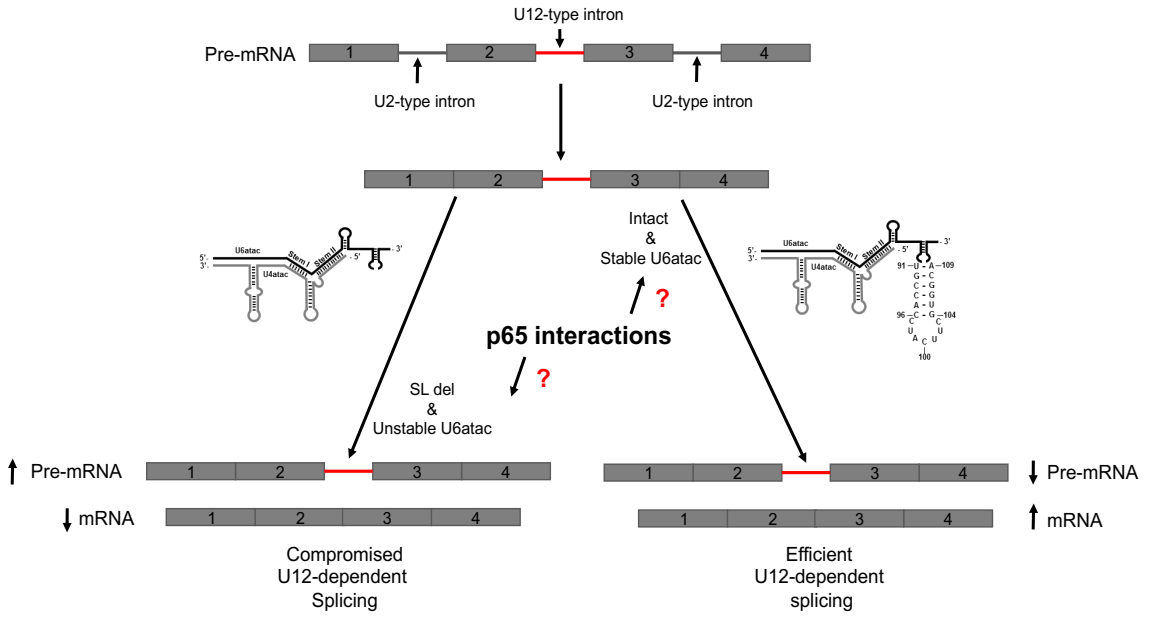


Figure 29: U6atac distal 3' stem-loop mediated fine-tuning of U12-dependent intron splicing. U6atac snRNA levels limit the splicing of U12-dependent intron. The distal 3' stem-loop of U6atac appears to play an important role in fine-tuning the splicing events. Right branch depicts the intact stem-loop element hence resulting in the normal splicing phenotype. However, the stem-loop deletion mutant results in compromised splicing (left panel) indicating that stability difference could be the possible factor. Whether or not the interactions of p65 to the distal 3' stem-loop plays any significant role in modulating the stability is still not known.

Mutations in snRNA that interfere with important RNA-protein interactions which ultimately lead to defective spliceosomal assembly and defective *in vivo* U12 dependent splicing have been reported (He H. et al., 2011; Daniele M. et al., 2015). So dissecting the mechanism of how and at what step the spliceosomal assembly is compromised when p65 interactions with U6atac are impaired, could be explored in future.

Based on the observation that 65K protein can interact with the other snRNA outside the U11/U12 di-snRNP complex, we want to perform a comprehensive, both *in vivo* as well as *in vitro*, RNA-protein interaction analyses

by different possible combinations of minor class snRNAs and the di-snRNP specific proteins. We have shown the expression of the full length fusion proteins in the mammalian cells (Figure 24). By performing immunoprecipitation assay followed by the RNA sequencing we want to test the genome wide RNA-protein interactions in the entire transcriptome. Since four out of seven proteins in the di-snRNPs do not have RNA recognition motifs but are still a part of the complex, we want to explore the possibility of protein-protein-RNA interactions by studying the *in vitro* super-shifts by incubating combination of two proteins with single minor class snRNA. For these *in vitro* studies we have already optimized the protocol of IPTG based induction and the purification process of these proteins are in progress (Figure 30).

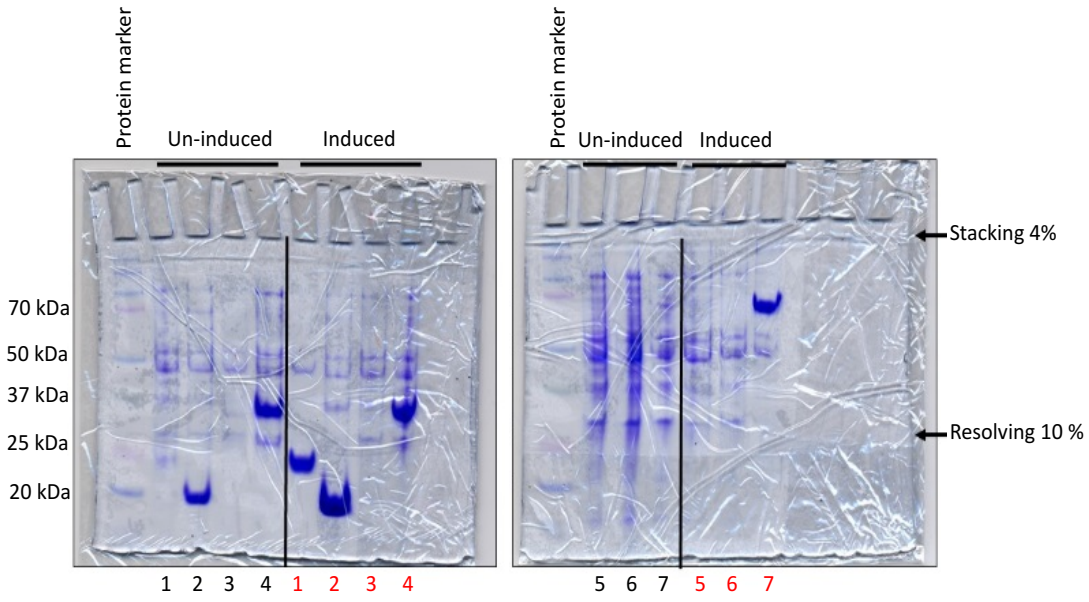


Figure 30: IPTG based induction of U11/U12 di-snRNP proteins in bacterial cells. The bacterial cells transformed with the constructs containing the gene of interest were induced using 0.5mM IPTG. The cells were lysed and the total cell lysates were resolved on the SDS-PAGE gels. In both the panels (Left and Right) the left most lane represents the protein marker. The lane number labeled from 1-7 (Black) shows the total protein profile from un-induced bacterial culture

and the lane 1-7 (Red) shows the induction profile. Lane 1 is 20K, lane 2 is 25K, lane 3 is 31K, lane 4 is 35K, lane 5 is 48K, lane 6 is 59K and lane 7 is 65K in both the color coded lane numbers.

BIBLIOGRAPHY

1. Baurén, G. & Wieslander, L. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. Cell. 1994; 76: 183–192.
2. Benecke, H., Luhrmann, R. & Will, C. L. The U11/U12 snRNP 65K protein acts as a molecular bridge, binding the U12 snRNA and U11-59K protein. The EMBO journal. 2005; 24: 3057-3069, doi:10.1038/sj.emboj.7600765.
3. Beyer, A.L. & Osheim, Y.N. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes and Development. 1988; 2: 754–765.
4. Braddock, M., Muckenthaler, M., White, M. R., Thorburn, A. M., Sommerville, J., Kingsman, A. J. & Kingsman, S. M. Intron-less RNA injected into the nucleus of Xenopus oocytes accesses a regulated translation control pathway. Nucleic Acids Research. 1994; 22: 5255–5264.
5. Brock, J. E., Dietrich, R. C. & Padgett, R. A. Mutational analysis of the U12-dependent branch site consensus sequence. RNA. 2008; 14: 2430-2439, doi:10.1261/rna.1189008.
6. Buchman, A.R. & Berg, P. Comparison of intron-dependent and intron-independent gene expression. Molecular and Cellular Biology. 1988; 8: 4395–4405.
7. Burge, C. B., Padgett, R. A. & Sharp, P.A. Evolutionary fates and origins of U12-type introns. Molecular Cell.1998; 2: 773–785.
8. Burge, C.B., Tuschl, T. & Sharp, P.A. Splicing of precursors to mRNAs by the spliceosomes. In Gesteland,R.F., Cech,T. and Atkins,J.F. (eds), The

RNA World II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999; 525–560.

9. Burgess, S. M. & Guthrie, C. Beat the clock: paradigms for NTPases in the maintenance of biological fidelity. Trends in Biochemical Sciences. 1993; 18: 381-3844.
10. Callis, J. et al. Introns increase gene expression in cultured maize cells. Genes and Development. 1987; 1: 1183–1200.
11. Candales, M. A. et al. Database for bacterial group II introns. Nucleic acids research. 2012; 40: D187-190, doi:10.1093/nar/gkr1043.
12. de la Mata, M., Lafaille, C. & Kornblihtt, A.R. First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. RNA. 2010; 16: 904–912.
13. Davila Lopez D., M., Rosenblad, M. A. & Samuelsson, T. Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. Nucleic acids research. 2008; 36: 3001-3010, doi:10.1093/nar/gkn142.
14. Delebecque, C. J., Silver, P. A. & Lindner, A. B. Designing and using RNA scaffolds to assemble proteins in vivo. Nature protocols. 2012; 7: 1797-1807, doi:10.1038/nprot.2012.102.
15. Dietrich, R. C., Incorvaia, R. & Padgett, R. A. Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. Molecular Cell. 1997; 1: 151–160.

16. Dietrich, R. C., Padgett, R. A. & Shukla, G. C. The conserved 3' end domain of U6atac snRNA can direct U6 snRNA to the minor spliceosome. RNA. 2009; 15: 1198-1207, doi:10.1261/rna.1505709.
17. Dietrich, R. C., Peris, M. J., Seyboldt, A. S. & Padgett, R. A. Role of the 3' splice site in U12-dependent intron splicing. Molecular and cellular biology. 2001; 21: 1942-1952, doi:10.1128/MCB.21.6.1942-1952.2001.
18. Dietrich, R. C., Shukla, G. C., Fuller, J. D. & Padgett, R. A. Alternative splicing of U12-dependent introns in vivo responds to purine-rich enhancers. RNA. 2001; 7: 1378-1388.
19. Doma, M. K. & Parker R. RNA Quality Control in Eukaryotes. Cell. 2007; 131: 660-668.
20. Duncker, B.P., et al. Introns boost transgene expression in Drosophila melanogaster. Molecular and General Genetics. 1997; 254: 291–296.
21. Frilander, M. J. & Meng, X. Proximity of the U12 snRNA with both the 5' splice site and the branch point during the early stages of spliceosome assembly. Molecular and Cellular Biology. 2005; 12: 4813-25.
22. Frilander, M. J. & Steitz, J. A. Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. Molecular Cell. 2001; 7: 217–226.
23. Frilander, M. J. & Steitz, J. A. Initial recognition of U12-dependent introns requires both U11/5′ splice-site and U12/branchpoint interactions. Genes & Development. 1999; 13: 851–863.

24. Gao, M., Fritz, D. T., Ford, L. P. & Wilusz, J. Interaction between a Poly(A)-Specific Ribonuclease and the 5' cap influences mRNA deadenylation rates in vitro. Molecular Cell. 2000; 5 (3): 479–488.
25. Gordon P.M., Sontheimer E.J., Piccirilli J.A. Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: Extending the parallels between the spliceosome and group II introns. RNA. 2000; 6: 199–205.
26. Griffiths-Jones S. Annotating noncoding RNA genes. Annual Review of Genomics and Human Genetics. 2007; 8: 279–298.
27. Guhaniyogi, J. & Brewer, G. Regulation of mRNA stability in mammalian cells. Gene. 2001; 265 (1–2): 11–23.
28. Guthrie, C. & Patterson, B. Spliceosomal snRNAs. Annual Review of Genetics. 1988; 22: 387–419.
29. Hall, S. L. & Padgett, R. A. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. Journal of molecular biology. 1994; 239, 357-365, doi:10.1006/jmbi.1994.1377.
30. Hall, S. L. & Padgett, R. A. Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. Science. 1996; 271: 1716-1718.
31. Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. Science. 2015; 349: 1191-1198, doi:10.1126/science.aac8159.
32. Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., & Lührmann, R. Protein composition of human prespliceosomes isolated by

a tobramycin affinity-selection method. Proceedings of National Academy of Sciences of United States of America. 2002; 99: 16719–16724.

33. Hastings, M. L. & Krainer, A. R. Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. RNA. 2001; 7: 471–482.
34. He, H. et al. Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. Science. 2011; 332: 238–240.
35. Incorvaia, R. & Padgett, R. A. Base pairing with U6atac snRNA is required for 5' splice site activation of U12-dependent introns in vivo. RNA. 1998; 4: 709-718.
36. Jackson, I. J. A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Research. 1991; 19: 3795–3798.
37. Jung, H.J. & Kang, H. The Arabidopsis U11/U12-65K is an indispensible component of minor spliceosome and plays a crucial role in U12 intron splicing and plant development. The Plant Journal. 2014; 5: 799-810.
38. Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. Molecular Cell. 2003; 12: 5–14.
39. Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N., & Moore, M. J. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA. 2002; 8: 426–439.
40. Keating, K. S., Toor, N., Perlman, P. S. & Pyle, A. M. A structural analysis of the group II intron active site and implications for the spliceosome. RNA. 2010; 16: 1-9, doi:10.1261/rna.1791310.

41. Kolossova, I. & Padgett, R. A. U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. RNA. 1997; 3: 227-233.
42. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. Annual review of genetics. 2004; 38: 1-35, doi:10.1146/annurev.genet.38.072902.091600.
43. Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., D evon K.,Dewar K., Doyle M., FitzHugh W., et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409: 860–921.
44. Lebo, K. J. & Zappulla, D. C. Stiffened yeast telomerase RNA supports RNP function in vitro and in vivo. RNA. 2012; 18: 1666-1678, doi:10.1261/rna.033555.112.
45. Levine, A. & Durbin, R. A computational scan for U12-dependent introns in the human genome sequence. Nucleic Acids Research. 2001; 29: 4006–4013.
46. Lewandowska, D., Simpson, C.G., Clark, G.P., Jennings, N.S., Barciszewska-Pacak, M., Lin, C.F., Makalowski, W., Brown, J.W., Jarmolowski, A. Determinants of plant U12-dependent intron splicing efficiency. Plant Cell. 2004; 16: 1340–1352.
47. Luo, H. R., Moreau, G. A., Levin, N., & Moore, M. J. The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. RNA. 1999; 5: 893–908.
48. Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., & Lührmann, R. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science. 2002; 298: 2205–2208.

49. Matsumoto, K., Wassarman, K. M. & Wolffe, A. P. Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. The EMBO Journal. 1998; 17: 2107–2121.
50. McNally, L.M., Yee, L. & McNally, M.T. Heterogeneous nuclear ribonucleoprotein H is required for optimal U11 small nuclear ribonucleoprotein binding to a retroviral RNA-processing control element: implications for U12-dependent RNA splicing. The Journal of Biological Chemistry. 2006; 5: 2478-88.
51. Meijer, H. A., Bushell, M., Hill, K., Gant, T. W., Willis, A. E., Jones, P., De Moor, C. H. A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. Nucleic Acids Research. 2007 35 (19): e132–e132.
52. Mewes, H. W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S.G. Overview of the yeast genome. Nature. 1997; 387: 7–65.
53. Michel, F., Costa, M. & Westhof, E. The ribozyme core of group II introns: a structure in want of partners. Trends in biochemical sciences. 2009; 34: 189-199, doi:10.1016/j.tibs.2008.12.007.
54. Modrek B., Lee C. A genomic view of alternative splicing. Nature Genetics. 2002; 30:13–19.
55. Montzka, K. A. & Steitz, J. A. Additional low-abundance human small nuclear ribonucleoproteins: U11, U12 etc. Proceedings of National Academy of Sciences of United States of America. 1988; 85: 8885–8889.

56. Nag, A. & Steitz, J. A. Tri-snRNP-associated proteins interact with subunits of the TRAMP and nuclear exosome complexes, linking RNA decay and pre-mRNA splicing. RNA Biology. 2012; 9: 334-342.
57. Niemela E. H. & Frilander M. J. Regulation of gene expression through inefficient splicing of U12-type introns. RNA Biology. 2014; 11: 1325-9.
58. Nilsen, T. W. RNA-RNA interactions in the spliceosome: unraveling the ties that bind. Cell. 1994; 78: 1-4.
59. Nilsen, T.W. RNA-RNA interactions in nuclear pre-mRNA splicing. In: RNA Structure and Function, R.W. Simons and M. Grunberg-Manago, eds. (Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press). 1998; 279–308.
60. Nottrott, S., Urlaub, H. & Lührmann, R. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. The EMBO Journal. 2002; 21: 5527–5538.
61. Padgett, R. A. & Shukla, G. C. A revised model for U4atac/U6atac snRNA base pairing. RNA. 2002; 8: 125-128.
62. Pandya-Jones, A. & Black, D.L. Co-transcriptional splicing of constitutive and alternative exons. RNA. 2009; 15: 1896–1908.
63. Papaemmanuil, E. et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. The New England Journal of Medicine. 2011; 365: 1384–1395.

64. Patel, A. A., McCarthy, M. & Steitz, J.A. The splicing of U12-type introns can be a rate-limiting step in gene expression. The EMBO Journal. 2002; 21: 3804–3815.
65. Patel, A.A & Steitz, J.A. Splicing Double: Insights from the second spliceosome. Nature Reviews Molecular Cell Biology. 2003; 12: 960-70.
66. Pessa, H. K., Ruokolainen, A. & Frilander, M. J. The abundance of the spliceosomal snRNPs is not limiting the splicing of U12-type introns. RNA. 2006; 12: 1883–1892.
67. Rappsilber, J., Ryder, U., Lamond, A.I., & Mann, M. Large-scale proteomic analysis of the human spliceosome. Genome Research. 2002; 8: 1231-1245.
68. Reed, R. Initial splice-site recognition and pairing during pre-mRNA splicing. Current. Opinion in Genetics and Development. 1996; 6: 215–220.
69. Reed, R. and Palandjian, L. Spliceosome assembly. In: Eukaryotic mRNA processing, A.R. Krainer, ed. (Oxford, UK: IRL Press). 1997; 103–129.
70. Russell, A. G., Charette, J. M., Spencer, D. F. & Gray, M. W. An early evolutionary origin for the minor spliceosome. Nature. 2006; 443: 863–866.
71. Santoro, B., De Gregorio, E., Caffarelli, E. & Bozzoni, I. RNA-protein interactions in the nuclei of Xenopusoocytes: complex formation and processing activity on the regulatory intron of ribosomal protein gene L1. Molecular and Cellular Biology. 1994; 14: 6975–6982.

72. Schneider, C., Will, C. L., Makarova, O. V., Makarov, E. M. & Lührmann, R. Human U4/U6.U5 and U4atac/ U6atac.U5 tri-snRNPs exhibit similar protein compositions. Molecular and Cellular Biology. 2002; 22: 3219–3229.
73. Sharp, P. A. & Burge, C. B. Classification of introns: U2 type or U12 type. Cell. 1997; 91: 875–879.
74. Sashital, D.G., Cornilescu, G., McManus, C.J., Brow, D.A., and Butcher, S.E. U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. Nature Structural and Molecular Biology. 2004; 11: 1237–1242.
75. Sebastian, M. et al. Minor class splicing shapes the zebrafish transcriptome during development. Proceedings of National Academy of Science. 2014; 8: 3062-67.
76. Shukla, G. C. & Padgett, R. A. Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. RNA. 1999; 5: 525-538.
77. Shukla, G. C. & Padgett, R. A. U4 small nuclear RNA can function in both the major and minor spliceosomes. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101: 93-98, doi:10.1073/pnas.0304919101.
78. Shukla, G. C., Cole, A. J., Dietrich, R. C. & Padgett, R. A. Domains of human U4atac snRNA required for U12-dependent splicing in vivo. Nucleic acids research. 2002; 30: 4650-4657.

79. Shukla, G.C. and Padgett, R.A. The intramolecular stem- loop structure of U6 snRNA can functionally replace the U6atac snRNA stem-loop. RNA. 2001; 7: 94–105.
80. Sikand, K. & Shukla, G. C. Functionally important structural elements of U12 snRNA. Nucleic acids research. 2011. 39, 8531-8543, doi:10.1093/nar/gkr530.
81. Simon, D. M., Kelchner, S. A. & Zimmerly, S. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. Molecular biology and evolution. 2009; 26: 2795-2808, doi:10.1093/molbev/msp193.
82. Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. Nature Structural & Molecular Biology. 2009; 16: 1128–1133.
83. Sonenberg, N. & Gingras, A. C. The mRNA 5′ cap-binding protein eIF4E and control of cell growth. Current Opinion in Cell Biology. 1998; 10 (2): 268–275.
84. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell. 1998; 92: 315-326.
85. Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., et al. A global
86. view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science. 2008; 321: 956–960.

87. Tarn, W. Y. & Steitz, J. A. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell. 1996; 84: 801-811.

88. Tarn, W. Y. & Steitz, J. A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. Science. 1996; 273: 1824-1832.

89. Tarn, W. Y. & Steitz, J. A. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. Trends in biochemical sciences. 1997; 22: 132-137.

90. Turunen, J. J., Niemela, E. H., Verma, B. & Frilander, M. J. The significant other: splicing by the minor spliceosome. Wiley interdisciplinary reviews. RNA. 2013; 4: 61-76, doi:10.1002/wrna.1141.

91. Valadkhan, S., Mohammadi, A., Jaladat, Y. & Geisler, S. Protein-free small nuclear RNAs catalyze a two-step splicing reaction. Proceedings of National Academy of Sciences of United States of America. 2009; 106: 11901–11906.

92. Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B. & Mattaj, I. W. A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. The Journal of Cell Biology. 1996; 133 (1): 5–14.

93. Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009; 136: 701–718.

94. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008; 456:470–476.
95. Wassarman, K. M. & Steitz, J. A. The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. Molecular and Cellular Biology. 1992; 12: 1276–1285.
96. Will, C. L. & Luhrmann R. Spliceosome structure and function. In: Gestland RF, Cech TR, Atkins JF, editors. The RNA World, Third
97. Edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006. pp. 369–400.
98. Will, C. L. & Luhrmann, R. Protein functions in pre-mRNA splicing. Current opinion in cell biology. 1997; 9: 320-328.
99. Will, C. L. & Luhrmann, R. Spliceosomal UsnRNP biogenesis, structure and function. Current opinion in cell biology. 2001; 13: 290-301.
100. Will, C. L. & Luhrmann, R. Spliceosome structure and function. Cold Spring Harbor perspectives in biology. 2011; 3, doi:10.1101/cshperspect.a003707.
101. Will, C. L. & Luhrmann, R. Splicing of a rare class of introns by the U12-dependent spliceosome. Biological chemistry. 2005; 386: 713-724, doi:10.1515/BC.2005.084.
102. Will, C. L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T. & Lührmann, R. The human 18S U11/U12 snRNP containsa

set of novel proteins not found in the U2-dependent spliceosome. RNA. 2004; 10: 929–941.

103. Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neu-bauer, G., Wilm, M., Lu¨hrmann, R., and Query, C.C. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. EMBO Journal. 2001; 20: 4536–4546.
104. Will, C. L., Schneider, C., Reed, R. & Lührmann, R. Identification of both shared and distinct proteins in the major and minor spliceosomes. Science. 1999; 284: 2003–2005.
105. Wu, Q. & Krainer, A. R. Splicing of a divergent subclass of AT-AC introns requires the major class spliceosomal snRNAs. RNA. 1997; 3: 586–601.
106. Wu, Q. & Krainer, A.R. AT-AC Pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. Molecular and Cellular Biology. 1999; 19 (5) 3225-36.
107. Wu, Q.& Krainer, A.R. Purine-rich enhancers function in the AT-AC pre-mRNA splicing pathway and do so independently of intact U1 snRNP. RNA. 1998; 4 (12): 1664-73.
108. Yan, C. et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. Science. 2015; 349: 1182-1191, doi:10.1126/science.aac7629.
109. Yu, Y. T. & Steitz, J. A. Site-specific crosslinking of mammalian U11 and U6atac to the 5′ splice site of an AT-AC intron. Proceedings of National Academy of Sciences of United States of America. 1997; 94: 6030–6035.

110. Yu Y.-T., Scharl E.C., Smith C.M., & Steitz, J.A. The Growing World of Small Nuclear Ribonucleoproteins. In: The RNA World, 2nd edition, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press).1999; 487–524.

111. Younis, I., Dittmar, K., Wang, W., Foley, S. W., Berg, M. G., Hu, K. Y., Wei, Z., Wan, L., & Dreyfus, G. Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. Published online 2013 Jul 30. doi: 10.7554/eLife.00780

112. Zhang, G., Taneja, K.L., Singer, R.H. & Green, M.R. Localization of pre-mRNA splicing in mammalian nuclei. Nature. 1994; 372: 809–812.

113. Zhou, Z., Licklider, L. J., Gygi, S. P., & Reed, R. Comprehensive proteomic analysis of the human spliceosome. Nature. 2002; 419: 182–185.

.

APPENDICES

Appendix 1. P120 Minigene sequence

The fragment of 987 base-pair containing four exons (Exon 5-8) and three introns (Intron E, F and G) was used a minigene. The wild type and the mutant construct constructs which were designed as previously described in Kolossova I. and Padgett R. A. (1997). The intron F is a U12-dependent intron (highlighted in red), with AT-AC as terminal dinucleotides, flanked by U2-dependent introns. Following is the complete sequence of the minigene which was cloned in pCB6 vector where the expression of this gene was driven by CMV promoter.

GGTACCTTGCTGCCCATTGAAAGAGCTGCTCGGAAGCAGAAGGCCCGGGAAGCTGCTGC
TGGGTGAGTTTTGGAAGCTATTGGGATGGAGCATAGGAACCCACTGGGGTGGGGAAGCT
GTGAGGAAGGAGGAGGTATCTGTTTTGGTCTGTGAATCACTCTGTTCCTGGACCTGTTT
CTAGGATCCAGTGGAGTGAAGAGGAGACCGAGGACGAGGAGGAAGAGAAAGAAGTGACC
CCTGAGTCAGGCCCCCCAAAGGTGGAAGAGGCAGATGGGGGCCTGCAGATCAATGTGGA
TGAGGAACCATTTGTGCTGCCCCCTGCTGGGGAGATGGAGCAGGatatccttgcagggc
agagtgaagagttaggaggaaggtggttgggagagggatttccaggccttaggacatca
tgacacagttccttaacaggcccacATGCCCAGGCTCCAGACCTGCAACGAGTTCACAA
GCGGATCCAGGATATTGTGGGAATTCTGCGTGATTTTGGGGCTCAGCGGGAGGAAGGGC
GGTCTCGTTCTGAATACCTGAACCGGCTCAAGAAGGATCTGGCCATTTACTACTCCTAT
GGAGACTTCCTGCTTGGCAAGCTCATGGACCTCTTCCCTCTGTCTGAGGTACTGGATTG
CCAGAATGCCTCTTTTGCTTTTCTTTTCTCGCCTCCTTACGCTGTGTAAGGAAGATGTT
CGGCCCCCTTGCTCCCCTCTTCTGACTGAGCTCCTTAGCCAGCCTCACTTATTCCCTGC
TGTCTTGCCCCGCAGCTGGTGGAGTTCTTAGAAGCTAATGAGGTGCCTCGGCCCGTCAC
CCTCCGGACCAATACCTTGAAAACCCGACGCCGAGACCTAAGCTTGCATGCCTGCAGGT
CGACTCTAGAGGATCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTCCTCTCCTGGCCT
TGGAAGTTGCCACTCCAGTGCCCACCAGCCTTGTCCTAATAAA

Appendix 2. Sequences and designing strategy of U11/U12 di-snRNP complex specific protein constructs for mammalian expression

The following are the cDNA sequences of open reading frames of the U11/U12 di-snRNP complex specific proteins explained in Material and Methods section 2.11. The genes were designed and codon optimized according to their expression in the mammalian cells. For all the sequences, Not I and BamH I restriction sites at the 5' and 3' ends respectively, are highlighted in yellow. Sequences in the green and red are the start and stop codons. FLAG tag sequence at the N-terminal is represented by the cyan color and the 6 × His tag is in magenta color at the C-terminal end of the sequence.

(Met- Flag-p65-6His: 1610 bp)

GCGGCCGCATGGACTATAAGGATGACGATGACAAGGCAGCTCCCGAGCAGCCGCTTGCG
ATATCAAGGGGATGCACGAGCTCCTCCTCGCTTTCCCCGCCTCGGGGCGACCGAACCCT
TCTGGTCAGGCACCTGCCGGCTGAGCTTACTGCTGAGGAGAAAGAGGACTTGCTGAAGT
ACTTCGGGGCTCAGTCTGTGCGGGTCCTGTCAGATAAGGGGCGACTGAAACATACAGCT
TTTGCCACATTCCCTAATGAAAAAGCAGCTATAAAGGCATTGACAAGACTCCATCAACT
GAAACTTTTAGGTCATACTTTAGTCGTTGAATTTGCAAAAGAGCAAGATCGAGTTCACT
CCCCATGTCCCACTTCAGGTTCTGAAAAAAAAAAAAGGTCTGATGACCCTGTCGAAGAT
GATAAAGAAAAAAAAGAACTTGGTTATTTAACAGTAGAAAATGGAATTGCACCAAACCA
TGGGCTGACTTTTCCTTTAAATTCATGCCTCAAGTATATGTACCCACCACCTTCCAGCA
CAATCCTAGCAAACATTGTAAATGCCTTGGCAAGCGTGCCTAAGTTCTATGTACAGGTC
CTTCATCTTATGAATAAAATGAATTTGCCCACACCTTTTGGACCAATTACTGCGCGACC
TCCCATGTATGAAGACTATATGCCATTGCATGCACCTCTTCCACCCACATCTCCTCAGC
CACCTGAGGAACCTCCTTTGCCAGACGAGGATGAGGAATTATCTAGTGAAGAATCAGAA
TATGAAAGCACTGATGATGAGGACCGACAGAGAATGAACAAATTAATGGAACTAGCAAA
TCTTCAGCCCAAAAGACCTAAAACAATAAAGCAGCGCCATGTGAGAAAAAAGAGAAAAA
TAAAGGATATGTTGAATACACCTTTGTGTCCTTCACACAGCAGTTTACATCCAGTGCTG
TTACCTTCAGATGTATTTGACCAACCACAACCTGTAGGTAACAAAAGAATTGAATTCCA
TATATCTACCGACATGCCAGCTGCATTTAAGAAAGATTTAGAAAAGGAACAAAATTGTG
AGGAAAAAAATCATGATTTACCTGCTACTGAAGTTGATGCATCCAATATAGGATTTGGA
AAAATCTTCCCCAAACCTAATTTGGACATCACAGAGGAGATTAAAGAAGACTCTGATGA
AATGCCTTCAGAATGTATTTCTAGAAGGGAATTGGAAAAGGGCAGAATTTCTAGAGAAG
AAATGGAAACACTTTCAGTTTTCAGAAGTTATGAACCGGGTGAACCAAACTGTAGAATT

TATGTAAAGAATTTAGCTAAACATGTTCAAGAAAAGGACCTTAAATATATTTTTGGAAG
ATATGTTGACTTTTCATCAGAAACACAGCGGATCATGTTTGATATACGTTTGATGAAAG
AAGGTCGTATGAAAGGACAAGCTTTCATTGGACTTCCTAATGAAAAAGCAGCAGCAAAA
GCCTTAAAGGAAGCTAATGGATATGTGCTTTTTGGAAAACCCATGGTGGTTCAGTTTGC
TCGATCTGCTAGACCAAAACAAGATCCTAAGGAAGGAAAAAGAAAGTGTCACCATCATC
ATCACCACTAAGGATCC

(Met- Flag-p59-6His: 1514bp)

GCGGCCGCATGGACTATAAGGATGACGATGACAAGGCCCTGCCACCATTCTTCGGCCAG
GGTCGCCCAGGCCCACCGCCCCCGCAGCCGCCGCCTCCTGCTCCTTTCGGCTGTCCGCC
ACCGCCGCTGCCCTCCCCGGCTTTCCCGCCGCCTCTCCCCCAGCGGCCCGGCCCTTTTC
CGGGGGCCTCCGCCCCCTTCCTTCAGCCTCCGCTGGCTCTGCAGCCCCGAGCCTCCGCG
GAGGCCTCCCGCGGCGGAGGCGGCGCTGGCGCCTTCTACCCGGTGCCACCACCGCCGCT
GCCTCCTCCGCCGCCCCAGTGTCGGCCCTTCCCGGGGACCGACGCCGGCGAGCGGCCGC
GGCCACCGCCTCCCGGCCCGGGGCCGCCCTGGAGCCCGCGGTGGCCTGAGGCGCCGCCG
CCGCCGGCCGACGTGCTCGGGGATGCGGCCCTCCAACGCCTGCGCGACCGGCAGTGGCT
GGAGGCGGTGTTCGGGACCCCGCGGCGGGCAGGCTGTCCGGTGCCCCAGCGCACGCATG
CCGGGCCCAGCCTTGGCGAAGTGCGCGCGCGATTGCTCCGGGCTCTGCGCCTGGTGCGG
CGGCTGCGCGGCCTGAGCCAGGCCCTGCGCGAGGCCGAAGCCGACGGCGCGGCCTGGGT
CCTGCTGTACTCCCAGACCGCGCCGCTGCGCGCGGAACTGGCCGAGCGGCTACAGCCGT
TGACCCAGGCTGCCTATGTGGGCGAGGCGCGGAGGAGGCTGGAGAGGGTCCGGCGCCGC
CGGCTGCGGCTTCGCGAGAGGGCCCGGGAACGCGAGGCCGAGCGGGAGGCAGAGGCCGC
GCGGGCAGTGGAACGCGAGCAGGAGATTGACCGCTGGAGGGTGAAGTGTGTGCAGGAGG
TGGAGGAGAAGAAGCGGGAGCAGGAACTCAAAGCAGCCGCTGATGGCGTACTATCTGAA
GTGAGGAAAAAACAAGCAGATACCAAAAGAATGGTGGACATTCTACGGGCTTTGGAGAA
ATTGAGGAAACTGAGGAAAGAGGCTGCAGCGAGGAAAGGGGTCTGTCCTCCAGCCTCAG
CAGATGAGACTTTTACGCATCATCTTCAGCGACTGAGAAAACTCATTAAAAAGCGCTCT
GAACTGTATGAAGCTGAAGAGAGAGCCCTCAGAGTTATGCTAGAAGGAGAACAAGAGGA
AGAGAGGAAAAGAGAATTAGAAAAGAAACAAAGAAAAGAAAAAGAGAAAATTTTACTTC
AGAAACGTGAAATTGAGTCCAAGTTGTTTGGGGATCCAGATGAGTTCCCACTTGCTCAC
CTCTTGGAGCCTTTCCGACAGTATTATCTCCAAGCCGAGCACTCCCTGCCAGCGCTCAT
CCAGATCAGGCATGATTGGGATCAGTACCTGGTGCCATCCGATCATCCCAAAGGCAACT
TCGTTCCCCAAGGATGGGTCCTTCCCCCGCTCCCCAGCAACGACATCTGGGCAACTGCT
GTTAAGCTGCATCACCATCATCATCACCACTAGGGATCC

(Met- Flag-p48-6His: 1076bp)

GCGGCCGCATGGACTATAAGGATGACGATGACAAGGAGGGCGAGCCTCCACCTGTGGAG
GAGCGGCGGCGGCTGCAGGAGGAGCTGAACGAGTTCGTGGAGAGCGGCTGCCGGACGTT
GGAGGAGGTGACCGCGTCCCTGGGCTGGGACCTAGATAGTCTGGATCCCGGGGAAGAGG
AGGCGGCGGAGGATGAAGTTGTGATATGTCCATACGATTCCAATCATCACATGCCTAAA
TCATCTTTGGCAAAGCACATGGCATCTTGTAGATTGAGGAAAATGGGCTATACCAAAGA
AGAAGAGGATGAAATGTATAATCCTGAGTTTTTCTATGAAAATGTGAAGATACCTTCGA
TTACTTTGAATAAGGACTCACAATTCCAGATAATTAAACAAGCTAGAACTGCAGTTGGG
AAAGACAGTGATTGTTATAATCAAAGAATTTATTCTTCATTGCCTGTTGAAGTTCCTTT
GAATCACAAACGGTTTGTTTGTGATCTAACTCAAGCTGATCGTCTTGCCCTCTATGATT
TCGTAGTTGAGGAGACAAAGAAAAAGCGCTCTGATTCTCAAATTATTGAAAATGACAGC
GATCTCTTTGTAGACTTGGCTGCCAAAATCAATCAAGACAATAGTCGAAAAAGTCCAAA
ATCCTACCTTGAAATCCTGGCAGAAGTACGAGATTATAAAAGAAGACGCCAGTCCTATA
GAGCCAAGAATGTTCACATAACCAAGAAATCATATACTGAGGTGATTCGAGATGTGATA
AATGTGCACATGGAAGAACTCAGCAATCATTGGCAAGAAGAGCAAGAGAAGGCAGAGGA
TGATGCCGAAAAGAATGAAGAAAGGCGATCAGCTTCAGTAGATTCACGGCAGTCTGGTG
GAAGCTATTTGGATGCTGAGTGTTCACGACATAGAAGGGATAGGAGTAGAAGCCCACAT
AAAAGAAAAAGAAACAAAGATAAGGATAAAAACTGTGAGTCGAGAAGAAGGAAAGAGAG
GGATGGGGAAAGACACCATAGTCATAAAAGAAGAAAGCAAAAAATACACCATCATCATC
ACCACTAAGGATCC

(Met- Flag-p35-6His: 797bp)

GCGGCCGCATGGACTATAAGGATGACGATGACAAGAACGATTGGATGCCCATCGCCAAG
GAGTATGATCCACTCAAAGCGGGCAGCATTGATGGCACCGATGAAGACCCACACGACCG
CGCGGTCTGGAGGGCAATGCTGGCACGATATGTCCCCAACAAAGGTGTCATAGGAGATC
CCCTCCTCACCCTGTTTGTGGCCAGACTAAACTTGCAGACCAAGGAGGACAAATTAAAG
GAAGTCTTTTCCCGCTATGGTGACATCCGGCGGCTTCGGCTGGTCAGGGACTTGGTCAC
AGGTTTTTCAAAGGGCTACGCCTTCATCGAATACAAGGAGGAGCGTGCCGTGATCAAAG
CTTACCGAGATGCTGATGGCCTGGTTATTGACCAGCATGAGATATTTGTGGACTACGAG
CTGGAAAGGACTCTCAAAGGGTGGATCCCTCGGCGACTTGGAGGCGGTCTTGGGGGAAA
AAAGGAGTCTGGGCAACTGAGATTTGGGGGACGGGACCGGCCTTTTCGAAAACCTATTA
ACTTGCCAGTTGTTAAAAACGACCTCTATAGAGAGGGAAAACGGGAAAGGCGGGAGCGA
TCTCGATCCCGAGAAAGACACTGGGACTCGAGGACAAGGGATCGAGACCATGACAGGGG
CCGGGAGAAGAGATGGCAAGAAAGAGAGCCGACCAGGGTGTGGCCCGACAATGACTGGG
AGAGAGAGAGGGACTTCAGAGATGACAGGATCAAGGGGAGGGAGAAGAAGGAAAGAGGC
AAGCACCATCATCATCACCACTAGGGATCC

(Met- Flag-p31-6His: 710bp)

GCGGCCGCATGGACTATAAGGATGACGATGACAAGAGTGGTGGATTGGCTCCAAGTAAG
AGCACAGTGTATGTATCCAACTTGCCTTTTTCCCTGACAAACAATGACTTGTACCGGAT
ATTTTCCAAGTATGGCAAAGTTGTAAAGGTTACCATCATGAAAGATAAAGATACCAGGA
AGAGTAAAGGGGTTGCATTTATTTTATTTTTGGATAAAGACTCTGCACAAAACTGTACC
AGGGCAATAAACAACAAACAGTTATTTGGTAGAGTGATAAAAGCAAGCATTGCTATTGA
CAATGGAAGAGCAGCTGAGTTCATCCGAAGGCGAAACTACTTTGATAAATCTAAGTGTT
ATGAATGTGGGGAAAGTGGACACTTAAGTTATGCCTGTCCGAAAAATATGCTCGGAGAA
CGTGAGCCTCCAAAGAAGAAAGAAAAAAAGAAAAAAAAGAAAGCTCCTGAACCAGAAGA
AGAAATTGAGGAAGTAGAAGAAAGTGAAGATGAAGGGGAGGATCCTGCTCTTGACAGCC
TCAGTCAGGCCATAGCATTCCAGCAAGCCAAAATTGAAGAAGAACAAAAAAAATGGAAA
CCCAGTTCAGGAGTCCCCTCAACATCAGATGATTCAAGACGCCCAAGGATAAAGAAAAG
CACATATTTCAGTGATGAGGAAGAACTTAGTGATCACCATCATCATCACCACTAAGGAT
CC

(Met- Flag-p25-6His: 455bp)

GCGGCCGCATGGACTATAAGGATGACGATGACAAGGACGTGTTCCAGGAGGGTCTGGCT
ATGGTGGTGCAGGACCCGCTGCTCTGCGATCTGCCGATCCAGGTTACTCTGGAAGAAGT
CAACTCCCAAATAGCCCTAGAATACGGCCAGGCAATGACGGTCCGAGTGTGCAAGATGG
ATGGAGAAGTAATGCCCGTGGTTGTAGTGCAGAGTGCCACAGTCCTGGACCTGAAGAAG
GCCATCCAGAGATACGTGCAGCTCAAGCAGGAGCGTGAAGGGGGCATTCAGCACATCAG
CTGGTCCTACGTGTGGAGGACGTACCATCTGACCTCTGCAGGAGAGAAACTCACGGAAG
ACAGAAAGAAGCTCCGAGACTACGGCATCCGGAATCGAGACGAGGTTTCCTTCATCAAA
AAGCTGAGGCAAAAGCACCATCATCATCACCACTGAGGATCC

(Met- Flag-p20-6His: 569bp)

GCGGCCGCATGGACTATAAGGATGACGATGACAAGGGGAAGCGATACTTCTGTGACTAC
TGCGACCGCTCCTTCCAGGACAACCTCCACAACCGCAAGAAGCACCTGAACGGGCTGCA
GCACCTCAAGGCCAAGAAGGTCTGGTACGACATGTTCCGAGATGCAGCTGCCATCTTGC
TGGATGAGCAGAACAAGCGGCCCTGCAGGAAGTTTCTACTGACAGGCCAGTGCGACTTT
GGCTCCAACTGCAGATTTTCCCACATGTCAGAGCGAGACCTGCAGGAGCTGAGCATCCA
GGTGGAGGAGGAGAGGCGAGCCAGGGAGTGGCTACTAGATGCTCCTGAGCTCCCCGAGG
GCCATCTGGAGGACTGGCTGGAGAAGAGAGCCAAGCGGCTGAGCTCAGCCCCAAGTAGC
AGGGCTGAACCCATCAGAACCACTGTCTTCCAGTACCCCGTGGGCTGGCCACCAGTTCA
GGAGCTGCCTCCATCCCTGCGGGCACCCCCACCTGGGGGGTGGCCTCTGCAGCCCAGAG
TCCAGTGGGGCCACCATCATCATCACCACTGAGGATCC

Appendix 3. Sequences and designing strategy of U11/U12 di-snRNP complex specific protein constructs for bacterial expression

Following are the cDNA sequences of open reading frames of the U11/U12 di-snRNP complex specific proteins explained in Material and Methods section 2.12. The genes were designed and codon optimized according to their expression in the bacterial (BL21) cells. The sequence highlighted in yellow are the flanking sequences at the 5' and 3' end. The green color sequence represents the Nde I and BamH I restriction sites at the 5' and 3' ends respectively. The FLAG tag and the the 6 × His tag are in cyan and magenta colors. The start codon and the stop codon are in green and red colors respectively.

(Met- Flag-p65-Gly-6His: 1629 bp)

ggaaatatacatATGGACTATAAGGATGACGATGACAAGGCTGCTCCGGAACAGCCCTT
GGCTATATCTAGGGGATGTACTAGCAGCAGCTCGTTGAGCCCCCCGCGGGGTGACAGAA
CCTTGCTGGTCAGGCATCTGCCCGCTGAATTGACTGCTGAAGAAAAGGAAGACTTGCTG
AAATACTTTGGAGCTCAGAGCGTTCGCGTCCTGTCTGACAAAGGAAGACTGAAGCACAC
AGCTTTCGCAACATTTCCGAACGAAAAGGCTGCTATAAAAGCTTTGACAAGATTACATC
AACTGAAGTTGCTCGGGCATACTCTCGTCGTAGAATTCGCTAAGGAACAAGACAGAGTA
CATAGCCCGTGTCCGACTTCTGGGAGCGAAAAGAAGAAGCGCAGCGACGACCCGGTCGA
AGACGACAAGGAAAAGAAGGAATTGGGGTACCTCACAGTAGAAAACGGAATAGCTCCGA
ACCATGGACTGACTTTCCCGCTCAACTCTTGTTTAAAATACATGTACCCGCCGCCGAGC
AGCACAATCCTAGCTAACATAGTAAACGCATTGGCTAGCGTTCCGAAATTTTACGTACA
GGTCTTGCACTTGATGAACAAAATGAACTTGCCGACACCGTTCGGACCGATAACTGCTA
GACCGCCGATGTACGAAGACTACATGCCGTTGCACGCTCCGTTGCCGCCGACAAGCCCG
CAGCCGCCGGAAGAACCGCCGTTGCCGGACGAAGACGAAGAACTCAGCTCTGAAGAATC
TGAATACGAAAGCACTGACGACGAAGACAGACAGCGCATGAACAAACTCATGGAACTAG
CTAACTTGCAGCCGAAACGCCCGAAAACAATAAAACAGCGGCATGTTCGCAAAAAACGC
AAAATAAAAGACATGTTGAACACACCGTTGTGTCCGTCTCATAGCTCTCTCCATCCGGT
TCTGCTCCCGTCTGACGTATTCGACCAACCGCAACCGGTAGGGAACAAACGCATAGAAT
TTCATATAAGCACCGACATGCCGGCTGCTTTCAAAAAGGACCTCGAAAAAGAACAAAAC
TGTGAAGAAAAGAACCATGACCTCCCGGCTACTGAAGTAGACGCTAGCAACATAGGATT
CGGAAAAATCTTCCCGAAACCGAACTTGGACATCACAGAAGAAATAAAGGAAGACAGCG
ACGAAATGCCGTCTGAATGTATCAGCCGCCGCGAATTGGAAAAAGGTCGCATCAGCCGC
GAAGAAATGGAAACATTGTCTGTATTTCGCTCTTACGAACCCGGGGAACCGAACTGTCG

CATCTACGTAAAAAACCTCGCTAAACATGTACAAGAAAAAGACTTGAAATACATATTCG
GACGCTACGTAGACTTCTCTTCTGAAACACAGCGCATCATGTTCGACATACGATTGATG
AAGGAAGGGCGAATGAAAGGACAAGCTTTTATAGGATTGCCGAACGAAAAGGCTGCTGC
TAAGGCACTCAAAGAAGCTAACGGATACGTTTTGTTCGGAAAACCGATGGTTGTACAGT
TCGCTAGAAGCGCTCGCCCGAAACAAGACCCGAAAGAAGGAAAACGCAAATGTgggcac
catcatcatcaccacTAATGAcggatccggctgcaa

(Met- Flag-p59-Gly-6His: 1533 bp)

ggaaatatacatATGGACTATAAGGATGACGATGACAAGGCACTGCCGCCGTTCTTCGG
CCAGGGGCGGCCGGGTCCGCCCCCGCCCCAGCCCCCCCCGCCGGCTCCGTTTGGTTGTC
CCCCGCCCCCCCTGCCGAGCCCCGCTTTTCCCCCCCCGTTACCGCAGCGCCCGGGTCCG
TTCCCCGGAGCAAGCGCACCGTTCCTCCAGCCGCCCCTGGCTCTGCAGCCGAGAGCAAG
CGCTGAAGCAAGCCGGGGTGGAGGTGGTGCTGGCGCCTTCTACCCCGTTCCGCCGCCCC
CCCTGCCGCCGCCCCCCCCCGCAGTGTCGCCCGTTTCCCGGAACCGACGCAGGTGAACGC
CCCCGCCCGCCCCCGCCGGGTCCCGGACCCCCGTGGAGCCCCCGCTGGCCGGAAGCTCC
CCCCCCCCCCGCAGACGTTTTAGGAGACGCTGCATTACAACGGCTGCGGGACCGCCAGT
GGCTGGAAGCTGTTTTCGGAACCCCCCGCCGCGCTGGTTGTCCCGTTCCGCAGCGGACT
CATGCAGGACCGAGCTTGGGTGAAGTTCGGGCTAGATTGTTACGGGCTCTGCGGCTGGT
TCGCCGCCTGCGGGGTCTGAGCCAGGCACTGCGGGAAGCAGAAGCAGACGGTGCTGCAT
GGGTCCTGCTGTACAGCCAGACCGCTCCCCTGCGGGCTGAACTGGCAGAACGCCTACAG
CCCTTGACCCAGGCTGCATACGTTGGCGAAGCTCGCCGCCGCCTGGAACGCGTCCGCCG
GCGGCGCCTGCGCTTGCGGGAACGCGCACGCGAACGGGAAGCAGAACGCGAAGCTGAAG
CAGCTCGCGCTGTTGAACGGGAACAGGAAATAGACCGGTGGCGCGTTAAATGTGTTCAG
GAAGTTGAAGAAAAGAAGCGGGAACAGGAATTAAAAGCTGCAGCTGACGGTGTACTAAG
CGAAGTTCGCAAAAAACAAGCTGACACCAAACGCATGGTTGACATACTACGCGCTTTGG
AAAAGTTGCGCAAGCTGCGCAAGGAAGCTGCTGCTCGCAAAGGAGTCTGTCCGCCGGCA
TCTGCTGACGAAACTTTTACTCATCATTTGCAGAGACTGCGCAAATTAATAAAGAAACG
GAGCGAACTGTACGAAGCTGAAGAACGCGCATTAAGAGTAATGCTAGAAGGAGAACAAG
AAGAAGAACGCAAACGCGAACTCGAAAAAAAGCAACGCAAGGAAAAGGAAAAGATACTC
TTGCAGAAGCGAGAAATAGAAAGCAAATTGTTCGGAGACCCGGACGAATTTCCGTTGGC
TCATTTATTGGAACCGTTTAGACAGTACTACTTACAAGCAGAACATAGCCTGCCGGCTT
TAATCCAGATCCGCCATGACTGGGACCAGTACCTGGTTCCGAGCGACCATCCGAAGGGT
AACTTTGTACCGCAAGGATGGGTCTTGCCGCCCTTACCGAGCAACGACATCTGGGCTAC
TGCTGTAAAACTGCATgggcaccatcatcatcaccacTAATGAcggatccggctgcaa

(Met- Flag-p48-Gly-6His: 1095 bp)

ggaaatatacatATGGACTATAAGGATGACGATGACAAGGAAGGTGAACCGCCGCCGGT
TGAAGAACGCCGCCGCCTGCAGGAAGAACTGAACGAATTCGTTGAAAGCGGTTGTCGCA
CTTTGGAAGAAGTTACCGCTAGCCTGGGTTGGGACCTAGACTCTCTGGACCCGGGAGAA
GAAGAAGCTGCTGAAGACGAAGTAGTTATATGTCCGTACGACAGCAACCATCACATGCC

GAAATCTAGCTTGGCTAAACACATGGCTAGCTGTCGCTTGCGCAAAATGGGTTACACCA
AGGAAGAAGAAGACGAAATGTACAACCCGGAATTCTTTTACGAAAACGTTAAAATACCG
TCGATAACTTTGAACAAAGACTCTCAATTTCAGATAATAAAACAAGCTCGCACTGCTGT
AGGAAAAGACTCTGACTGTTACAACCAACGCATCTACAGCTCTTTGCCGGTAGAAGTAC
CGTTGAACCATAAACGCTTCGTATGTGACCTAACTCAAGCTGACCGATTGGCATTATAC
GACTTTGTAGTAGAAGAAACAAAAAAGAAACGGAGCGACAGCCAAATAATAGAAAACGA
CAGCGACTTATTCGTAGACTTGGCTGCAAAAATCAACCAAGACAACTCTAGAAAATCTC
CGAAAAGCTACTTGGAAATCCTGGCTGAAGTAAGAGACTACAAACGCCGCCGGCAGAGC
TACCGCGCAAAAAACGTACATATAACCAAAAAGTCTTACACTGAAGTTATAAGAGACGT
TATAAACGTTCACATGGAAGAATTAAGCAACCATTGGCAAGAAGAACAAGAAAAAGCTG
AAGACGACGCAGAAAAAAACGAAGAACGCAGATCTGCTTCTGTAGACTCTCGCCAGAGC
GGGGGAAGCTACTTGGACGCTGAATGTTCTAGACATCGCCGCGACCGCTCTCGCAGCCC
GCATAAACGCAAACGCAACAAAGACAAAGACAAAAACTGTGAATCGCGCCGCCGCAAGG
AACGCGACGGAGAACGCCATCATTCTCATAAGCGCCGCAAACAAAAGATAgggcaccat
catcatcaccacTAATGAcggatccggctgcaa

(Met- Flag-p35-Gly-6His: 816 bp)

ggaaatatacatATGGACTATAAGGATGACGATGACAAGAACGACTGGATGCCGATCGC
AAAAGAATACGACCCGTTAAAAGCTGGTAGCATAGACGGTACCGACGAAGACCCGCATG
ACCGGGCTGTCTGGCGCGCTATGCTGGCTAGATACGTCCCGAACAAAGGGGTCATAGGA
GACCCGTTATTAACCCTGTTCGTTGCAAGACTAAACTTGCAGACCAAAGAAGACAAGCT
CAAAGAAGTCTTCAGCCGGTACGGGGACATCCGCCGCTTGCGCCTGGTCCGCGACTTGG
TCACAGGGTTCTCTAAAGGTTACGCATTCATCGAATACAAAGAAGAACGAGCAGTTATC
AAAGCTTACAGAGACGCTGACGGTCTGGTAATAGACCAGCATGAAATATTCGTTGACTA
CGAACTGGAACGCACTTTAAAAGGATGGATTCCGCGCAGATTGGGAGGTGGGTTGGGAG
GAAAGAAAGAAAGCGGACAACTGCGCTTCGGAGGACGCGACCGCCCGTTCAGAAAACCG
ATAAACTTGCCGGTAGTAAAAAACGACTTATACCGCGAAGGAAAACGCGAACGCCGCGA
AAGAAGCAGAAGCAGAGAACGCCATTGGGACTCGCGCACACGCGACAGAGACCATGACA
GGGGTCGGGAAAAACGCTGGCAAGAACGCGAACCCACCCGCGTTTGGCCGGACAACGAC
TGGGAACGCGAACGCGACTTTCGCGACGACCGCATCAAAGGAAGGGAAAAAAAAGAAAG
AGGTAAAgggcaccatcatcatcaccacTAATGAcggatccggctgcaa

(Met- Flag-p31-Gly-6His: 729 bp)

ggaaatatacatATGGACTATAAGGATGACGATGACAAGTCTGGGGGATTGGCTCCGTC
TAAAAGCACAGTTTACGTAAGCAACTTGCCGTTCAGCCTGACAAACAACGACCTCTACC
GGATATTCAGCAAATACGGTAAAGTAGTAAAAGTAACCATCATGAAAGACAAAGACACC
CGCAAATCTAAAGGAGTAGCTTTCATACTCTTCCTCGACAAAGACAGCGCTCAAAACTG
TACCCGCGCTATAAACAACAAGCAGCTCTTCGGGCGCGTTATAAAAGCTAGCATAGCTA
TAGACAACGGACGCGCTGCTGAATTCATCAGACGCAGAAACTACTTCGACAAGAGCAAA
TGTTACGAATGTGGAGAATCTGGACATCTCTCTTACGCATGTCCCAAAAACATGTTAGG
AGAACGAGAACCGCCGAAAAAAAAGGAAAAGAAAAAGAAGAAAAAGGCTCCGGAACCGG

AAGAAGAAATAGAAGAAGTAGAAGAATCTGAAGACGAAGGAGAAGACCCGGCTTTGGAC
AGCTTATCTCAGGCAATAGCTTTTCAGCAAGCAAAGATAGAAGAAGAACAAAAGAAGTG
GAAGCCGTCTTCTGGAGTCCCGTCTACATCTGACGACTCTCGCCGGCCGCGCATAAAAA
AGAGCACATACTTTTCTGACGAAGAAGAACTCTCTGACgggcaccatcatcatcaccac
TAATGAcggatccggctgcaa

(Met- Flag-p25-Gly-6His: 474 bp)

ggaaatatacatATGGACTATAAGGATGACGATGACAAGGACGTTTTCCAGGAAGGGCT
GGCTATGGTTGTTCAGGACCCCCTGTTATGTGACCTGCCCATCCAGGTAACTCTGGAAG
AAGTCAACAGCCAAATAGCACTAGAATACGGTCAGGCTATGACTGTCAGAGTTTGTAAA
ATGGACGGAGAAGTAATGCCGGTTGTAGTAGTTCAGTCTGCAACAGTCCTGGACCTGAA
AAAAGCAATCCAGCGCTACGTTCAGTTAAAACAGGAACGAGAAGGAGGTATACAGCATA
TCAGCTGGAGCTACGTTTGGCGCACTTACCACCTGACCAGCGCTGGAGAAAAGTTAACT
GAAGACCGCAAAAAATTAAGAGACTACGGTATCCGCAACAGAGACGAAGTAAGCTTCAT
CAAGAAACTGCGCCAAAAAgggcaccatcatcatcaccacTAATGAcggatccggctgc
aa

(Met- Flag-p20-Gly-6His: 588 bp)

ggaaatatacatATGGACTATAAGGATGACGATGACAAGGGAAAAAGATACTTCTGTGA
CTACTGTGACCGGAGCTTCCAGGACAACTTACATAACCGGAAAAAACATCTGAACGGAC
TGCAGCATTTAAAAGCAAAAAAAGTCTGGTACGACATGTTTAGAGACGCTGCTGCAATC
TTGCTGGACGAACAGAACAAACGCCCGTGTCGCAAATTCCTACTGACAGGTCAGTGTGA
CTTCGGTAGCAACTGTCGCTTCAGCCACATGTCTGAAAGAGACCTGCAGGAACTGAGCA
TCCAGGTTGAAGAAGAACGCAGAGCACGCGAATGGCTACTAGACGCTCCGGAATTACCG
GAAGGTCATCTGGAAGACTGGCTGGAAAAGAGAGCAAAACGCCTGAGCTCTGCACCGTC
TAGCAGGGCTGAACCGATCCGCACCACTGTCTTCCAGTACCCGGTTGGTTGGCCGCCGG
TACAGGAACTGCCGCCGAGCCTGCGCGCTCCGCCGCCGGGAGGATGGCCGCTGCAGCCG
AGAGTCCAGTGGGGTgggcaccatcatcatcaccacTAATGAcggatccggctgcaa