

5-2013

## Structure-Function Analysis Of Human Integrator Subunit-4

Anupama Sataluri

Follow this and additional works at: [https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations](https://digitalcommons.library.tmc.edu/utgsbs_dissertations)



Part of the [Biochemistry Commons](#), and the [Molecular Biology Commons](#)

---

### Recommended Citation

Sataluri, Anupama, "Structure-Function Analysis Of Human Integrator Subunit-4" (2013). *Dissertations and Theses (Open Access)*. 360.

[https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations/360](https://digitalcommons.library.tmc.edu/utgsbs_dissertations/360)

This Thesis (MS) is brought to you for free and open access by the MD Anderson UTHealth Houston Graduate School at DigitalCommons@TMC. It has been accepted for inclusion in Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact [digcommons@library.tmc.edu](mailto:digcommons@library.tmc.edu).

# STRUCTURE-FUNCTION ANALYSIS OF HUMAN INTEGRATOR

## SUBUNIT 4

By:

Anupama Sataluri

APPROVED:

---

Eric Wagner, Ph.D. (Supervisory Advisor)

---

Vasanthi Jayaraman, Ph.D.

---

Phillip Carpenter, Ph.D.

---

Joel Neilson, Ph.D.

---

Ambro van Hoof, Ph.D.

APPROVED:

---

Dean, The University of Texas  
Graduate School of Biomedical Sciences at Houston

# **STRUCTURE-FUNCTION ANALYSIS OF HUMAN INTEGRATOR**

## **SUBUNIT 4**

**A**

### **THESIS**

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

M.D. Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

**MASTER OF SCIENCE**

by

Anupama Sataluri

May 2013



## **ACKNOWLEDGEMENTS**

I would like to express sincere gratitude to my advisor, Dr. Eric J. Wagner for his mentoring throughout my thesis work. Your leadership, hard work, attention to detail and passion for science has set high standards that I will always strive to achieve. I would also like to thank all the committee members: Drs. Michael Blackburn, Vasanthi Jayaraman, Phillip Carpenter, Joel Neilson, and Ambro van Hoof for their constructive suggestions and ideas, which were critical in shaping my project.

I would like to thank all the Wagner lab members both past and present: Todd Albrecht, Natoya Peart, Scott Collum, Dr. David Baillat, Dr. Patience Masamha, Dr. Nader Ezzeddine, and Dr. Jiandong Chen for creating a friendly environment in the lab. I have learned so many small scientific tips and tricks, which cannot be found in textbooks from all of you that I will remember for a long time. Also, I would like to thank the BMB program staff, students, and faculty for making BMB graduate program great.

My stay in Houston over the past three years was made joyful and interesting thanks to my friends. I feel glad to have met Swaty Singh, Pooja Dhupkar, Seema Mukherjee, Deepthi Kothapalli, Aditi Thakkar and Kavitha Reddy. Thank you for celebrating my happiness and cheering me up through tough times. I will miss all the great times we've shared while in Houston.

Finally, a big thank you to my husband, Dhishan Kande for all the encouragement and for always being there for me during good times and bad. Last but not the least, I feel grateful for having Uma and Badri Narayan Sataluri as my parents. Thank you for all the love and support and for always having confidence in my abilities.

# **Structure-function analysis of human Integrator subunit 4**

Anupama Sataluri

Advisor: Eric. J. Wagner, Ph.D.

Uridine-rich small nuclear RNAs (U snRNA) are RNA Polymerase-II (RNAPII) transcripts that are ubiquitously expressed and are known to be essential for gene expression. snRNAs play a key role in mRNA splicing and in histone mRNA expression. Inaccurate snRNA biosynthesis can lead to diseases related to defective splicing and histone mRNA expression. Although the 3' end formation mechanism and processing machinery of other RNAPII transcripts such as mRNA has been well studied, the mechanism of snRNA 3' end processing has remained a mystery until the recent discovery of the machinery that mediates this process. In 2005, a complex of 14 subunits (the Integrator complex) associated with RNA Polymerase-II was discovered. The 14 subunits were annotated Integrator 1-14 based on their size. The subunits of this complex together were found to facilitate 3' end processing of snRNA. Identification of the Integrator complex propelled research in the direction of understanding the events of snRNA 3' end processing. Recent studies from our lab confirmed that Integrator subunit (IntS) 9 and 11 together perform the endonucleolytic cleavage of the nascent snRNA 3' end to generate mature snRNA. However, the role of other members of the Integrator complex remains elusive. Current research in our lab is focused on deciphering the role of each subunit within the Integrator complex.

This work specifically focuses on elucidating the role of human Integrator subunit 4 (IntS4) and understanding how it facilitates the overall function of the complex.

IntS4 has structural similarity with a protein called “Symplekin”, which is part of the mRNA 3’end processing machinery. Symplekin has been thoroughly researched in recent years and structure-function correlation studies in the context of mRNA 3’end processing have reported a scaffold function for Symplekin due to the presence of HEAT repeat motifs in its N-terminus. Based upon the structural similarity between IntS4 and Symplekin, ***we hypothesized that Integrator subunit 4 may be behaving as a Symplekin-like scaffold molecule that facilitates the interaction between other members of the Integrator Complex.*** To answer this question, the two important goals of this study were to: 1) identify the region of IntS4, which is important for snRNA 3’ end processing and 2) determine binding partners of IntS4 which promote its function as a scaffold. IntS4 structurally consists of a highly conserved N-terminus with 8 HEAT repeats, followed by a nonconserved C- terminus. A series of siRNA resistant N and C-terminus deletion constructs as well as specific point mutants within its N-terminal HEAT repeats were generated for human IntS4 and, utilizing a snRNA transcriptional readthrough GFP-reporter assay, we tested their ability to rescue misprocessing. This assay revealed a possible scaffold like property of IntS4. To probe IntS4 for interaction partners, we performed co-immunoprecipitation on nuclear extracts of IntS4 expressing stable cell lines and identified IntS3 and IntS5 among other Integrator subunits to be binding partners which facilitate the scaffold like function of hIntS4. These findings have established a critical role for IntS4 in snRNA 3’ end processing, identified that both its N and C termini are essential for its function, and mapped putative interaction domains with other Integrator subunits.

Keywords: UsnRNA 3’end processing, Integrator subunit-4, Symplekin, HEAT repeats

## TABLE OF CONTENTS

	Page No.
Approval Sheet.....	i
Title page.....	ii
Acknowledgements.....	iv
Abstract.....	v
Table of contents.....	vii
List of Figures.....	ix
List of Tables.....	xi
<b>Chapter - 1 Introduction.....</b>	<b>1</b>
INTRODUCTION AND BACKGROUND.....	2
Uridine rich non-coding transcripts – small nuclear RNAs.....	2
snRNA transcription.....	4
Comparison of RNAPII transcript 3' end processing mechanisms.....	8
The Integrator complex.....	12
Current model of snRNA 3'end formation by the Integrator complex.....	15
The role of Integrator complex beyond snRNA 3'end formation.....	18
SIGNIFICANCE OF THIS WORK.....	18
<b>Chapter - 2 Materials and Methods.....</b>	<b>20</b>
Cell culture and stable cell line generation.....	21



Generation of hIntS4 deletion mutants.....	21
hU7-GFP reporter assay, RNA interference and transfections.....	22
Western Blot Analysis.....	25
Immunofluorescence.....	25
Site-directed Mutagenesis.....	26
Nuclear extract preparation.....	27
Immunoprecipitation.....	28
 <b>Chapter - 3 Results and Discussion.....</b>	 <b>34</b>
Introduction.....	35
<i>In silico</i> analysis of IntS4.....	<b>38</b>
Human IntS4 is required for snRNA 3' end formation in HeLa cells.....	44
Development of siRNA resistant hIntS4 to study the role of IntS4 in snRNA 3' end processing.....	46
hIntS4 and Cajal body structural integrity.....	48
Deletion analysis of IntS4 demonstrates that Integrator function is highly intolerant to disruptions in IntS4 structure.....	51
Identification of N-terminal HEAT repeats essential to the function of IntS4.....	57
Identification of binding partners of IntS4.....	63
 <b>Chapter - 4 Conclusions and Future Directions.....</b>	 <b>67</b>
 <b>References.....</b>	 <b>72</b>
 <b>Vita.....</b>	 <b>83</b>

## LIST OF FIGURES

### Chapter-1 Introduction

Figure 1.1: Classification of snRNAs.....	6
Figure 1.2: snRNA gene structural features.....	7
Figure 1.3: 3' end formation of RNAPII transcripts.....	11
Figure 1.4: Integrator complex – the 3' end processing machinery of snRNA.....	13
Figure 1.5: Model of Integrator complex mediated snRNA 3'end formation.....	17

### Chapter-2 Materials and Methods

Figure 2.1: Schematic of U7-GFP reporter assay.....	24
---	----

### Chapter-3 Results and Discussion

Figure 3.1: Model of IntS4 - Structural similarities with Symplekin.....	37
Figure 3.2: Sequence alignment and PONDR analysis of Integrator subunit 4.....	40
Figure 3.3: Model prediction for HEAT repeat region of hIntS4.....	43
Figure 3.4: IntS4 depletion results in significant snRNA misprocessing in HeLa cells.....	45
	47

Figure 3.5: siRNA resistant hIntS4 completely rescues the misprocessing Phenotype in snRNA.....	
Figure 3.6: IntS4 depletion results in disruption of Cajal body structural integrity.....	50
Figure 3.7: Generation of siRNA resistant deletion constructs for hIntS4.....	52
Figure 3.8: Deletion analysis of full-length hIntS4 reveals a possible scaffold-like property of hIntS4.....	53
Figure 3.9: Immunofluorescence images of full-length IntS4 and each truncation construct to determine localization.....	55
Figure 3.10: Prediction of amino acids within the IntS4 N-terminal HEAT repeat region that protrude from the concave surface.....	58
Figure 3.11: Identification of HEAT repeats important for IntS4 mediated snRNA 3' end processing.....	61
Figure 3.12: Summary of results from Figure 3.11.....	62
Figure 3.13: Identification of potential candidates for IntS4 binding by co- Immunoprecipitation assays.....	65

## LIST OF TABLES

### Chapter - 2 Materials and Methods

Table 2.1: Primers for generating deletion mutants from siRNA resistant full Length hIntS4.....	29
Table 2.2: Primers for quickchange site-directed mutagenesis for introducing point mutations in the N-terminus of hIntS4.....	30
Table 2.3: Sequencing oligonucleotides for quickchange mutations in N-terminus of hIntS4.....	32
Table 2.4: Primers for generating pET49b IntS4 clones for recombinant protein expression in E.coli.....	33

### Chapter - 3 Results and Discussion

Table 3.1: HEAT repeat sequence showing the residues chosen for Alanine mutations.....	60
--	----





# **Chapter-1**

## **Introduction**

### **Uridine rich non-coding transcripts - small nuclear RNAs:**

Small nuclear RNAs (snRNAs) are a species of essential non-coding RNAs that are ubiquitously expressed, uridine rich and non-polyadenylated transcripts that predominantly localize and function in the nucleus. This class of RNAs is widely understood to play a role in spliceosome mediated pre-mRNA splicing. snRNAs are classified as either Sm-class snRNAs or Lsm-class based on sequence features of their genes, binding partners, and protein co-factors (Chwedenczuk et al. 2006; Matera et al. 2007).

The Sm-class (U1, U2, U4, U5, U7, U11, U12 and U4atac) are RNA polymerase-II transcripts and are the major category of U snRNAs that form the spliceosome, which facilitates the removal of introns and splicing of pre-mRNA. The exception to this group is the U7 snRNA, which, like the other Sm-class snRNAs assembles into a snRNP but plays a role in histone pre-mRNA processing (described in more detail below)(Schumperli and Pillai 2004; Marzluff 2005; Marzluff et al. 2008; Yang et al. 2009). Structurally, the Sm-class of snRNAs is characterized by the presence of a 5'-trimethylguanosine cap, a 3' stem-loop, and an internal Sm site that functions as a binding site for a heteroheptameric ring of proteins called the "Sm proteins" (Figure 1). These Sm proteins are assembled onto the snRNA Sm site in the cytoplasm through the action of the SMN protein complex and then are shuttled back into the nucleus where they receive further base modifications prior to their role in gene expression.

In contrast, the Lsm class of snRNAs (U6 and U6atac) are RNAPIII transcripts and differ structurally in that they have a monomethylphosphate cap, a 3' stem-loop followed



by a stretch of uridines at the end which serve both as a termination site and subsequently facilitate the binding of Lsm proteins (a distinct heptameric ring). These two Lsm-class snRNAs reside in the nucleus and are devoid of the shuttling into the cytoplasm (Matera et al. 2007) (Figure 1.1).

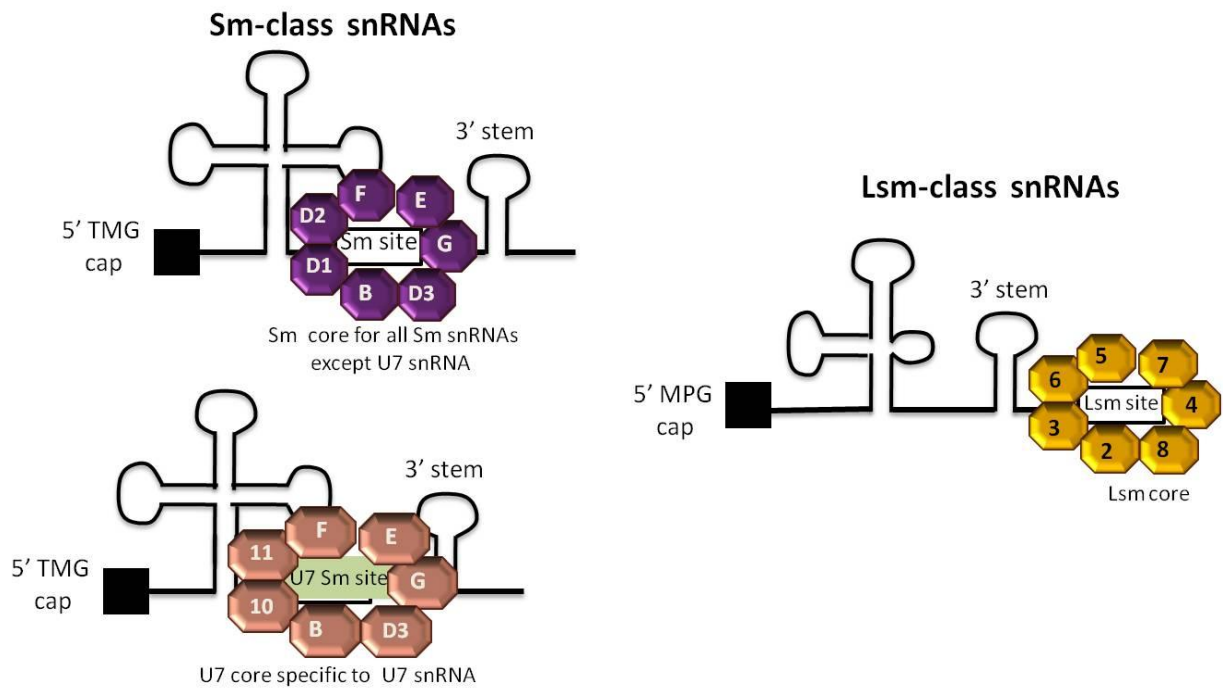
The Sm class snRNAs are widely expressed and are essential to all eukaryotic cells, yet they have relatively simple gene architecture unlike their mRNA counterparts. As stated above, the genes encoding these RNAs are transcribed by RNA Polymerase II, have TATA-less promoters, lack introns and an open reading frame (reviewed in (Hernandez et al. 1989; Hernandez 2001)). Their promoters are characterized by the presence of unique and highly conserved elements called the PSE (Proximal Sequence Element) that has been found to be essential for snRNA transcription and the DSE (Distal Sequence Element) that functions as a transcriptional enhancer (Cheung et al. 1993; Jensen et al. 1998) (Figure 1.2). They also lack classical cleavage and polyadenylation sites and are thus, not 3' end polyadenylated, which is consistent with a lack of protein translation potential. Instead, cleavage of the nascent pre-snRNAs is directed by a downstream element called the 3' box, which is located 9-19 bp downstream of the coding region (Hernandez 1985; Ezzeddine et al. 2011). This 3' box is snRNA gene specific and is necessary for accurate snRNA 3' end processing (reviewed in (Egloff et al. 2008)) (Figure 2).

**snRNA transcription:**

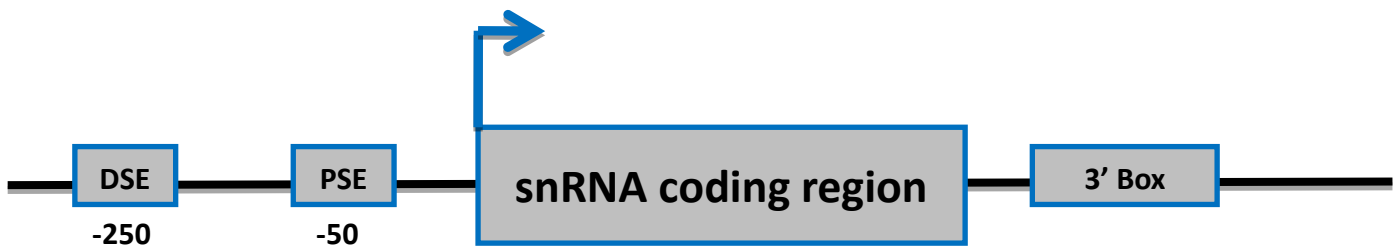
This area has been relatively well researched over the past few years and a basic model has emerged. Briefly, transcription at the snRNA genes is initiated by a set of snRNA gene specific proteins, which recognize the aforementioned PSE and ultimately recruit RNAPII. These factors are present in a five-membered complex that has gone by several names: the SNAPc (snRNA activator protein complex), PBP (PSE-binding protein), and PTF (PSE – binding transcription factor) (Buratowski 2009; Hung and Stumph 2011). Subsequent to recruitment of RNAPII, a series of events involving the sequential phosphorylation of Serines 5, 2 and 7 (in that order) located within the C-terminal domain (CTD) of the heptad repeat (YS<sup>2</sup>PTS<sup>5</sup>PS<sup>7</sup>) of the largest subunit of RNAPII (Rpb1) occur. Specifically, following transcription initiation Ser5 is phosphorylated (hereafter referred to as Ser5P) by cyclin-dependent kinase 7 (Cdk7)(Egloff and Murphy 2008) This step has been well studied to be important for promoter clearance and is essential to recruit the capping enzyme to the nascent snRNA to facilitate addition of a 7mG cap. Following Ser5 phosphorylation is Ser2 phosphorylation by Cdk9, which normally stimulates the elongation capacity of RNAPII but in the case of the short snRNA transcripts is thought to be recognized by a yet-to-be identified protein (Medlin et al. 2003; Uguen and Murphy 2003; Jacobs et al. 2004; Medlin et al. 2005; Egloff et al. 2008; Egloff et al. 2010). Following Ser5 phosphorylation is Ser7 phosphorylation by Cdk9, which normally stimulates the elongation capacity of RNAPII but in the case of the short snRNA transcripts is thought to be recognized by a yet-to-be identified protein (Medlin et al. 2003; Uguen and Murphy 2003; Jacobs et al.

2004; Medlin et al. 2005; Egloff et al. 2008; Egloff et al. 2010). Finally, the Ser5P is thought to be dephosphorylated by a phosphatase and then Ser7 phosphorylation transpires but the kinase that conducts this and the biochemical meaning of this mark is still not clear (Egloff et al. 2007). snRNA 3' end processing has been known to be co-transcriptional and although much is known about the overall process beginning at the 5' end transcription initiation events, the mechanism of the 3' end processing events that lead up to maturation and release of the nascent RNA transcript is an emerging field of research. The research focus of the Wagner lab is to decipher the detailed mechanism and series of events involved in the 3' end processing of snRNA genes.

**Figure 1.1. Classification of snRNAs.** The Sm-class snRNAs (U1, U2, U4, U5, U11, U12, U4atac) form the majority of the major and minor spliceosomes have an Sm binding site, and are transcribed by RNAPII. The exception is U7 snRNA, which has a few deviations from the Sm consensus site (underlined) and is involved in the processing of histone pre-mRNA. The Lsm-class snRNAs (U6 and U6atac) are RNAPIII transcribed, have a Lsm binding site, and are integrated with the Lsm heptameric ring proteins.



**Figure 1.2. snRNA gene structural features.** snRNA genes are characterized by the presence of a TATA-less promoter, which comprises of a distal sequence element (DSE) and a proximal sequence element (PSE). They also lack an ORF, are nonpolyadenylated, and have a 3' box that is important for 3' end processing.



## **Comparison of 3' end processing mechanisms and machinery for major RNA polymerase-II transcripts:**

The cis-acting elements that direct the 3' end formation of snRNA have been previously well-characterized through research conducted from 1985 through ~2004 (Mandel et al 2008). During that period, no protein factors were identified that actually carry out this process and it was assumed that members of the other two RNAPII 3' end RNA processing complexes were involved. Although poly(A)+ mRNA and the replication-dependent histone mRNA are structurally and functionally very distinct, their 3' end processing pathways converge in that a partially overlapping set of machinery is utilized. Below is a short review of what is known about the other two parallel 3' end processing pathways to give perspective to potential function of snRNA 3' end processing complex, which is described afterward.

mRNA 3' end processing – the Cleavage and Polyadenylation machinery: The detailed mechanism and factors involved 3' end formation of the mRNA has been well established over years of research (Figure 1.3, left). The cleavage and polyadenylation machinery is known to perform the 3' end processing event. Poly(A) mRNAs are characterized by the presence of two cis-regulatory elements which define their 3' end (Dominski et al. 2005; Mandel et al. 2008). The first being the AAUAAA polyadenylation signal (PAS) and the second one is the G/U rich downstream element (DSE). The recognition of the PAS and the DSE occurs by initial binding of the Cleavage and Polyadenylation Specificity Factor (CPSF) complex and the Cleavage Stimulation Factor (CstF) complex, respectively. The CPSF complex consists of five members

referred to, by their molecular mass. The largest member of this complex is CPSF160, which is the only member with a known RNA binding domain and recognizes the PAS (Murthy and Manley 1995). The DSE is bound using a similar RNA binding domain present in CstF64, a member of the trimeric CstF complex (MacDonald et al. 1994).

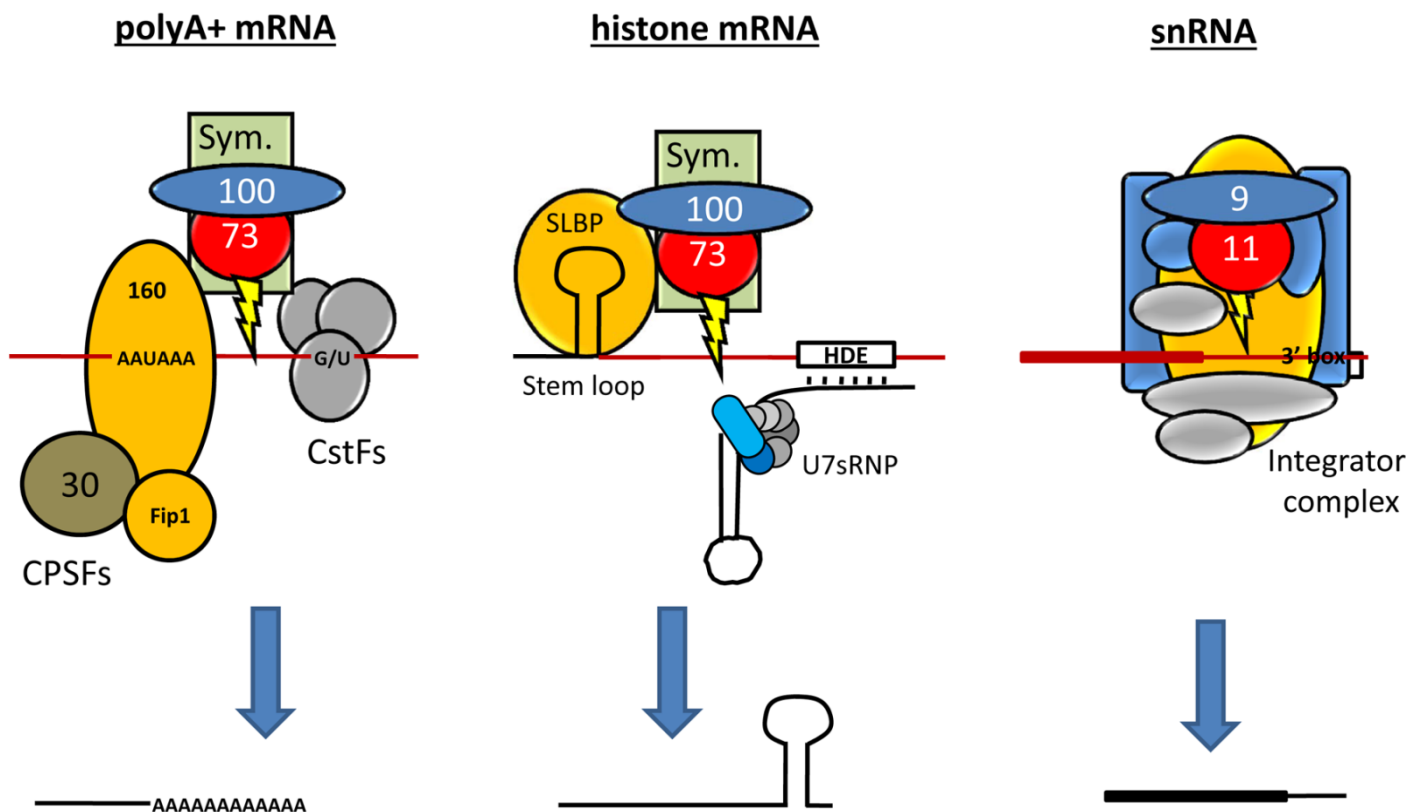
Following the recognition of the PAS and DSE via the binding of the CPSF and CstF complexes, the next step is the recruitment of the cleavage factor to the cleavage site, which results in the cleavage and generation of the nascent mRNA. A CA dinucleotide, which serves as the optimal cleavage site is located between the PAS and the DSE and is the target site of the endonuclease CPSF73 (Mandel et al 2006). Symplekin, a large scaffold molecule along with CPSF100 and CPSF73 form the core cleavage factor and are thought to be positioned to the cleavage site through protein-protein interactions with the CPSF and CstF bound to the PAS and DSE.. CPSF73 and CPSF100 have each been crystallized and provide evidence that CPSF73 is the actual RNA endonuclease (Mandel et al. 2006). The structures demonstrate that both of these factors are members of the metallo- $\beta$ -lactamase (MBL) family of proteins, which belong to the category of zinc-dependent hydrolases known to have a property of catalyzing  $\beta$ -lactam bonds commonly seen in antibiotic compounds (reviewed in (Callebaut et al. 2002). In addition to this fold, both CPSF subunits contain an additional  $\beta$ -CASP domain (CPSF, Artemis, SNM1, PSO), which alters the target bond specificity away from  $\beta$ -lactam bonds to the cleavage of single stranded RNA. CPSF100 is structurally very similar to CPSF73, but was found to be incapable of coordinating zinc ions at the enzyme active site thereby rendering it enzymatically inactive with no RNA endonuclease activity (Mandel et al. 2006). Two crystal structures of Symplekin exist

and they reveal that it has a series ARM/HEAT repeat domains at its N-terminus (Kennedy et al. 2009; Xiang et al. 2010). HEAT repeats evolutionarily have been considered to be domains that act as surfaces for protein-protein interactions and hence function as scaffolds (Infeld et al. 2006). A detailed account of structural details of HEAT repeats, their properties and functions with respect to Symplekin and Integrator subunit 4 are discussed in Chapter 3.

Histone mRNA 3' end processing: Histone mRNA genes are similar to snRNA genes in that they lack introns and hence are not spliced (reviewed in (Marzluff et al. 2008)). Furthermore, they don't terminate in a poly(A) tail. Owing to this simple architecture, processing of this species is fairly straightforward and requires only one endonucleolytic cleavage reaction to generate the mature transcript. In contrast to mRNA 3' end processing signals, histone mRNA genes contain a unique and highly conserved stem loop sequence in addition to a histone downstream element (HDE) (Mowry and Steitz 1987). While the elements at the 3' end of the histone pre-mRNA are distinct from their poly(A)+ mRNA counterparts, binding of machinery to the 3' end processing signals is also an essential and requisite step for cleavage of the nascent RNA. The stem loop element is recognized and bound by the Stem Loop Binding Protein (SLBP) and the U7 snRNP binds to the HDE (Wang et al. 1996). Once these two elements are recognized, the same cleavage factor used for pre-mRNA, which contains Symplekin, CPSF100, and CPSF73 gets recruited to the cleavage site and generates the mature histone mRNA (Sullivan et al. 2009). Both the poly(A) and histone pre-mRNA complexes are depicted in Figure 1.3



**Figure 1.3. 3' end formation of RNAPII transcripts.** This schematic describes that mRNA and histone mRNA share many of the components of the 3' end processing machinery (Symplekin, CPSF 73 and 100). In contrast, snRNA utilizes an entirely different set of proteins called the Integrator complex which performs the 3' end formation in this RNA species.

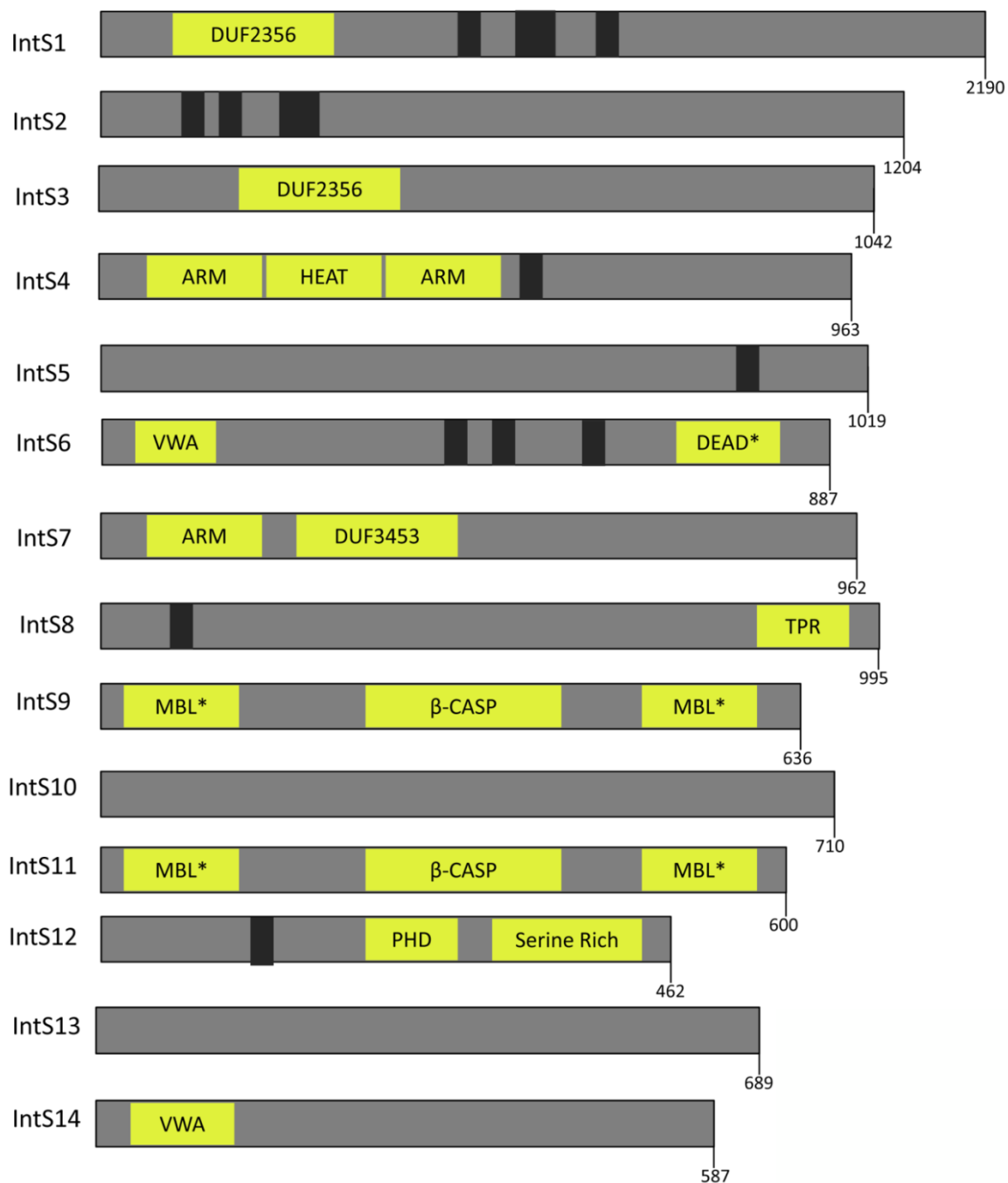


The Integrator Complex: The Sheikhattar lab in 2005 (Baillat et al. 2005), serendipitously identified and purified a 12 subunit complex that was found to be associated with CTD of RNAPII. This complex termed the “Integrator complex” was found to function as the processing machinery for RNAPII-transcribed U snRNAs. The 12 different subunits identified as part of the Integrator complex were annotated (Integrator 1-12) according to their predicted molecular weight (Figure 1.4). These proteins are absent in *Saccharomyces cerevisiae* but orthologues in other metazoans are readily detectable.

In order to gather clues about this new protein complex and possibly draw comparisons to the CPSF processing machinery, Baillat et al and other labs (Chen and Wagner 2010) analyzed secondary structural features of the 12 member complex. Detailed domain analysis (Chen and Wagner 2010) demonstrated that Integrator 6 (IntS6) had von Willebrand factor type A domain (VWA), IntS4 and 7 were revealed to have Armadillo and HEAT repeats, while IntS12 has an internal Plant Homeodomain finger (PHD). IntS9 and IntS11 both contained the MBL/ $\beta$ -CASP domains and have a high homology with CPSF 73 and 100. Work by Albrecht (Albrecht and Wagner 2012) established that IntS9/11 form a heterodimer similar to CPSF73/100. Subsequently, IntS11 which has the catalytically active zinc-dependent hydrolase activity performs the endonuclease function to cleave the snRNA at its 3' end, while IntS9 similar to CPSF73 is enzymatically inactive. Recent work from our lab has revealed the role of the PHD finger containing IntS12 in Integrator complex mediated 3' end processing of snRNA. It is now known that an N-terminal microdomain of IntS12 contributes to the overall function of IntS12 (Chen et al. 2013). Furthermore, in addition to 12 Integrator subunits, a recent RNAi screen identified two additional subunits (Asunder/IntS13, IntS14) to be part of the existing Integrator complex (Chen et al. 2012a).

### Figure 1.4. Integrator complex – the 3' end processing machinery of snRNA

This figure shows the different domains present in the 14 subunits of the recently discovered Integrator complex. The yellow regions represent domains identified by PFAM analysis and the black regions are areas of high homology.



Although most of the above mentioned Integrator subunits have domains which correspond to a particular known evolutionary function, there are a few Integrator subunits (IntS1, IntS2, IntS3, IntS5) which have domains of unknown function annotated as DUF regions. PFAM analysis performed on these subunits resulted in no correlation to any other family of existing known domains. Following the discovery of the complex, multiple studies to understand the biological and functional role of each Integrator subunit have emerged. Depletion of IntS1 in mouse embryos resulted in embryonic lethality suggesting that IntS1 is essential for survival (Hata and Nakayama 2007). IntS3 has been found to be involved in detecting DNA double stranded breaks by functioning as part of the sensor of single-stranded DNA (SOSS) heterotrimeric complex (Huang et al. 2009; Li et al. 2009; Skaar et al. 2009). Zebrafish treated with antisense morpholinos to IntS5 demonstrated red blood cell defects (Tao et al. 2009). IntS4 null flies exhibited early developmental defects where the flies did not survive beyond second instar larvae (Ezzeddine et al. 2011). Additionally, RNAi mediated IntS4 depletion in HeLa cells significantly disrupted Cajal body formation indicating that IntS4 is required for the integrity of Cajal bodies (Albrecht et al unpublished data). A similar result was also noticed for IntS11 suggesting its importance in Cajal body formation as well. IntS6 (or DICE1 – deleted in cancer 1) known to be a tumor suppressor gene is an important player and is overexpressed in prostate cancer cell lines and other non-small cell lung carcinomas; CpG hypermethylation of IntS6 promoter elements was observed to be responsible for the reduced levels of the protein (Wieland et al. 1999; Filleur et al. 2009).

### Current model of snRNA 3' end formation by the Integrator Complex.

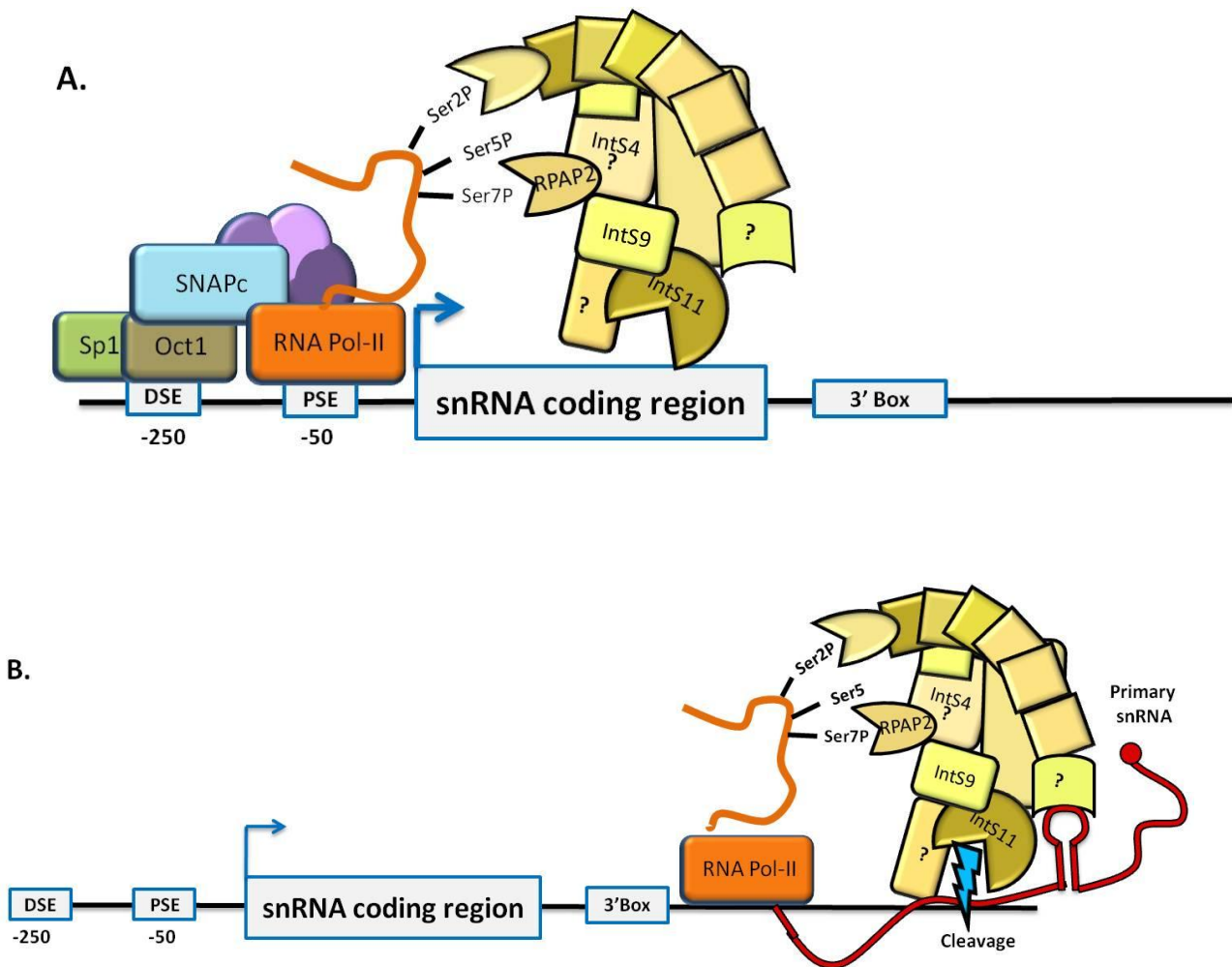
The step wise mechanism of Integrator complex mediated snRNA 3' end processing remains to be completely understood; nonetheless recent research (Albrecht and Wagner 2012; Egloff et al. 2012; Xiang et al. 2012) has helped a basic model to emerge (Figure 1.5). Central to this model, is the reversible phosphorylation pattern of the serines (2, 5 and 7) of the RNAPII CTD heptamer repeats (YSPTSPS). These events, catalyzed by important CTD kinases and phosphatases, are believed to be responsible for the recruitment of a multitude of protein factors known to play a role in the co-transcriptional 3' end processing of snRNAs (Buratowski 2009; Bartkowiak et al. 2011). Phosphorylation of Ser5 and Ser7, most likely by CDK7, dictates initiation and promoter clearance by RNAPII. According to a recent report (Egloff et al. 2012), subsequent to Ser5 phosphorylation, Ser7 phosphorylation mark is recognized by the putative Ser5 phosphatase, RNAPII associated protein-2 (RPAP2). RPAP2, a 612 amino acid protein, as the name suggests was found to be a component associated with the RNAPII machinery (Bah et al. 2012) and additionally was found to interact with snRNA genes using chromatin immunoprecipitation (Egloff et al. 2012). The deletion of the yeast homolog Rtr1 was reported to cause an accumulation of Ser5 phosphorylated RNAPII in mRNA coding regions (Mosley et al. 2009). Similarly, Egloff et al 2012 demonstrated RPAP2 to be a Ser5 phosphatase. Like Rtr1, Ssu72 is another Ser5 phosphatase that is an essential part of the yeast pre-mRNA cleavage machinery. The Manley group in 2010 (Xiang et al. 2010) showed evidence of a crystal structure of human Symplekin and Ssu72 where Ssu72 was found to bind the 6<sup>th</sup> HEAT repeat of human Symplekin.

RPAP2 does not share homology with Ssu72, but just as Ssu72 binds to Symplekin at the N-terminus, RPAP2 could bind the N-terminal HEAT repeats within IntS4.

According to the currently understood model for Integrator-mediated snRNA 3'end formation, RPAP2 facilitated phosphatase activity drives the recruitment of a subset of Integrator subunits, likely IntS4 and IntS5. Following this step, Ser2 is phosphorylated by the catalytic subunit of pTEFb complex, CDK9 kinase. During the transcription cycle, RPAP2 mediated dephosphorylation of Ser5 could be a vital step in the recruitment of the rest of the members of Integrator complex. With all the Integrator subunits being recruited onto the snRNA gene, the complex is ready to accomplish its function and IntS9 and 11 as discussed earlier cleave the 3' end of snRNAs leading to generation of the snRNA primary transcript.

The model presented here is an attempt at converging what we understand of the RNAPII CTD phosphorylation pattern and the role of Integrator complex. This model is preliminary and many facets of still need to be developed. For example, contradictory to the emerging model described here, recent structural and biochemical studies (Xiang et al. 2012) reported that RPAP2 is devoid of ser5 phosphatase activity *in vitro*. This raises questions about the role of RPAP2 in recruitment of Integrator subunits to the snRNA genes. In addition, the mechanism of promoter specificity of Integrator complex to snRNA promoter region remains to be deciphered. However, this model emphasizes the role of importance of CDK9 and other kinases and manages to elucidate the overall co-transcriptional 3'end processing event.

**Figure 1.5. Model of Integrator complex mediated snRNA 3'end formation.** A) Transcription initiation at snRNA genes by recruitment of transcription factors and recruitment of RNAPII. B) Schematic of change in phosphorylation state of Ser5 of RNAPII CTD and subsequent events that ultimately lead to IntS11 mediated 3'end cleavage of the snRNA 3'end



The Integrator Complex beyond snRNA 3' end formation. The function of the Integrator complex as the processing machinery for snRNA biogenesis is well established. However, there are many emerging roles for the Integrator complex and its individual subunits outside of snRNA biogenesis. For instance, the SOSS complex is completely distinct from the Integrator complex (Huang et al. 2009). Barring the overlap of IntS3, the subunits of these two complexes appear to be mutually exclusive. Although, we have a preliminary idea about the role IntS3 in DNA damage control, details of the mechanism and other key players involved remains to be deciphered. Additionally, a recent report (Cazalla et al. 2011) demonstrated that a primate specific herpes virus utilizes the Integrator complex to generate viral microRNAs instead of the canonical microprocessor complex. These studies clearly reveal the role of the Integrator complex far beyond just snRNA biogenesis.

#### **SIGNIFICANCE OF THIS WORK:**

In order to fully appreciate the capacity of this highly complex set of proteins, many basic questions first need to be answered. For example, what is the function of each subunit of Integrator complex and in what way does each subunit contribute to the overall function of the Integrator complex? What is the step wise mechanism and events involved in the 3' end formation of snRNA? What is the genome wide profile of genes affected by the depletion of individual Integrator subunits and snRNA 3' end misprocessing? In the context of disease state, what are the implications of Integrator complex depletion? Gaining a basic understanding in this direction will be a step further



in realizing the complete potential of the Integrator complex and currently our lab is investigating many of the aforementioned questions.

This study investigates the structure-function correlation of human Integrator subunit-4 and seeks to decipher how the structural aspects of IntS4 affect the overall snRNA 3' end processing event. The rationale for focusing on IntS4 is based upon observations made in the Wagner laboratory that RNAi-mediated depletion of IntS4 in either human or *Drosophila* cells results in the most significant amount of snRNA misprocessing. This suggests that IntS4 is playing a highly critical function to the complex. Based upon this observation and the structural similarity between IntS4 and Symplekin, **we hypothesized that Integrator subunit 4 contributes to the Integrator complex mediated snRNA 3' end processing by behaving as a Symplekin-like scaffold molecule that facilitates the interaction between other members of the complex.** To address the stated hypothesis, the two important goals of this study were to 1) identify the region of IntS4 which is important for snRNA 3' end processing and 2) determine binding partners of IntS4 which promote its function as a scaffold.

# **Chapter-2**

## **Materials and Methods**

### **Cell culture and stable cell line generation:**

HEK 293T (hereafter referred to as 293T cells) and HeLa cell lines were maintained at 37°C with 5% CO<sub>2</sub> in Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen) with 10% Fetal Bovine Serum (Phenix Research) and 1% penicillin-streptomycin (Invitrogen). 293T cells were used for stable line generation using Blastocidin selection following the transfection with a pcDNA6-FLAG vector, which contains a Blastocidin resistance-encoding gene. Into pcDNA6-FLAG, either full-length human IntS4, IntS4 deletion/point mutant constructs, or full length human IntS11 was cloned (cloning scheme described in the following sections). Selection following transfection was conducted through supplementing standard media conditions with 5 µg/ml of Blastocidin (Invitrogen). Stable integration of the various pcDNA6 plasmids was maintained using 5 ug/ml of Blastocidin. Frozen stocks of these lines were created in standard media supplemented with 10% DMSO and stored in liquid N<sub>2</sub>.

### **Generation of hIntS4 deletion mutants:**

Full-length, siRNA-resistant hIntS4 cDNA was cloned into a pcDNA3.1 3x myc vector previously generated in the Wagner laboratory. The siRNA-resistance was created through the introduction of wobble mutations using site-directed mutagenesis and this plasmid was used as a template to generate nine different truncation mutants. Briefly, the oligonucleotides described in Table 2.1 [designed in collaboration with another graduate student in the laboratory, Ms. Natoya Peart] were used for PCR amplification

of the intended fragments. These PCR amplified products were assessed for purity on a 1% agarose gel with ethidium bromide. The remaining PCR product was purified over a PCR purification column and with 1 µg of pcDNA3.1 myc vector was subjected to restriction digestion using KpnI and XbaI (Thermo Scientific) for 15 min at 37°C. After vector digestion, FastAP alkaline phosphatase (Thermo Scientific) was added to the vector digestion reaction mixture and incubated at 37°C for an additional 10 min in order to dephosphorylate the vector preventing its re-circularization. The digested products were gel purified and the concentration of the purified products was determined by Nanodrop spectrophotometer (ND-1000 series). A 10 µL ligation reaction with both the vector and PCR insert was set up at a 2:1 molar ratio of PCR product to vector and the reaction mix was allowed to ligate at room temperature for 4 hrs. Following the ligation, 5 µL was transformed into chemically competent XL-1 blue E.coli cells. As a control, a “vector only” ligation reaction was set up and used in the manner described above. Positive clones were screened by restriction digestion and confirmed by sequencing.

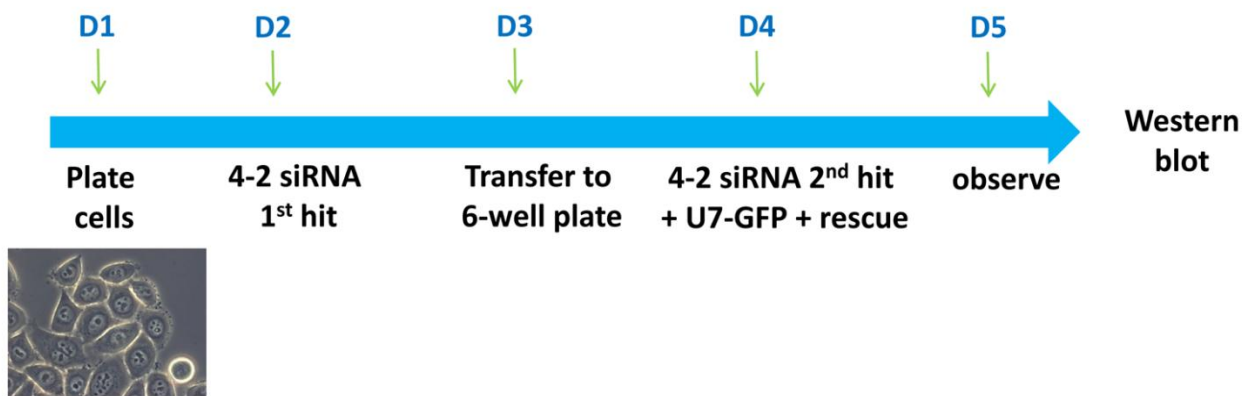
#### **hU7-GFP reporter assay, RNA interference and transfections:**

The pcDNA3 hU7-GFP reporter was made previously in the Wagner lab (Albrecht and Wagner, MCB et.al. 2012). Design and synthesis of the siRNA targeting human IntS4 ORF (4-1 and 4-2 siRNA), was conducted and purchased from Sigma Aldrich. RNA interference assays were conducted as described previously (Wagner et al, 2002 Mol Cell) and are shown pictorially in Figure 2.1. Briefly,  $8.5 \times 10^5$  HeLa cells/well were plated in a 24 well plate and 24hrs later were transiently transfected with 4-1 and 4-2

siRNA using Lipofectamine 2000 (Invitrogen) using a slightly modified version of the manufacturer's protocols. Specifically, 3  $\mu$ L/well of siRNA and 47  $\mu$ L/well of Opti-MEM medium (Life Sciences) were mixed in tube A and incubated at room temperature for 7 minutes. In tube B, 3  $\mu$ L/well of Lipofectamine 2000 and 12  $\mu$ L/well of Opti-MEM were mixed and incubated for 7 min at room temperature. The tubes were then combined and incubated further for 25 min. After this incubation, 38  $\mu$ L/well of Opti-MEM was supplemented to the transfection reaction mix to bring up the volume to a total of 103  $\mu$ L which was then added to 24-well of cells. The day after the RNAi treatment, cells were transferred without dilution into a single well of a 6-well plate containing 2 mLs total of standard 10% FBS DMEM growth media. The following day, cells were allowed to attach and were transfected again with a second hit of siRNA as described above along with 500 ng of hU7-GFP reporter plasmid in every well and 800 ng of siRNA-resistant hIntS4 full length or deletion mutant constructs. 48 hrs later after this transfection, GFP signal was visualized using fluorescence microscopy and cells were harvested for Western blot analysis to measure the levels of GFP in total cell lysates (described below).

### Figure 2.1. Schematic of U7-GFP reporter assay.

The U7-GFP reporter assay is designed to detect misprocessing of snRNA upon siRNA mediated Integrator 4 knockdown. The assay has been standardized to optimal conditions in HeLa cells and shown in this figure is the experimental design and the sequence of siRNA transfections administered to the cells during the course of the assay.



### **Western Blot Analysis:**

293T or HeLa cells were harvested 48hrs post transfection by washing with 1X PBS (137mM NaCl, 2.7mM KCl, 10mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8mM KH<sub>2</sub>PO<sub>4</sub>) followed by resuspension in 1X SDS Loading buffer (50mM Tris-HCl pH 6.8, 2%SDS, 5% glycerol, 1% β-mercaptoethanol, 100mM DTT). The samples were denatured at 95°C for 5 min and 20 µl of a lysate/SDS loading buffer mixture was resolved on a 12.5% SDS polyacrylamide gel at 150V until the dye front ran off the gel. The gel was then transferred under power and overnight at 35V onto a PVDF membrane (Millipore). This membrane was then blocked for 30 min in blocking buffer (1X PBS, 5% nonfat dry milk, 0.1%TWEEN 20) and then probed for 1 hr with primary antibody (1:5000 α-Tubulin [Abcam], 1:1000 α-myc, 1:5000 α-GFP [Clontech], 1:2000 α-HA [Covance], 1:5000 α-FLAG [Sigma] and 1:1000 α-hIntS1-12 [Bethyl labs]). After primary antibody probing, the membrane was washed 3 times with blocking buffer and then incubated in secondary antibodies conjugated to horseradish peroxidase at a dilution of 1:5000. The signal from the secondary antibody was detected using hyperfilm (Phenix Research) subsequent to a 5 minute exposure to enhanced chemiluminescence substrate (ECL).

### **Immunofluorescence:**

HeLa cells were plated at  $2.5 \times 10^5$  cells/well in 6-well plates containing poly-D-lysine coated cover slips and transiently transfected the next day with 500ng of pcDNA3-myc tagged hIntS4 plasmids using Lipofectamine 2000 transfection reagent. The cells were

allowed to express the transfected proteins for 24 hrs and then were fixed for 10 min with 4% paraformaldehyde (PFM)(Electron Microscopy Sciences). Following fixation, cells were permeabilized with 0.5% Triton X-100 in PBS for 5 min. The cells were washed thrice with copious amounts of 1X PBS and blocked for 30 min at 37°C with 10% Normal goat serum solution (NGS). Mouse anti-myc antibody was added at 1:1000 dilution in 10% NGS and incubated for 30 min at 37°C after which the cells were washed twice with 1X PBS. Fluorescently conjugated secondary anti-mouse antibody (AlexaFluor 555-Invitrogen) was added at 1:1000 and incubated for 30 min at 37°C. During and following this incubation all subsequent steps were conducted under low light and/or protection with aluminum foil. Nuclei were visualized with 4',6-diamidino-2-phenylindole (DAPI) staining using a 1:10,000 dilution for 10 min at 37°C. The cells were then mounted on microscopic slides with Fluorogel (Electron Microscopy Sciences) and stored at -20°C for long term storage.

### **Site-directed Mutagenesis:**

Site-directed mutagenesis was used to introduce point mutations into select residues in each of the eight HEAT repeats of the N-terminus of hIntS4 by utilizing the QuickChange protocol (Stratagene). Briefly, a pair of Quickchange oligonucleotides (Table 2.2) was used in a PCR reaction using 25 ng of pCDNA-FLAG hIntS4 template plasmid in a 50 µl reaction mixture followed by DpnI restriction digestion with 1 µl of DpnI FastDigest enzyme (Thermo Scientific) at 37°C for 30 min to eliminate methylated template DNA strand. Reactions lacking mutagenic oligonucleotides were run as



controls. PCR products digested in such a manner were then transformed with *E. coli* XL-1 blue competent cells, colonies were grown and DNA was isolated and ultimately all clones were screened by sequencing (Table 2.3).

### **Nuclear extract preparation:**

Large scale nuclear extracts were prepared for 293T stable lines as per the previously described protocol (Dignam et al. 1983) with minor changes. All tools and buffers used for this protocol were maintained at 4°C and the entire procedure was performed on ice. The cells were scraped off Greiner 145mm x 20mm culture dishes and washed twice with cold 1X PBS and resuspended in 5 times packed cell volume of hypotonic Buffer-A (1.5mM MgCl<sub>2</sub>, 10mM KCl, 10mM HEPES pH 7.9, 0.5mM phenylmethylsulfonyl fluoride [PMSF], 0.5mM DTT) for 20 min to allow for swelling. Swelling was monitored using bright field microscopy every 5-10 min. After sufficient swelling, cells were lysed using a glass dounce for 20 strokes. Cell lysis was also confirmed using brightfield microscopy. The dounced solution was then centrifuged at 2000 rpm for 10 min to collect the nuclear pellet which was then resuspended in ½ volume Buffer C (20mM Tris pH 7.9, 420mM KCl, 0.2mM EDTA, 1.5mM MgCl<sub>2</sub>, 0.5mM DTT, 0.5mM PMSF) and the nuclear content was extracted by incubating for 1 hr with rotation. Post buffer-C treatment, the solution was centrifuged for 30 min at 4°C, 14,000 rpm and the final nuclear extract supernatant was collected and dialyzed overnight in Buffer-D (20mM HEPES pH 7.9, 20% glycerol, 100mM KCl, 0.2mM EDTA, 0.5mM PMSF, 0.5mM DTT). Protein concentration in the

dialyzed nuclear extracts was assessed using a Bradford assay reagent (Biorad) and then were flash frozen on dry ice and stored at -80°C for future use.

### **Immunoprecipitation:**

Nuclear extracts from stable lines of FLAG tagged hInts4 full length, N and C-terminus fragments, and point mutants were subjected to immunoprecipitation studies with anti-FLAG antibody conjugated agarose beads (anti-FLAG M2 affinity gel, Sigma). Briefly, 125 µg nuclear extract was incubated with 20 µl of anti-FLAG agarose beads in Buffer-D with rotation at 4°C for 4 hrs. Buffer-D used for this purpose was devoid of DTT to prevent damage to the beads. Next, the protein bound beads were washed thrice with 1X Tris Buffered Saline (TBS-50mM Tris HCl pH 7.4, 150mM NaCl) after which 1X SDS loading buffer was added and the beads were denatured at 95°C to prepare them for loading on SDS-PAGE gel. For the control beads alone sample, the same procedure was followed for 20 µl of Protein A/G beads (Santa Cruz, Biotechnology Inc.) pre-incubated with 2 µg antigen purified antibody (Mouse IgG) in Buffer-D for 1 hr at 4°C. The presence of IP or co-IP of the proteins of interest was then detected by Western blot analysis as described above.

**Table 2.1. Primers for generating deletion mutants from siRNA resistant full length hIntS4**

<b>Primer set</b>	<b>Primer sequence 5'-3'</b>
<b>N-term truncation primers</b>	
WO389-R-XbaI	GGCCTCTAGATTAGCGCCGTGCAGGTTTGGGC
NO11-F-KpnI ( $\Delta$ N1)	GGCCGGTACCAAAGATTATAGGGGATTACTTC
NO12-F-KpnI ( $\Delta$ N2)	GGCCGGTACCACATTTCTTGGAGCAGACCCTT
NO13-F-KpnI ( $\Delta$ N3)	GGCCGGTACCACATGGGTTGGAAGATGAGATG
NO15-F-KpnI ( $\Delta$ N4)	GGCCGGTACCAGAGCTGCTGGAATTCACCATC
<b>C-term truncation primers</b>	
WO391-F-KpnI	GGCCGGTACCGATGGCGGCGCACCTTAAGAAG
NO6-R-XbaI ( $\Delta$ C1)	GGCCTCTAGATTAAAGGAGTTTGTCCACAAAGCT
NO7-R-XbaI ( $\Delta$ C2)	GGCCTCTAGATTACAATGCTGGCATTGTTGGACA
NO8-R-XbaI ( $\Delta$ C3)	GGCCTCTAGATTAGAGTTCATGAAGAGCCTCTCG
NO9-R-XbaI ( $\Delta$ C4)	GGCCTCTAGATTAATACATCTCATCTTCCAACCC
NO10-R-XbaI( $\Delta$ C5)	GGCCTCTAGATTACTCCATAGAGCCCAACAGTTT

**Table 2.2. Primers for quickchange site-directed mutagenesis for introducing point mutations in the N-terminus of hIntS4**

<b>Primer sets</b>	<b>Primer sequence 5'-3'</b>
<b>Helix1</b> Int4-hcr1a- F Int4-hcr1a- R Int4-hcr1b- F Int4-hcr1b- R	CCATCTGTGAGACTGGCAATTGCATCAGCGTTGGGTTTATTATCAAAG CTTTGATAATAAACCCAACGCTGATGCAATTGCCAGTCTCACAGATGG ATTGCATCAGCGTTGGGTGCATTATCAAAGACAGCAGCA TCCTGCTGTCTTTGATAATGCACCCAACGCTGATGCAAT
<b>helix1A-SET-1</b> Int4-hcc1A1- F Int4-hcc1A1- R	AACATCCTGCAGAATGCAAAGTCTCATGCAGTCCTAGCTCAACTG CAGTTGAGCTAGGACTGCATGAGACTTTGCATTCTGCAGGATGTT
<b>helix1A-SET-2</b> Int4-hcc1A2- F Int4-hcc1A2- R	GTCCTAGCTCAACTGCTGGCTACTTTGCTTGCAATTGGC GCCAATTGCAAGCAAAGTAGCCAGCAGTTGAGCTAGGAC
<b>helix1A-SET-3</b> Int4-hcc1A3- F Int4-hcc1A3- R	GCTCAACTGCTGGATACTGCGCTTGCAATTGGCACTAAG CTTAGTGCCAATTGCAAGCGCAGTATCCAGCAGTTGAGC
<b>helix2</b> Int4-hcg2- F Int4-hcg2- R	AGAAATAAGTGCCTGGCGTTACTTGGCGCTCTTGGCTCTTTGGAG CTCCAAAGAGCCAAGAGCGCCAAGTAACGCCAGGCACTTATTTCT
<b>helix3</b> Int4-hcb3a- F Int4-hcb3a- R Int4-hcb3b- F Int4-hcb3b- R	GTCAGAACAGCAGCTATAGCAGCCATGTTGCAGCTCCAT ATGGAGCTGCAACATGGCTGCTATAGCTGCTGTTCTGAC GCCATGTTGCAGCTCCATGCAAGAGGACTGAAATTACAC GTGTAATTTCAAGTCCTCTTGCATGGAGCTGCAACATGGC
<b>helix4</b> Int4-hcg4a- F Int4-hcg4a- R Int4-hcg4b- F Int4-hcg4b- R	CGCAGTGCTGCAGTCGCGCTTATCGCGGTCGTGAGTCAGCTC GAGCTGACTGACGACCGCGATAAGCGCGACTGCAGCACTGCG CTTATCGCGGTCGTGAGTCGCGCTCTATCCTGAAAGCATT AATGCTTTCAGGATAGAGCGCACTGACGACCGCGATAAG
<b>helix5</b> Int4-hco5- F Int4-hco5- R	AGTGATGGCTCTTGGGCGGTTCTGTGTTGCGGCAGCAAACTGTTG CAACAGTTTTGCTGCCGCAACACGAACCGCCCAAGAGCCATCACT
<b>helix5A SET-1</b> Int4-hcc5A1- F Int4-hcc5A1- R	CGTACTGCACATGAGGCTGCCAAGGAAGCTTACAGTTCGGGGGAG CTCCCCCGAACTGTAAGCTTCCTTGGCAGCCTCATGTGCAGTACG

<b>helix5A SET-2</b> Int4-hcc5A2- F Int4-hcc5A2- R	GCCAAGGAACTTTACAGTGCGGGGGCGTTTTCCAGTGGCAGAAAG CTTTCTGCCACTGGAAAA CGCCCCCGCACTGTAAAGTTCCTTGGC
<b>helix6</b> Int4-hcw6- F Int4-hcw6- R	TTGGAAGATGAGATGTATGCGGTTCGTATTGCTGCTGTG CACAGCAGCAATAGCAACCGCATACATCTCATCTTCCAA
<b>helix7</b> Int4-hcy7a- F Int4-hcy7a- R Int4-hcy7b- F Int4-hcy7b- R	TTCAACGATGAAATTGAGGCAGTACGTCTGCAGTCTATA TATAGACTGCAGACGTACTGCCTCAATTTTCATCGTTGAA GTACGTCTGCAGTCTATAGCTACCATGGCAAAAATCTCTAACACATC GATGTTGTTAGAGATTTTTGCCATGGTAGCTATAGACTGCAGACGTAC
<b>helix8</b> Int4-hcp8a- F Int4-hcp8a- R Int4-hcp8b- F Int4-hcp8b- R	CTAGAGGATTCATCCAGAGCTATTCGAGCGGCTCTTCATGAACTCTTA TAAGAGTTCATGAAGAGCCGCTCGAATAGCTCTGGATGAATCCTCTAG ATTCGAGCGGCTCTTCATGCACTCTTATGCTGTACTAAT ATTAGTACAGCATAAGAGTGCATGAAGAGCCGCTCGAAT

**Table 2.3. Sequencing oligonucleotides for the quickchange mutations in N-terminus of hIntS4**

Oligo name	Oligo sequence
seq-ANU-A (for helices 1,1A1,1A2, 1A3,2)	GCAATACTTGCTCCAGTTTGCC
seq-ANU-B (for helices 3,4,5,5A1,5A2)	GGAGAAAAGTGTCACAAAAGATG
seq-ANU-C (for helices 6,7,8)	GGCAGAAAGTGGGGAGATGAT

**Table 2.4. Primers for generating pET49b IntS4 clones for recombinant protein expression in E.coli**

Primer set	Primer sequence 5'-3'
<b>Full-length IntS4</b>	
ANU34-FL-delC-fwd-NotI-ATG	GGCCGCGGCCGCGAGCGGCGCACCTTAAGAAGCGG
ANU33-delN-rev-Xho	GGCCCTCGAGTTAGCGCCGTGCAGGTTTGGGCATTAT
<b>N-term truncation (delN)</b>	
ANU32-delN-fwd-NotI	GGCCGCGGCCGCAACCAAGTACCCTACTGATAGG
<b>C-term truncation (delC)</b>	
ANU31-delC-rev-Xho	GGCCCTCGAGTTACAGCTCCACCAATGCAAGATGAAT

# **Chapter-3**

## **Results and Discussion**



## Introduction:

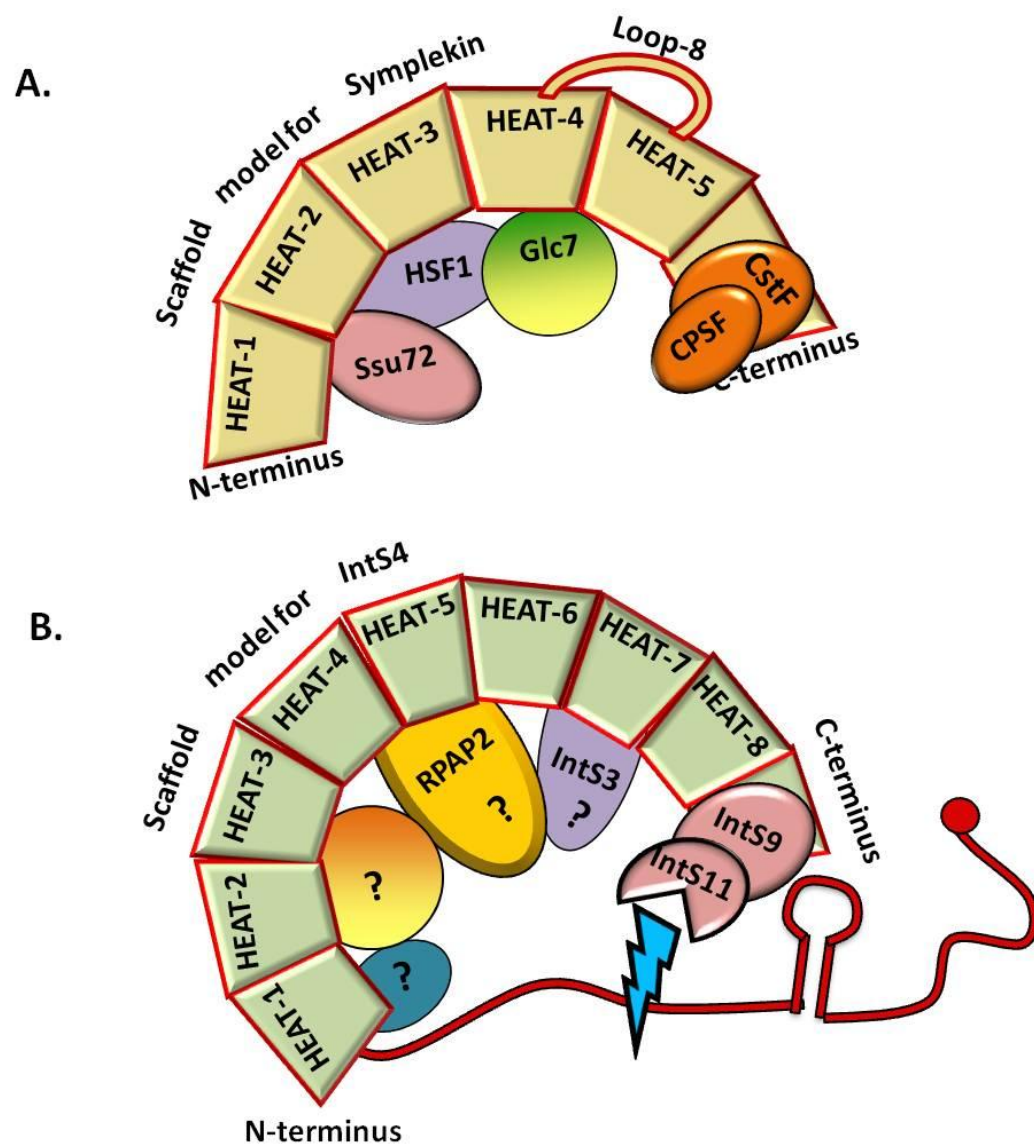
Previous work from our lab (Ezzeddine et al. 2011) has identified that RNAi-mediated IntS4 depletion in *Drosophila* S2 cells results in significant increase in snRNA misprocessing as detected by the levels of accumulated misprocessed snRNAs tested by RT-PCR. The level of snRNA misprocessing after IntS4 depletion was found to be highest among all the other Integrator subunits. In addition, Mr. Todd Albrecht has performed analogous experiments in human cells using siRNA and also found that human IntS4, when depleted, leads to most dramatic levels of misprocessed snRNA (T.A. Albrecht, personal communication). Thus, with the reagents available to us, the observation of IntS4 being the most critical for snRNA 3' end processing is conserved in humans and flies. Furthermore, we have reported that the loss of IntS4 expression results in larval lethality in flies (Ezzeddine et al. 2011). These IntS4 null flies exhibited early developmental defects which compromised their survival beyond second instar larvae. Taken together, these findings all support a model where IntS4 is critical to the snRNA processing machinery and therefore understanding its role is imperative.

Symplekin is a 1160 residue scaffold molecule known to be part of the macromolecular processing machinery that is responsible for 3' end processing of histone mRNAs and canonical pre-mRNAs (Takagaki and Manley 2000; Hofmann et al. 2002; Xing et al. 2004; Kolev and Steitz 2005; Sullivan et al. 2009; Ruepp et al. 2011). Kennedy et al in 2009, crystallized the structure of the HEAT repeat domain of *Drosophila melanogaster* Symplekin (Kennedy et al. 2009). Their work provided evidence for the presence of five canonical HEAT repeats along with an extended loop (loop8). This loop structure was found to dampen movement in the HEAT domain and is

thought to be responsible for providing a neutral surface for protein-protein interactions. The five HEAT repeat motifs form a hydrophobic concave core which contributes to the scaffold function of Symplekin (Kennedy et al. 2009). Symplekin is thought to associate with other mRNA 3' end processing factors through a scaffold that exists at the N-terminus using the aforementioned HEAT repeats but also using a scaffold at its C-terminus with a yet-to-be identified domain(s). The N-terminal HEAT repeats have been shown through crystallization to interact with the serine 5 phosphatase SSu72 as well as HSF1 and Glc7, which are two proteins also found to be involved in 3' end formation. The HEAT repeat region of Pta1 (yeast homologue of Symplekin) was reported to bind Ssu72 and Glc7, which are both known to be involved in 3' end processing (Ghazy et al. 2009). In addition, the HEAT repeat domain of mouse Symplekin was demonstrated to bind to the transcription factor, HSF1. Symplekin via its interaction with HSF1 is thought to bridge transcription regulation and 3' end processing in pre-mRNAs (Xing et al. 2004). Furthermore, The C-terminal domain of Symplekin is thought to associate with CPSF100 and CPSF73 to facilitate in recruiting them to the cleavage site of histone and poly(A)+ pre-mRNA (Kennedy et al. 2009; Sullivan et al. 2009).

Integrator subunit 4 and Symplekin are not considered homologs by any database; however, due to the presence of N-terminal HEAT repeats and the critical role of both these proteins in the context of RNA 3' end processing, we believe that they are likely to have structural and functional similarities. The model that I am testing here is that IntS4 behaves analogously to Symplekin using its N-terminal HEAT repeats to bind RPAP2 and possibly other Integrator subunits and uses its C-terminus to bind to IntS9/11 to position them for cleavage of the pre-snRNA. This model is depicted in Figure 3.1

**Figure 3.1. Model of IntS4 - Structural similarities with Symplekin.** A. Crystal structure of the HEAT domain of Symplekin (Kennedy et al 2009) displaying five canonical HEAT repeats with a loop 8. According to this model the hydrophobic concave core formed by the repeats provides a surface for proteins to bind and interact. B. A hypothesized model of the N-terminal HEAT repeat region of Integrator subunit 4 with similar binding characteristics depicted.



*In silico* analysis of IntS4. There have been no crystal structures reported for any of the Integrator subunits published to date therefore we have only limited access to proven structural information for IntS4. Therefore, as a first step towards understanding the role of IntS4 in snRNA 3' end formation we utilized four computer-assisted tools to glean information about IntS4's structure: CLUSTALW alignment to identify conserve regions (using Vector NTI software, Vector NTI Advance 11), PFAM analysis (<http://pfam.sanger.ac.uk/>) to identify predicted domains based upon primary amino acid sequence, PONDR analysis (<http://www.pondr.com/index>) to predict regions of order/disorder, and Swiss-Model (<http://swissmodel.expasy.org/>) to generate theoretical structures of IntS4. These three tools will aid in the rationale for the structure/function analysis that is described afterwards.

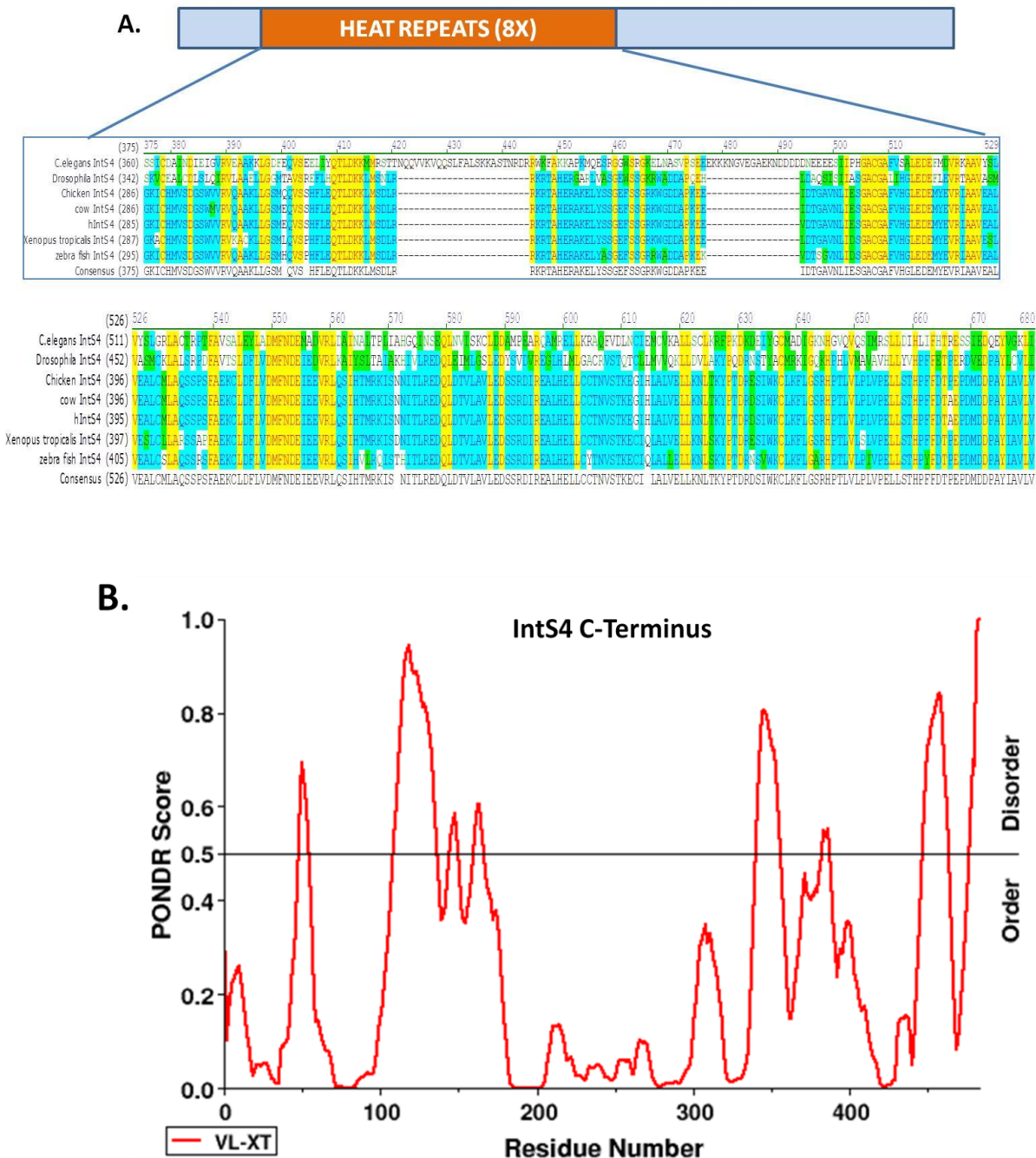
Full-length hIntS4 is 963 residues in length and has an N-terminus region with 8 predicted HEAT repeats using PFAM analysis. Further, CLUSTALW alignment of Integrator 4 amino acid sequences from several diverse species identified that the greatest amount of conservation lies within the N-terminus (Figure 3.2A). In contrast, such structural information about the C-terminus is limited. The C-terminus of hIntS4 is highly non-conserved and does not exhibit any homology with any known proteins. PFAM analysis for domain identification has revealed the presence of no known domains in this region thereby greatly restricting any structural correlations and evidence which would better our understanding about this region of IntS4.

We reasoned that lack of any significant domains and homology is perhaps due to the presence of natural disorder in the protein. To investigate this rationale, the C-terminus region was subjected to PONDR analysis. PONDR stands for Prediction Of

Naturally Disordered Regions and, as the name suggests, this method of analysis helps identify regions of disorder in proteins. Disordered proteins in their native states are mostly devoid of a definitive 3D structure; nevertheless they are known to undergo disorder-order transitions upon ligand binding (Hewig et al. 2007). Figure 3.2B shows a representation of the PONDR output where, propensity of disorder is predicted for each residue and plotted as a PONDR score. As seen, any residue that scores  $> 0.5$  is considered to have significant disorder. The score and threshold value of 0.5 was validated based upon a combination of prediction methods such as hydropathy index, VL-XT, XL-1 and CaN. Of these, the VL-XT method predicts disorder in a stretch of 14 residues at a time and provides a score (Trippe et al. 2006). From this analysis, 40% of the C-terminus was found to be disordered and the rest of the residues were found to show ordered behavior. Collectively, the combination of PFAM analysis, conservation at the N-terminus, and PONDR analysis suggests that IntS4 may adopt a defined structure at its N-terminus through its eight predicted HEAT repeats and possesses some level of inherent disorder/flexibility at its C-terminus.

**Figure 3.2 Sequence alignment and PONDR analysis of Integrator subunit 4. (A)**

Sequence alignment of the HEAT repeat region of IntS4 demonstrates evolutionary conservation. (B) PONDR analysis of the C-terminus region of hIntS4 shows presence of disorder in the region.



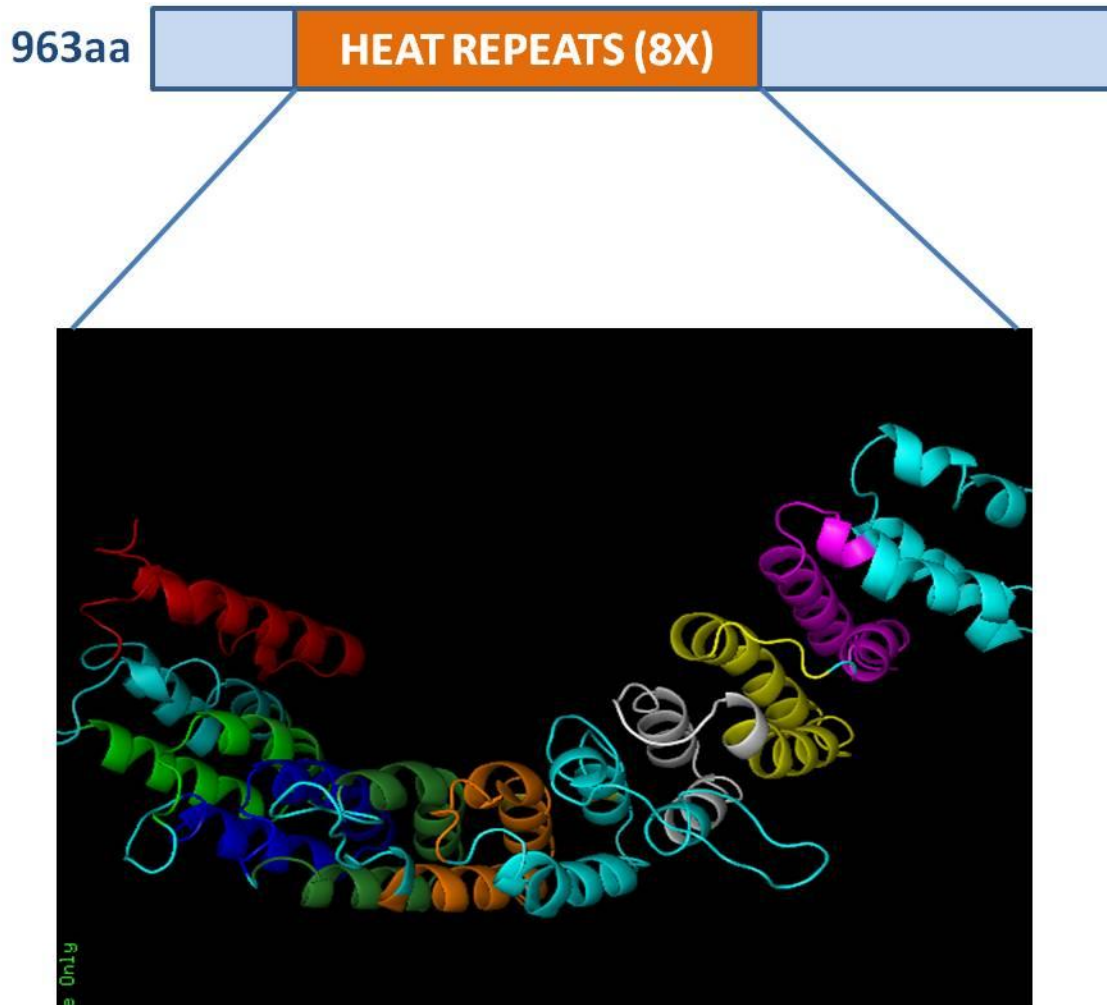
The Swiss-model website utilizes a proprietary software application that uses homologous and solved structures as a template to generate a novel/unknown structure, which in this case is human IntS4. Initial attempts to model the full-length IntS4 were unsuccessful and this was very likely due to the lack of any predicted motifs within the C-terminus. Therefore, we tried to model the N-terminal half of IntS4 only. Because IntS4 lacks significant homology with other proteins that have been crystalized, multiple sequence alignments with other known HEAT repeat proteins (PP2A,  $\beta$ 1-importin,  $\beta$ -catenin) were therefore employed to predict the N-terminal IntS4 structure. It should be noted that Swiss-model only generates predicted structures if they have met minimal criteria of thermodynamic stability. The generated model (Figure 3.3) predicts the presence of ten HEAT repeats within the N-terminus and each repeat is represented with a different color. This predicted structure is in good agreement with PFAM prediction of IntS4 having eight HEAT repeats, but it surprisingly also predicted two additional repeats. These two additional repeats exist despite high thermodynamic restrictions instituted by the Swiss-model program and were likely not predicted by PFAM given that they deviate slightly from the consensus HEAT repeat amino acid sequence.(Shukla and Trippett 2006). In light of this, I hereafter refer to these additional two repeats as “pseudo HEAT repeats”.

HEAT repeats were first discovered in a distinct set of proteins devoid of any functional similarity which included Huntington, Elongation factor 3, the PR65/A subunit of protein phosphatase 2A (PP2A) and the TOR protein (target of rapamycin) and hence the name HEAT repeat. HEAT and ARM (Armadillo) repeats are motif repeats which are functionally known to provide a large scaffolding surface for protein-protein interactions

(Matuszewski et al. 2006). A single HEAT repeat is 30-40 residues in length and is formed when two helices fold in a way to form a helical hairpin like structure. This is readily apparent in the predicted model of IntS4 presented in figure 3.3. The ability to stack up with neighboring repeats and form a continuous concave hydrophobic core is one of the unique characteristics of this domain family which explains its appearance as an elongated super-helix (Casanova and Trippe 2006). This hydrophobic core formed by HEAT repeat domains in transport protein families such as  $\beta$ -importins and karyopherins has been documented to provide a ligand binding surface and is thought to be responsible for the ability of transport proteins to efficiently bind multiple proteins simultaneously (Casanova et al. 2007). Evolutionarily, HEAT repeats have been found in diverse proteins with no necessary functional correlation, yet all these proteins perform the basic function of scaffolding and mediating multiple protein-protein interaction. Given that IntS4 has potentially 10 HEAT repeats, there is a significant surface that can accommodate multiple interactions with other Integrator subunits as well as RPAP2.



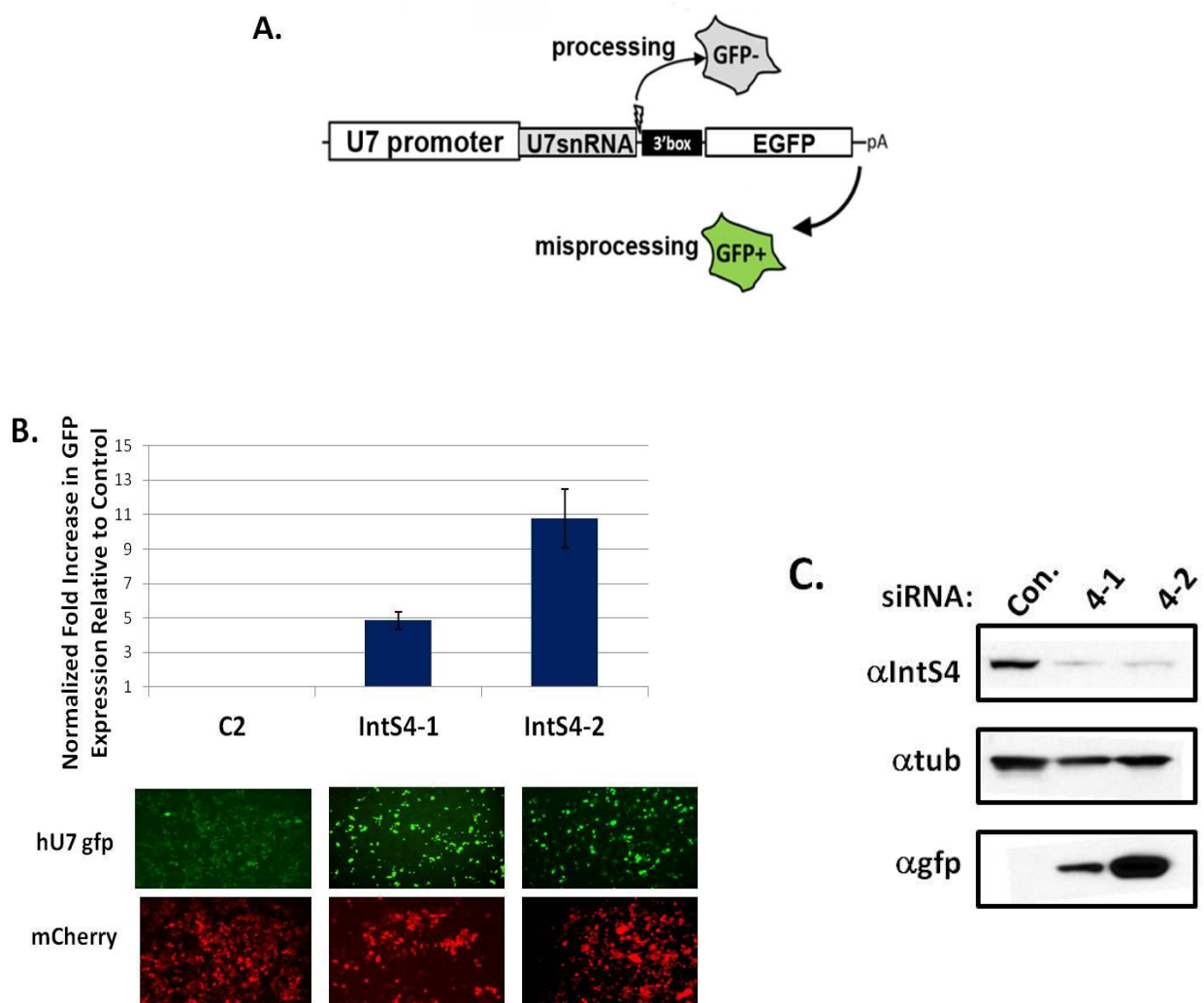
**Figure 3.3. Model prediction for HEAT repeat region of hIntS4.** Swiss-model based IntS4 HEAT repeat structure predicts eight HEAT repeat and two pseudo HEAT repeats (colored in cyan).



Human IntS4 is required for snRNA 3' end formation in HeLa cells. Previously, the Wagner laboratory has developed a fluorescence-based reporter system capable of monitoring snRNA 3' end formation (Ezzeddine et al. 2011; Albrecht and Wagner 2012; Chen et al. 2012b; Chen et al. 2013). This transcriptional readthrough GFP reporter assay has the ability to monitor *in vivo* U7 snRNA misprocessing through a gain in GFP expression readout (Figure 3.4A). This reporter was adapted for human cell use subsequently used for functional analysis of IntS9/11 in HeLa cells (Albrecht and Wagner 2012).

To deplete IntS4 from HeLa cells, hIntS4 was targeted using sequence specific siRNAs (termed 4-1 and 4-2). The siRNA treatment was then followed by transfection with the U7-GFP reporter and cells were subsequently analyzed two days later for evidence of readthrough as a surrogate readout for snRNA misprocessing. Both a fluorescence plate reader and a Western blot analysis were used to quantify the GFP signal. A plasmid encoding mCherry was co-transfected and served as a transfection control. Using this approach, we observed that compared to control siRNA treatment in HeLa cells, both 4-1 and 4-2 siRNAs targeting IntS4 generated significantly greater intensity (5- and 10-fold respectively) of GFP signal. The 4-2 siRNA was more effective in generating misprocessing as observed by both the plate reader data and Western blot analysis and therefore was chosen for all further experiments involving IntS4 RNAi depletion. Overall RNAi mediated endogenous IntS4 knockdown resulted in significant misprocessing in HeLa cells which is in consensus to similar results previously observed in a *Drosophila* S2 cell line (Ezzeddine et al. 2011).

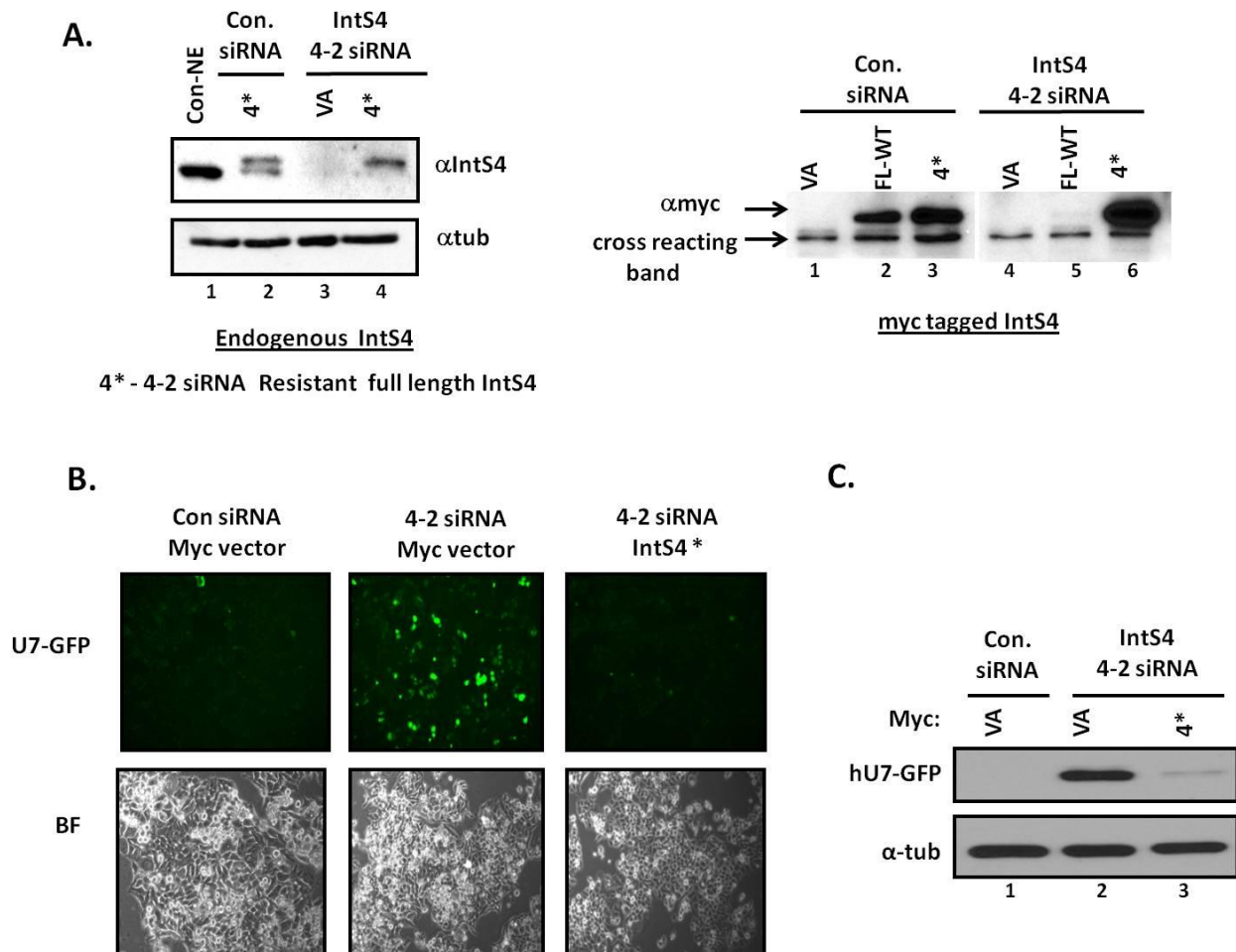
**Figure 3.4. IntS4 depletion results in significant snRNA misprocessing in HeLa cells.** (A) Schematic showing design of the hU7-GFP reporter gene which is used for the transcriptional readthrough GFP based reporter assay to detect misprocessing in snRNA (B) Graphical representation of quantified GFP signal obtained from the GFP reporter assay in HeLa cells. Also shows fluorescence images of siRNA treated HeLa cells where mCherry was used for normalization of cell count (C) Western blot of HeLa cell lysates from the above assay using anti-GFP, anti-Tubulin or anti-hIntS4 antibodies.



Development of siRNA resistant hIntS4 to study the role of IntS4 in snRNA 3' end processing: The data shown in Figure 3.4 demonstrate that the U7-GFP can readily measure significant and reproducible amounts of snRNA misprocessing in response to knockdown of IntS4 in HeLa cells. In order to perform a structure/function analysis of IntS4, however, an RNAi-resistant IntS4 cDNA is necessary to allow for re-transfection into depleted cells in order to rescue the misprocessing phenotype (i.e. GFP expression). The 4-2 siRNA targets nucleotides 1685–1706 of the human IntS4 so point mutations were introduced by site-directed mutagenesis in select nucleotides of this sequence to obtain a 4-2 siRNA resistant full-length IntS4 construct (IntS4\*). The introduced point mutations were designed in such a way to be translationally silent so the IntS4\* cDNA encodes the same IntS4 protein as the endogenous IntS4 mRNA. Further, the RNAi-resistant cDNA was cloned to also contain three N-terminal myc tags to facilitate further analysis using Western blotting or immunofluorescence.

To test the functionality of the newly generated construct thus generated, HeLa cells were transfected with the mutant construct and subjected to the previously described GFP-reporter assay after siRNA transfection and GFP signal was assessed. This assay is based on the rationale that despite siRNA mediated depletion of endogenous IntS4 levels supplementation with the siRNA resistant mutant construct is expected to rescue the phenotype. As expected, the cells with IntS4 depletion and vector alone treatment display strong GFP signal whereas the mutant construct supplemented cells exhibit GFP signal in similar intensity to the control (Figure 3.5). We **I** therefore conclude that the misprocessing of U7-GFP reporter can be fully rescued through expression of the RNAi-resistant IntS4 cDNA.

**Figure 3.5. siRNA resistant hIntS4 completely rescues the misprocessing phenotype in snRNA.** The region of hIntS4 that base pairs to 4-2 siRNA was identified and specific nucleotides were mutated to generate a 4-2 siRNA resistant hIntS4 construct. (A) Western blot cell lysates from IntS4 depleted and non-depleted HeLa cells transfected with myc-tagged IntS4\* plasmid using anti-myc and anti-hIntS4 antibodies. Lane 1 (left panel) indicates HeLa nuclear extract used as a control (B) HeLa cells treated with 4-2 siRNA were subsequently transiently transfected with hU7-GFP reporter construct along with IntS4\* or empty vector. Bright field and fluorescence images are represented. (C) Western blot of cell lysates collected from the above mentioned reporter assay probed for anti-GFP. Tubulin was used as loading control.



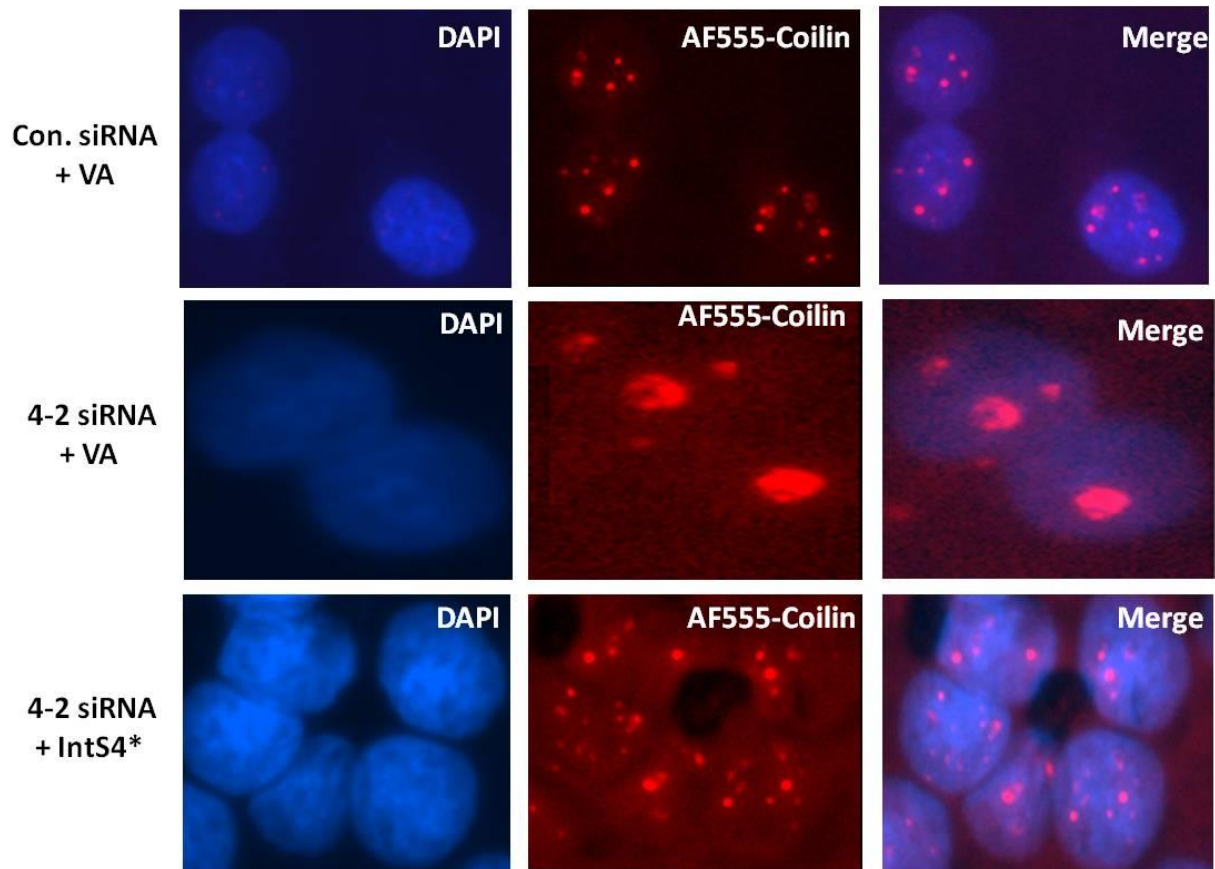
IntS4 and Cajal body structural integrity: Having established that IntS4 is essential for snRNA 3' end processing using the U7-GFP reporter assay and now generating an RNAi-resistance rescue strategy to restore processing, we wanted to also examine the effects of IntS4 depletion on snRNA related events and processes. An example of such a process is the formation and maintenance of nuclear structures that have high levels of snRNAs called Cajal bodies (CBs). snRNAs localize transiently to these structures for their final steps of their maturation process which includes snRNA base modification and snRNA-protein (snRNP) assembly ((Dundr et al. 2004)). Cajal bodies are characterized by the presence of Coilin, a protein that is widely accepted as a marker, although its cellular function within the Cajal body remains unknown ((Liu et al. 2009)).

To investigate the effect of IntS4 depletion on Cajal body formation and structure, control or IntS4 specific siRNAs were transfected into HeLa cells plated on poly-D-lysine coated cover slips followed by supplementation with either empty vector (myc-tagged pcDNA3.1) or a siRNA resistant myc-tagged full-length hIntS4 construct (described in previous section). These cells were not transfected with the U7-GFP reporter to simplify cell biological analysis. Rather, the transfected cells were fixed and probed with DAPI and Alexafluor555-tagged fluorescent secondary antibody specific to Coilin prior to imaging with a fluorescence microscope. Control siRNA treated cells are expected to have normal localization of Coilin in Cajal bodies owing to an intact endogenous IntS4 and Integrator complex which would imply accurate snRNA 3' end processing and maturation. As expected, control siRNA treated cells exhibited normal Coilin and Cajal body behavior (Figure 3.6). In striking contrast, IntS4 depletion supplemented with an empty vector demonstrated diffuse Coilin localization throughout the nucleus and

relocalization in large punctate structures that resemble nucleoli suggesting that an additional phenotype associated with snRNA misprocessing compromised Cajal body structure. Importantly, proper Coilin localization to the CBs can be restored after transfection of the RNAi-resistant IntS4 cDNA (Figure 3.6, bottom panel). This demonstrates that IntS4 depletion disrupts Cajal bodies and hence is important to their structural integrity. These set of results shed light on the overall functional significance of IntS4. It should be noted that during the course of these experiments, data were published from another laboratory that confirmed this result. The Shibahara lab demonstrated strikingly similar results that when IntS4 or IntS11 are knocked down in cells, CBs are redistributed into the nucleolus (Takata et al. 2012).

**Figure 3.6. IntS4 depletion results in disruption of Cajal body structural integrity.**

HeLa cells were plated on poly-D-lysine coated cover slips and treated with control or 4-2 siRNA followed by supplementation with empty vector or siRNA resistant IntS4 plasmid (IntS4\*). These were fixed and further probed with DAPI for staining nuclei, anti-coilin antibody, and Alexafluor555 fluorescent antibodies followed by imaging with fluorescence microscopy. VA in all panels indicates “vector alone” transfections.



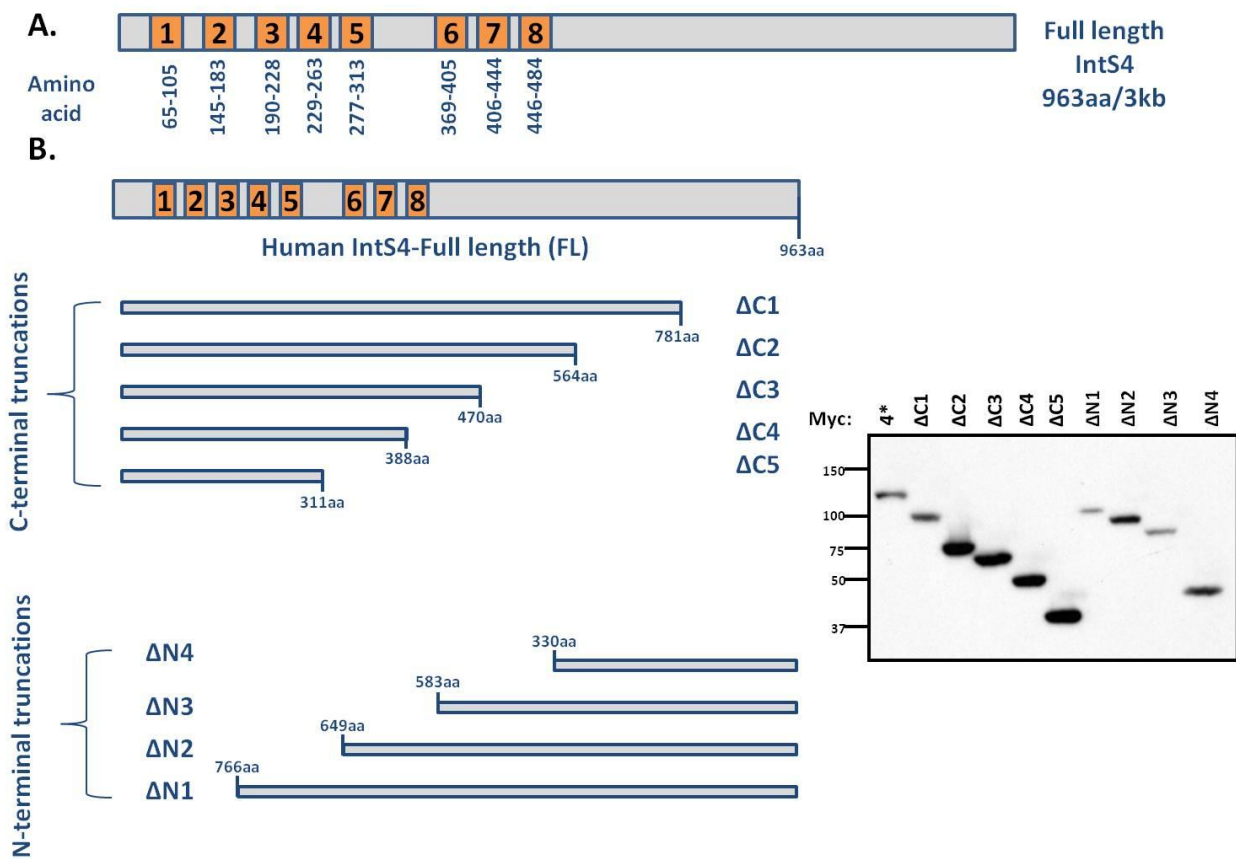


**Deletion analysis of IntS4 demonstrates that Integrator function is highly intolerant to disruptions in IntS4 structure.** Human Integrator subunit 4 is a 963aa protein with a highly conserved N-terminus consisting of 8 HEAT repeats (as predicted by PFAM) and a relatively nonconserved C-terminus (Figure 3.7A). To enable structure-function analysis of IntS4, it is important to gain an understanding of which region (N or C-terminus) of IntS4 mediates the snRNA 3' end processing. For this purpose, a series of siRNA resistant N and C-terminal hIntS4 deletion constructs were generated by truncating 100 amino acids at a time from the full length mutant IntS4\* (Figure 3.7B). Each of these deletion mutants were cloned with an N-terminal myc tag and were first transfected along with the full length IntS4\* into HeLa cells that were depleted of endogenous IntS4 to determine their expression potential. Using Western blot analysis probing for the myc tag, we observed that each of the deletion mutants was expressed but also that there was some degree of variation between their levels of expression (Figure 3.7B, figure inset).

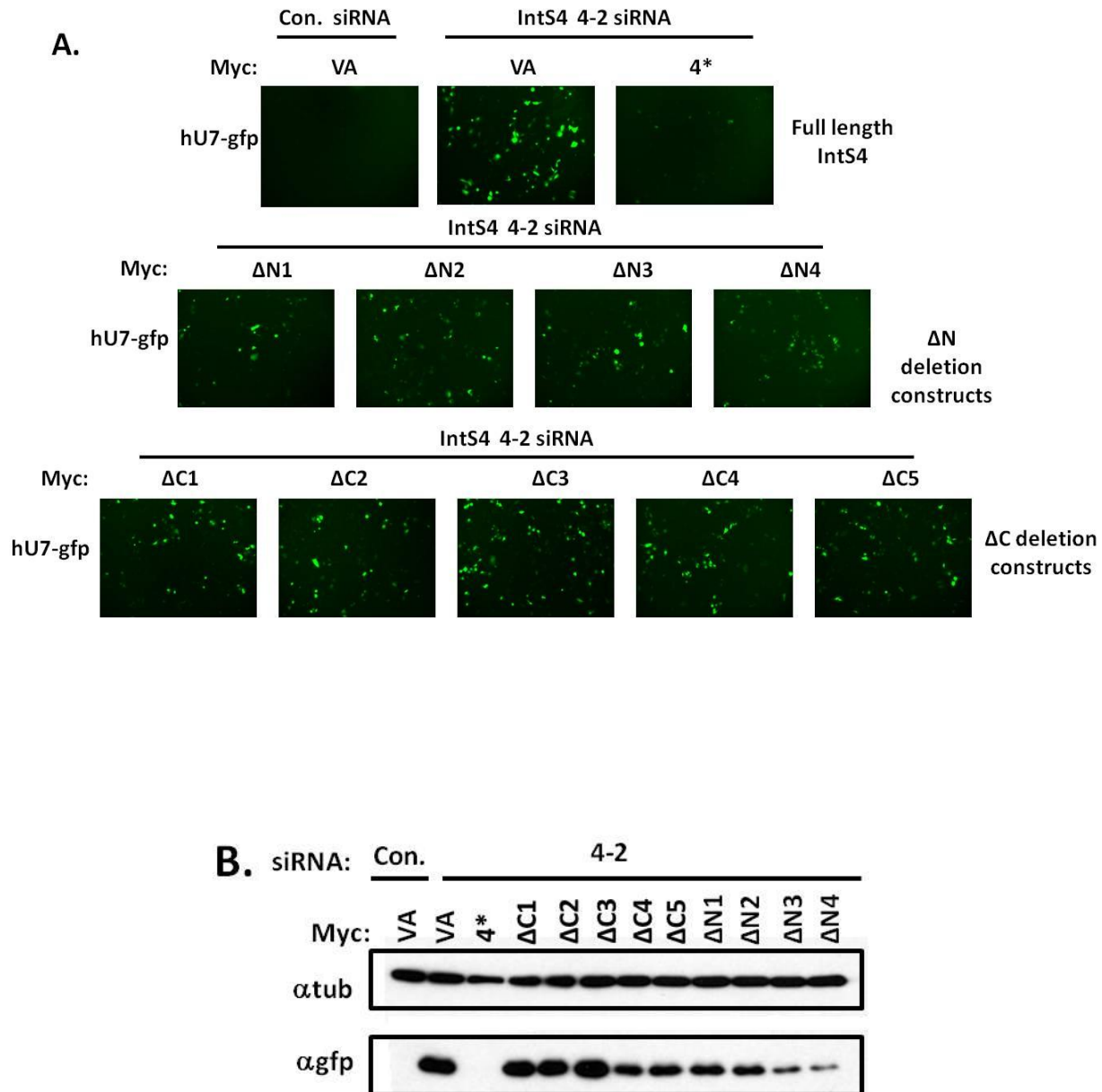
Having established that the IntS4 RNAi-resistant deletion fragments are expressed after transfection into cells depleted of IntS4, we next determined their ability to restore the U7-GFP reporter processing. The purpose was to perform an initial qualitative assessment of the essential region of IntS4 and identify truncations which displayed an ability to rescue misprocessing. We analyzed the U7-GFP reporter both using fluorescence microscopy and by Western blotting using anti-GFP antibodies (Figure 3.8). While the full-length IntS4\* was fully capable of restoring the processing of the U7-GFP reporter by virtue of a loss of GFP expression, none of the fragments were capable of rescue. This result is unlikely due to some design flaw in the approach as a recent

**Figure 3.7. Generation of siRNA resistant deletion constructs for hIntS4. (A)**

Schematic of the full-length hIntS4 displaying 8 HEAT repeats in the highly conserved N-terminus and the non-conserved C-terminus (B) (left panel) Schematic of hIntS4 full-length along with the N and C-terminal truncations (Right panel) Western blot of cell lysates from HeLa cells transfected with 500ng of truncation construct and probed with an anti-myc antibody.



**Figure 3.8. Deletion analysis of full-length hIntS4 reveals a possible scaffold-like property of hIntS4.** (A) Fluorescence images from GFP-reporter assay using siRNA resistant full-length and truncation constructs of hIntS4 (B) western blots of cell lysates obtained from the reporter assay probed for anti-GFP antibody. Tubulin was used as loading control.

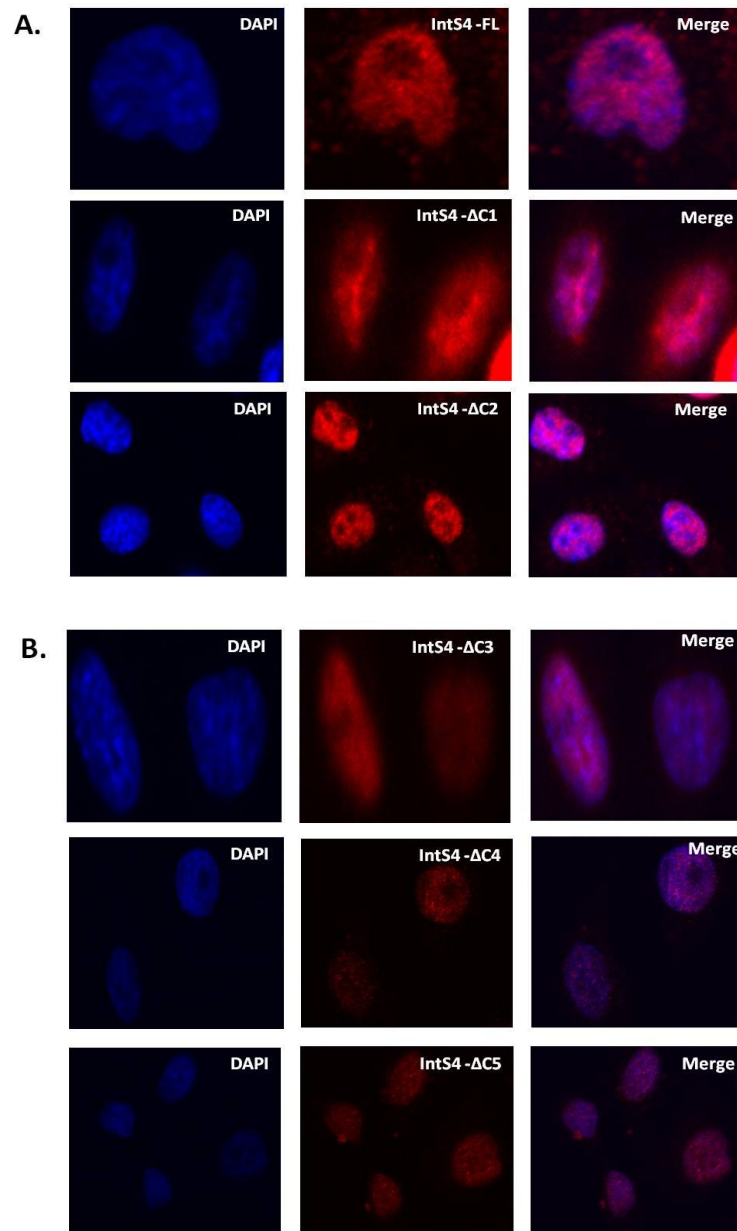


publication from our laboratory demonstrated that a small region of IntS12 is sufficient to restore U7-GFP processing but rather supports a conclusion that IntS4 is intolerant to deletion. This conclusion supports, although not proves, a model where IntS4 is acting as a scaffold for other Integrator subunits.

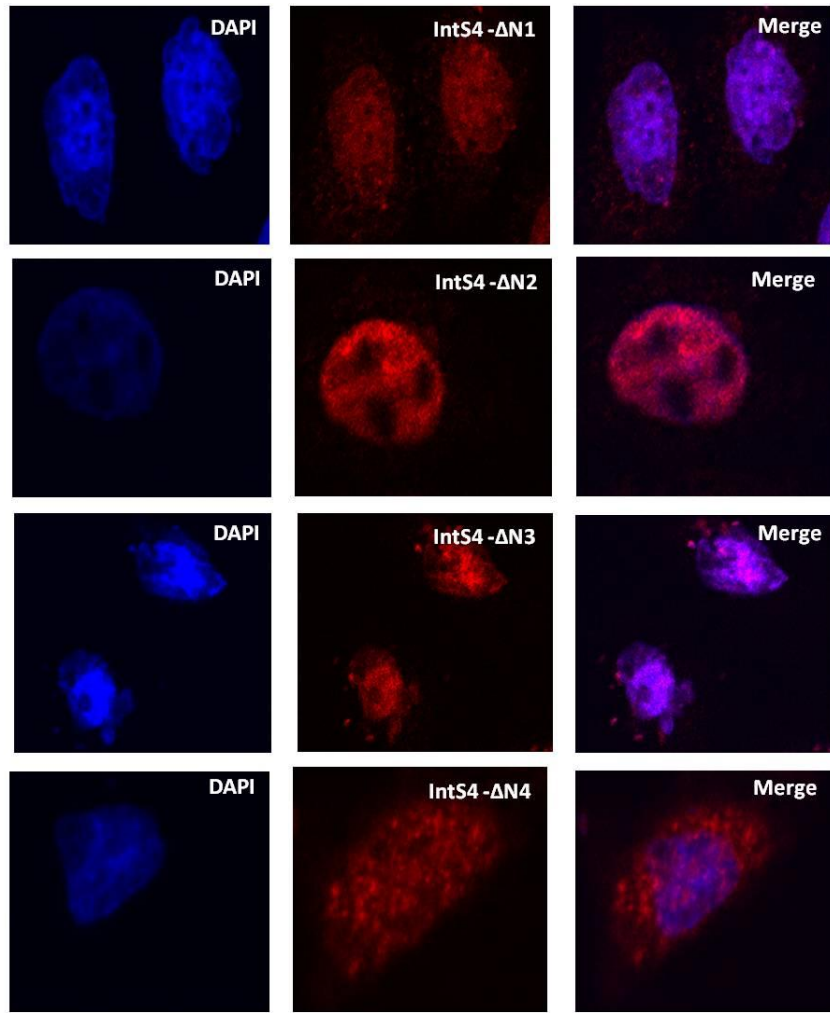
Subcellular Localization does not explain failure of IntS4 fragment to rescue snRNA processing of the U7-GFP reporter.

Integrator subunits, in particular full-length IntS4 are essentially nuclear proteins as their primary role is to participate in co-transcriptional processing. Although there were no apparent anomalies in expression of the fragments, it was necessary to ensure that truncations to the full-length protein did not affect localization. To address this concern, the truncation constructs were transfected into HeLa cells plated on cover slips which were subsequently fixed and imaged by Fluorescence and confocal microscopy. Full-length IntS4 as expected was observed to be localized to the nucleus (Figure 3.9A). Most N and C-terminal fragments were nuclear according to expectation except the smallest C-terminal fragment ( $\Delta N5$ ). The immunofluorescence images helped conclude that the unexpected GFP signal from the fragments is due to the inherent property of hIntS4 alone and not due to anomalous artifacts. These results support the model that both the N and C-terminus of hIntS4 were important and were contributing to the Integrator complex mediated snRNA 3' end formation. Such behavior could be possible if there are multiple binding partners for hIntS4 and each of these partners individually was contributing to the function of the Integrator complex.

**Figure 3.9. Immunofluorescence images of full-length IntS4 and each truncation construct to determine localization.** HeLa cells were transfected with 500ng of each plasmid and they were fixed and imaged using confocal and fluorescence microscopy. (A) Images of full-length,  $\Delta C1$ , and  $\Delta C2$  fragments of IntS4 (B) Images of  $\Delta C3-5$  fragments of IntS4 (C) Images of  $\Delta N1-4$  fragments of IntS4.



C.

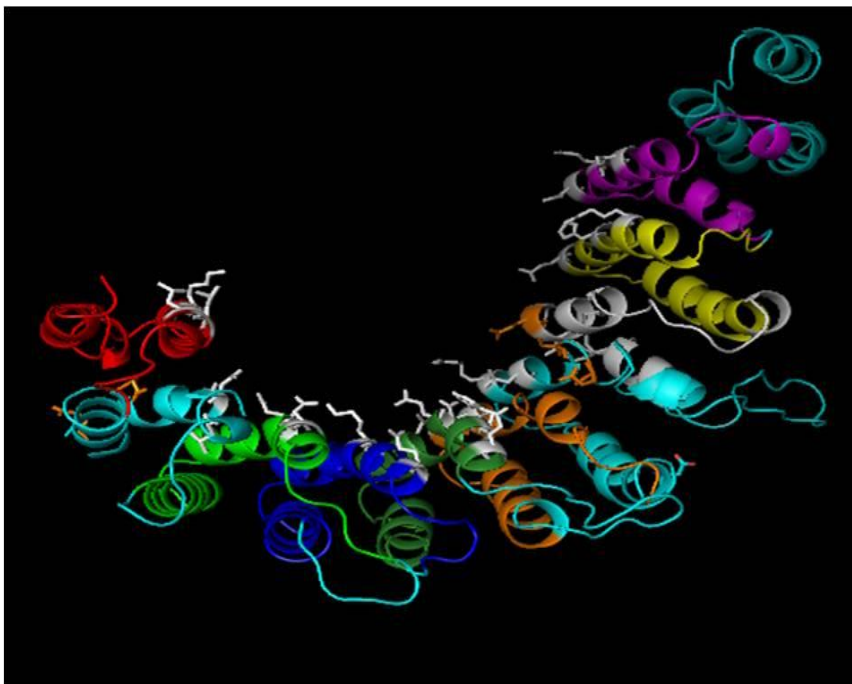


## **Identification of N-terminal HEAT repeats essential to the function of IntS4.**

The previous set of experiments involving deletion analysis of IntS4 indicated that both the N and C-terminus are important to its function. This interpretation is consistent with the results that the GFP-reporter assay utilizing the truncation fragments, failed to isolate a single region responsible for the function of IntS4. As an alternative approach to identify loss of function mutations within IntS4, we used the predicted structural model of the IntS4 N-terminus that was generated using Swiss-model. This *in silico* model of hIntS4 predicts a structure for the HEAT repeat containing N-terminal region of hIntS4. To generate point mutants, we used this model in combination with the other known structural models of human and Drosophila Symplekin that are present in the literature.

The IntS4 structural model predicts canonical concave surface as seen in Symplekin which is known to provide a surface for multiple protein interactions. Based on our hypothesis, if Symplekin and IntS4 are structurally similar, the concave surface of IntS4 would be expected to behave in a similar manner as well. Careful examination of the sequence on the concave side of each repeat revealed the following features: 1) Each repeat is linked to its neighboring repeat by a small and rigid loop, which consists mostly of arginine and glutamate residues that in the majority of HEAT repeats was highly conserved. It is noteworthy that this is the structural feature lacking in the two pseudo HEAT repeats but the fold may be retained through alternative means. 2) The repeats forming the concave surface have a few residues, which significantly protruded into the concave pocket (Figure 3.10). The amino acids that extend out from the concave surface are important for other known HEAT repeat binding proteins and were

**Figure 3.10. Prediction of amino acids within the IntS4 N-terminal HEAT repeat region that protrude from the concave surface.** Using Swiss-model, we highlighted the amino acid side chains (in gray) that are predicted to extend from the concave alpha helix surface. These protruding amino acids have been shown to be critical to mediate HEAT domain protein-protein interactions with other proteins.





previously shown to be essential for Symplekin's HEAT repeat 6 to bind to Ssu72. In fact, a single point mutation at one these types of amino acids in Symplekin completely abrogated association with Ssu72. Based upon this reasoning, these residues were selected and point mutations were introduced. (Table 3.1) Given the high level of uncertainty associated with a predicted model of the IntS4 HEAT repeat region, we generated mostly multiple mutations within each HEAT repeat (usually 2-3) and we gave an emphasis to amino acids that were not only predicted to be protruding but also were highly conserved. The loop region between each repeat however was left unaltered in order to not disrupt overall structure of a given HEAT repeat.

The rationale for this approach was that residues with predicted protrusion into the hydrophobic pocket were possibly involved in binding other Integrator subunits or RPAP2 and mutations to these amino acids would provide evidence regarding their importance in mediating snRNA 3' end processing. The mutations were introduced into the full-length IntS4 RNAi-resistant cDNA and tested for their ability to restore processing of the U7-GFP reporter after knock down of the endogenous IntS4. HEAT repeat mutants which fail to rescue would then be used in biochemical experiments to determine which Integrator subunits' interaction was disrupted.

The results of this experiment are shown in Figure 3.11 and summarized in Figure 3.12. We observed that all of the HEAT point mutants were expressed as well as the wild-type protein. Moreover, the majority of the mutants were as functional as the wild-type at restoring the U7-GFP reporter processing as none were as deleterious as the empty vector rescue. However, we did observe that four mutants did have some limited ability to rescue suggesting that these HEAT repeats are important for IntS4 function.

**Table 3.1. HEAT repeat sequence showing the residues chosen for Alanine mutations.** Pseudo HEAT repeats are denoted by 1A, 5A. Mutants 1A1, 1A2, 1A3, 5A1 and 5A2 represent different sets of mutations to the pseudo repeats.

HEAT repeat mutant	HEAT repeat sequence and residues mutated to Alanine
1	EAESVEGVVRILLEHYKENDPSVRLK*IASL*LGLL*SKTAGFSP
2	IQMRLVDVACKHLTDTSHGVRNKCLQ*LLGNLGSLEKSVT
3	AARDVQKIIGDYFSDQDPRVRTAAIK*AMLQLHE*RGLKLHQ
4	TIYNQACKLLSDDYEQVRSAAVQ*LIW*VVSQ*LYPESIVPIPSNNEI
5	RLVDDAFGKICHMVSDGSWVVRVQ*AAKLLGSMEQVSSH
6	AFVHGLEDEMYE*VRIAAVEALCMLAQSSPS
7	FAEKCLDFLVDMFNDEIEE*VRLQSIH*TMR*KISNNITLRED
8	QLDTVLAVLEDSSRD*IRE*ALHE*LLCCTNVSTKEG
1A1	DCIMDDAINILQNE*KSHQ*VLAQLLDTLLAIGTKLPENQA
1A2	DCIMDDAINILQNEKSHQVLAQLLD*TLLAIGTKLPENQA
1A3	DCIMDDAINILQNEKSHQVLAQLLDTL*LAIGTKLPENQA
5A1	FLEQTLDDKKLMSDLRRKRTAHER*AKEL*YSSGEFSSGRKWGDDAPKEEVDTGAVNLIESGACG
5A2	FLEQTLDDKKLMSDLRRKRTAHERAKELYSS*GE*FSSGRKWGDDAPKEEVDTGAVNLIESGACG

\* Conserved residue

**Figure 3.11. Identification of HEAT repeats important for IntS4 mediated snRNA 3' end processing.** The point mutant constructs developed as shown in Figure 14 were used for the GFP reporter assay. (A) Fluorescence images showing GFP signal (B) western blot from reporter assay derived cell lysates. The blots were probed for anti-GFP and anti-Tubulin. Also shown is the western blot probed for anti-myc to test expression of the HEAT repeat mutants (C) *In silico* predicted HEAT repeat model for hIntS4 showing the HEAT repeats identified to be important for IntS4 mediated snRNA 3' end processing. The important residues are shown in red.

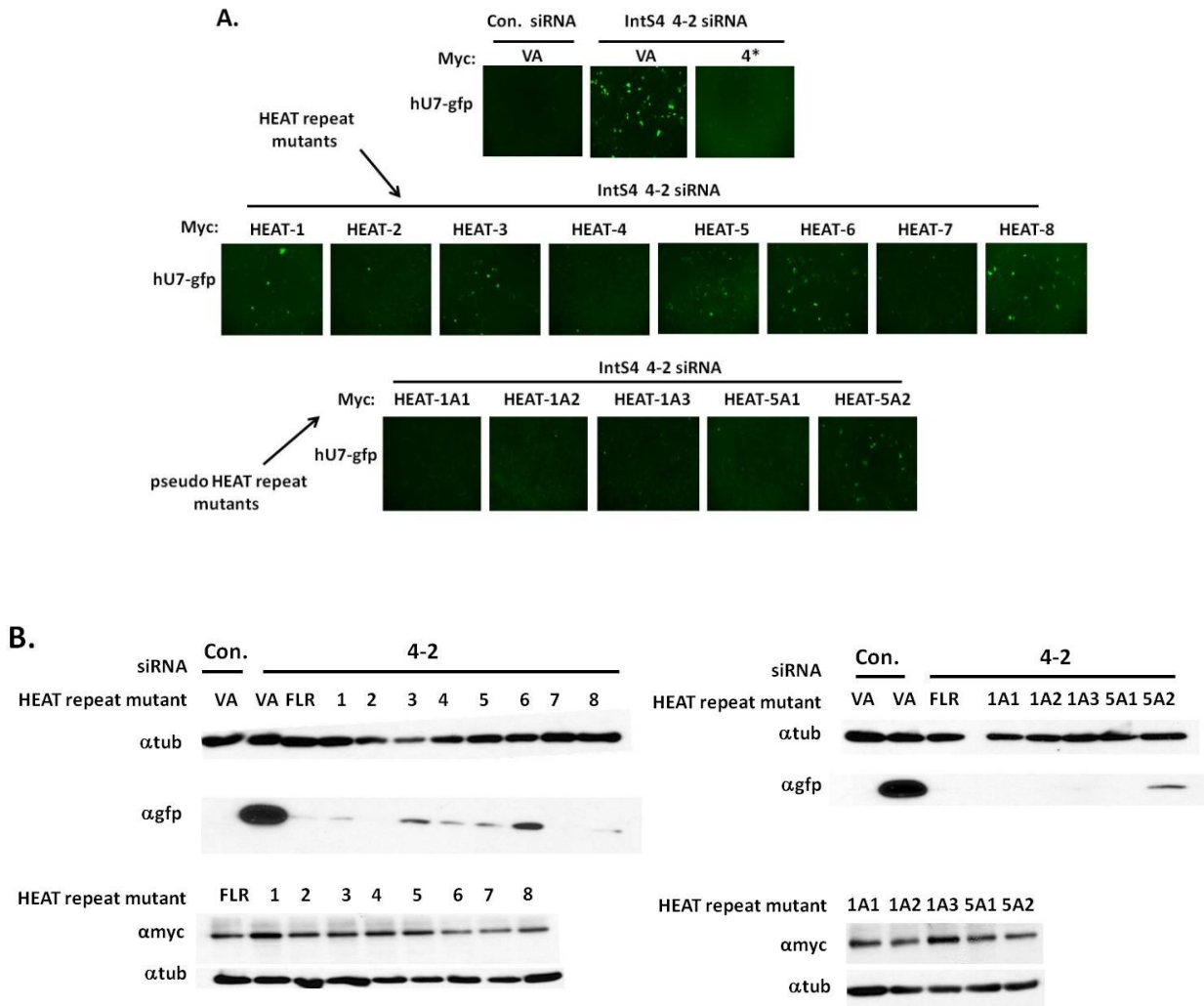
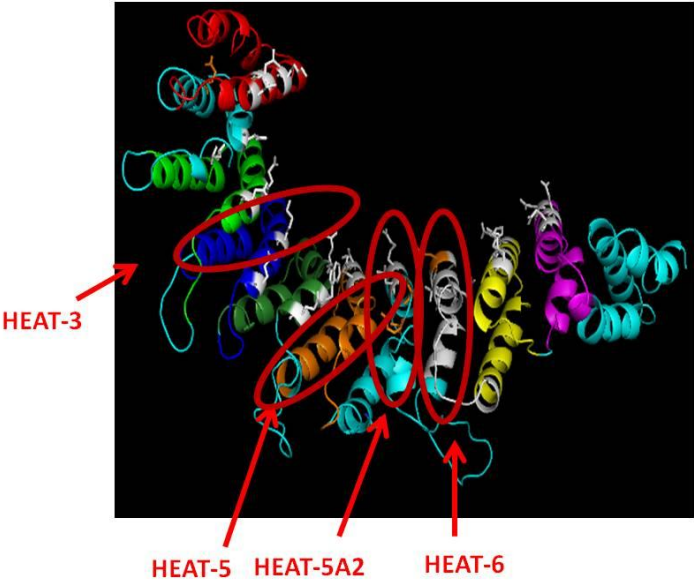


Figure 3.12. Summary of results from Figure 3.11.

HEAT repeat mutant	Rescue of misprocessing (yes/no)
1	Yes
2	Yes
3	No
4	Yes
5	No
6	No
7	Yes
8	Yes
1A1	Yes
1A2	Yes
1A3	Yes
5A1	Yes
5A2	No



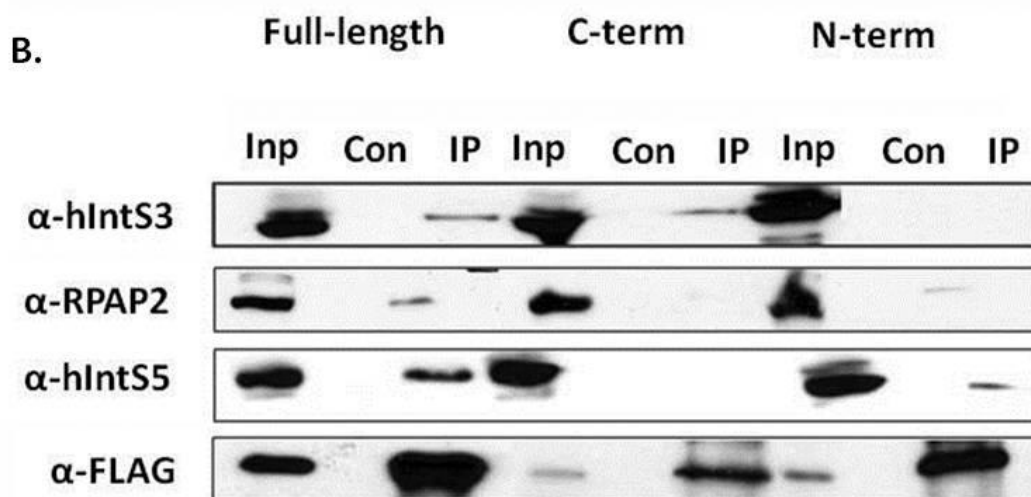
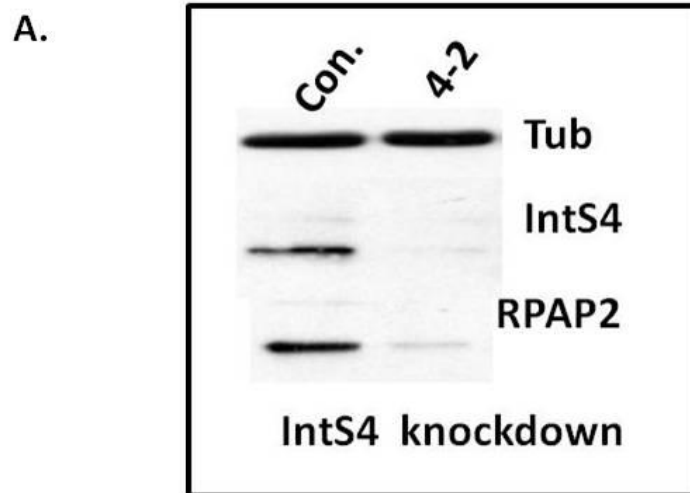
Identification of binding partners of IntS4: Determining binding partners of IntS4 is necessary to substantiate the hypothesis that IntS4 behaves as a scaffold. Given that IntS4 is part of the 14 subunit Integrator complex, the most obvious binding partners are expected to be other Integrator subunits. To identify IntS4 binding partners, we developed a co-Immunoprecipitation assay using nuclear extracts derived from 293T stable cell lines transfected with plasmids encoding either Flag-tagged hIntS4 full-length, N, or C-terminal proteins. Following immunoprecipitation using anti-FLAG agarose beads, eluted proteins were resolved using SDS-PAGE and then were probed for either Integrator subunits or RPAP2. It is important to note for interpretation that these experiments will not identify proteins that directly bind to IntS4 but rather those that are associated either directly or indirectly with fragments of IntS4.

Using our pulldown assay, we observed that RPAP2 is pulled down with the N-terminal fragment of IntS4, but not the C-terminal fragment. Compared to the input, the amount pulled down with both the full-length and N-terminal fragments is not robust. This reflects either substoichiometric association with RPAP2 or low affinity. Given that RPAP2 is thought to not be exclusively associated with Integrator this is the expected results. In addition, siRNA mediated IntS4 knockdown in HeLa cells resulted in co-depletion of RPAP2. Co-depletion although does not provide absolute evidence of interaction, it gives a hint about protein associations in a complex. The observation that RPAP2 co-depletes upon IntS4 knockdown is in consensus with the co-IP data and helps substantiate that there is a possible interaction between IntS4 and RPAP2.

Among the 14 subunits of the integrator complex, we observed an IP pattern similar to that of RPAP2 for IntS5 but the reverse pattern of association for IntS3 with

the C-terminal fragment having the interaction domain. These data give an indication of potential candidates for binding; however precise residues of IntS4 which bind these candidates are currently unknown and will need to be examined in the future.

**Figure 3.13. Identification of potential candidates for IntS4 binding by co-Immunoprecipitation assays. A)** Western blot showing co-depletion of RPAP2 upon siRNA mediated knockdown of IntS4. **B)** Western blot IP and co-IP on Nuclear extracts from HEK 293T stable cell lines expressing Flag-tagged IntS4, N and C-terminal fragments. Input lane represents 10% of IP. Agarose beads conjugated with FLAG-antibody were used as control.



# **Chapter-4**

## **Conclusions and Future Directions**



## CONCLUSIONS:

The purpose of this study was to perform structure-function analysis on human Integrator subunit 4 with the intent of gaining an understanding of how IntS4 contributes to the Integrator complex overall function. Taken into account the reported structure of Symplekin and its binding interactions with the mRNA processing machinery, we hypothesized that Integrator 4 similar to Symplekin, behaves as a scaffold molecule and facilitates binding among other Integrator subunits. Initial functional analysis of IntS4 demonstrated its importance in snRNA 3' end processing and in maintaining Cajal body structure. RNAi mediated depletion of IntS4 resulted in significant amount of misprocessing from which we gathered that IntS4 is critical to the functioning of the Integrator complex. These data although provided an insight into the functional characteristics of IntS4, did not reveal a mechanism or other key players involved. Understanding the mechanism by which IntS4 contributes to the function of Integrator complex will bring us a step closer to deciphering the series of events in snRNA 3' end processing.

Neither, the crystal structure of Integrator 4 nor any other subunit of the Integrator complex has been resolved to date. Taking this limitation into consideration, structural analysis of IntS4 performed in this study was based on sequence and secondary structure predictions. A GFP based transcriptional readthrough reporter assay previously developed in the lab was the basis of the most investigations in this study. Initial domain analysis performed by generating siRNA resistant N and C-terminal truncation constructs of the full-length IntS4 demonstrated that IntS4 deletions from

either the N or the C-terminus drastically affects its function. This non-tolerance to truncations led us to believe a possible scaffold like function for IntS4 with the involvement of multiple binding partners each having a distinct role. An alternative interpretation is that IntS4 may not behave as a scaffold but rather any of the deletions that were made disrupted the entire structure of the protein. We disfavor this interpretation as the deletion mutants were shown to accumulate to levels near that of the wild type protein.

### **Structural analysis of IntS4:**

Structurally, IntS4 is 963 residues in length with a conserved N-terminus and a nonconserved C-terminus. The C-terminus has no homology to any known proteins. Additionally it also lacks known domains making its structural interpretation nontrivial. PONDR analysis on the C-terminal region of IntS4 revealed the presence of disorder but surprisingly not to the extent we expected. This region is interesting merely based on the fact that no structure or functional information can be gathered from its sequence.

The N-terminal region on the other hand is far from disordered and is known to contain a HEAT repeat domain family. Secondary structure predictions were performed on IntS4 with multiple templates for homology modeling due to lack of homology with any one single protein. This prediction helped generate an *in silico* model for the N-terminal HEAT repeat region of hIntS4. The HEAT repeat containing *in silico* model has a canonical hydrophobic concave pocket which, similar to Symplekin is thought to aid in

ligand binding. The predicted eight HEAT repeats seen in the model form a superhelical structure which provides a surface that enables the binding of multiple other Int subunits.

Next, the predicted model was used as a basis for further N-terminal analysis. Careful examination of the repeat motif sequence on the concave side revealed the presence of residues with polar side chains protruding into the hydrophobic pocket. Such residues were found on every concave HEAT repeat and were also found to be conserved across 4 species (Figure 3.3). These were mutated to alanine and the mutant constructs thus generated were used for the GFP reporter assay. This assay helped determine the HEAT repeats important for IntS4 function and snRNA 3' end processing. Having identified these HEAT repeats, in the future it will be interesting to identify their binding partners.

### **Identification of IntS4 binding partners:**

Specific aim 2 of this project was to identify candidate proteins that bind to IntS4 and help validate its function as a scaffold molecule. For this purpose, Flag-tagged IntS4 full length, N and C-terminal constructs were transfected into HEK 293T cells and stable cell lines were generated with Blastocidin selection. Nuclear extracts from these cell lines was obtained and used for IP and co-IP assays. Co-IP's revealed that IntS3 and 5 were pulled down with stronger affinity with IntS4. Co-depletion analysis and co-IP's are an indication of proteins interaction; however they do not provide direct evidence of binding. In the case of a macromolecular complex such as the Integrator complex, due

to the possibility of subcomplex formation, it becomes necessary to show evidence of direct binding. Far western analysis and *in vitro* translation and pull down assays to confirm the binding candidates are currently underway.

Given its highly conserved sequence and important contribution to snRNA processing, gaining a complete understanding of the step wise events mediated by IntS4 in facilitating the function of the Integrator Complex will shed light on the mechanism of this important group of proteins.

#### **FUTURE DIRECTIONS:**

This study provides a preliminary evidence of the scaffold like function of IntS4. However, in order to understand how IntS4 contributes to the function of the Integrator complex, many questions will need to be investigated. First, what are the direct binding partners of IntS4? While the coimmunoprecipitation assay was effective to define interactions with the complex present in nuclear extract, it does not reveal any direct interactions. Efforts were made during the course of this thesis research to clone and express recombinant IntS4 fragments to use in direct pulldown assays. These experiments will need to be continued to answer the question of direct binding. Another important consideration when interpreting the results present in this thesis is that while the results are supportive of a scaffold function for IntS4, they do not rule out other potential critical functions for IntS4 including binding to the snRNA or the CTD of Rpb1. If IntS4 does behave like Symplekin and acts as a platform for IntS9 and 11 to interact, it could also position the cleavage complex to cleave RNA using an intrinsic RNA

binding domain, or interact with another subunit that binds RNA. Other questions that remain are: How and when is IntS4 recruited? Once it contributes to the Integrator complex function, how is it recycled and reused to facilitate additional snRNA 3' end formation events? What are the possible targets and functions of IntS4 outside of snRNA biogenesis? These questions will be interesting to answer in the future and will help in gaining a complete picture of the mechanism of IntS4 mediated snRNA 3' end formation.

## References:

- Albrecht TR, Wagner EJ. 2012. snRNA 3' end formation requires heterodimeric association of integrator subunits. *Mol Cell Biol* **32**: 1112-1123.
- Bah A, Wischniewski H, Shchepachev V, Azzalin CM. 2012. The telomeric transcriptome of *Schizosaccharomyces pombe*. *Nucleic Acids Res* **40**: 2995-3005.
- Baillat D, Hakimi MA, Naar AM, Shilatifard A, Cooch N, Shiekhatair R. 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**: 265-276.
- Bartkowiak B, Mackellar AL, Greenleaf AL. 2011. Updating the CTD Story: From Tail to Epic. *Genet Res Int* **2011**: 623718.
- Buratowski S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**: 541-546.
- Callebaut I, Moshous D, Mornon JP, de Villartay JP. 2002. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* **30**: 3592-3601.
- Casanova MF, Trippe J, 2nd. 2006. Regulatory mechanisms of cortical laminar development. *Brain Res Rev* **51**: 72-84.

- Casanova MF, Trippe J, 2nd, Switala A. 2007. A temporal continuity to the vertical organization of the human neocortex. *Cereb Cortex* **17**: 130-137.
- Cazalla D, Xie M, Steitz JA. 2011. A primate herpesvirus uses the integrator complex to generate viral microRNAs. *Mol Cell* **43**: 982-992.
- Chen J, Ezzeddine N, Albrecht TR, Waltenspiel B, Warren WD, Marzluff WF, Wagner EJ. 2012a. An RNAi Screen Identifies Additional Members of the Drosophila Integrator Complex and a Requirement for Cyclin C/Cdk8 in snRNA 3' End Formation. *RNA* **in press**.
- Chen J, Ezzeddine N, Waltenspiel B, Albrecht TR, Warren WD, Marzluff WF, Wagner EJ. 2012b. An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA* **18**: 2148-2156.
- Chen J, Wagner EJ. 2010. snRNA 3' end formation: the dawn of the Integrator complex. *Biochem Soc Trans* **38**: 1082-1087.
- Chen J, Waltenspiel B, Warren WD, Wagner EJ. 2013. Functional analysis of the integrator subunit 12 identifies a microdomain that mediates activation of the Drosophila integrator complex. *J Biol Chem* **288**: 4867-4877.

- Cheung CH, Fan QN, Stumph WE. 1993. Structural requirements for the functional activity of a U1 snRNA gene enhancer. *Nucleic Acids Res* **21**: 281-287.
- Chwedenczuk J, Zin P, Rzazewski K, Trippenbach M. 2006. Simulation of a single collision of two bose-einstein condensates. *Phys Rev Lett* **97**: 170404.
- Dominski Z, Yang XC, Purdy M, Wagner EJ, Marzluff WF. 2005. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol* **25**: 1489-1500.
- Dundr M, Hebert MD, Karpova TS, Stanek D, Xu H, Shpargel KB, Meier UT, Neugebauer KM, Matera AG, Misteli T. 2004. In vivo kinetics of Cajal body components. *J Cell Biol* **164**: 831-842.
- Egloff S, Murphy S. 2008. Role of the C-terminal domain of RNA polymerase II in expression of small nuclear RNA genes. *Biochem Soc Trans* **36**: 537-539.
- Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S. 2007. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**: 1777-1779.
- Egloff S, O'Reilly D, Murphy S. 2008. Expression of human snRNA genes from beginning to end. *Biochem Soc Trans* **36**: 590-594.



- Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S. 2010. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem* **285**: 20564-20569.
- Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. 2012. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell* **45**: 111-122.
- Ezzeddine N, Chen J, Waltenspiel B, Burch B, Albrecht T, Zhuo M, Warren WD, Marzluff WF, Wagner EJ. 2011. A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation. *Mol Cell Biol* **31**: 328-341.
- Filleur S, Hirsch J, Wille A, Schon M, Sell C, Shearer MH, Nelius T, Wieland I. 2009. INTS6/DICE1 inhibits growth of human androgen-independent prostate cancer cells by altering the cell cycle profile and Wnt signaling. *Cancer Cell Int* **9**: 28.
- Ghazy MA, He X, Singh BN, Hampsey M, Moore C. 2009. The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing. *Mol Cell Biol* **29**: 2296-2307.

- Hata T, Nakayama M. 2007. Targeted disruption of the murine large nuclear KIAA1440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death. *Biochim Biophys Acta* **1773**: 1039-1051.
- Hernandez G, Pirela J, Siniscalchi M, Lobo J, Grases P. 1989. [A case of peritoneal histoplasmosis]. *G E N* **43**: 301-305.
- Hernandez N. 1985. Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *Embo J* **4**: 1827-1837.
- Hernandez N. 2001. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem* **276**: 26733-26736.
- Hewig J, Trippe R, Hecht H, Coles MG, Holroyd CB, Miltner WH. 2007. Decision-making in Blackjack: an electrophysiological analysis. *Cereb Cortex* **17**: 865-877.
- Hofmann I, Schnolzer M, Kaufmann I, Franke WW. 2002. Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in *Xenopus laevis* oocytes. *Mol Biol Cell* **13**: 1665-1676.
- Huang J, Gong Z, Ghosal G, Chen J. 2009. SOSS complexes participate in the maintenance of genomic stability. *Mol Cell* **35**: 384-393.

- Hung KH, Stumph WE. 2011. Regulation of snRNA gene expression by the *Drosophila melanogaster* small nuclear RNA activating protein complex (DmSNAPc). *Crit Rev Biochem Mol Biol* **46**: 11-26.
- Infeld E, Zin P, Gocalek J, Trippenbach M. 2006. Statics and dynamics of Bose-Einstein condensates in double square well potentials. *Phys Rev E Stat Nonlin Soft Matter Phys* **74**: 026610.
- Jacobs EY, Ogiwara I, Weiner AM. 2004. Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. *Mol Cell Biol* **24**: 846-855.
- Jensen RC, Wang Y, Hardin SB, Stumph WE. 1998. The proximal sequence element (PSE) plays a major role in establishing the RNA polymerase specificity of *Drosophila* U-snRNA genes. *Nucleic Acids Res* **26**: 616-622.
- Kennedy SA, Frazier ML, Steiniger M, Mast AM, Marzluff WF, Redinbo MR. 2009. Crystal structure of the HEAT domain from the Pre-mRNA processing factor Symplekin. *J Mol Biol* **392**: 115-128.
- Kolev NG, Steitz JA. 2005. Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. *Genes Dev* **19**: 2583-2592.

- Li Y, Bolderson E, Kumar R, Muniandy PA, Xue Y, Richard DJ, Seidman M, Pandita TK, Khanna KK, Wang W. 2009. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem* **284**: 23525-23531.
- Liu JL, Wu Z, Nizami Z, Deryusheva S, Rajendra TK, Beumer KJ, Gao H, Matera AG, Carroll D, Gall JG. 2009. Coilin is essential for Cajal body organization in *Drosophila melanogaster*. *Mol Biol Cell* **20**: 1661-1670.
- MacDonald CC, Wilusz J, Shenk T. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol* **14**: 6647-6654.
- Mandel C, Bai Y, Tong L. 2008. Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences* **65**: 1099-1122.
- Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**: 953-956.
- Marzluff WF. 2005. Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Curr Opin Cell Biol* **17**: 274-280.

- Marzluff WF, Wagner EJ, Duronio RJ. 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9**: 843-854.
- Matera AG, Terns RM, Terns MP. 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8**: 209-220.
- Matuszewski M, Rosberg CR, Neshev DN, Sukhorukov AA, Mitchell A, Trippenbach M, Austin MW, Krolikowski W, Kivshar YS. 2006. Crossover from self-defocusing to discrete trapping in nonlinear waveguide arrays. *Opt Express* **14**: 254-259.
- Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, Murphy S. 2005. P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *Embo J* **24**: 4154-4165.
- Medlin JE, Uguen P, Taylor A, Bentley DL, Murphy S. 2003. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *Embo J* **22**: 925-934.
- Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL, Washburn MP. 2009. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell* **34**: 168-178.

- Mowry KL, Steitz JA. 1987. Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNA's. *Science* **238**: 1682-1687.
- Murthy KG, Manley JL. 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* **9**: 2672-2683.
- Ruepp MD, Schweingruber C, Kleinschmidt N, Schumperli D. 2011. Interactions of CstF-64, CstF-77, and symplekin: implications on localisation and function. *Mol Biol Cell* **22**: 91-104.
- Schumperli D, Pillai RS. 2004. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. *Cell Mol Life Sci* **61**: 2560-2570.
- Shukla NN, Trippett TM. 2006. Non-Hodgkin's lymphoma in children and adolescents. *Curr Oncol Rep* **8**: 387-394.
- Skaar JR, Richard DJ, Saraf A, Toschi A, Bolderson E, Florens L, Washburn MP, Khanna KK, Pagano M. 2009. INTS3 controls the hSSB1-mediated DNA damage response. *J Cell Biol* **187**: 25-32.

- Sullivan KD, Steiniger M, Marzluff WF. 2009. A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. *Mol Cell* **34**: 322-332.
- Takagaki Y, Manley JL. 2000. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* **20**: 1515-1525.
- Takata H, Nishijima H, Maeshima K, Shibahara K. 2012. The integrator complex is required for integrity of Cajal bodies. *J Cell Sci* **125**: 166-175.
- Tao S, Cai Y, Sampath K. 2009. The Integrator subunits function in hematopoiesis by modulating Smad/BMP signaling. *Development* **136**: 2757-2765.
- Trippe R, Guschina E, Hossbach M, Urlaub H, Luhrmann R, Benecke BJ. 2006. Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA* **12**: 1494-1504.
- Uguen P, Murphy S. 2003. The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *Embo J* **22**: 4544-4554.

- Wang ZF, Whitfield ML, Ingledue TC, 3rd, Dominski Z, Marzluff WF. 1996. The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev* **10**: 3028-3040.
- Wieland I, Arden KC, Michels D, Klein-Hitpass L, Bohm M, Viars CS, Weidle UH. 1999. Isolation of DICE1: a gene frequently affected by LOH and downregulated in lung carcinomas. *Oncogene* **18**: 4530-4537.
- Xiang K, Manley JL, Tong L. 2012. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. *Nat Commun* **3**: 946.
- Xiang K, Nagaïke T, Xiang S, Kilic T, Beh MM, Manley JL, Tong L. 2010. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**: 729-733.
- Xing H, Mayhew CN, Cullen KE, Park-Sarge OK, Sarge KD. 2004. HSF1 modulation of Hsp70 mRNA polyadenylation via interaction with symplekin. *J Biol Chem* **279**: 10551-10555.
- Yang XC, Torres MP, Marzluff WF, Dominski Z. 2009. Three proteins of the U7-specific Sm ring function as the molecular ruler to determine the site of 3'-end processing in mammalian histone pre-mRNA. *Mol Cell Biol* **29**: 4045-4056.



## **Vita**

Anupama Sataluri was born on June 3<sup>rd</sup> 1986 in Hyderabad, Andhra Pradesh, India where she grew up. After graduating from Keyes High School, she went to Osmania University where she graduated with a Bachelor of Science degree in Pharmacy in 2007. During the years of 2007-2010, she was enrolled in the Graduate School of Pharmaceutical Sciences, Duquesne University in a MS track in Pharmacology and Toxicology under the mentorship of Dr. David A. Johnson. In August 2010, she joined the Graduate School of Biomedical Sciences at the University of Texas Health Science Center at Houston.