Biochemical and Genetic Characterization of Factors that Influence Branchsite Selection during Pre-mRNA Splicing in Yeast

By

Tucker Joe Carrocci

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2017

Date of final oral examination: 5/1/2017

The dissertation is approved by the following members of the Final Oral Committee:

Aaron A. Hoskins, Assistant Professor, Biochemistry
Samuel E. Butcher, Professor, Biochemistry
Marvin Wickens, Professor, Biochemistry
Ronald T. Raines, Professor, Biochemistry and Chemistry
David A. Wassarman, Professor, Medical Genetics

ABSTRACT

RNA splicing plays many critical roles in eukaryotic gene expression, yet even 40 years after its discovery, the mechanism by which the spliceosome recognizes and removes introns is still poorly understood. In my thesis, I describe methods to produce RNAs for studies of the spliceosome and my work towards understanding an early step in splicing, branchsite selection.

In my thesis, I discuss methods to label RNA fluorescently. I used MS2 fusion proteins tags to the label RNAs in living yeast using small molecule fluorophores. These tags allow for visualization of localized RNAs and demonstrate that organic fluorophores can be targets to RNAs in living yeast. I then discuss work performed in collaboration with Dr. Jiacui Xu on the development of aptamers that can bind fluorescent molecules directly, obviating the need for protein. Finally, I show that the 10DM24 deoxyribozyme can be used to prepare long, site-specifically modified fluorescent pre-mRNAs and demonstrate methods that can circumvent debranching by the splicing factor, Dbr1.

In the second portion of my thesis, I discuss work on the role of splicing factors that contribute to branchsite choice. I explored the role of cancer causing mutations in a protein called SF3b1 using the yeast spliceosome. Mutations of SF3b1 result in changes in branchsite usage by the spliceosome. These mutations can impair and improve nonconsensus intron splicing, which we propose derives from a change in the conformation of the protein. I also explored differences between the human and yeast proteins and define regions that cannot be exchanged despite high sequence conservation. These differences explain the lack of efficacy of the spliceosome inhibitor, pladienolide B on the yeast spliceosome.

Finally, I used smFRET to observe dynamics in the splicing factor Prp5, which ensures proper branchsite usage by the spliceosome. I showed that Prp5 primarily adopts an open state in the absence of ATP and RNA but transitions to a closed state upon binding. This work provides a strong foundation to define the mechanism of Prp5 proofreading of the branchsite. Overall, my

work has significantly contributed to our understanding of spliceosome and the mechanisms of branchsite selection.

ACKNOWLEDGEMENTS

I am extremely fortunate to have had an extensive support network during my graduate career. First, thank you to my advisor Aaron Hoskins. I have learned many lessons from you over the years, including how to build a lab. You been instrumental to my growth as a scientist and I will carry those lessons with me in my future career. I also need to thank the past and current members of my committee for their support and advice throughout my time in graduate school. I would also like to thank Melissa Harrison and Sam Butcher for their support over the years. I owe a huge debt to the Biotechnology Training Program that allowed me to spend a summer working at New England Biolabs under the supervision of Ming Xu. I learned a great deal that I used during my thesis work.

Thank you to the past and present members of the Hoskins Lab. I am so grateful to have had your emotional and scientific support over the years. Thank you to Josh and Maggie, the other two-thirds of the tri-snRNP. I have learned so much from both of you and I could not have done this without your support. You have both become family. Thank you Sandy for being such a happy and positive force and thank you Alex for being the best gym buddy possible. You both made working here more joyful. Thank you Josh Paulson for everything that you helped me accomplish. You are been a constant help. Thank you to all the other members of the lab – Sarah, Ian, Chris, Xin, Jiacui, and more – for your help and emotional support.

I have been the graduate mentor to a number of talented undergraduate students. Thank you Matt, Doug, and David. You each worked on very different things and it was great to have a constant collaboration on our work. All of you have been amazing and working with you was one of my favorite parts of graduate school.

The environment at UW-Madison has been a terrific one to find lifelong friends and colleagues. Thank you to Allison for being the best conference buddy a person could ask for and for being a constant source of inspiration. Your dedication to scientific excellence and rigor is something that I strive to emulate in my own work. Thank you Mike, Brian, and Harriet for being

some of the best lab neighbors and trivia partners possible. I have always relied on your willingness to offer feedback or advice.

Finally, thank you to my friends and family. I met some of my closest friends through my graduate program. Thank you Molly for always being my best friend. You are honestly the best person that I know. Thank you Greg for seeing every terrible movie that we have seen together. Thank you Andrew for being the best roommate and Laura for being a terrific friend. Thank you to everyone else. I love you all and you have made this work possible.

Table of Contents

List of Tables and Figures

Chapter One

Thesis Overview

My graduate work is best divided into two areas of RNA research and therefore, my thesis is also divided into two broad sections. The first section focuses on developing fluorescent labeling methods for RNA (**Chapters 2 and 3**) and the second focuses on understanding the function of components involved in branchsite recognition and usage during pre-mRNA splicing (**Chapters 4-7**).

Chapter two discusses my work on fluorescent RNA labeling *in vivo* using a combination of organic fluorophores and MS2 RNA-binding protein tags. This chapter also includes my work in collaboration with Dr. Jiacui Xu on the development of a benzylguanine aptamer that be combined with commercially available benzylguanine dyes (marketed as SNAP dyes by New England Biolabs). Both of these topics have been published.

Chapter three discusses my unpublished data studying the 10DM24 deoxyribozyme as a tool for RNA modification *in vitro*. In this work, I evaluated structure-activity relationships of 10DM24 and used the deoxyribozyme as a tool to prepare long, site-specifically modified, fluorescent pre-mRNAs. My work shows that the linkage formed by 10DM24 is vulnerable to cleavage by a spliceosome-related protein called Dbr1. Finally, I discuss methods to bypass cleavage by Dbr1 through the use of phosphorothioate modified nucleotides.

Chapter four provides a broad overview of the process of pre-mRNA splicing in a detailed review. This chapter was published as a component of the Encyclopedia of Cell Biology.

Chapter five examines the impact of cancer-associated mutations in a protein called SF3b1 on the yeast spliceosome (using the yeast homolog Hsh155). I showed that these mutations alter the fidelity of branchsite usage by the spliceosome and proposed a mechanism by which these mutations cause changes in pre-mRNA splicing. Notably, this branchsite fidelity mechanism is independent of the splicing factor Prp5. This work was published in *Nucleic Acid Research*.

Chapter six extends the work on the yeast SF3b complex that I began in chapter five. I sought to validate contacts to Hsh155 and Rds3 observed in recently published cryo-electron

microscopy structures and understand their impact on branchsite choice. I also used chimeric Hsh155/SF3b1 proteins in yeast understand differences between the two homologs.

Chapter seven explores conformational dynamics in the splicing factor Prp5. I developed a protein construct of Prp5 that is amenable for single molecule Förster resonance energy transfer studies (smFRET). Using smFRET, I monitored changes in protein conformation upon binding of ATP and RNA and compared the dynamics in the isolated helicase core to those in the full length protein.

Chapter eight discusses the path forward on areas that I have worked on in graduate school. This includes a discussion on the future of deoxyribozymes as RNA labeling tools, proposed work to better understanding species-specific differences between Hsh155/SF3b1, and thoughts on the mechanism of Prp5.

Chapter Two

Imaging of RNAs with Spectrally Diverse Small Molecule Fluorophores

This chapter has been published in two parts:

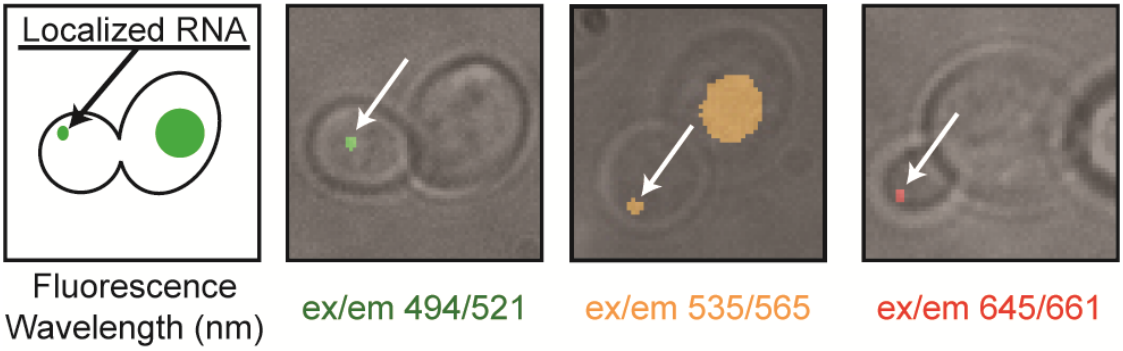
1. Carrocci,T.J. and Hoskins,A.A. (2014) Imaging of RNAs in live cells with spectrally diverse small molecule fluorophores. *Analyst*, **139**, 44–47.

2. Xu, J., Carrocci, T. J., and Hoskins, A. A. (2016). Evolution and characterization of a benzylguanine-binding RNA aptamer. *Chem. Commun*. **52**, 549–552.

ABSTRACT

We have developed new protein-based tags for use in imaging of RNAs in living cells. These tags are based on the MS2 phage coat protein fused to either *E. coli* dihydrofolate reductase or the SNAP tag. Labeled RNAs can be visualized with fluorophores spanning the visible spectra range.

GRAPHICAL ABSTRACT



INTRODUCTION

Fluorescence light microscopy (e.g., epi-fluorescence, confocal, or STED imaging) is a powerful tool for studying the dynamics and localization of biomolecules in living cells. Proteins are routinely imaged by fusion to fluorescent proteins (FPs).[1] On the other hand, RNA imaging *in vivo* is less straightforward due to the lack of intrinsically fluorescent motifs.[2 3]

Fusion of the MS2 bacteriophage coat protein to FPs (e.g., MS2-eGFP) has proven useful for studying the cellular dynamics of RNAs encoding MS2 stem loops.[2,4,5] Binding of multiple MS2-FPs to the RNA results in a detectable fluorescence signal that is colocalized with the RNA. However, the MS2-FP fluorophores are often constitutively active resulting in high background fluorescence.[4,5]

Targeting organic fluorophores to RNAs represents a promising alternative strategy for fluorescent labeling. The fluorescence signal can be controlled by the addition of small molecules to the growth media. Furthermore, many organic dyes have higher quantum yields and extinction coefficients than FPs, making them brighter fluorophores.[6]

To extend the benefits of organic fluorophores to RNA imaging in live cells, we have developed small molecule-based RNA tags by fusion of the MS2 coat protein to either the *Escherichia coli* dihydrofolate reductase (MS2-eDHFR) or the SNAP tag (MS2-SNAP). We have used these fusions to image RNAs in live *Saccharomyces cerevisiae* (yeast) cells with a diverse range of fluorophores. The use of MS2-eDHFR or MS2-SNAP fusions significantly expands the repertoire of fluorophores available for imaging RNAs in live cells.

MATERIALS AND METHODS

Unless otherwise noted, Herculase II DNA polymerase (Stratagene) was used for PCR. All plasmids (**Table 2.1**) were confirmed by sequencing. Yeast transformation and growth was carried out using standard techniques and media.[1] Fluorescent TMP derivatives were obtained from ActiveMotif. Fluorescent ligands for the SNAP tag were obtained from New England Biolabs. Ivan Correa, Jr. (New England Biolabs) generously provided bG-SiRhod and SNAP-Cell® 505-Star.

RNA-binding Protein Plasmid Construction:

The eDHFR coding sequence was fused to the C-terminus of the MS2 coat protein by first amplifying the eDHFR gene from plasmid pMONDHFR[2] with PCR primers TJC001 and TJC002 (**Table 2.2**). The resulting PCR product contained the eDHFR gene along with an N-terminal, six amino acid linker (AGSGSG) and NotI and HpaI restriction enzyme recognition sites at the 5' and 3' ends of the gene, respectively. The pGFP-MS2/Leu plasmid[3] (gift of R. Singer, **Table 2.1**) containing the dlFG mutant of the MS2 coat protein along with a N-terminal nuclear localization signal (NLS) and a hemagglutinin (HA) tag was digested with the NotI and HpaI restriction enzymes to remove the GFP coding sequence. The eDHFR fragment was then ligated into this plasmid at these same sites to yield MS2-eDHFR (pAAH0099, **Table 2.1**). MS2-SNAP was created using primers TJC003 and TJC0004 (**Table 2.2**) to amplify the SNAP gene from the $SNAP_f$ Vector (New England Biolabs). The PCR product was digested with NotI/HpaI and ligated into the digested pGFP-MS2/Leu plasmid to yield pAAH0096 (**Table 2.1**). To create the tdMS2-eDHFR plasmid (pAAH0176, **Table 2.1**) containing the gene for eDHFR fused to a tandem dimer of the MS2 protein, three gene blocks (IDT) were ligated together by Gibson Assembly (New England Biolabs), amplified by PCR using *Taq* DNA Polymerase (Promega) and primers ah103_tdMS2_DHFR_F and ah104_tdMS2_DHFR_R (**Table 2.2**). The product was then ligated into plasmid p425 GPD (www.atcc.org) at the SpeI and HindIII restriction sites. The full gene and amino acid sequence of this construct is shown below:

tdMS2-eDHFR Gene Sequence:

GGGACTAGTATGCCAAAGAAGAAAAGGAAAGTTGGCTACCCCTACGACGTGCCCGACTACGCCATCGAAG
GCCGCCATATGCTAGCCGTTAAAATGGCATCAAATTTCACACAATTTGTACTAGTTGATAACGGCGGAAC
TGGCGACGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGATCGCTGAATGGATCAGCTCTAACTCGCGT
TCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAG
TCGAGGTGCCTAAAGGCGCCTGGCGTTCGTACTTAAATATGGAACTAACCATTCCAATTTTCGCCACGAA
TTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCA
ATCGCAGCAAACTCCGGCATCTACATCTACGCCATGGCTTCTGCTTCTAACTTTACTCAGTTCGTTCTCG
TCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGATCGCTGAATGGAT
CAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAATCGC
AAATACACCATCAAAGTCGAGGTGCCTAAAGGCGCCTGGCGTTCGTACTTAAATATGGAACTAACCATTC
CAATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAA
CCCGATTCCCTCAGCAATCGCAGCAAACTCCGGCATCTACGCGGCCGCTGGTTCTGGTTCTGGTATGATC
AGTCTGATTGCGGCGTTAGCGGTAGATCGCGTTATCGGCATGGAAAACGCCATGCCGTGGAACCTGCCTG
CCGATCTCGCCTGGTTTAAACGCAACACCTTAAATAAACCCGTGATTATGGGCCGCCATACCTGGGAATC
AATCGGTCGTCCGTTGCCAGGACGCAAAAATATTATCCTCAGCAGTCAACCGGGTACGGACGATCGCGTA
ACGTGGGTGAAGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGACGTACCAGAAATCATGGTGATTGGCG
GCGGTCGCGTTTATGAACAGTTCTTGCCAAAAGCGCAAAAACTGTATCTGACGCATATCGACGCAGAAGT
GGAAGGCGACACCCATTTCCCGGATTACGAGCCGGATGACTGGGAATCGGTATTCAGCGAATTCCACGAT
GCTGATGCGCAGAACTCTCACAGCTATTGCTTTGAGATTCTGGAGCGGCGGTAAAAGCTTGGG

tdMS2-eDHFR Protein Sequence:

MPKKKRKVGYPYDVPDYAIEGRHMLAVKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQA
YKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAA
NSGIYIYAMASASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT
IKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAAAGSGSGMISLI
AALAVDRVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHTWESIGRPLPGRKNIILSSQPGTDDRVTWV
KSVDEAIAACGDVPEIMVIGGGRVYEQFLPKAQKLYLTHIDAEVEGDTHFPDYEPDDWESVFSEFHDADA
QNSHSYCFEILERR.

Reporter RNA Plasmid:

Reporter RNAs bearing multiple MS2 stemloops were expressed by transformation of yeast strains (**Table 2.3**) with pGAL-lacZ-MS2-ASH1/TRP as described elsewhere.[3]

Reporter Induction and *In Vivo* labeling:

Prior to fluorescent labeling, yeast were transformed by plasmids expressing MS2 fusion proteins either in the presence or absence of a second plasmid expressing the reporter RNA. Single colonies of transformed yeast were then grown at 30^{o}C to early mid-log phase in the appropriate synthetic dropout media (25 mL) containing 2% w/v raffinose and 0.02% w/v glucose. To induce RNA expression, a portion of this culture (100 µL) was removed, placed into a sterile 1.5 mL microfuge tube, pelleted by centrifugation, and resuspended in media (100 µL) containing 3% w/v galactose. For labeling of RNAs with trimethoprim, cells were treated with trimethoprim

dyes (10μM) for 0.75-3h at 30°C while rotating. Care was taken to avoid excessive light exposure during and after staining. For labeling of RNAs with SNAP dyes, a portion of an overnight culture (100 μL) was removed, placed into a sterile 1.5 mL microfuge tube, and incubated with SNAP dyes (5 μM) for 7 h. Cells were then collected by centrifugation and fresh, galactose-containing media with the SNAP dye (5 μM) was added. Cells were then grown for an additional 3h, washed with fresh media (1 mL), and plated for imaging.

Image Acquisition and Analysis:

Yeast were imaged in glass bottom microwell petri dishes (MatTek) pretreated with concanavalin A (1 μg/ml, Sigma) for surface adherence. Images were acquired using a Nikon TE-2000 microscope (100X 1.45NA oil immersion objective, 100ms exposure time) and either a CoolSnap CCD (0.017μm^2 per pixel, Photometrics) or EM-CCD (0.027μm^2 per pixel, Photometrics) camera. The microscope was controlled using MetaMorph imaging software. Fluorophores were excited using a mercury arc lamp (Exfo). eGFP, fluorescein, and SNAP-Cell® 505-Star were imaged using 488/30x and 535/30m excitation and emission filters, respectively. Hexachlorofluorescein was imaged using 530/30x and 575/40m excitation and emission filters, respectively. SiRhod was imaged using 654/24x and 675lp excitation and emission filters, respectively.

Images were first analyzed by visually determining the proportion of cells in a population possessing a fluorescent particle localized within the budding daughter cell. Integrated intensity values for individual particles within the bud were determined using ImageJ software. These particles were enclosed in a 13×13 pixel box for measurement and background was subtracted using an identically-sized region adjacent to the fluorescent particle.

Designation	RNA-Binding Protein	Fusion Tag
pAAH0096	MS2	SNAP
pAAH0099	MS2	eDHFR
pAAH0176	tdMS2	eDHFR
pGFP-MS2/Leu[a]	MS2	eGFP

[a]Plasmid described in Reference 3.

Table 2.1 Protein expression plasmids.

Primer Designation	Sequence (5' to 3')
TJC001	GA**GCGGCCGC**TGGTTCTGGTTCTATGATCAGTCTGATTGCGGCG
TJC002	CC**GTTAAC**TTACCGCCGCTCCAGAATCTC
TJC003	GA**GCGGCCG**CAGGAGGATCAGGAGGATCAATGGACAAAGACTGCGAAATG
TJC004	GAG**TTAAC**TTAACCCAGCCCAGGCTT
ah103_tdMS2_DHFR_F	GGG**ACTAGT**ATGCCAAAAAAGAAAAGAAAAGTTGG
ah104_tdMS2_DHFR_R	CCC**AAGCTT**TTACCGCCGCTCC

Table 2.2 Primers used in this chapter. Restriction enzyme recognition sites are shown in bold.

Strain Designation	Plasmid(s)	Reporter RNA	Tag Expressed
yAAH0143	pAAH0096	-	MS2-SNAP
yAAH0152	pAAH0096, pGAL-lacZ-MS2-ASH1/TRP	+	MS2-SNAP
yAAH0144	pAAH0099	-	MS2-eDHFR
yAAH0155	pAAH0099, pGAL-lacZ-MS2-ASH1/TRP	+	MS2-eDHFR
yAAH0146	pAAH0176	-	tdMS2-eDHFR
yAAH0157	pAAH0176, pGAL-lacZ-MS2-ASH1/TRP	+	tdMS2-eDHFR
yAAH0142	pGFP-MS2/Leu	-	MS2-eGFP
yAAH0153	pGFP-MS2/Leu, pGAL-lacZ-MS2-ASH1/TRP	+	MS2-eGFP

[a]All yeast used in this study are transformation products of the yeast strain K699 (Mat**a**, *his3-11, leu2-3, ade2-1, trp1-1, ura3, ho; can1-100)*.

Table 2.3 Strains used in this chapter: the yeast strain K699[a] was used for all experiments

RESULTS

The eDHFR tag tightly binds fluorescent analogs of trimethoprim (TMP)[7] while the SNAP tag reacts with fluorescent benzyl guanine (bG) or benzylchloropyrimidine (CP) derivatives to form a covalent adduct to the protein.[8,9] Since both the eDHFR and SNAP tags have been used to image proteins in live cells[7,10,8], we reasoned that fusion of these tags to MS2 coat protein should also permit RNA imaging in the presence of the respective small molecules.

We coexpressed the MS2-eDHFR or MS2-SNAP proteins along with a reporter mRNA in yeast. The reporter mRNA encodes (5' to 3') a fusion of the RP51A and *lacZ* transcripts, 6 copies of the MS2 stem loop, and the 3' untranslated region of the *ASH1* mRNA (**Figure 2.1A**).[4] Because the MS2 coat protein binds its target RNA stem loop as a dimer, the reporter mRNA can be bound by up to twelve copies of the MS2 fusion protein. In budding yeast cells, the reporter mRNA localizes to the bud and can be observed as a punctate fluorescent particle using MS2-eGFP (**Figure 2.1B**).[4] The observation of a similar fluorescent signal in the presence of the MS2-eDHFR or MS2-SNAP tags and appropriate ligands would be indicative of RNA labeling.

MS2 Coat Protein Systems can Label RNA In Vivo

In the presence of the reporter RNA and the MS2-eDHFR tag, fluorescent particles were observed at the yeast bud tip after incubation with fluorescein-TMP for 45 min. Fluorescein-TMP particles were observed in 29±3% (±SD) of budding cells whereas 34±5% of budding cells expressing MS2-eGFP showed particles (**Table 2.4**). Thus, MS2-eDHFR/fluorescein-TMP appears similarly as efficient in detecting the reporter RNA as MS2-eGFP.

Cells lacking the reporter mRNA failed to produce localized particles when treated with fluorescein-TMP **(Figure 2.1B, Table 2.4**). Consistent with previous observations,[4] a small portion (~10%) of budding cells expressing either MS2-eGFP or MS2-eDHFR was found to possess punctate fluorescence within the mother cell.

RNA particles labeled with fluorescein-TMP were readily detected by visual inspection of

the microscopic images. Interestingly, the average integrated particle intensity for MS2-eDHFR labeled RNAs was lower than that for MS2-eGFP particles (1,870 vs 7,435 AU, respectively). However, the distribution of MS2-eDHFR particle intensities was much sharper than that observed for MS2-eGFP (**Figure 2.2**).

It has previously been observed that two MS2 monomers can be fused to create a tandem dimer (tdMS2).[11] tdMS2-eGFP fusions can display an increase in mRNA brightness and labeling uniformity when compared with MS2-eGFP in mammalian cells.[12] To determine if these observations extended to TMP labeling of RNAs, we expressed a tdMS2-eDHFR fusion protein along with the reporter mRNA in yeast. We predicted that this reporter mRNA would be capable of binding up to six copies of tdMS2-eDHFR (**Figure 2.3**) potentially resulting in half as many eDHFR molecules localized to each transcript when compared to MS2-eDHFR.

Cells expressing tdMS2-eDHFR also displayed fluorescent particles in the presence of the reporter mRNA and fluorescein-TMP (**Figure 2.3**). The brightness of the fluorescent particles observed with tdMS2-eDHFR (1593 AU) was near to those observed with MS2-eDHFR and showed a similar distribution (**Figure 2.3**). However, tdMS2-eDHFR particles potentially have only half the number of fluorescent TMP molecules bound to each transcript. This result may be indicative that TMP dye accumulation in yeast is limiting for the labeling reaction or intermolecular quenching between fluorescein-TMP molecules may be reducing the total fluorescence output. Further analysis of these observations will be required to extend the use of tdMS2-eDHFR fusions for quantitative imaging of RNA abundance.

Given the success of RNA labeling with MS2-eDHFR fusions, we next attempted RNA labeling by covalent attachment of fluorescein-bG to MS2-SNAP proteins. Based upon previous studies,[13] we treated yeast cells expressing MS2-SNAP with fluorescein-bG for 6-10h. After this time, we observed fluorescent particles in the bud tip (**Figure 2.4, Table 2.4**). No particles were observed in the absence of the reporter RNA (**Figure 2.4, Table 2.4**). Incubation for shorter amounts of time (45min) did not result in fluorescent particles (**Figure 2.4**). These results indicate

that while both MS2-eDHFR and MS2-SNAP can both be used to image RNAs in live yeast, longer incubation times with the small molecule fluorophores are required for MS2-SNAP imaging.

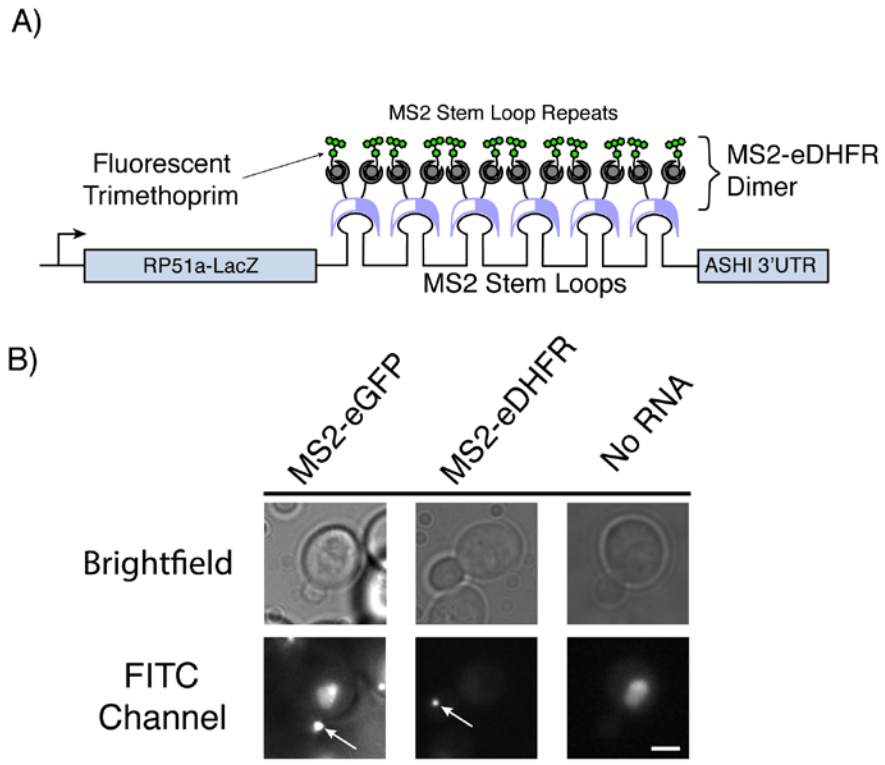


Figure 2.1 RNA labeling using MS2-based systems (A) Illustration depicting the reporter mRNA containing six copies of MS2 protein-binding stem loops. Fluorescent trimethoprim ligands are targeted to the mRNA by fusion of eDHFR to the MS2 protein. Each stem loop can bind one MS2-eDHFR dimer. This illustration represents the maximum possible number of proteins and fluorophores that can be targeted to the RNA. (B) Micrographs of living yeast cells. Visualization of the mRNA (white arrows) can be achieved with both MS2-eGFP as described in reference 3 or by MS2-eDHFR after incubation with fluorescein-TMP. No particles were observed in the absence of the reporter mRNA after incubation with fluorescein-TMP and expression of MS2-eDHFR (far right). Images in the FITC channel have been adjusted to identical contrast settings. Images are of equal magnification and a scale bar (3 μm) is shown in the lower right image.

Expressed Protein	Reporter RNA	Number of Budding Cells	Number of Localized Fluorescent Particles	%(±SD)[a]
MS2-eGFP	-	80	0	0
MS2-eGFP	+	87	30	34±5
MS2-eDHFR[b]	-	311	0	0
MS2-eDHFR[b]	+	307	88	29±3
MS2-SNAP[c]	-	103	0	0
MS2-SNAP[c]	+	97	37	38±5

[a]The error was calculated as the variance for a binomial distribution. [b]RNA particles were visualized after incubation with fluorescein-TMP.[c]RNA particles were visualized after incubation with fluorescein-bG.

Table 2.4 Number of Detected RNA Particles using Various MS2 Fusions

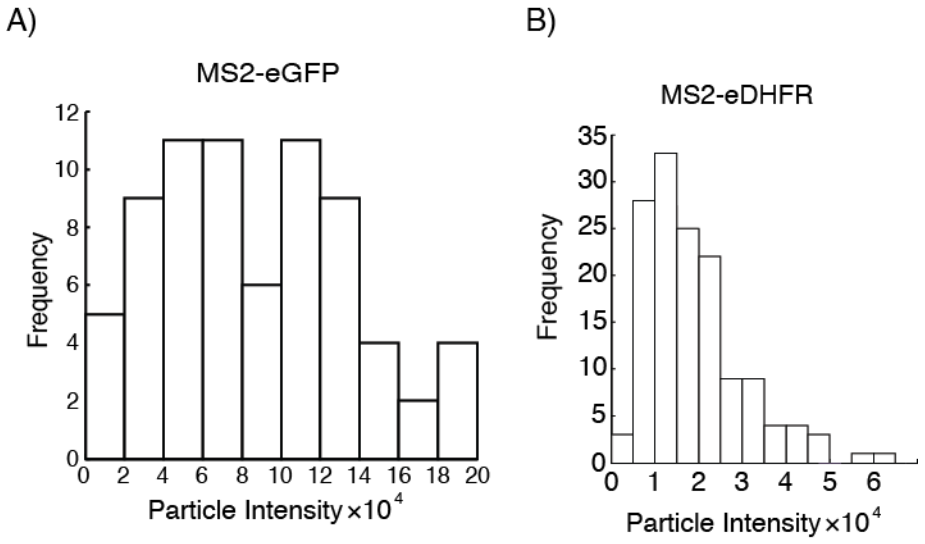


Figure 2.2 Histograms of Particle Intensities using MS2 systems (A) Histogram of particle intensity using MS2-eGFP. (B) Histogram of particle intensity using MS2-eGFP.

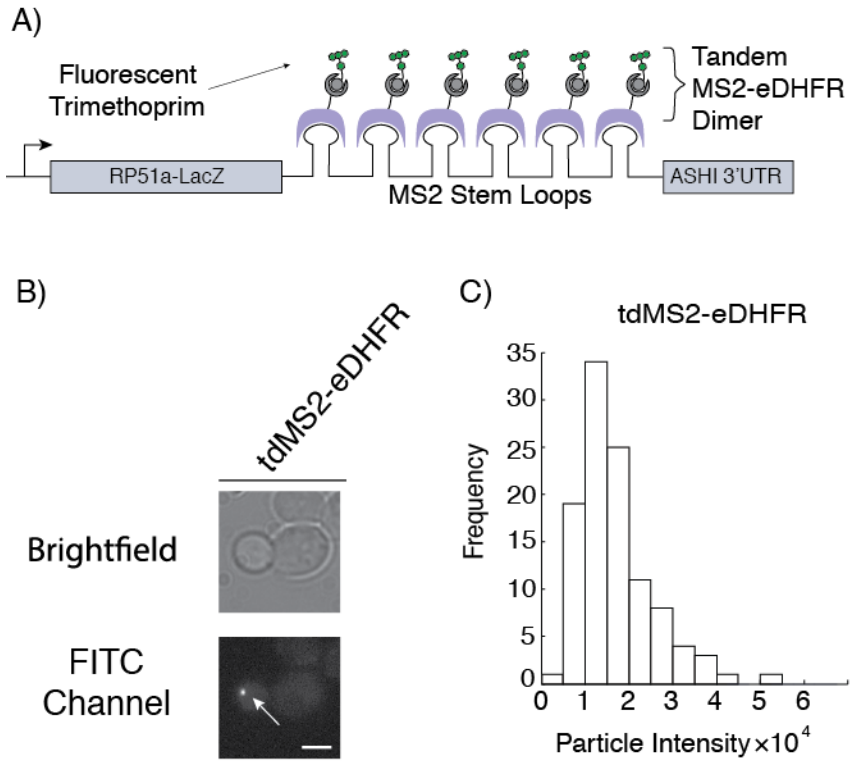


Figure 2.3 RNA localization using the tdMS2-eDHFR system (A) Illustration depicting RNA localization using a tandem MS2 dimer. Each tdMS2-eDHFR protein will bind a stem loop present in the reporter mRNA, resulting in a maximum of 6 copies of eDHFR per mRNA. (B) Micrographs of living yeast cells. mRNAs can be localized to the bud tip using fluorescein-TMP in conjunction with the tdMS2-eDHFR tag. The scale bar represents a length of 3 µm. (C) Histogram of fluorescent particle intensities. The overall distribution of tdMS2-eDHFR is similar to MS2-eDHFR (Supplemental Figure 1).

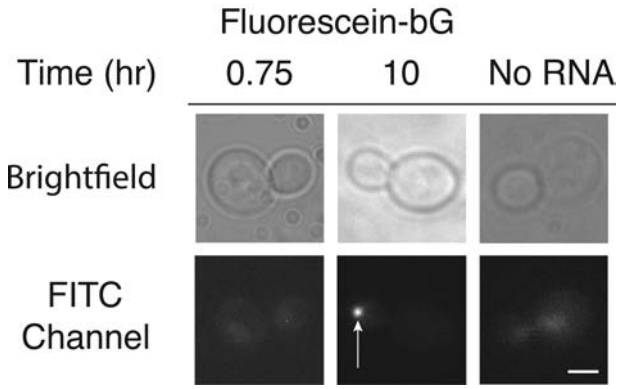


Figure 2.4 RNA visualization using the MS2-SNAP tag after incubation (10hr) with fluorescein-bG. Incubation times shorter than 1hr did not result in mRNA labeling. The white arrow indicates mRNAs localized to the bud tip. No particles were observed in the absence of the reporter mRNA after incubation with fluorescein-bG and expression of MS2-SNAP (far right). Images in the FITC channel have been adjusted to identical contrast settings. Images are of equal magnification and a scale bar (3 µm) is shown in the lower right image.

Organic Fluorophores can Target RNAs *in vivo*

A major advantage of the MS2-eDHFR and MS2-SNAP tags is the ability to spectrally tune the fluorescent signal by switching the derivative of TMP or bG used for labeling. Incubation of cells expressing MS2-eDHFR and the reporter RNA with a red-shifted fluorescein-TMP derivative (hexachlorofluorescein-TMP, hex-TMP) produced fluorescent particles in the bud tip (**Figure 2.5**). Control experiments also showed that some yeast cells exhibited fluorescence in the bud neck in the presence of hex-TMP but absence of the reporter mRNA (data not shown). Despite similar chemical structures, fluorescein-TMP and hex-TMP appear to have unique localization properties in live yeast.

RNAs could also be labeled using the MS2-SNAP tag with red-shifted fluorophores that are structurally distinct from the dihydroxyxanthene/fluorescein scaffold. The rhodamine derivatives Dy505 (SNAP-Cell® 505-Star) and silicon rhodamine (SiRhod)-bG[14] both produced fluorescent particles localized to the bud tip in yeast expressing the MS2-SNAP tag and reporter mRNA (**Figure 2.5**). Neither SNAP-Cell® 505-Star nor SiRhod-bG produced particles in the absence of the reporter mRNA (data not shown). SNAP-Cell® 505-Star is recognized by MS2-SNAP via a CP ligand rather than bG. This indicates that fluorescent derivatives of both CP and bG can be used to label MS2-SNAP proteins for RNA visualization in living yeast. In sum, our results show that the MS2-eDHFR and MS2-SNAP proteins can be used to fluorescently label RNAs *in vivo* using a set of ligands that span from blue-green to near-infrared wavelengths.

Small molecule fluorophore labeling with MS2-eDHFR and MS2-SNAP tags offers several practical advantages for visualizing RNAs in vivo. One obvious benefit is that the fluorescent signal is uncoupled from the MS2-eDHFR or MS2-SNAP protein: fluorescence can be controlled in these experiments by addition of a fluorophore. Cellular RNAs are not fluorescent until the fluorophore is added, and unlike fluorescent proteins, no additional cloning or molecular biology steps or personnel hours are necessary to change the wavelength at which the RNA is monitored. As we have demonstrated, RNAs can be visualized at different fluorescence wavelengths simply

by incubating the same cells with different small molecules. The microscopic observation wavelength is theoretically only limited by the ability to chemically synthesize cell-permeable small molecule conjugates for intracellular labeling.

The use of red-shifted fluorophores such as SiRhod for RNA imaging is also experimentally advantageous. These fluorophores can be excited with limited interference from cellular components (e.g., flavin) that contribute to background autofluorescence when GFP is imaged (**Figure 2.6**). In addition, long wavelength excitation may potentially induce fewer phototoxic effects that result from intense and prolonged illumination with photodamaging blue light.[15] These attributes will make the MS2-eDHFR and MS2-SNAP proteins valuable additions to the RNA imaging toolkit and will enable future microscopy experiments in RNA transport, localization, and turnover. In particular, we believe that these methods will provide a foundation for multi-color pulse-chase experiments of RNA dynamics during cell growth. Such experiments will prove useful for understanding RNA-based regulation.

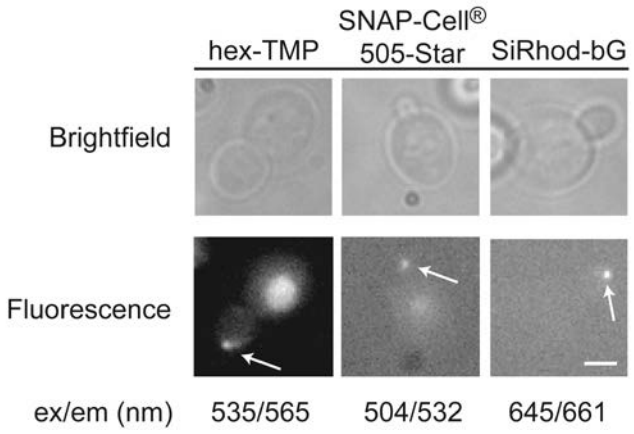


Figure 2.5 Fluorescent observation of localized mRNAs using spectrally different dyes. Incubating yeast cells expressing MS2-eDHFR with HEX-TMP ligand results in fluorescent particles in the daughter cell. The MS2-SNAP tag can be labeled using derivatives of Dy505 (SNAP-Cell® 505-Star) and Si-Rhod. Wavelengths for fluorescence excitation and emission maxima for each fluorophore are shown below the images. Images are of equal magnification and a scale bar (3 μm) is shown in the lower right image.

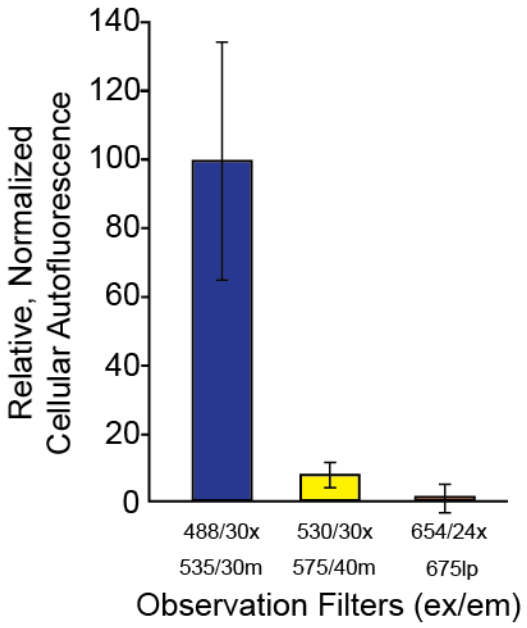


Figure 2.6 Cellular autofluorescence of live *S. cerevisiae* when imaged at three different wavelengths. The cellular background is reduced ~10-fold when imaging is performed at wavelengths capable of exciting dyes such as hex-TMP (530/30x) versus shorter wavelengths commonly used to excite GFP (488/30x). A further ~10-fold reduction is observed at wavelengths that can be used to excite far-red dyes such as SiRhod-bG (654/24x).

ACKNOWLEDGEMENTS

We would like to thank Rob Singer (Albert Einstein College of Medicine) and Ivan Correa (New England Biolabs) for donating reagents and helpful discussions. We thank the Tom Martin laboratory (U. Wisconsin) for microscope access. We thank Jacquelyn Turri for preliminary work on this project. This work was made possible by startup funds from the University of Wisconsin-Madison, WARF, and the Department of Biochemistry. AAH is funded by a K99/R00 award (GM086471) and the Beckman Young Investigators program. TJC is supported by the NIH Biotechnology Training Program at U. Wisconsin-Madison (5T32GM08349).

REFERENCES

1. B. N. G. Giepmans, S. R. Adams, M. H. Ellisman, and R. Y. Tsien, *Science*, 2006, **312**, 217–224.

2. S. Tyagi, *Nat Meth*, 2009, **6**, 331–338.

3. J. S. Paige, K. Y. Wu, and S. R. Jaffrey, *Science*, 2011, **333**, 642–646.

4. E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long, *Mol Cell*, 1998, **2**, 437–445.

5. D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, *Science*, 2011, **332**, 475–478.

6. Z. Chen, V. W. Cornish, and W. Min, *Curr Opin in Chem Biol*, 2013.

7. L. W. Miller, Y. Cai, M. P. Sheetz, and V. W. Cornish, *Nat Meth*, 2005, **2**, 255–257.

8. A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, and K. Johnsson, *Nat Biotechnol*, 2003, **21**, 86–89.

9. X. Sun, A. Zhang, B. Baker, L. Sun, A. Howard, J. Buswell, D. Maurel, A. Masharina, K. Johnsson, C. J. Noren, M.-Q. Xu, and I. R. Corrêa Jr., *Chembiochem*, 2011, **12**, 2217-2226.

10. M. A. McMurray and J. Thorner, *Curr Biol*, 2008, **18**, 1203–1208.

11. D. S. Peabody, *Arch Biochem Biophys*, 1997, **347**, 85–92.

12. Bin Wu, J. A. Chao, and R. H. Singer, *Biophys J*, 2012, **102**, 2936–2944.

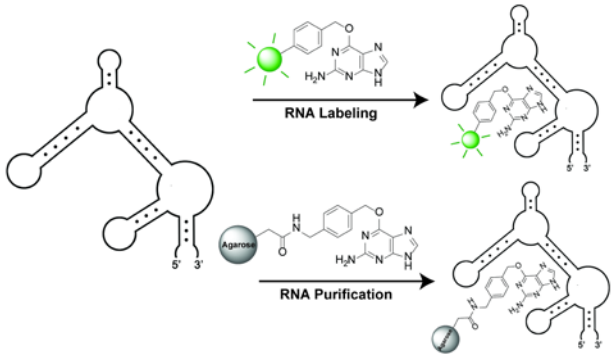
13. C. Chidley, H. Haruki, M. G. Pedersen, C. Fellay, S. Moser, and K. Johnsson, *Chimia (Aarau)*, 2011, **65**, 720–724.

14. G. Lukinavičius, K. Umezawa, N. Olivier, A. Honigmann, G. Yang, T. Plass, V. Mueller, L. Reymond, I. R. Corrêa, Z.-G. Luo, C. Schultz, E. A. Lemke, P. Heppenstall, C. Eggeling, S. Manley, and K. Johnsson, *Nat Chem*, 2013, **5**, 132–139.

15. D. R. Rines, D. Thomann, J. F. Dorn, P. Goodwin, and P. K. Sorger, in *Live Cell Imaging A Laboratory Manual*, eds. R. D. Goldman, J. R. Swedlow, and D. L. Spector, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2nd edn. 2010, pp. 333–350.

The second portion of this chapter was performed in collaboration with Jiacui Xu, a former postdoc in the Hoskins lab. My major contribution was in the design of the library and the selection strategies used in the generation of the benzyl guanine RNA aptamer. This work has been published:

Xu, J., Carrocci, T. J., and Hoskins, A. A. (2016). Evolution and characterization of a benzylguanine-binding RNA aptamer. *Chem. Commun*. **52**, 549–552.



INTRODUCTION

One of the defining features of a chemical approach for studying biology is the specific targeting of small molecules to biological targets (tagging). Protein biochemists have an excellent toolkit for tagging proteins of interest for studying cellular localization, biophysical characterization, or for use in protein purification. Tagging proteins with green fluorescent protein (GFP) or derivatives has been used for a wide-range of microscopy experiments[1] in addition to detection or isolation of GFP-tagged proteins with anti-GFP antibodies.[2] More recently, a number of protein tags have been developed that can be used to target small molecules to proteins of interest. Some of these tags, like the SNAP and Halo tags, are self-labelling, meaning that they covalently react with a suicide substrate such as benzylguanine (bG) or chloroalkane derivatives for the SNAP or Halo tag, respectively.[3,4] Other protein tags can tightly bind small molecules such as the strong interaction between dihydrofolatereductase and trimethoprim conjugates (Ligand Link).[5] Finally, some protein tags act as substrates for other enzymes which can be used to install small molecules (i.e., sortase, acyl carrier protein, or biotin ligase).[6] All of these tags have found use both *in vitro* and *in vivo* and a significant amount of work has provided information on the cell permeability and non-specific interactions of the small molecule substrates of these protein tags.[7] Many of the small molecules that show desirable properties (*e.g.*, cell permeability and low non-specific binding) are commercially available and widely used by protein biochemists.

In contrast with protein labelling technologies, RNA labelling is an emerging field. By far the most common methods for labelling RNAs in cells are hybridization of fluorescent oligos (fluorescence *in situ* hybridization)[8] or targeting GFP-tagged proteins to the RNA with specific RNA sequences (e.g., MS2-labeling)[9,10]. More recently, RNA aptamers have been developed that tightly bind small molecules, including fluorophores and antibiotics (e.g., the Spinach, Mango, or tobramycin aptamers)[11-13]. Some of these aptamers have been used to image RNAs inside cells or to purify RNA complexes. However, many of the small molecule derivatives that interact with these RNA aptamers are not yet commercially available.

We wondered if commercially available, small molecule tools developed for protein labelling and already vetted for cell permeability and non-specific interactions could be repurposed as RNA labelling reagents. Specifically, we sought to develop RNA aptamers that tightly associate with small molecule protein tags. Here we have carried out Systematic Evolution of Ligands by Exponential Enrichment (SELEX)[14] to evolve a tight-binding RNA aptamer against bG—the key functional group recognized by the protein-based SNAP tag (**Figure 2.7A**). Functional and structural probing of the evolved aptamer revealed that it binds with high affinity to bG, guanine, related metabolites are poor competitors for bG binding, and bG appears to induce a change in RNA aptamer structure. Finally, the aptamer can bind several commercially available fluorescent derivatives of bG and be purified from a complex pool of total cellular RNA using commercial reagents—key features for development and wide-spread adoption of an RNA affinity tag.

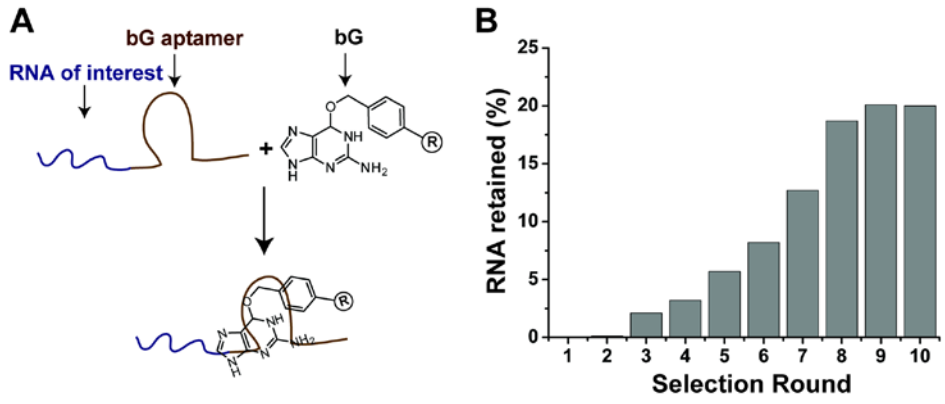


Figure 2.7 Cartoon of a bG-binding RNA aptamer and results from SELEX. (A) A bG-binding RNA aptamer can be used to target small molecule derivatives of bG to RNAs of interest. R represents a functional group such as a fluorophore or biotin. (B) Quantification of the fraction of RNA retained on bG agarose following each round of SELEX.

MATERIALS AND METHODS

Preparation of Agarose Resin for SELEX

For SELEX experiments using O^6-benzylguanine (bG) derivitized agarose (SNAP capture pull-down resin, New England Biolabs), a mock resin was prepared to deplete RNAs that would non-specifically bind to the agarose. Dry *N*-hydroxysuccinimidylester (NHS)-activated agarose resin (33 mg, Life Technology) was first resuspended in 500 µL of phosphate buffered saline (PBS) by mixing end-over-end at 4°C overnight. Tris buffer was then added (500 µL 1 M Tris-HCl, pH 7.4) and mixing was continued for an additional 2 h at room temperature. The resin was then washed with PBS and stored at 4°C.

Preparation of Dynabeads for SELEX

For SELEX experiments using magnetic Dynabeads both bG-derivitized and mock dynabeads were prepared. For synthesis of the bG resin, tosyl-activated Dynabeads (600 µL, Life Technology) were prewashed and equilibrated in borate buffer (600 µL; 0.1 M sodium borate, pH 9.5). Amine-derivitized bG (360 µg; BG-PEG-NH2, New England Biolabs) was dissolved in dimethylsulfoxide (DMSO) to a final concentration of 20 µg/µL. This was then diluted by addition of 432 µL of borate buffer. The resulting solution was then combined with the Dynabeads and incubated with end-over-end mixing at 37°C for 48 h.

After incubation, the beads were then isolated with a magnet and washed twice with 600 µL of DMSO and once with 600 µL of 0.2 M Tris, pH 8.5. The beads were then resuspended in an additional 600 µL of 0.2 M Tris, pH 8.5, and incubated with mixing at 37°C for an additional 4 h. Beads were then isolated with a magnet and washed five times with PBS buffer. The derivitized beads were stored in 600 µL of PBS at 4°C.

Mock Dynabeads were prepared using the same protocol as above except that an amine-derivitized PEG molecule was used in place of bG. NH2-PEG_3-OH (360 µg, ChemPep Inc.) was dissolved in DMSO to a final concentration of 60 µg/µL before dilution with 444 µL of borate buffer.

Subsequent steps for preparation of the mock Dynabeads were identical to those used for preparation of the bG-derivitized Dynabeads.

Preparation of the Starting RNA Library for SELEX

The initial pool of RNAs for SELEX experiments was prepared using a partially structured N24-2 DNA library, in which two 24-base stretches of random nucleotides were separated by a fixed sequence of 12 bases. The fixed sequence (5'-CTG CCG AAG CAG-3') encoded a 4-base pair RNA stem topped by a UUCG tetraloop. The DNA library containing approximately 10^{14} molecules was prepared with a Klenow extension reaction (166 µL) that contained 1 µM of template DNA oligo (5'-GTG ACG CGA CTA GTT ACG GA (N:25:25:25;25)-N24-CTG CCG AAG CAG-N24-TTC ATT CAG TTG GCG CCT CC-3', HPLC purified; IDT), 1 µM of N24-2 reverse primer containing a T7 RNA polymerase promoter sequence (5'-ATG TAA TAC GAC TCA CTA TAG GAG GCG CCA ACT GAA TGA A-3', HPLC purified; IDT), 0.5 mM dNTPs, 0.1 mg/mL BSA, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 1 mM DTT, and 0.05 U/µL of Klenow Fragment (exo-, New England Biolabs).

The entire Klenow extension reaction was then further amplified by large-scale polymerase chain reaction (PCR, 12 mL total volume) in a reaction containing 1 µM of the reverse primer, 1 µM of forward primer (5'-GTG ACG CGA CTA GTT ACG GA-3', HPLC purified; IDT), 0.2 mM dNTP, 1x GoTaq buffer (Promega), and 0.025 U/µL GoTaq DNA Polymerase (Promega). PCR was carried out in 50 µL aliquots with five cycles of 95°C denaturation for 1 min, 54.5°C annealing for 1 min, and 72°C extension for 1 min. The resulting PCR products were combined before being extracted twice with chloroform:phenol:isoamyl alcohol (25:24:1), precipitated with 2 volumes of ethanol, and resuspended in water.

The initial RNA pool was prepared from 100 µg of the PCR-amplified DNA product in a 400 µL T7 RNA polymerase transcription reaction using the T7-flash Kit (Epicentre Biotechnologies). 120 µCi of [^{32}P]-α-UTP (10 µCi/µl , PerkinElmer) was included to aid in RNA quantification and

selection monitoring. Subsequent to transcription, RNase-free DNase I (100 U, New England Biolabs) was added to digest the DNA templates for 1 h at 37°C. The RNA products were then extracted twice with chloroform:phenol:isoamyl alcohol (25:24:1), precipitated with 3 volumes of ethanol and further purified on a 8% denaturing polyacrylamide (PAGE) gel. The RNA was excised from the gel and eluted in 0.3 M NaOAc (pH 5.2) overnight before being ethanol precipitated. RNA pellets were washed with 70% ethanol and dissolved in water before SELEX.

SELEX using bG Agarose Resin

RNAs prepared from the random DNA library were diluted to a final concentration of ~1 µM and volume of 1 mL in Selection Buffer (SB; 40 mM HEPES pH 7.4, 125 mM KCl, 5 mM $MgCl_2$, and 5% DMSO). The RNAs were then heat denatured by incubation at 75°C for 5 min before cooling at room temperature for 10 min.

RNAs were then incubated with 1 mL of the mock agarose resin for 30 min at room temperature with end-over-end mixing. The mixture was then placed into a Micro Bio-Spin column (Bio-Rad) and the flowthrough collected by gravity flow. The flowthrough was then incubated with 1 mL of SNAP capture pull-down resin (New England Biolabs) for 30 min at room temperature with mixing. The mixture was then placed into a second Micro Bio-spin column and washed with 6 x 1 mL of Selection Buffer. RNAs were eluted by incubation of the resin with 1 mL of 5 mM bG (Matrix Scientific) overnight at room temperature with end-over-end mixing and collecting the eluate. The resin was subsequently incubated with another 1 mL of 5 mM bG for 1 h. Both eluates from the bG incubation were then combined.

The eluted RNAs were then ethanol precipitated using glycogen (1 µg, Fermentas) as a coprecipitant before reverse transcription (RT). cDNAs were prepared using the Maxima H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) in a reaction containing 2.5 µM of the N24-2 forward primer and 0.5 mM dNTP. This mixture was then heated to 65°C for 5 min. After cooling on ice, the kit buffer was added followed by 1 µL of the enzyme mix. The reaction was then incubated at 50°C for 45 min before being terminated by heat denaturation at 85°C for 5 min. RT

products were then amplified by PCR in total volumes of 200-400 µL using GoTaq DNA polymerase before subsequent *in vitro* transcription in the presence of 10 µCi of [^{32}P]-α-UTP for the next round of SELEX. In total, ten rounds of RNA transcription, bG agarose binding and elution, cDNA synthesis, and PCR were carried out.

SELEX using bG DynaBeads

SELEX using magnetic bG DynaBeads was carried out essentially as described for bG agarose except that 1 mL of mock or bG-derivatized DynaBeads were used.

Cloning and Sequencing of DNAs Encoding RNA Aptamers

PCR-amplified DNAs obtained after 10 rounds of SELEX using either the bG agarose or bG Dynabeads were ligated into the pJET1.2/blunt cloning vector (Thermo scientific CloneJET PCR Cloning Kit). Sequencing results were aligned using Clustal Omega (EMBL). Prior to SELEX samples of the initial DNA pool were also sequenced to confirm library diversity.

Preparation of Aptamers JX1-6 and JX1 derivatives

Aptamers for structural and binding studies were prepared by *in vitro* transcription. Templates for transcription of aptamers JX1-6 were prepared by PCR amplification of the corresponding regions from the pJET1.2 plasmids used for sequencing. The resulting DNAs were then used as templates for transcription with T7 RNA polymerase. RNAs were then purified by HiTrap Q column (GE Healthcare), precipitated with ethanol, resuspended in RNase free H_2O, and quantified by measuring the UV absorbance at 260 nm. Transcription templates for derivatives of JX1 were prepared by PCR amplification of ssDNA templates (IDT).

Column binding assay

[^{32}P]-labeled RNA (100 pmol) was added to SB (50 µL) and then incubated with either bG agarose or mock resin (50 µL) for 30 min. The resin was then washed 20 times with SB (50 µL for each wash). Bound RNA was then eluted by addition of 10 mM bG. RNA concentration in the wash and elution buffers were determined by scintillation counting.

Fluorescence Polarization Assays to Determine Aptamer Binding Affinity

Binding assays (60 µL) were carried out with a fixed concentration of SNAP-Surface ® 488 (Atto488-bG, 50 nM, New England Biolabs) or SNAP-Surface Dy549 ® (Dy549-bG, 50 nM, New England Biolabs) and addition of RNA aptamers from 0.1-10 µM. Prior to each assay, the RNA aptamers were diluted into SB, unfolded by heating to 75^{o}C for 5 min, and then cooled at room temperature for 10 min. Binding reactions were carried out at room temperature by addition of the fluorophore followed by incubation at room temperature for 30 min prior to collection of fluorescence data.

Fluorescence polarization (FP) data were collected using a Infinite M1000 Pro plate reader (Tecan) in 96-well flat bottom, black polystyrol plates (Corning Incorporated). For monitoring binding of Atto488-bG, excitation and emission wavelengths were set to 470 and 520 nm, respectively, with a 5 nm bandwidth. Binding of Dy549-bG was monitored with excitation and emission wavelengths set to 530 and 575 nm, respectively, with a 5 nm bandwidth. The instrument gain was set to 100 and 10 flashes were collected from each well with a 30 ms settle time between wells. Data were background corrected using a buffer blank and collected in triplicate. Reported values represent the mean and standard deviation. FP data were analyzed and fit to Equation (1) using Origin 8.0 (OriginLab Corporation) where FP_{max} represents the maximum change in fluorescence polarization, FP_{min} the FP value in the absence of aptamer, E_o represents the aptamer concentration, and ΔFP represents the measured change in FP at each aptamer concentration.

$$\Delta FP = \frac{(FP_{max} - FP_{min}) \times E_0}{K_d + E_0} \quad (1)$$

Competitive Binding Assay

Binding of non-fluorescent molecules was measured by their ability to compete with Atto488-bG (50 nM) in the FP assay. FP experiments were carried out as described above except for addition of the competitor during binding. Guanine (Tokyo Chemical Industry Co., LTD.) and guanosine

(Chem-IMPEX International, Inc.) were added to a final concentration of 250 µM due to their limited solubility.

RNA Aptamer Structural Prediction

Prediction of RNA secondary structure by free energy minimization was done by the MFold program.

RNase T1 digestion assay

1 pmol of 5′-[^{32}P] labeled JX1 RNA were dissolved in a volume of 10 µl containing 40 mM Hepes buffer (pH 7.4), 125 mM KCl, 5 mM $MgCl_2$, in the presence of 1µg Yeast tRNA (Life Technology). After unfolding for 5 min at 75 ℃, the RNA was left at room temperature for 10 min. Then the binding reaction was performed at room temperature for 30 min with the addition of 1 µL DMSO or 1 µl of 2 mM BG. The RNase T1 (Thermo Scientific) digestion reactions were performed at 37 ℃ for 10 min by adding 0.1 U (1 µL) RNase T1 and subsequent immediate quenching of the reaction on ice and addition of gel loading buffer.

G-specific ladder:

G-specific sequencing ladders were generated by incubation of 1pmol of 5'- ^{32}P labeled JX1 RNA with 1 µg *yeast* tRNA in 10 µl 12.5 mM sodium citrate buffer (pH 4.5) containing 3.5 M urea and 0.5 mM EDTA. After denaturation of the RNA for 10 min at 55°C, 0.1 U (1µl) RNase T1 were added and the sample was incubated for additional 10 min at 55°C. Then the reaction was terminated by addition of gel loading buffer.

Native Purification of JX1 from Yeast Total RNA

JX1 (10 nM) was combined with yeast total RNA (1 µg, prepared by hot phenol extraction of *S. cerevisiae*) in Selection Buffer in a final volume of 50 µL. The mixture was then heated to 75ºC for 5 min and then cooled at room temperature for 5 min. This mixture was then incubated while mixing end-over-end at room temperature for 30 min with bG agarose (50 µL) that had been prewashed with selection buffer. The flowthrough was collected and the resin washed five times

with selection buffer in column format. Bound RNAs were eluted by incubating the resin with 50 µL of Selection Buffer containing 5 mM benzyl guanine with end-over-end mixing at room temperature for 1 h and then collecting the eluate. Samples were combined with an equal volume of 95% deionized formamide before being loaded onto a denaturing 10% acrylamide gel. After electrophoresis, bands were identified by ethidium bromide staining.

Results

To prepare RNAs for SELEX, we created a double stranded DNA transcription template library by PCR amplification of an oligonucleotide containing two, 24-nucleotide (nt) regions of random sequence (**Figure 2.8**). The random sequence regions were separated by a small, 12 nt constant region that would encode a hairpin topped by a UUCG tetraloop in the transcribed RNAs. A similar, partially-structured library design has been used to evolve other aptamers and can lead to high binding affinities.[11,15] The final library was predicted to contain approximately 10^{14}unique sequences and was used to prepare RNAs by *in vitro* transcription. Sequencing of samples taken from this library confirmed that it contained the expected random sequences separated by the constant region.

Selection of bG aptamers using SELEX

We then carried out SELEX using two different solid supports in parallel: commercially available bG-derivitized agarose (SNAP-Capture Pull-Down Resin, New England Biolabs) and bG-deritivized magnetic DynaBeads that were prepared by incubation of an amine derivative of bG with *p*-toluene sulfonyl-activated beads. In both cases, RNAs were incubated with the support, eluted with high concentrations of free bG, and reverse transcribed to prepare cDNA for subsequent rounds of selection. We followed the progress of the selection by incorporation of [^{32}P] into the RNAs to accurately quantify how much RNA was being retained on the support.

Upon quantification of RNA binding during each round of SELEX, we obtained very different results using bG agarose compared with bG Dynabeads. With bG agarose, the fraction of RNA (as measured by [^{32}P] counts per minute) retained on the resin increased from background levels during the first round of SELEX to ~20% after the ninth round (**Figure 2.7B**). Since no additional enrichment was observed between rounds nine and ten, we did not carry out further rounds of selection. In contrast, we observed no significant enrichment of retained RNAs using bG Dynabeads after ten rounds of selection (data not shown). Nonetheless, we submitted DNAs prepared from reverse transcription of eluted RNAs from both bG agarose and bG Dynabeads for

sequencing.

Surprisingly, some of the same RNA sequences were found from samples obtained from both bG agarose and bG Dynabeads (**Table 2.5**). The JX1 sequence was found in 38% (23 of 60) of sequenced samples obtained from the bG agarose SELEX and 14% of sequenced samples obtained from bG Dynabeads (5 of 36). Since we obtained the same RNA sequence from two different selections, it strongly suggested that the JX1 sequence bound bG. Four other sequences (JX2-5) were obtained from bG agarose that were less abundant than JX1, and of these JX2 was also observed when bG Dynabeads were used. The most common sequence selected using bG Dynabeads (JX6) was present in 25% of those samples (9 of 36) and was not found in the sequencing results from the bG agarose selection.

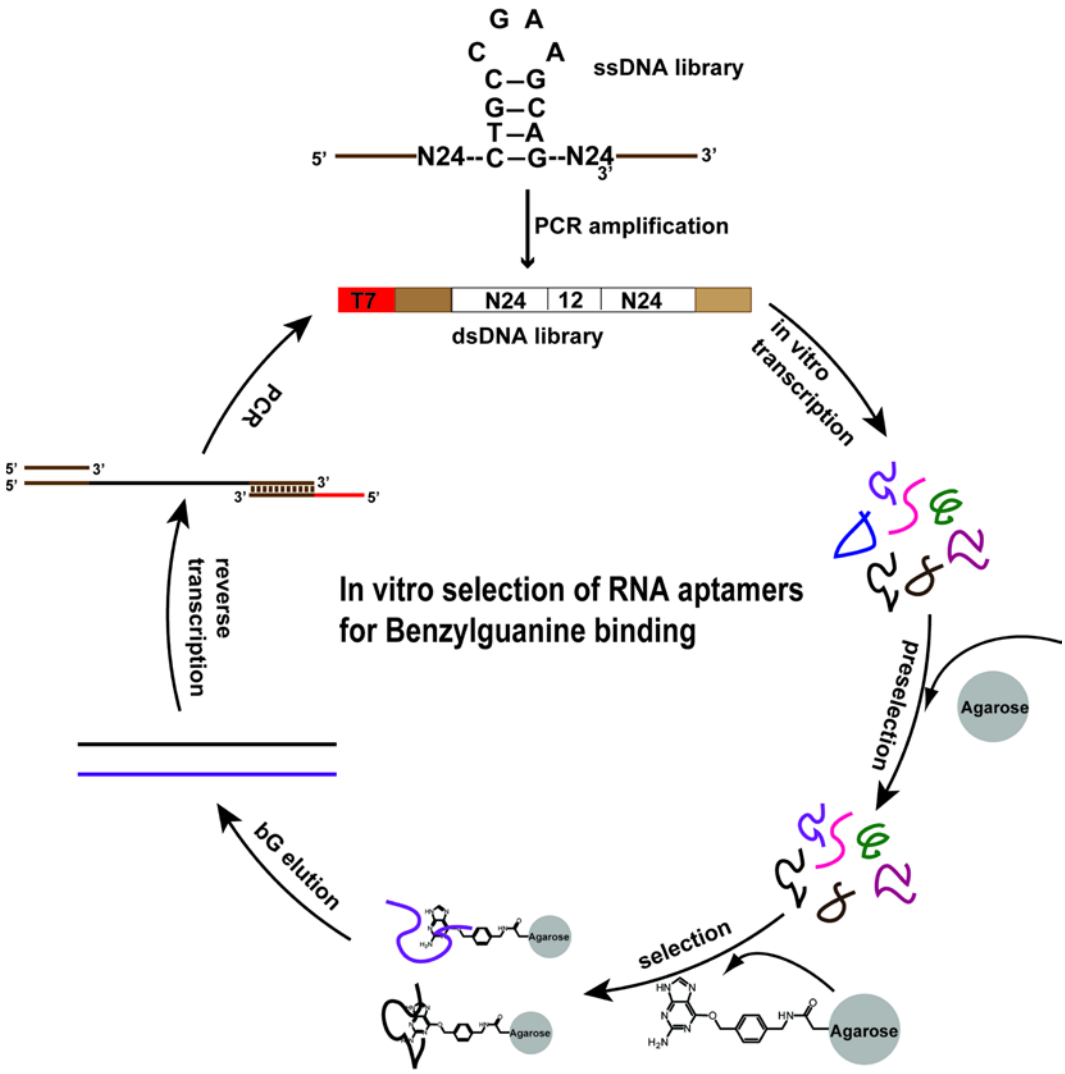


Figure 2.8 Scheme for carrying out SELEX for evolution of the bG binding aptamer. The SELEX experiment began by PCR amplication of a ssDNA library containing two, 24-nt regions of random sequence separated by a common hairpin-forming region. RNAs were then prepared by transcription, non-specific interactors removed by incubation with a mock agarose resin, and incubated with bG-agarose. RNAs were eluted by incubation with bG before cDNAs were prepared and amplified. A similar protocol was used when DynaBeads were used in place of agarose.

RNA Name	Sequences	AE	BE
JX1 RNA	GGAGGCGCCAACUGAAUGAACGAGUUGAAAUGCGGAUCAAGUUACUGCUUCGGCAGUACCUAGUGG GAAAAGUUCUUGGGUCCGUAACUAGUCGCGUCAC	23	5
JX2 RNA	GGAGGCGCCAACUGAAUGAAACCGUGCUAAGGAGUUGGACUCCGCUGCUUCGGCAGAGCGAUAUAG UGCCAGCUGAUACAUCCGUAACUAGUCGCGUCAC	20	7
JX3 RNA	GGAGGCGCCAACUGAAUGAAAGGUUACGCUUCAACGCCGUGCAGCUGCUUCGGCAGGCUUGUGAG UACAUGGCCCAUUGAUCCGUAACUAGUCGCGUCAC	3	
JX4 RNA	GGAGGCGCCAACUGAAUGAAUGUUCUGGAUCUGCUAGAGACUAUCUGCUUCGGCAGCACUCUCCGU UAACAACACGUACAUCCGUAACUAGUCGCGUCAC	5	
JX5 RNA	GGAGGCGCCAACUGAAUGAAAGAGGUAACGAAAGAUGAUUGUGCCUGCUUCGGCAGUACAACUUCU CGGAUAAAUGAUUAUCCGUAACUAGUCGCGUCAC	2	
JX6 RNA	GGAGGCGCCAACUGAAUGAAGGUGUAUUUCAGUUAAAGAACUUGCUGCUUCGGCAGCGUUCUGAGU UACUAGUCCAAUAAUCCGUAACUAGUCGCGUCAC		9

Table 2.5 Nucleotide sequences of the JX1-6 RNA aptamers. AE and BE indicate the number of occurrences of each aptamer sequence in the sequencing results from RNAs eluted from agarose (AE) or DynaBeads (BE). Red nucleotides indicate the primer binding regions common to all aptamers, while green nucleotides indicate the common UUCG tetraloop region.

Characterization of Identified Aptamers

To directly test that the evolved sequences were capable of binding bG independent of the SELEX protocol, we constructed a transcription template encoding the most frequently observed aptamer sequence (JX1) and transcribed the DNA to produce a [^{32}P]-labeled product. We then tested the RNA for specific binding to bG agarose and improved binding to bG agarose when compared with the starting RNA pool. The JX1 aptamer strongly associated with bG agarose, which was able to capture ~70% of the input RNA **(Figure 2.9A**). The bG agarose resin was ~300-fold more selective for JX1 compared to the mock agarose resin. Additionally, we observed little binding of the initial RNA pool to bG agarose suggesting that only certain RNA sequences are capable of tightly binding bG agarose **(Figure 2.9B)**. Together these results indicate that JX1 is binding specifically to bG rather than non-specifically associating with the resin.

To measure the affinity of the selected RNA aptamers for bG, we used a fluorescence polarization (FP) assay[16] and the fluorescent bG derivative Atto488-bG (SNAP-Surface 488®, New England Biolabs). In this assay, FP of Atto488-bG would increase when the dye is bound by the aptamer (**Figure 2.10A**). We transcribed, purified, and measured the binding affinity of aptamers JX1-6 for Atto488-bG. As expected, JX1, the most frequently observed aptamer sequence, also bound Atto488-bG the most tightly (K_d = 219 nM) compared to JX2-6. No change in FP was observed when the initial RNA pool was incubated with the fluorophore **(Figure 2.11).** The other aptamers displayed a range of binding affinities from 0.4-6.6 μM (**Figure 2.10C and 2.12)**. Based on these results, we selected JX1 for further analysis.

To test how well JX1 binds bG derivatives other than Atto488-bG we carried out the FP assay with a bG derivative containing a structurally and spectrally different fluorophore (Dy549-bG, SNAP-Surface 549®, New England Biolabs). We were able to observe a change in FP upon incubation of Dy549-bG with the JX1 aptamer; however, analysis of the binding data revealed a decrease in binding affinity (K_d= 1070 nM, **Figure 2.11**). The decrease in affinity may be due to

unfavorable interactions between the tetrasulfonated cyanine fluorophore of Dy549 and the RNA. Less negatively charged molecules, like the disulfonated fluorone derivative Atto488-bG, may interact more favorably with RNA and lead to higher affinities.

To test how selective JX1 is for bG over other guanine metabolites we used a FP competition assay[17]. In this experiment, we would predict that competitors would result in an increase in free Atto488-bG and a corresponding loss in FP (**Figure 2.13A**). In all cases, we introduced the competitor and Atto488-bG simultaneously to JX1. As a proof-of-concept, non-fluorescent bG (20 µM) was able to effectively compete with Atto488-bG (50 nM) and reduce the observed FP to near background levels (**Figure 2.13B**). This result also indicates that bG lacking PEG linkers or functional groups is able to be bound by JX1, in agreement with isolation of RNAs during SELEX by bG elution.

We next analyzed changes in FP upon addition of guanine, guanosine, guanosine 5' – monophosphate (GMP), or guanosine 5'-triphosphate (GTP) (**Figure 2.13B**). Unlike with bG, we observed no change in fluorescence polarization with any of these metabolites when they were introduced in 400-fold excess of the Atto488-bG. At even higher concentrations of competitor (100 µM), we observed a slight decrease in FP with guanine but not the other molecules. GMP and GTP did not change the FP even when included at 1 mM concentration (20,000-fold excess). At 10 mM GMP and GTP we did observe a loss in FP; however, at these high concentrations it is possible that the changes observed could also be due to quenching of the fluorophore by GMP or GTP[18]in addition to competition for the ligand binding site. From these data, we conclude that common guanine metabolites are poor competitors of Atto488-bG for binding to JX1. These data also may indicate that JX1 recognizes bG through interaction with the benzyl moiety and/or through interactions with the guanine N9 position, which is available for hydrogen bonding in bG but part of the glycosidic linkage in guanosine, GMP, and GTP.

Structural analysis JX1 by Mfold[19] suggests that the RNA can fold into a complex structure containing multiple stem loops (**Figure 2.14A**). To further analyze the structure, we carried out

enzymatic cleavage assays using RNase T1[20] both in the presence and absence of bG (**Figure 2.14B**). Analysis of the T1 digest showed significant changes between the free aptamer and after complex formation with bG. The data support a conformational change occurring in the RNA that leads to shielding of guanines in putative stem loops I and III after bG binding. In contrast, guanines located between these regions become more susceptible to RNase T1 cleavage. The latter observation was corroborated by increased cleavage at U64 during in line probing (data not shown).

Interestingly, we also observed changes in JX1 in the presence of bG that correspond to locations found within the primer binding site used in the SELEX protocol. This suggests that these regions may also be important for the structure of JX1 and bG binding. In support of this, we truncated JX1 to remove 30 nt from the 5' and 13 nt from the 3' end of the RNA (JX1, Δ1-30, Δ87-100) and analyzed Atto488-bG binding by FP. The results showed little evidence for Atto488-bG association (**Figure 2.14C**)—confirming our prediction that regions at the very 5' and 3' ends of JX1 are important for aptamer function.

We next attempted to minimize JX1 by removal of the constant region encoding a UUCG-tetraloop that was present in the initial DNA library (**Figure 2.8**). Deletion of this region resulted in an aptamer (JX1, Δ41-60) that was still functional for Atto488-bG binding. However, the binding affinity was reduced ~4-fold (K_D = 957 nM, **Figure 2.14C**).This indicates that not only are primer binding regions used in the SELEX protocol essential for binding, but the UUCG tetraloopregion also plays an important functional role. In sum, these experiments suggest that JX1 is a conformationally complex and potentially dynamic RNA aptamer.

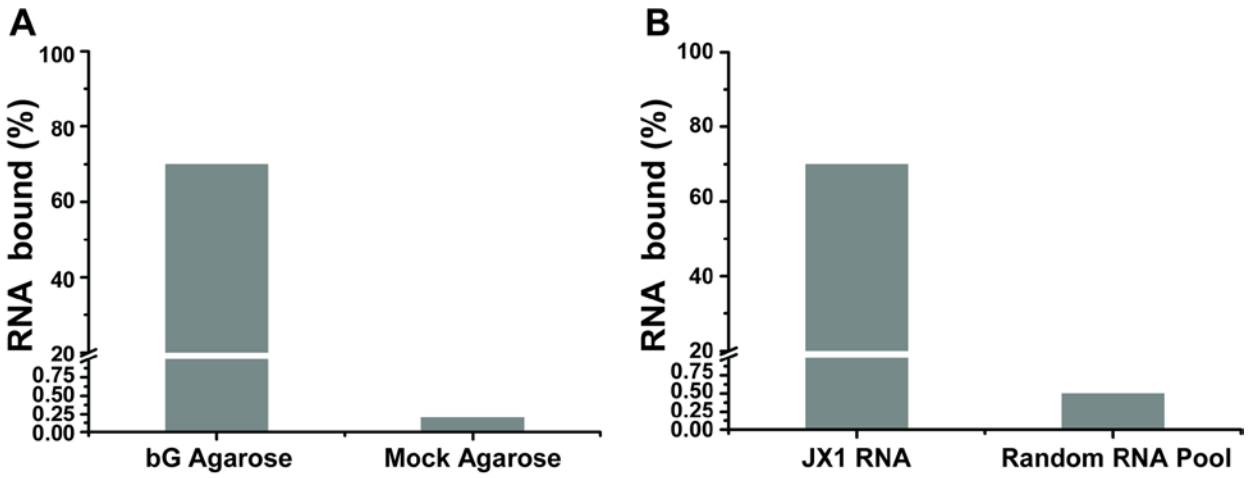
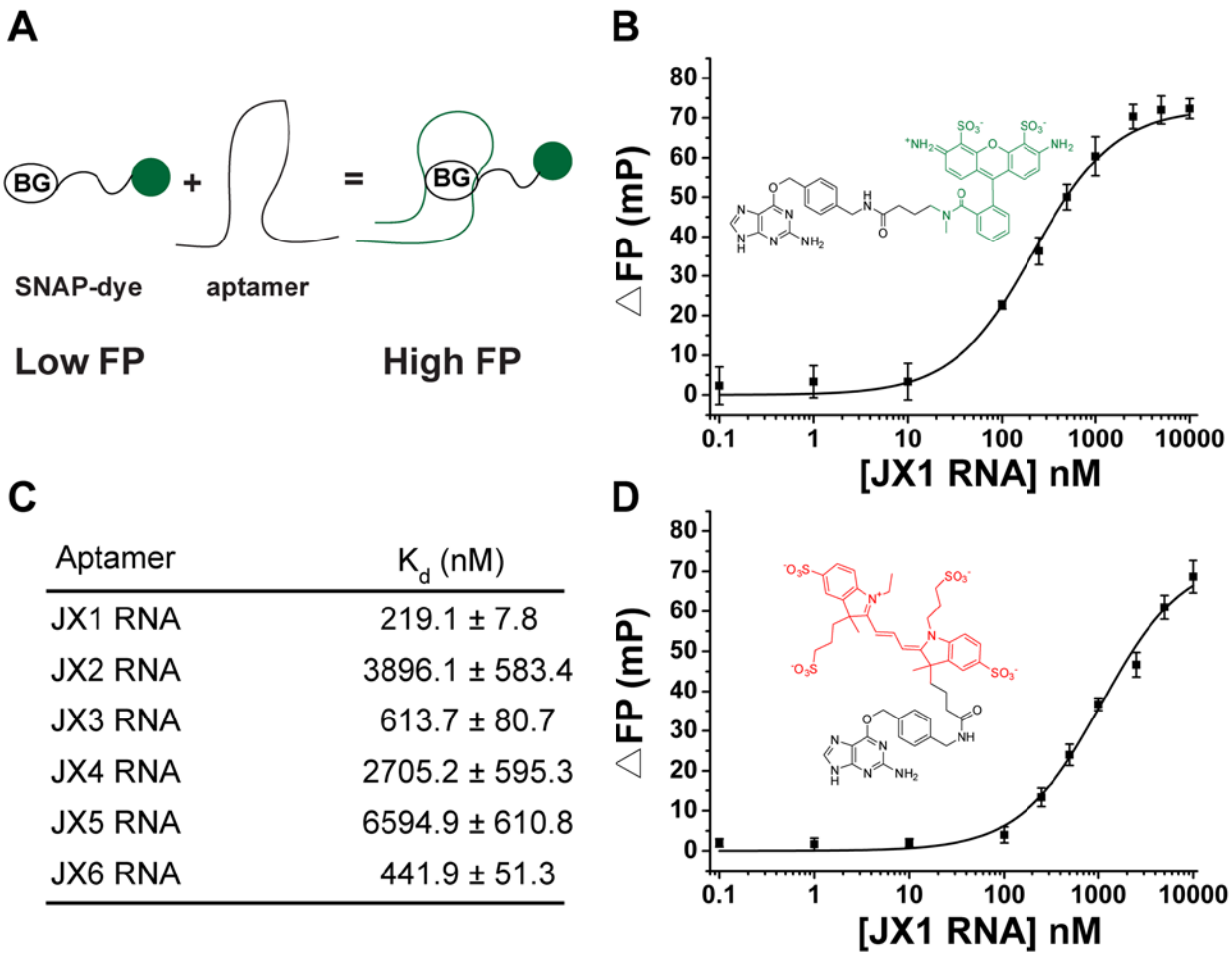


Figure 2.9 Binding of the JX1 aptamer to bG agarose. (A) The JX1 aptamer specifically binds to bG agarose but not to a mock agarose resin lacking bG. (B) JX1 binds to bG agarose much more efficiently than the starting random RNA pool used in SELEX.



C

Aptamer	K_d (nM)
JX1 RNA	219.1 ± 7.8
JX2 RNA	3896.1 ± 583.4
JX3 RNA	613.7 ± 80.7
JX4 RNA	2705.2 ± 595.3
JX5 RNA	6594.9 ± 610.8
JX6 RNA	441.9 ± 51.3

Figure 2.10 FP assay for measuring affinity of evolved aptamers to fluorescent bG derivatives. (A) Cartoon depicting the FP assay. An increase in FP of the bG derivative is expected upon aptamer binding. (B) Results from FP analysis of Atto488-bG (shown) binding to JX1. (C) Results of FP analysis of Atto488-bG binding to aptamers JX1-6. (D) Results from FP analysis of Dy549-bG (shown) binding to JX1 showing a calculated K_d of 1070±160 nM. In (B) and (C), each data point represents the average of three measurements ± S.D.

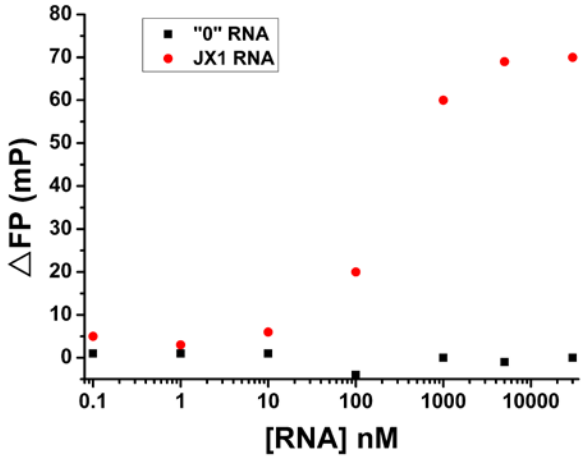


Figure 2.11 The starting, random RNA pool (black squares) shows no change in FP when incubated with Atto488-bG compared with the JX1 aptamer (red circles).

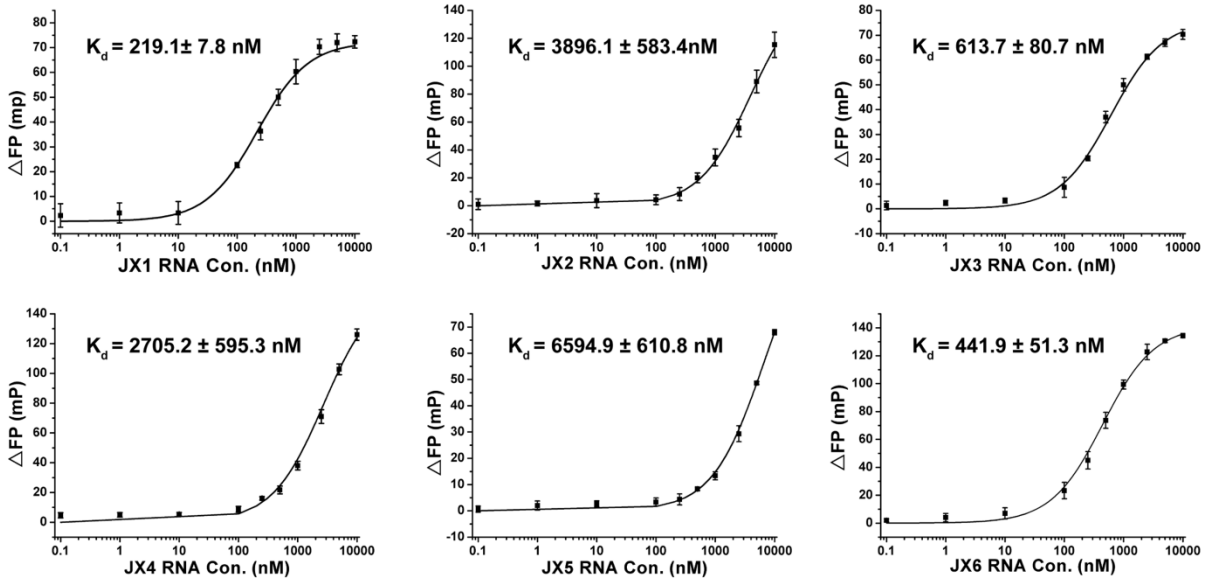


Figure 2.12 Binding affinity of the JX1-6 aptamers for Atto488-bG as measured by FP. Each point represents the average of three measurements ±S.D.

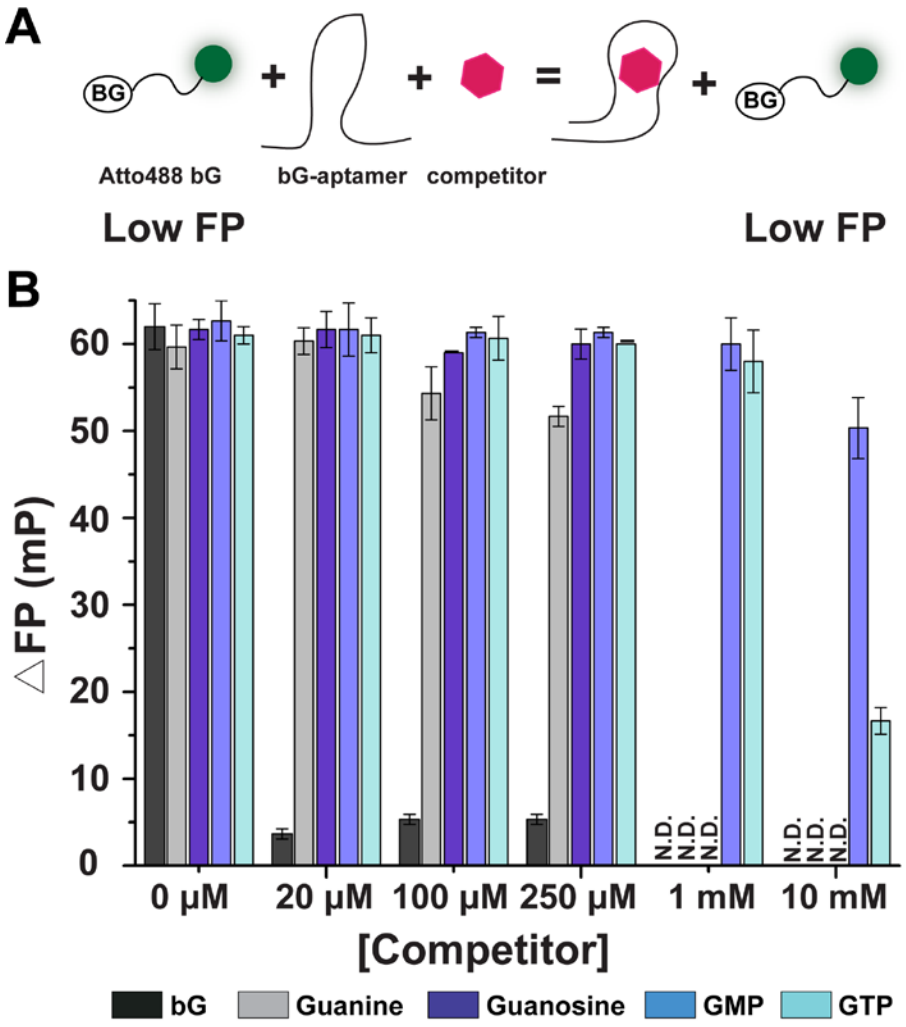


Figure 2.13 FP competition assay for monitoring specificity of JX1 for bG. (A) Cartoon showing how an effective competitor of Atto488-bG for JX1 binding prevents binding of the RNA to the fluorophore results in low FP. (B) Results from the competition assay for selected guanine-containing molecules. High ΔFP values result from Atto488-bG binding to JX1 and lack of inhibition from the competitor. Each bar graph represents the average from three separate experiments and error bars represent ± S.D. N.D., Not Determined.

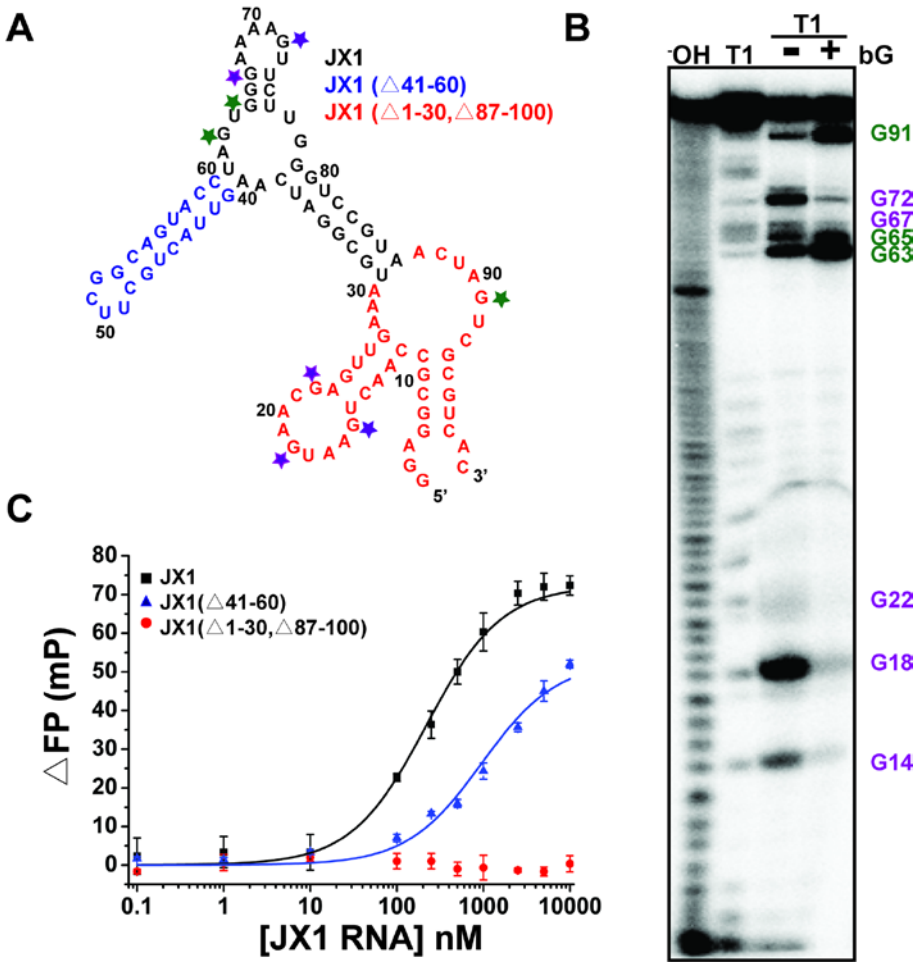


Figure 2.14 Structural characterization of JX1. (A) Secondary structure prediction of JX1 based on Mfold analysis. Regions truncated in efforts to minimize JX1 are shown in blue and red. Purple and green stars represent nucleotides that showed decreased or increased RNase T1 cleavage after addition of bG. (B) Results from RNase T1 digestion of JX1. Lane 1, RNA ladder produced by alkaline hydrolysis. Lane 2, RNase T1 digestion under denaturing conditions. Lanes 3 and 4, RNase T1 digestion under native conditions in the presence of absence of bG. Positions of increased or decreased RNase T1 cleavage seen upon addition of bG are noted. (C) FP assay of minimized JX1 RNAs (red and blue) compared with full-length JX1 (black). No Atto488-bG binding was observed with the JX1(Δ1-30, Δ87-100) RNA (red). The JX1(Δ41-60) RNA was able to bind Atto488-bG (K_d = 957±52).

JX1 Purifies RNA from Heterogeneous Mixtures

Two key features of the JX1 aptamer are the commercial availability of bG-deritivitized reagents and the ability to elute the JX1 aptamer from bG agarose under native conditions without RNA unfolding. We wondered if these features could be used to purify JX1 from a complex RNA mixture. To test this hypothesis, we combined JX1 (10 nM) with unfractionated yeast cellular RNA (total yeast RNA, 1 µg). The resulting RNA mixture was then heated to denature the RNAs and allowed to cool before incubation with commercially available bG agarose. The resin was then placed into a column, and the flow through was collected. Bound RNAs were eluted by first washing the column with five volumes of buffer and then eluting with a solution of commercially available bG (5 mM). RNAs present at each stage were then visualized by denaturing polyacrylamide gel electrophoresis and ethidium bromide staining.

Under these conditions, JX1 was readily purified from total yeast RNA and eluted from bG agarose without the use of a denaturant (*e.g.,* urea, formamide, or EDTA). Recovery of JX1 was ~50% based upon analysis of band intensities from the input and eluate samples (**Figure 2.15A**). No band of similar size to JX1 was present in either the input or eluate from a control experiment in which the aptamer was not included with the yeast total RNA (**Figure 2.15B**). These results reveal not only can JX1 be purified from cellular RNAs under native conditions but also that JX1 can fold into a structure capable of binding bG agarose after first being denatured in the presence of cellular RNAs.

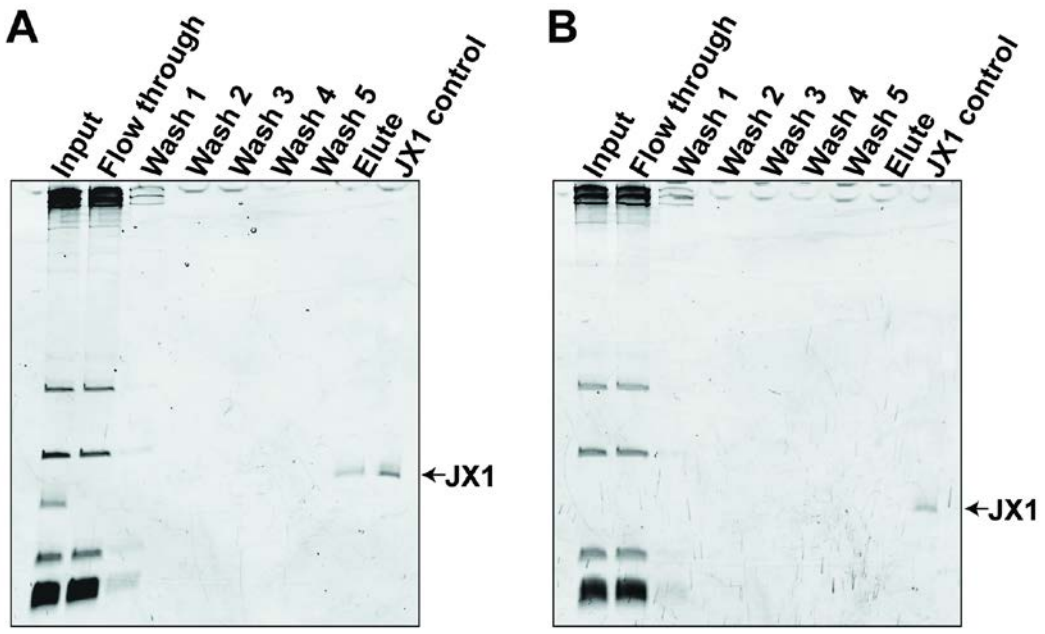


Figure 2.15 Purification of JX1 from total yeast RNA. (A) JX1 can be purified and eluted with bG under native conditions in the presence of cellular RNAs. (B) In a control experiment, no band of similar size to JX1 could be observed in either the input or eluate if JX1 was omitted from the reaction mixture.

Conclusions

We have evolved an RNA aptamer that specifically and tightly binds bG ligands that currently in wide-spread use as substrates for the protein SNAP tag. We have shown that these ligands can be used to both target fluorphores to the aptamer as well as purify the aptamer from a complex mixture of nucleic acids under native conditions. Development of a bi-functional small molecule library—whose members can be used to label either proteins or RNAs—could represent a valuable addition to the chemical biology toolkit. Further engineering of the JX1 aptamer may lead to other, more diminutive bG binding or reactive aptamers that can function analogously to the protein SNAP tag.

ACKNOWLEDGEMENTS

We thank the Dan Stevens and the Biophysical Instrumentation Facility (BIF, Dept of Biochemistry, U. Wisconsin-Madison) for access to microplate fluorometers for FP measurements. JX, TJC, and AAH are supported by a Beckman Young Investigator award to AAH and startup funds from the Dept. of Biochemistry, WARF, and the U. Wisconsin-Madison. AAH is also supported by a Shaw Scientist Award from the Greater Milwaukee Foundation. TJC is also supported by the NIH Biotechnology Training Program at U. Wisconsin-Madison 5T32GM08349) and a William H. Peterson Fellowship.

REFERENCES

1. D.M. Chudakov, M.V. Matz,S. Lukyanovand K.A. Lukyanov, *Physiol. Rev.*,2010,**90,**1103.

2. R.Zhuang, Y.Zhang, R. Zhang, C.J.Song, C.H.Song, K. Yang, A. Yang and B.Q. Jin, *ProtExpr&Pur.*, 2008, **59**, 138.

3. A. Juillerat, T. Gronemeyeret, A. Keppler, S. Gendreizig, H. Pick, H. Vogel and K. Johnsson, *Chem Biol.,* 2003,**10**, 313.

4. G.V.Los and K. Wood, *Methods Mol Biol.,* 2007, **356**, 195.

5. L.W. Miller, Y.F. Cai, M.P.Sheetz and V.W. Cornish. *Nature Method. 2005,***2**, 255.

6. M. Rashidian, J.K. Dozier and M.D. Distefano, *Bioconjugate Chemistry*, 2013,**24**, 1277.

7. P. J. Bosch, I.R. CorrêaJr, M. H. Sonntag, J. Ibach, L. Brunsveld, J. S. Kanger and V. Subramaniam. *Biophys. J.*, 2014, **107**, 803.

8. A.M. Femino, F.S. Fay, K. Fogarty and R.H. Singer, *Science*. 1998,**280,** 585.

9. E. Bertrand, P. Chartrand, M. Schaefer, S.M. Shenoy, R.H. Singer and R.M. Long. *Mol. Cell.,*1998.**2**.437.

10. T. J. Carrocci and A. A. Hoskins, *Analyst*, 2014, **139**, 44.

11. J.S. Paige, K.Y. Wu, S.R. Jaffrey, *Science*, 2011,**333**, 642.

12. E.V. Dolgosheina, S.C. Jeng, S.S. Panchapakesan, *R.*Cojocaru, P.S. Chen, P.D. Wilson, N. Hawkins, P.A. Wiggins and P.J. Unrau. *ACS Chem. Biol.,* 2014, **9,** 2412.

13. I. Shin, J. Ray, V. Gupta, M. Ilgu, J. Beasley, L. Bendickson, S. Mehanovic,G.A. Kraus and H.M. Nilsen, *Nucleic Acids Res.,* 2014,**42**,e90.

14. C. Tuerk and L. Gold, *Science*, 1990, 249, 505.

15. J.H.Davis, and J.W.Szostak, *Proc. Natl.Acad.Sci. U.S.A.,* 2002, **99,** 11616.

16. J.M. Pagano, C.C. Clingman and S.P. Ryder, *RNA,* 2010, **17,** 14.

17. X. Geng, D.P. Zhang, H.L. Wang and Q. Zhao. *Anal Bioanal Chem*., 2013, **405**, 2443.

18. I.Nazarenko, R. Pires, B. Lowe, M. Obaidy and A. Rashtchian, *Nucleic Acid Res*., 2002, **30**, 2089.

19. M. Zuker. *Nucleic Acid Res*., 2003, **31**, 3406.

20. M.F. Carey, C.L. Peterson, S.T. Smale, *Cold Spring HarbProtoc.* 2013, **3,** doi: 10.1101/pdb.prot071910.

Chapter Three

Deoxyribozymes as Tools for RNA Labeling

Significant portions of this work were performed directly by or would not have been possible without the contribution of Matt Ashton. Modified RNAs used for debranching assays were prepared by Dr. Lea Büttner and all 10DM24 ligation assays were performed according to advice from Dr. Claudia Höbartner.

ABSTRACT

Deoxyribozymes are useful tools for RNA modification. We have studied the activity of the 10DM24 deoxyribozyme which modifies RNA substrates with small GMP moieties to understand the structure of the enzyme and optimize conditions for catalysis. We define non-essential regions and show that molecular crowding agents can improve labeling kinetics. We show that 10DM24 can be used to prepare fluorescently modified pre-mRNA splicing substrates by ligation of fluorescent GMP moieties to internal 2′OH of targeted adenosines. Unfortunately, the harsh reaction conditions required for 10DM24 activity result in rapid degradation of the RNA target. Preliminary work using the small amount of fluorescent RNA that can be prepared has suggested that modified RNAs can be immobilized on a slide and visualized using a TIRF Prism fluorescence microscope, suggesting they can be used as substrates in single-molecule fluorescence experiments. *In vivo* 2′ to 5′ linkages similar to those generated by 10DM24 are resolved by action of the debranchase known as Dbr1. We show that the linkage to fluorescent GMP is readily cleaved by Dbr1 and propose methods to circumvent removal in cellular extract, including the use of fluorescent phosphorothioate GTP to prepare fluorescent RNAs.

INTRODUCTION

Deoxyribozymes are DNA oligonucleotide-based catalysts evolved *in vitro* capable of performing a variety of chemical reactions(1). DNA-based catalysts have a number of advantages over traditional protein or RNA enzymes (ribozymes). Generation of novel DNA enzymes is relatively facile, as functional variants can be readily selected from random sequence populations. Similar approaches would be impractical with a protein enzyme. Additionally, DNA is considerably more stable and easier to produce synthetically than RNA. The specificity of deoxyribozymes is somewhat modifiable, as most bind their targets through Watson-Crick base pairing in regions outside of the catalytic center. In principal, altering the primary sequence of the DNA can readily change the target of the deoxyribozyme; however, in practice, most exhibit some sequence preference. Since their initial discovery, the incredibly versatility of DNA as enzymes has been demonstrated, as deoxyribozymes have been developed that carry out a wide variety of functions ranging from RNA cleavage to protein modification(2).

Deoxyribozymes are capable of ligating two strands of RNA together to form both linear (with 3′-5′ linked) and branched (2′-5′ linked) products(3). For example, the 7S11 deoxyribozyme was selected to catalyze the formation of a 2′-5′ linkage in a Mg^{2+}-dependent manner between a donor RNA strand and the 2′ hydroxyl of a targeted internal adenosine(4). It targets specific internal regions through base pairing with the target strand and modifies the RNA with fairly rapid kinetics(5). Despite this success, 7S11 was relatively intolerant of sequence diversity surrounding the adenosine target and has since been replaced in practical use with the 10DM24 deoxyribozyme(6).

10DM24 is a deoxyribozyme of related sequence and structure to 7S11 that catalyzes an identical reaction between the 2′ hydroxyl of a targeted adenosine and a 5′-triphosphorylated RNA substrate beginning with a guanine nucleotide(7). Importantly, 10DM24 displayed improved tolerance for different substrates than previous deoxyribozymes and could also target positions other than adenosine, albeit with decreased labeling kinetics and efficiency(6, 7). Subsequent

work characterizing 10DM24 identified that the terminal nucleotide of the donor RNA could be deleted, altering the activity resulting in the ligation of only a single GMP molecule to the target (**Figure 3.1A**)(7). This represented a facile mechanism to label RNA, as fluorescently labeled variants of GTP are commercially available or can be synthesized relatively easily. This method also installs a significantly smaller label than is used with the original 10DM24 system or the DECAL method(8).

Despite the improved kinetics and reactivity of 10DM24, the practically of using these enzymes to modify RNAs is limited by harsh reaction conditions, lengthy reaction times, and high concentrations of fluorescent nucleotide required. The original selection conditions for 10DM24 required 100 mM $MgCl_2$ at pH 9 and overnight incubation at elevated temperatures to achieve significant labeling(7). Reduction of either variable limits yield. In collaboration with an undergraduate researcher, Matt Ashton, we sought to improve 10DM24 labeling by addition of molecular crowding agents. We also sought to understand the contribution of various regions of 10DM24 through structure-function studies of the deoxyribozyme. As an alternative approach, addition of lanthanide cofactors and their ability to accelerate reaction rates for 10DM24 has recently been demonstrated(9). Additionally, upon addition of terbium chloride, the reaction proceeds under milder conditions (pH 7.5) and at lower concentrations of GTP(10).

In this chapter, we have used deoxyribozymes to generate fluorescently modified RNAs for *in vitro* studies, focusing on two model pre-mRNA substrates for use in single molecule assays. Ultimately, we are interested in combining these modified RNAs in yeast splicing extracts which contain the lariat debranching enzyme, Dbr1(11). Dbr1 naturally resolves 2′-5′ linkages created as a product of splicing. To this end, we sought to understand whether these modified nucleotides would be substrates for Dbr1 and developed new methods to circumvent debranching .

MATERIALS AND METHODS

DNA and RNA oligonucleotides were purchased from IDT (Coralville, IA) and used without further purification. GTP stocks were purchased from Promega (Madison, WI) and fluorescent EDA-GTP was purchased from JenaBiosciences (Germany). A complete list of oligonucleotides used in this study is located in **Table 3.1**.

10DM24 Ligation Assays

The ligation ability of 10DM24 was tested using the originally selected RNA substrate sequence unless otherwise stated(7). The ratio of L substrate (L), deoxyribozyme (E) and R cofactor (R) was 1:10:30 to ensure that all L was bound to the deoxyribozyme and all deoxyribozyme was bound to R. Linked E-R constructs were used in 5-fold excess relative to L. The L RNA substrate was modified at the 5′ end with a Cy5 fluorophore or 5′ end-labeled using α-[^{32}P] -ATP and T4 polynucleotide kinase (New England Biolabs) to monitor the reaction. The oligonucleotides were annealed in 5 mM HEPES-NaOH pH 7.5, 15 mM NaCl, 0.1 mM EDTA by heating to 95°C for 3 min and cooling to ~22°C for 15 min. The reaction buffer, metal cofactor(s), and GTP were then added to initiate the reaction. Crowding agents were added prior to the addition of divalent cofactors and GTP. Reactions were performed at 37°C. Timepoints were removed and quenched using denaturing stop solution (95% deionized formamide, 50 mM EDTA, 0.025% dextran blue) and samples were analyzed by 20% denaturing PAGE (19:1, 7M urea, 1xTBE) and imaged using a Typhoon Imager (GE Healthcare). Gel images were analyzed using FIJI Image Analysis Software and the data were fit using MatLab software.

10-23 Deoxyribozyme Cleavage Assay

10-23 was used to determine the extent of modification of long RNAs (pre-RP51A and pre-UBC4)(12). Five hundred fold excess cleaving deoxyribozymes targeting regions flanking the modification were annealed to the target RNA by heating to 95°C and subsequent cooling to 23°C over 30 min. RNAs were annealed in 50 mM Tris-HCl pH 7.5, 150 mM NaCl and then divalent metals (10mM $MgCl_2$, 10 mM MnCl2) were added to initiate the reaction. Reactions proceeded

for 3 hr at 37°C followed by quenching and analysis of cleavage products by denaturing PAGE (20%; 19:1, 7M urea, 1xTBE). Gels were then dried and exposed to a phosphorimager screen. Screens were imaged using a Typhoon Imager (GE Healthcare) and quantified using ImageJ software.

Single Molecule Experiments

Fluorescently modified UBC4 was modified at the 3′ end to include a biotin for surface immobilization using a Pierce biotin 3′ end DNA labeling kit and purified by denaturing PAGE or immobilized using a biotinylated capture oligo. Modified RNAs were immobilized on a polyethylene glycol passivated quartz slide and imaged using a homebuilt prism-based TIRF microscope. RNAs were imaged in 50 mM Tris–HCl pH 7.5, 400 mM NaCl, 4.5 mg/ml glucose, 150 U/ml catalase, 40 U/ml glucose oxidase, and 3 mM Trolox. Images were acquired using Metamorph software.

Purification of Dbr1

Dbr1 was expressed and purified essentially as previously described(11). Briefly, BL21 (DE3) competent cells were transformed with pET16b-Dbr1 (gift of B. Schwer) and grown overnight at 37°C with antibiotic selection. The following day a single colony was picked and grown in 10 mL LB for 4 hours at 37°C with shaking (220rpm) prior to inoculation of 1L LB media. The culture was grown until reaching OD_{600} = 0.6-0.8, cooled at 4°C on ice for 30 min, and then absolute ethanol was added to a final concentration of 2%. Cells were then induced with 0.4mM IPTG overnight at 16°C, and the cell pellet was stored at -80°C until use.

The cell pellet was resuspended in 50 mM Tris pH 7.4, 250 mM NaCl, 10% (w/v) sucrose, and 0.2mg/ml lysozyme and mixed for 30 min at room temperature. The lysate was adjusted to 0.1% (v/v) Triton X-100 and sonicated thrice to reduce viscosity. Insoluble material was removed by centrifugation and the resulting soluble protein was purified using a 1mL HisTrap HP column (GE Healthcare) that had been equilibrated with lysis buffer lacking lysozyme and containing glycerol instead of sucrose. Bound protein was washed with 50 mL buffer containing 25 mM

imidazole and eluted using a linear gradient to a final concentration of 500 mM imidazole. Fractions containing Dbr1 were pooled, diluted 10-fold using low salt buffer (150mM NaCl) and subsequently purified using a 1 mL Heparin FF (GE Healthcare) column. Protein was eluted using a linear gradient to 750 mM NaCl and stored at -80°C until use.

Dbr1 Cleavage Assays

Branched RNA substrates used for debranching assays were prepared and purified by Lea Büttner in the laboratory of Claudia Höbartner. RNAs were radiolabeled using T4 polynucleotide kinase and α-P^{32} ATP for use in debranching experiments. Debranching reactions were performed in 50 mM Tris-HCl (pH 7.0), 4 mM $MnCl_2$, 2.5 mM DTT, 25 mM NaCl, 0.01% Triton X-100, 0.1 mM EDTA, 0.15% (v/v) glycerol, 200 fmol RNA, using 10 ng Dbr1 for 1 h at 23°C(11). Aliquots were removed at specified time points, quenched by addition of stop buffer (95% deionized formamide, 10mM EDTA) and frozen immediately in liquid nitrogen until analysis. RNAs were resolved on 20% denaturing gels (19:1) and the gels were subsequently dried and exposed overnight to a phosphorimager screen. The screen was scanned using a Typhoon Imager (GE Healthcare) and band intensities quantified using ImageJ software.

Name	Sequence (Red = RNA; Black = DNA)	Use
TJC_MA_RNA L	GGA UAA UAC GAC UCA C	Original RNA Target used for ligation assays
TJC_MA_Cy5_RNAL	/3Cy5Sp/GGA UAA UAC GAC UCA C	Cy3 labeled RNA Target used for ligation assays
TJC_MD_DNAR	GAA GAG ATG GCG ACG G	DNA cofactor
TJC_MA_ΔRNA	GA AGA GAU GGC GAC GG	RNA cofactor used structure function studies
TJC_ΔR_5P_RNA	/5Phos/rGrA rArGrA rGrArU rGrGrC rGrArC rGrG	RNA cofactor used in mRNA labeling experiments
TJC_RDNA_DL_P3_13	GAA GAG ATG GCG ACG GGA AAC CGT CGC CAT CTC CCG TAG GTG AAG GGC ***GTG AGG*** GTT CCA TTC CCG TAT TAT CC	Linked 10DM24-DNAR with 13nt P3 Helix
TJC_RDNA_DL_P3_11	GAA GAG ATG GCG ACG AAA GTC GCC ATC TCC CGT AGG TGA AGG GCG ***TGA GGG*** TTC CAT TCC ***CGT ATT ATC C***	Linked 10DM24-DNAR with 11nt P3 Helix
TJC_RDNA_DL_P3_9	GAA GAG ATG GCG GAA ACG CCA TCT CCC GTA GGT GAA GGG C***GT GAG*** GGT TCC ATT CCC GTA TTA TCC	Linked 10DM24-DNAR with 9nt P3 Helix
TJC_RDNA_DL_P3_7	GAA GAG ATG GGA AAC CAT CTC CCG TAG GTG AAG GGC ***GTG AGG*** GTT CCA TTC CCG TAT TAT CC	Linked 10DM24-DNAR with 7nt P3 Helix
TJC_RDNA_DL_P3_5	GAA GAG ATG AA AA TCT CCC GTA GGT GAA GGG C***GT GAG*** GGT TCC ATT CCC GTA TTA TCC	Linked 10DM24-DNAR with 5nt P3 Helix
TJC_RDNA_DL_P3_3	GAA GAG GAA ACT CCC GTA GGT GAA GGG C***GT GAG*** GGT TCC ATT CCC GTA TTA TCC	Linked 10DM24-DNAR with 3nt P3 Helix
TJC_RDNA_DL_P3_0	GAA CCG TAG GTG AAG GGC GTG AGG GTT CCA TTC CCG TAT TAT CC	Linked 10DM24-DNAR with 0nt P3 Helix
CH D699	CCG TCG CCA TCT CCC GTA GGT GAA GGG CTA CGT GGT TCC ATT CCT AGA AGA AGA C	10DM24 variant used to modifiy UBC4
CH D700	CCG TCG CCA TCT CCC GTA GGT GAA GGG CAG TTC GGT TCC ATT CCT TAG CAA TAC G	10DM24 variant used to modifiy UBC4
CH D701	CCG TCG CCA TCT CCC GTA GGT GAA GGG CTT AGT GGT TCC ATT CCC TTT AGC AAT A	10DM24 variant used to modifiy UBC4
TJC_UBC4_I_Pos103_10DM24	CCG TCG CCA TCT CCC GTA GGT GAA GGG CTA AGT GGT TCC ATT CCT TAA AAA AAA AGC ACG	10DM24 variant used to modifiy UBC4
TJC_RP51a_E1_Pos24_10DM24	CCG TCG CCA TCT CCC GTA GGT GAA GGG CTA GTC GGT TCC ATT CCC GAG ATC CCA AG	10DM24 variant used to modifiy RP51a
TJC_RP51a_I_Pos32_10DM24	CCG TCG CCA TCT CCC GTA GGT GAA GGG CTG TCC GGT TCC ATT CCT AAA AGC CTC CT	10DM24 variant used to modifiy RP51a
TJC_RP51a_I_Pos48_10DM24	CCG TCG CCA TCT CCC GTA GGT GAA GGG CTC GAC GGT TCC ATT CCC AAT ATT ACG TG	10DM24 variant used to modifiy RP51a
TJC_RP51a_E1_Pos29_10DM24	CCG TCG CCA TCT CCC GTA GGT GAA GGG CAT TGC GGT TCC ATT CCA GTC TCG AGA TC	10DM24 variant used to modifiy RP51a
TJC_RP51a_E2_Pos218_10DM24	CCG TCG CCA TCT CCC GTA GGT GAA GGG CAA GCC GGT TCC ATT CCT AGA AGC ACG C	10DM24 variant used to modifiy RP51a
TJC_UBC4_10-23_pos25	TCT TTA GCA AGG CTA GCT ACA ACG AAC GTT TAG	10-23 Variant that cleaves UBC4 at position 25
TJC_UBC4_10-23_pos45	ATA CCT TTC TAG AGG CTA GCT ACA ACG ACA CTT AGT	10-23 Variant that cleaves UBC4 at position 45
TJC_Rp51a_10-23_pos11	GAG ATC CCA AGG GCT AGC TAC AAC GAT AGC TTT C	10-23 Variant that cleaves UBC4 at position 11
TJC_Rp51a_10-23_pos36	CAT ACC ATT TTG GGC TAG CTA CAA CGA TAT TGC TAG	10-23 Variant that cleaves UBC4 at position 36
TJC_Rp51a_10-23_pos58	GCC TCC TTT AGG GCT AGC TAC AAC GAC CAT ATT AA	10-23 Variant that cleaves UBC4 at position 58
TJC_Rp51a_10-23_pos82	GAC TCA ATA TTA GGC TAG CTA CAA CGA GTG TCC T	10-23 Variant that cleaves UBC4 at position 82
TJC_Rp51a_10-23_pos101	ACT GAT ATC AAA AAG GCT AGC TAC AAC GAA CGG TC	10-23 Variant that cleaves UBC4 at position 101
TJC_UBC4	GGGAAACAUGUCUUCUUCUAAACGUAUUGCUAAAGAACUAAGUGAUCUAGAAAGGUAUGUCUAAAGUUAUGGCCACGUUUCAAAUGCGUGCUUUUUUUUUAAAACU UAUGCUCUUAUUUACUAACAAAAUCAACAUGCUAUUGAACUAGAGAUCCACCUACUUCAUGUUCAGCCGGUCCAGUCGGCGAUGAUCUAUAUCGGAAUUCGCG	UBC4 sequence used in this study. Underline denotes the intron
TJC_RP51a	GGGAAAGCUAGCUUGGGAUCUCGAGACUAGCAAUAACAAAAUGGUAUGUUAAUAUGGACUAAAGGAGGCUUUUAAGGACACGUAAUAUUGAGUCGACCGUGUUUUU GAUAUCAGUAUACUAACAAGUUGAAUUGCAUUGUUUGUAUUUUUAUUUUGUAUUGCUUUUCGUCAUUUUAAUAGGGUAGAGUUAGAACCAAGACCGUCAAGCGUGC UUCUAAGGCUUUGAUUGAACGUUACUAUCCAAAGUUGACUUUGGAUUUCCAAACCAACAAGAGACUUUGUGAUGAAAUCGCCACUAUCCAAUCCAAGAGAUUGAGA AACAAGAUUGCUGGUUACACCACCCAUUUGAUGAAGAGAAUCCAAAAGGGUCGAUCCGGAGCGAGUAGAUCCACCUACUUCAUGUU	Rp51A sequence used in this study. Underline denotes the intron
R243_BsA	GGA UAA UAC GAG UCA C	RNA sequence used for debranching assay. GMP is attached to the underlined position
R243_BsG	GGA UAA UAC GGG UCA C	RNA sequence used for debranching assay. GMP is attached to the underlined position
R243_BsC	GGA UAA UAC GCG UCA C	RNA sequence used for debranching assay. GMP is attached to the underlined position
R243_BsU	GGA UAA UAC GUG UCA C	RNA sequence used for debranching assay. GMP is attached to the underlined position
R005_GMP	GGA UAA UAC GAC UCA C	RNA sequence used for debranching assay. GMP is attached to the underlined position
R005_Rp	GGA UAA UAC GAC UCA C	RNA sequence used for phosphorothiate assay. Rp-GMP is attached to the underlined position
R005_Sp	GGA UAA UAC GAC UCA C	RNA sequence used for phosphorothiate assay. Sp-GMP is attached to the underlined position
R005_Rp_Cy3	GGA UAA UAC GAC UCA C	RNA sequence used for phosphorothiate assay. Rp-GMP is attached to the underlined position
R005_Sp_Cy3	GGA UAA UAC GAC UCA C	RNA sequence used for phosphorothiate assay. Sp-GMP is attached to the underlined position

Table 3.1 List of oligomers used in this study

RESULTS AND DISCUSSION

10DM24 Modifies RNA Internally at Targeted 2′ Hydroxyls

Structural characterization of 10DM24 and related deoxyribozymes demonstrated that it forms four helices (termed P1 to P4) between 10DM24, the target RNA, and an RNA R cofactor (termed R because it binds the right binding arm) to generate the architecture necessary to form the active site (**Figure 3.1A and 3.1B**)(5). Interestingly, it was observed that separating the terminal guanine nucleotide from the RNA cofactor would generate an enzyme that modifies specific RNA positions with a small molecule in a somewhat generalizable way(7). Indeed, 10DM24 readily modifies RNA substrates over the to completion course of 24 h (**Figure 3.1C**). Fluorescent modification of long RNAs is difficult, requiring ligation of individual oligomers prepared by in vitro transcription or solid-phase synthesis. 10DM24 labeling is an appealing strategy to replace traditional methods to modify RNAs of interest by using fluorescently labeled GTP nucleotides (**Figure 3.1D**). Despite the given functionally of the system and the existing body of literature on the reactivity of 10DM24, a thorough investigation of the regions of the enzyme that affect catalysis has not been performed. To this end, we investigated the role of different structures within 10DM24 and how they influence catalysis.

We began by investigating the role of the P3 helix and whether it would impact catalysis. We introduced a DNA R cofactor in place of the traditional RNA oligonucleotide and repeated the ligation assay using 10DM24 and a short model substrate. Interestingly, substitution of these nucleotides with deoxyribose bases did not block modification of the target (**Figure 3.2A**). This lead to the hypothesis that R cofactor could function in *cis* as a 5′ extension of the deoxyribozyme, generating a bimolecular system in place of the initially developed more complicated trimolecular system. Testing of this construct revealed that it could also catalyze addition of GMP (**Figure 3.2B**). Interestingly, further work on the activity of the system using a DNA cofactor revealed more sluggish kinetics than those that exist with the RNA R cofactor (**Figure 3.2C**); however, the difference between the bimolecular and trimolecular systems using the DNA cofactor was

minimal. This demonstrates that the R cofactor does not need to be an independent strand to function efficiently but the formation of an RNA/DNA hybrid contributes to rate. It is likely that the major role of the P3 helix is to contribute to the formation of the P4 helix that interacts with the free GTP nucleotide to orient it for catalysis.

We next tested whether the length of the P3 helix contributes significantly to the rate of catalysis. The initial R cofactor forms a 13 base pair helix with the P3 region of 10DM24. We shortened the length of helix in two base pair increments using the linked 10DM24-R sequence and tested them for activity (**Figure 3.3A**). Complete removal of the helix eliminated activity and only very short P3 variants had defects in the rate (**Figure 3.3B**). These data suggest that the P3 region of 10DM24 likely plays little role in catalysis beyond helping to position the GTP in the P4 helix.

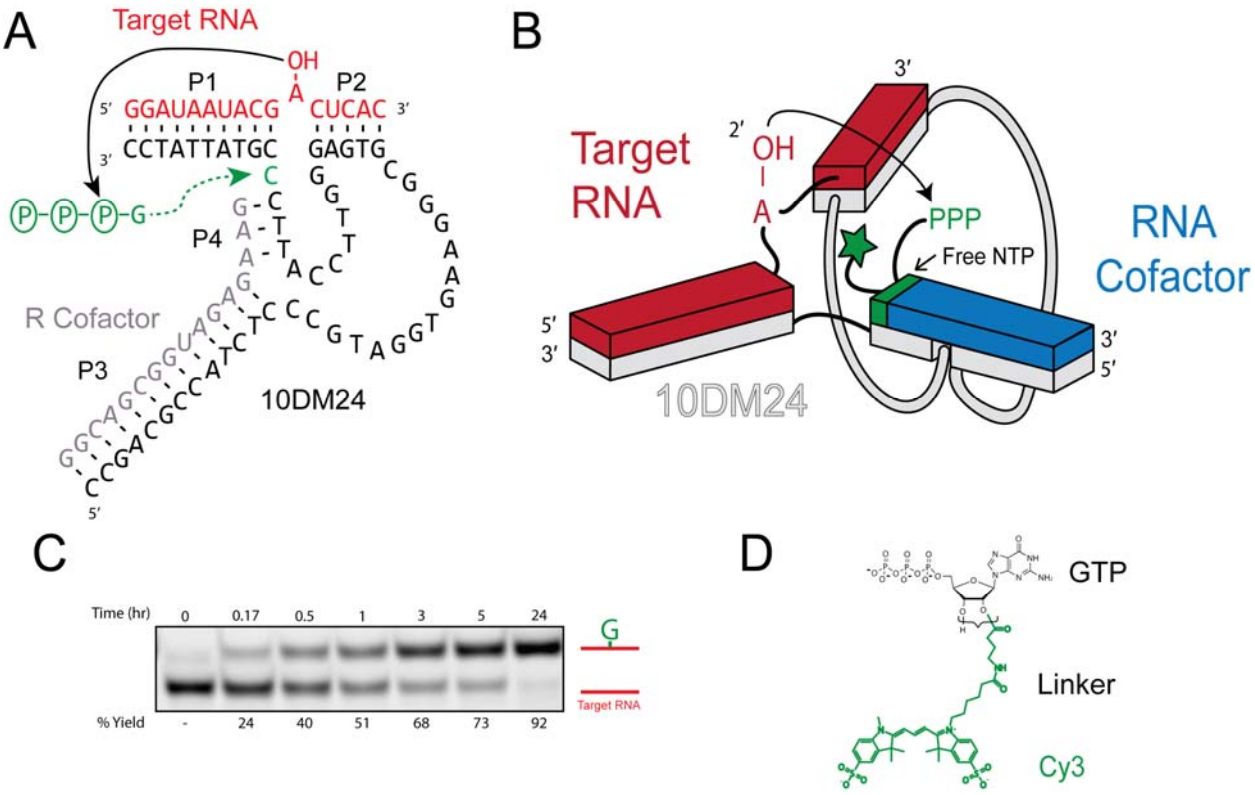


Figure 3.1 10DM24 modifies the 2′-OH of internal adenosine nucleotides. (A) Cartoon schematic of RNA modification by 10DM24. Basepairing between the target RNA (red), the R cofactor (grey) and the 10DM24 DNA (black) promotes addition of a single GMP moiety to the target adenosine. The specificity of this addition is governed by the ability of the free nucleotide to base pair with a specific residue in 10DM24 (shown in green). (B) Kinetic assay demonstrating 10DM24 activity. Upon 10DM24 catalyzed addition of GMP, the target RNA bands migrate more slowly. Samples were taken at specific time points throughout the reaction to show the progression of GMP incorporation. Percent yield was calculated as the fraction of RNA shifted to the higher band. The reaction is specific for free GTP (not shown). (C) Ligation assay using 10DM24 shows incorporation of GMP to [32P]-labeled substrate over the course of 24 hr. (D) Structure showing commercially available GMP-EDA-Cy3.

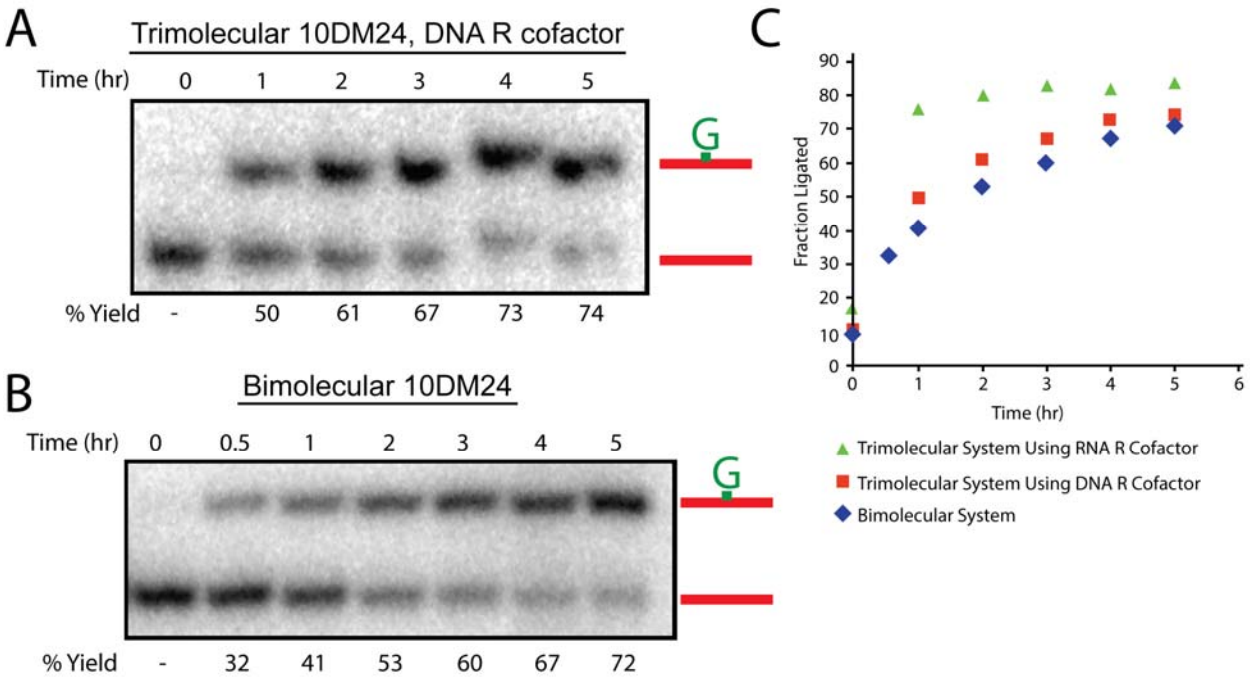


Figure 3.2 Kinetic assays of tri- and bimolecular 10DM24 system demonstrating activity. (A) Assay showing that the 10DM24 catalyzes the addition of GMP under the trimolecular system which uses DNA for the R cofactor. (B) Kinetic assay using the bimolecular system and DNA R. Neither the shortened P3 helix region or the attachment of the R cofactor to the 10DM24 prevent the addition of GMP onto the target RNA. (C) Graph comparing the kinetics of the tri- and bimolecular systems.

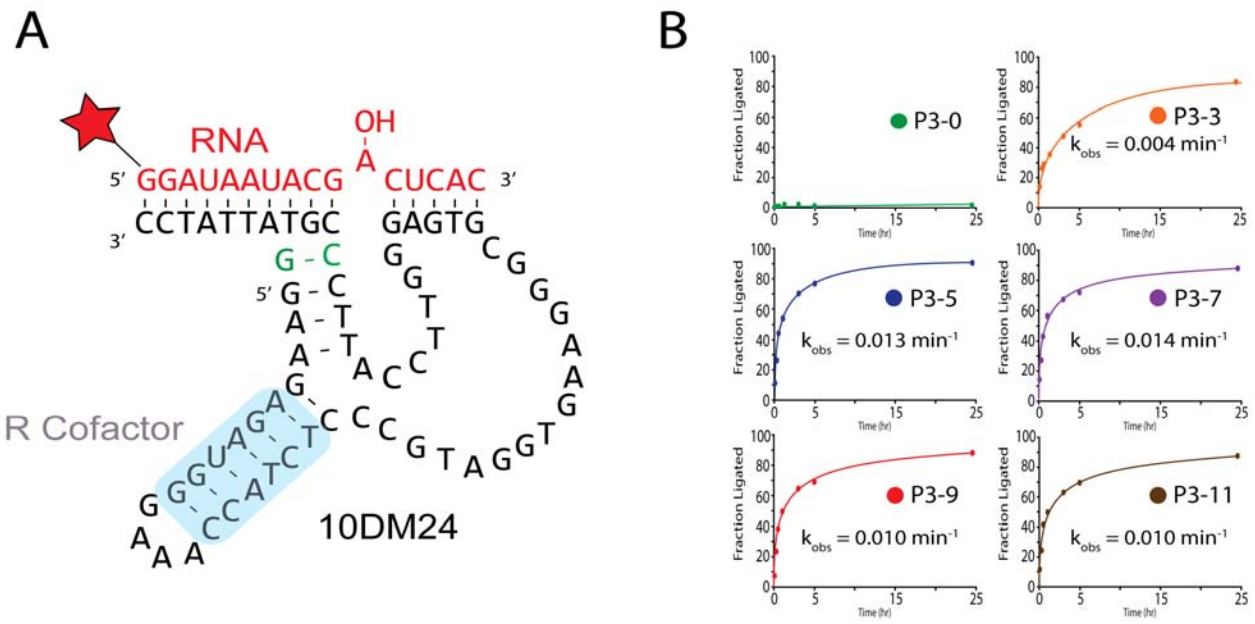


Figure 3.3 Kinetic assays of the bimolecular 10DM24 system with varying P3 helix lengths. (A) Cartoon schematic depicting the modified bimolecular 10DM24 deoxyribozyme. The P3 helix is highlighted in blue and has been shortened from 13 to 7 base pairs. The target RNA strand is red with a 5′ Cy5 label. (B) Graphs comparing the kinetics of the different P3 helix lengths for the bimolecular system.

Modulation of 10DM24 Activity by Addition of Molecular Crowding Agents

Molecular crowding agents can have a profound impact on the activity of ribozymes, likely by affecting the folding of the enzyme into catalytically active conformations (13). Prior work had established that the folding of 10DM24 into a catalytically active conformation occurs in a Mg^{2+} dependent manner(14). However, high Mg^{2+} concentrations are linked with RNA degradation at high temperatures. The reactivities of deoxyribozymes can be modulated by the addition of organic cosolvents(15). We therefore hypothesized that molecular crowding agents could impact the folding of 10DM24 and lessen the Mg^{2+} requirement for catalysis. To test this, we used the molecular crowding agents PEG 8000 and Ficoll 400 and monitored for an effect on the rate of product formation. Addition of Ficoll 400 had only a modest effect whereas the addition of PEG 8000 sharply increased the activity of 10DM24 (**Figure 3.4A**). We next sought to determine the impact of Ficoll 400 and PEG 8000 at various concentrations of Mg^{2+}. PEG 8000 appears to perform better than Ficol 400, as it increased the rate at both high and lower concentrations of Mg^{2+} (**Figure 3.4B**). Ficoll 400 had little effect at 20 mM $MgCl_2$ and a modest effect at higher concentrations. These data suggest that molecular crowding agents may represent a useful method to maintain deoxyribozyme activity while reducing Mg^{2+}.

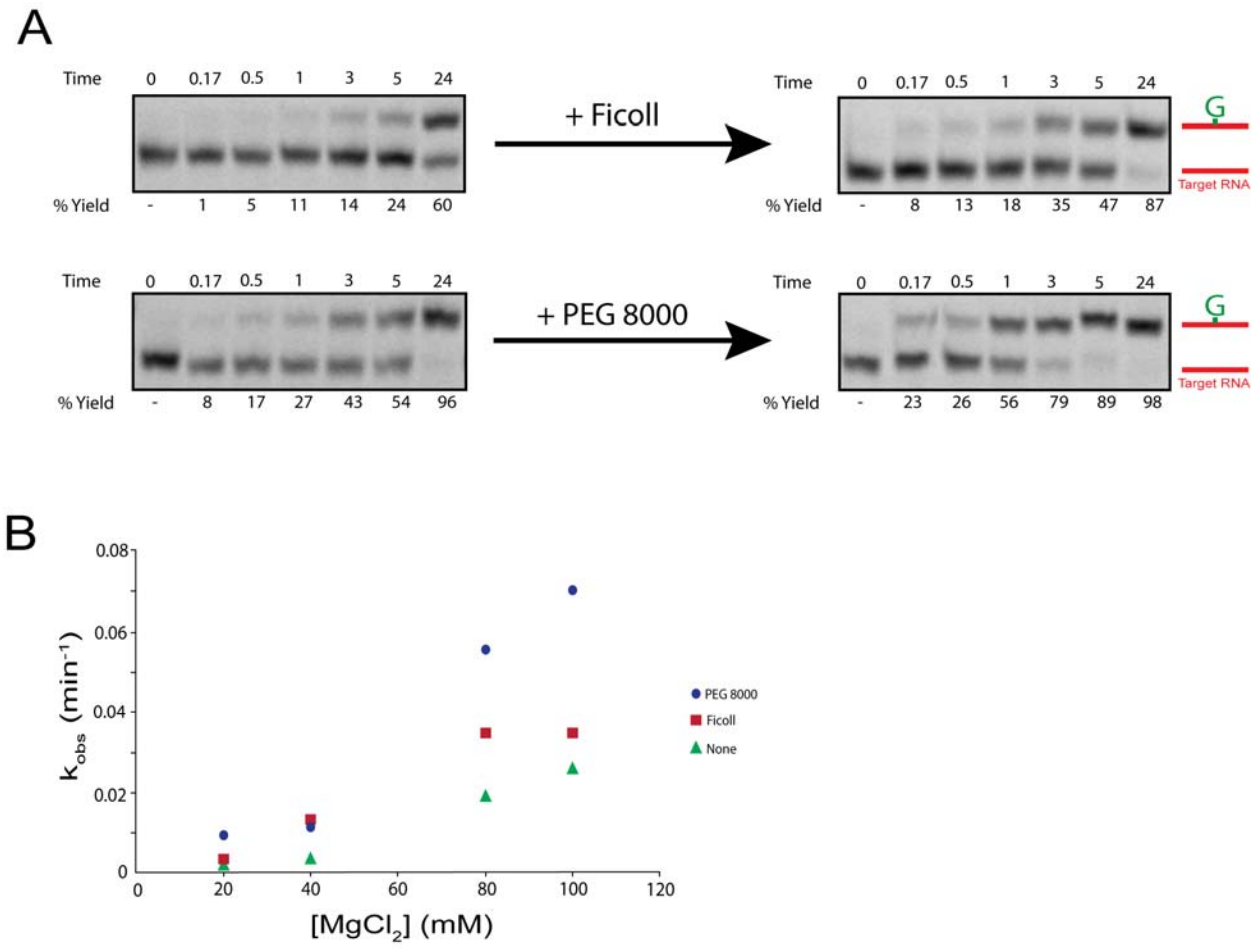


Figure 3.4 Molecular crowding agents improve 10DM24 activity (A) Gel images of the kinetics of the trimolecular system before and after the addition of Ficoll or PEG 8000. (B) Graph displaying the k_{obs} values for multiple magnesium concentrations with and without the molecular crowding agents.

Labeling of Model pre-mRNAs using 10DM24

The addition of lanthanide cofactors also increases the reactivity of 10DM24 and drastically lowers the concentration of GTP required to modify short RNA substrates(9, 10). We sought to apply these conditions to longer RNAs used for *in vitro* splicing assays. Specifically, we began by modifying the ~200 nucleotide UBC4 pre-mRNA using 10DM24 variants targeting specific regions in the intron (**Figure 3.5A**)(16, 17). The complex was assembled by annealing 10DM24, the R cofactor and the pre-mRNA and the reaction performed by incubating the complex at 37°C for 1 hour. Timepoints were removed and quenched to monitor for fluorescent GTP incorporation. Unsurprisingly, significant degradation of the RNA was observed during the labeling protocol to the point where almost no RNA was detectable after 1 h (**Figure 3.5B**). This is consistent with prolonged incubation of the long RNA at high temperatures in the presence of high concentrations of Mg^{2+} (100mM) resulting in nucleolytic cleavage. However, reduction of the Mg^{2+} concentration to 80 mM and lowering the temperature to 22°C reduced degradation. However, this likely occurs at the cost of 10DM24 modification rate, as concentrations of $MgCl_2$ greater than 80 mM were shown to be optimal for activity.

To determine the extent of UBC4 modification by 10DM24 under these conditions, I utilized an assay that resulted in cleavage of the pre-mRNA into smaller fragments to allow for resolution and visualization of the modified and unmodified populations by PAGE. Multiple attempts to use 2′ OMe oligo-directed RNaseH cleavage were attempted but were ultimately unsuccessful (data not shown). Because of this, variants of the 10-23 RNA cleaving deoxyribozyme were used(12). 10-23 cleaves RNA strands adjacent to purines in a sequence generalizable fashion. The modified transcript was purified after labeling, annealed to 10-23 variants and then cleaved (**Figure 3.5C**). Fragments generated by 10-23 cleavage were identified by a series of controlled reactions using only a single 10-23 variant. The band migrating with aberrant mobility corresponds to fluorescent labeled substrate which could be visualized in the gel (**Figure 3.5D**). Under these conditions, modification of UBC4 to ~40% was possible. Longer

incubation times would likely result in increased modification at the cost of reducing yields due to RNA degradation.

Given the success of labeling conditions with UBC4, we next sought to extend the labeling assays to include longer pre-mRNA substrates such as RP51A. RP51A has been used extensively for single molecule studies of the spliceosome and is nearly double the length of UBC4 (**Figure 3.6A**)(18). We began by surveying active positions in the RP51A intron using GTP-Cy3 and monitoring for the incorporation of fluorescent after 1 hr at 22°C. Multiple positions in the RP51A intron can be labeled albeit with different efficiencies (**Figure 3.6B**). The most active position (RP51A 2) can be labeled with both GTP-Cy3 and GTP-Cy5, underscoring the modular design of the system (**Figure 3.6C**). However, given the differences in reactivity observed with RP51A positions (**Figure 3.6B**) and the previously discussed sequence preference that deoxyribozymes display, we expanded our search for reactive positions in both UBC4 and RP51A.

Modification of the intron would allow for release of the intron after splicing, but labeling of the exons would generate a labeled RNA that can be observed even after spliceosome release. For this reason, we examined positions in both the exons and introns of the substrates. Highly reactive positions could be found in both RP51A and UBC4 as well as positions that could only be labeled very poorly (**Figure 3.6D**). Evaluation of the sequence context did not reveal an obvious preference. Previous work on model substrates observed that positions flanked by adenosine or guanine nucleotide display the best labeling kinetics. Long RNA positions that are flanked by purines are likely to be reactive (RP51A-2, UBC4-8), but can also display limited reactivity (RP51A 4, UBC4 9). These results suggest that the sequence preference observed for short, model RNA substrates might not translate to long RNAs, possibly due to competing structures within the RNA. Despite this, positions in RP51A that are modified well in our conditions show similar labeling efficiency to those observed for UBC4 (35% vs 38%; **Figures 3.5D and 3.6D**) demonstrating that these RNAs can be modified in reasonable quantities.

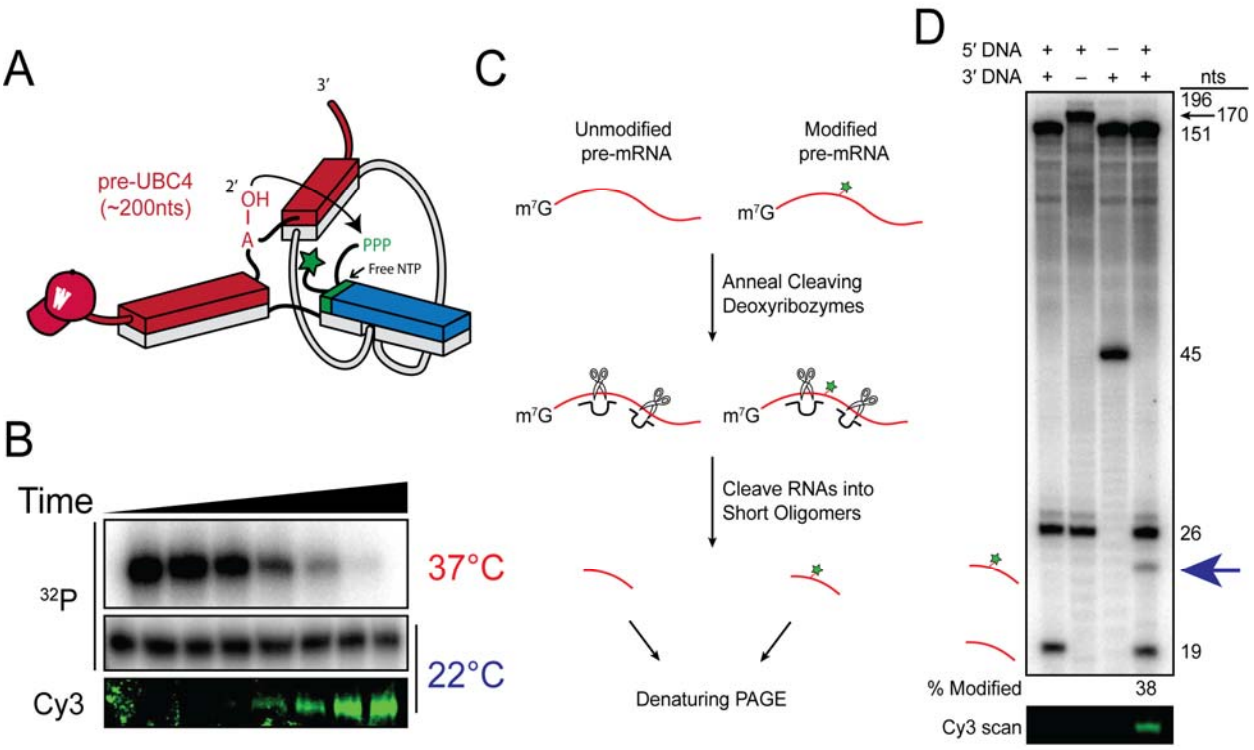
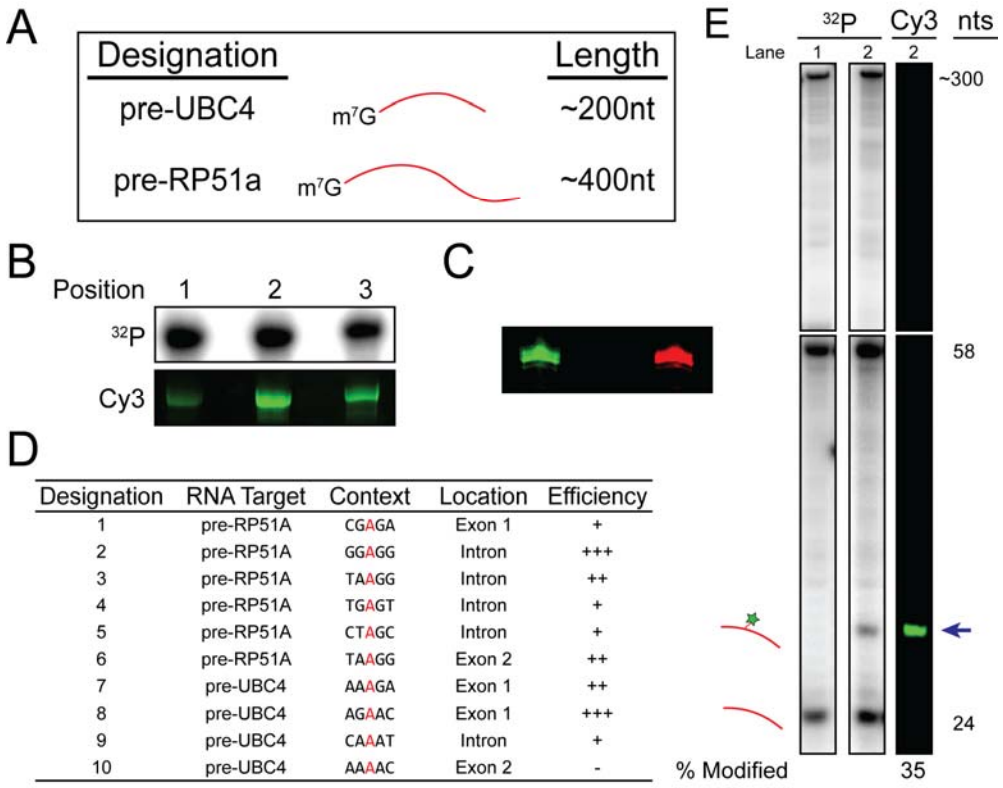


Figure 3.5 Modification of the UBC4 pre-mRNA using the 10DM24 deoxyribozyme. (A) Cartoon schematic of post-transcriptional modification of capped UBC4 using fluorescent GTP derivatives. (B) Fluorescent modification of UBC4 at different temperatures. Degradation of the target RNA was rapid in the reactions at 37°C. Reducing the temperature to 22°C prevented the rapid degradation but still allowed the reaction to proceed. (C) Cartoon schematic of deoxyribozyme mediated cleavage assay to determine the extent of pre-UBC4 modification. The purified RNA transcript is annealed to the variants of the 10-23 deoxyribozyme. 10-23 cleaves the RNA site-specifically to generate a mixture of short oligomers. The GMP-Cy3 modified oligomer can be resolved from the unmodified oligomer to allow for the determination of the extent of labeling. (D) Demonstration of a cleavage assay. Pre-UBC4 was annealed to oligomers that cleave the 5′ and 3′ portion of the transcript and the reaction was allowed to proceed for 5hr at 37°C. The resulting oligomers were separated by denaturing PAGE. The addition of GMP-Cy3 results in a clear gel shift. The fluorescently modified band is indicated with the blue arrow. A Cy3 scan of the gel is shown at the bottom.



Designation	RNA Target	Context	Location	Efficiency
1	pre-RP51A	CGAGA	Exon 1	+
2	pre-RP51A	GGAGG	Intron	+++
3	pre-RP51A	TAAGG	Intron	++
4	pre-RP51A	TGAGT	Intron	+
5	pre-RP51A	CTAGC	Intron	+
6	pre-RP51A	TAAGG	Exon 2	++
7	pre-UBC4	AAAGA	Exon 1	++
8	pre-UBC4	AGAAC	Exon 1	+++
9	pre-UBC4	CAAAT	Intron	+
10	pre-UBC4	AAAAC	Exon 2	-

Figure 3.6 Modification of the ~400nt RP51A pre-mRNA. (A) Cartoon illustrating the size difference between pre-UBC4 and pre-RP51A. Pre-RP51A is twice the length. (B) 10DM24 ligation assay demonstrating that multiple positions in the pre-mRNA can be targeted using different deoxyribozyme variants. Different positions have different reactivities. (C) The fluorescence of the modified RNA can be modified by using different GTP derivatives. Fluorescent GTP derivatives that span the visible spectrum are commercially available. Pre-RP51A has been modified with GMP-Cy3 and GMP-Cy5 (D) Qualitative description of the incorporation efficiency at multiple position in pre-RP51A. The target adenosine is shown in red. The key for the Table 3.is: +++ = very efficient; ++ = efficient; + = efficient; - = unreactive. (E) Representative cleavage assay showing site-specific modification. Lane 1 is unmodified pre-RP51A and lane 2 contains pre-RP51A modified with GMP-Cy3. The GMP-Cy3 oligo is indicated by the blue arrow.

Imaging of Model Pre-mRNAs Labeling using 10DM24

We also evaluated these modified RNAs as candidates for single molecule studies. Fluorescent UBC4 was annealed to a biotinylated tether oligo to facilitate surface immobilization on a quartz slide using a biotin-streptavidin sandwich (**Figure 3.7A**). Unbound RNA was washed away with buffer and the surface immobilized molecules were visualized. Hundreds of fluorescent particles were observed on the surface of the slide (**Figure 3.7B**). Movies monitoring the fluorescent intensity over time were recorded and analyzed using Matlab software. Fluorescent signals displayed the expected behavior for molecules singly labeled with only Cy3 (**Figure 3.7C**). Molecules blinked and photobleached in a single step. Furthermore, analysis of the fluorescent intensities of these particles over the population reveal a uniform distribution, suggesting only single fluorophores were installed (**Figure 3.7D**). Therefore, the fluorescently labeled RNAs generated using deoxyribozymes are suiTable 3.for single molecule experiments.

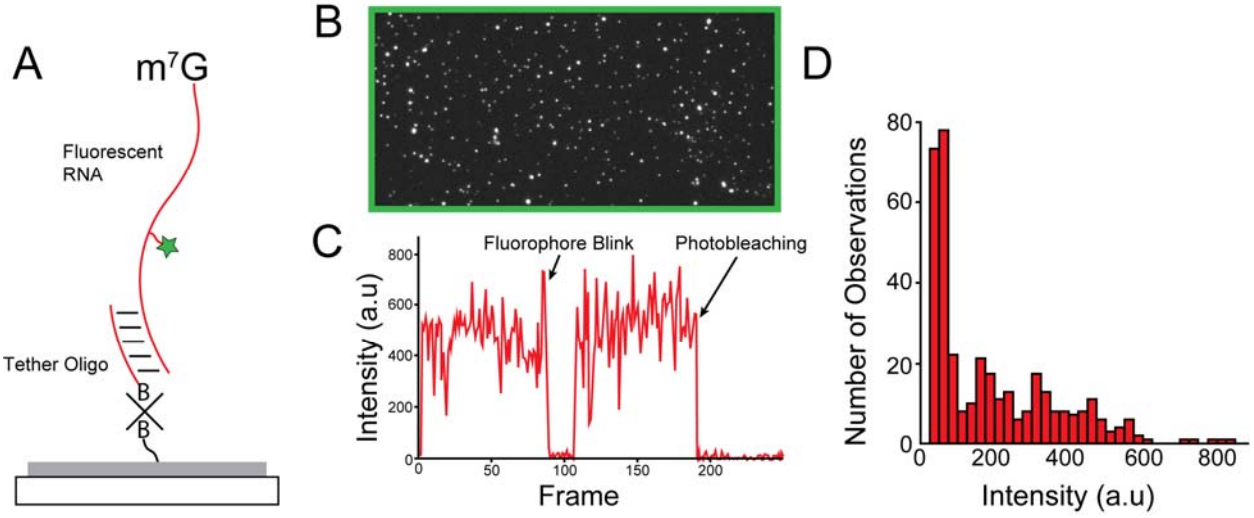


Figure 3.7 Single-molecule experiments using fluorescent RNAs prepared using the 10DM24 deoxyribozyme. (A) Cartoon illustrating the immobilization scheme in use with the prepared RNAs. A biotinylated 2′OMe tether oligo is used to immobilize the fluorescent RNA. (B) A representative field of view showing immobilized GMP-Cy3 modified pre-UBC4. (C) Intensity trace taken from a single molecule from the field of view in section B. (D) Histogram of intensity values. The intensity of 268 molecules was determined for over 971 observations.

Dbr1 Rapidly Removes 10DM24 Linkages Generated by 10DM24

The goal in applying this labeling method to long RNAs is to be able to produce substrates that can be combined with yeast whole cell extracts competent for splicing for single molecule studies investigating the spliceosome. 10DM24 generates a 2′-5′ linkage that is reminiscent of the product of splicing, the lariat intron. *In vivo* a debranchase known as Dbr1 is responsible for cleaving the lariat 2′-5′ linkage, allowing for degradation or further processing of the intron RNA(11). We examined whether the presence of the fluorophore would be sufficient to block Dbr1 cleavage or whether cleavage would occur too slowly to be relevant for *in vitro* splicing assays (which occurs over tens of minutes; **Figure 3.8A**). Recombinant Dbr1 expressed and purified from *E. coli* lysate (**Figure 3.8B**) was combined with UBC4 labeled with GTP-Cy3. During the course of 1 hr, the fluorescent signal is lost whereas the autoradiography signal remains, consistent with cleavage of the GMP-Cy3 from UBC4 (**Figure 3.8C**). Quantification of the fluorescent signal relative to the [^{32}P] signal shows that the half-life for the fluorescent modification on the long substrate is 8 min (**Figure 3.8D**). This is similar to the timescale required for *in vitro* splicing to occur, suggesting that the debranching reaction would obfuscate analysis of splicing. This complication could be avoided however by deleting Dbr1 (non-essential in *S. cerevisiae* but essential in metazoans) or by generating linkages that cannot be cleaved by Dbr1.

Adenosine is an inherently favored nucleophile for generating 2′-5′ linkages, displaying better kinetics than other branchsites with both the spliceosome and *in vitro* selected deoxyribozymes(6). Despite this, 10DM24 can target other nucleotides for modification with the GMP moiety(7). Because the primary function of Dbr1 is to resolve linkages generated by the spliceosome, we hypothesized that it might display altered kinetics when cleaving linkages derived from other branchsite nucleotides. To test this, we generated short 16-mer branched RNA substrates identical in sequence except for the branched position and monitored debranching of these substrates by Dbr1. Surprisingly, the lifetime of the branched population was considerably shorter on the 16-mer substrate than for branched UBC4 (**Figure 3.9A**). This can be explained

by at least two possible mechanisms: either the fluorophore interferes with debranching by Dbr1, or it takes Dbr1 longer to find the branched positions in longer RNAs. We favor the latter mechanism, as debranching of fluorescently labeled substrates discussed below showed fast kinetics, with half-lives on the order or seconds.

Different branchsite nucleotides showed differences in the rate at which they were processed by Dbr1 (**Figure 3.9A**). Clear preference for purine over pyrimidine branchsites can be observed and adenosine is most readily debranched. Quantification of the fraction debranched over time shows that Dbr1 will efficiently process in the following order A>G>U>C (**Figure 3.9B**). It is interesting to note that this is also the inverse of the previously characterized preference of what 10DM24 modifies, suggesting a relationship between ease of modification and ease of cleavage(10). Taken together, these data suggest that altering the identity of the nucleotide modified by 10DM24 would likely not prevent fluorophore removal by Dbr1 during an *in vitro* splicing assay.

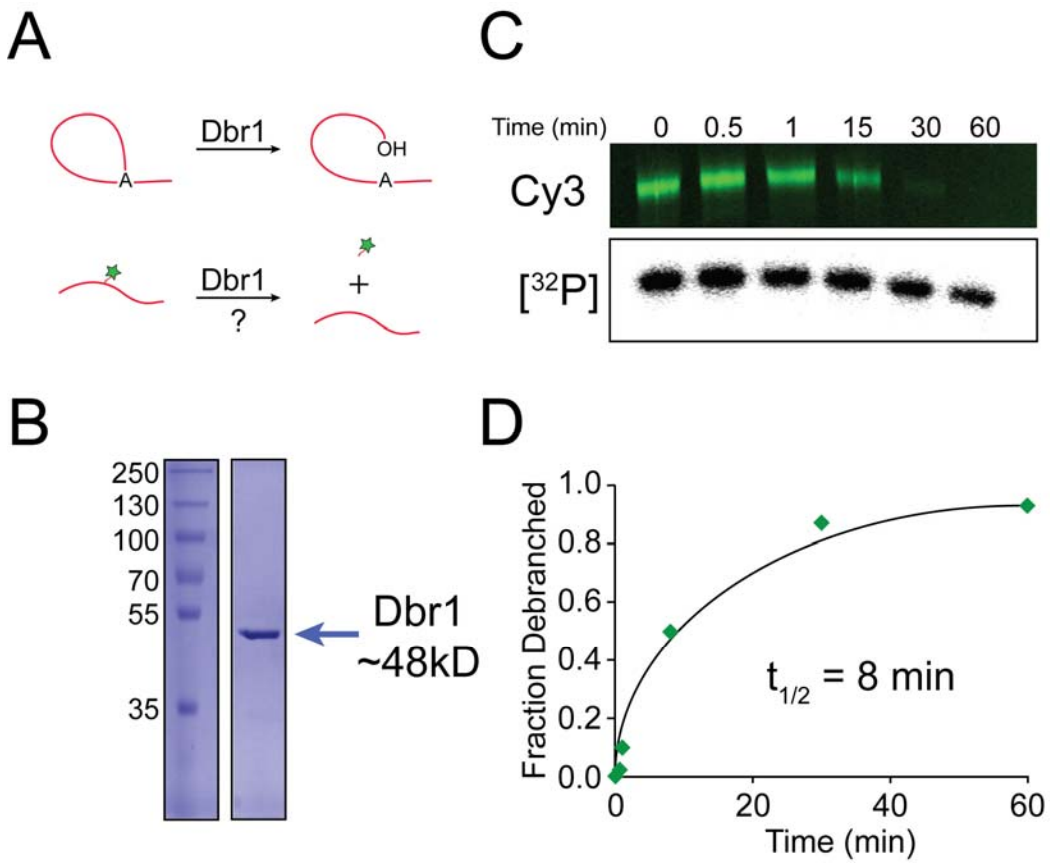


Figure 3.8 Debranching of modified RNAs by Dbr1. (A) Cartoon schematic to illustrate debranching activity by Dbr1. Dbr1 cleaves the 2′-5′ linkage generated as a product of splicing (top). It was unclear whether the fluorophore would sterically inhibit Dbr1 activity (bottom). (B) Purification of Dbr1. Dbr1 from *S. cerevisiae* was recombinantly expressed and purified as described by Khalid *et al*. The purified Dbr1 protein prep was analyzed by SDS-PAGE and the blue arrow denotes the purified protein of the correct size. (C) Debranching assay using GMP-Cy3 modified UBC4 pre-mRNA and purified Dbr1. Time points were removed from the debranching reaction and quenched at 0, 0.5, 1, 8, 30, and 60 minutes. Samples were analyzed using PAGE (12%, 8M Urea). (D) Quantitative analysis of debranching activity. The Cy3 fluorescent intensity was normalized to the [^{32}P] signal and the fraction debranched was

calculated relative to the fluorescent intensity at t = 0. The half-life of the modified fluorophore is approximately 8 minutes.

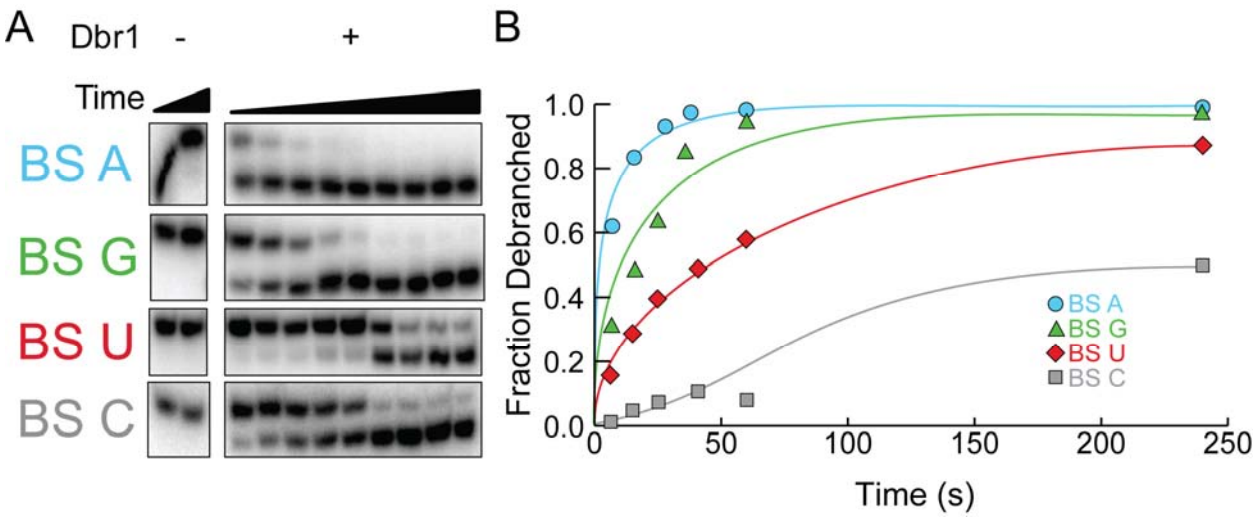


Figure 3.9 Debranching of alternative branchsites. (A) Debranching of RNA substrates with different branchpoint nucleotides by Dbr1. Branchpoint identity is indicated on left and all other sequence in the RNA is identical. (B) Quantification of debranching in A. Dbr1 has preference for substrates in the following order: A>G>U>C.

Phosphorothioate Modified Linkages Inhibit Dbr1 Activity

An alternative approach to block debranching is to change the linkage generated by 10DM24 to something that can not be cleaved by Dbr1. 10DM24 can also incorporate both isomers of phosphorothioate GTP and many phosphodiesterases are inhibited by specific configurations of the phosphorothioate. We therefore reasoned that one of the resulting linkages would be resistant to cleavage by Dbr1. To test this, we generated RNA substrates modified with samples of GTP-α-S with either the Sp and Rp configurations. An inversion of stereochemistry occurs when 10DM24 forms the linkage such that RNAs modified with Sp-GTP-α-S have an Rp configuration after formation of the 2′-5′ linkage. Incubation of RNAs modified with GTP, Rp-GTP-α-S, or Sp-GTP-α-S showed that the linkage was stable in the presence of Dbr1 after 1 h only if the linkage has an Rp configuration (**Figure 3.10**). This result unambiguously demonstrates that Dbr1 is unable to cleave 2′-5′ linkages with an Rp configuration.

We next developed a scheme to produce fluorescent phosphorothioate modified GTP with which to modify pre-mRNA substrates. The method developed relies on Cu(I)-catalyzed azide-alkyne cycloaddition of alkyne-modified phosphorothioates nucleotides and azide functionalized fluorophores (**Figure 3.11**). This was then followed by resolution of the mixture into pure Rp and Sp samples. RNAs can be modified using these nucleotides. Debranching of these substrates by Dbr1 showed that RNAs with Sp-G-Cy3 could be cleaved in a manner similar to GMP modified RNAs (**Figure 3.12**). Surprisingly, substrates with the Rp configuration also showed debranching over time, albeit to a lower extent. This observation is consistent with the presence of GMP-Cy3-modified RNAs (i.e., RNAs containing phosphate but not phosphorothioate linkages) contaminating the sample. Mass spectrometry was used to confirm this hypothesis (data not shown). Despite significant effort, the point of contamination was not identified or remedied.

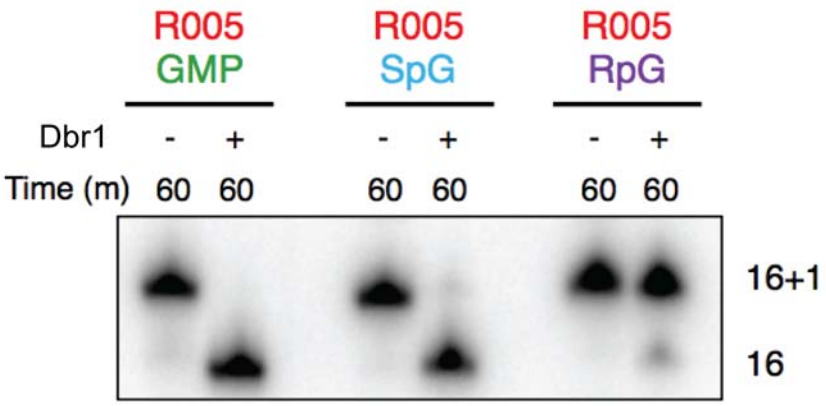


Figure 3.10 Demonstration that phosphorothioate modification of the linkage generated by 10DM24 blocks debranching by Dbr1. RNAs with an Rp configuration of the phosphorothioate inhibit Dbr1 debranching even after 1 hr incubation.

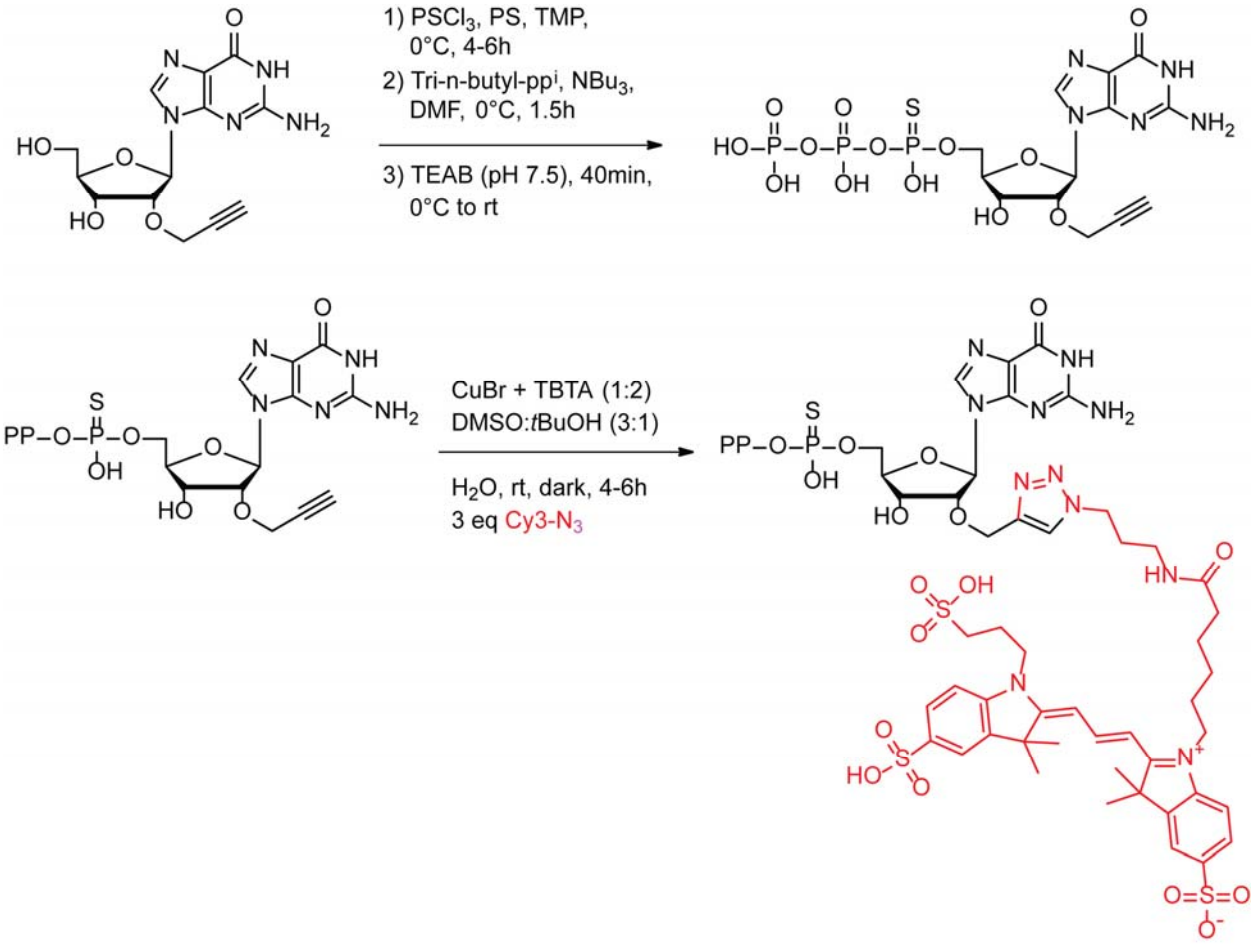


Figure 3.11 Scheme used to synthesis fluorescent, phosphorothioate modified GTP analogs for use with 10DM24. Scheme was developed and labeled phosphorothioate GTP molecules were prepared by Lea Büttner in the lab of Claudia Höbartner.

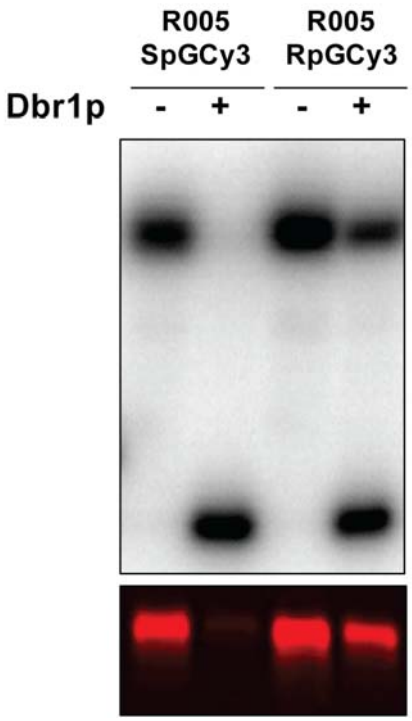


Figure 3.12 Modification of the target RNA with fluorescent phosphorothioate GTP results in uncleavable linkages. RNAs were incubated in the presence of Dbr1 under cleaving conditions and the resulting products were analyzed by PAGE. The Rp configuration is resistant to cleavage. Partial debranching is likely the results of contaminating fluorescent GTP (phosphate containing) used to label the RNAs.

CONCLUSIONS AND PERSPECTIVE

In summary, this work sought to use deoxyribozyme-based labeling methods to prepare long, site-specifically modified RNAs for single-molecule studies. Much of this work was motivated by the desire to employ these RNAs in assays investigating the mechanisms of pre-mRNA splicing. Unfortunately, the linkage is poorly compatible with these studies because they must be performed in cell extract containing a protein that readily removes the label. The work presented here on generating linkages not cleaved by Dbr1 represent an initial step towards overcoming these issues. However, RNAs labeled using the current generation of deoxyribozymes are better suited for experiments employing purified components free of debranching enzyme. Alternatively, selection of deoxyribozymes capable of labeling RNAs in other ways would represent a significant advance and will be necessary if these catalysts are to replace splinted ligation as the method of choice to produce fluorescently modified RNAs. Ideally, this selection of new catalysts would be performed in the presence of molecular crowding agents and low Mg^{2+} concentrations to circumvent degradation of the target during labeling.

REFERENCES

1. Höbartner,C. and Silverman,S.K. (2007) Recent advances in DNA catalysis. *Biopolymers*, **87**, 279–292.

2. Silverman,S.K. (2016) Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem. Sci.*, **41**, 595–609.

3. Baum,D.A. and Silverman,S.K. (2008) Deoxyribozymes: useful DNA catalysts in vitro and in vivo. *Cell. Mol. Life Sci.*, **65**, 2156–2174.

4. Coppins,R.L. and Silverman,S.K. (2004) A DNA enzyme that mimics the first step of RNA splicing. *Nat. Struct. Mol. Biol.*, **11**, 270–274.

5. Coppins,R.L. and Silverman,S.K. (2005) A deoxyribozyme that forms a three-helix junction complex with its RNA substrates and has general RNA branch-forming activity. *J. Am. Chem. Soc.*, **127**, 2900–2907.

6. Zelin,E., Wang,Y. and Silverman,S.K. (2006) Adenosine is inherently favored as the branch-site RNA nucleotide in a structural context that resembles natural RNA splicing. *Biochemistry*, **45**, 2767–2771.

7. Höbartner,C. and Silverman,S.K. (2007) Engineering a selective small-molecule substrate binding site into a deoxyribozyme. *Angew. Chem. Int. Ed.*, **46**, 7420–7424.

8. Baum,D.A. and Silverman,S.K. (2007) Deoxyribozyme-catalyzed labeling of RNA. *Angew. Chem. Int. Ed.*, **46**, 3502–3504.

9. Javadi-Zarnaghi,F. and Höbartner,C. (2013) Lanthanide cofactors accelerate DNA-catalyzed synthesis of branched RNA. *J. Am. Chem. Soc.*, **135**, 12839–12848.

10. Büttner,L., Javadi-Zarnaghi,F. and Höbartner,C. (2014) Site-specific labeling of RNA at internal ribose hydroxyl groups: terbium-assisted deoxyribozymes at work. *J. Am. Chem. Soc.*, **136**, 8131–8137.

11. Khalid,M.F., Damha,M.J., Shuman,S. and Schwer,B. (2005) Structure–function analysis of yeast RNA debranching enzyme (Dbr1), a manganese-dependent phosphodiesterase. *Nucleic Acids Res.*, **33**, 6349–6360.

12. Santoro,S.W. and Joyce,G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. USA*, **94**, 4262-4266.

13. Paudel,B.P. and Rueda,D. (2014) Molecular crowding accelerates ribozyme docking and catalysis. *J. Am. Chem. Soc.*, **136**, 16700–16703.

14. Turriani,E., Höbartner,C. and Jovin,T.M. (2015) Mg^{2+}-dependent conformational changes and product release during DNA-catalyzed RNA ligation monitored by Bimane fluorescence. *Nucleic Acids Res.*, **43**, 40–50.

15. Alila,K.O. and Baum,D.A. (2011) Modulation of an RNA-branching deoxyribozyme by a small molecule. *Chem. Commun.*, **47**, 3227.

16. Abelson,J., Blanco,M., Ditzler,M.A., Fuller,F., Aravamudhan,P., Wood,M., Villa,T., Ryan,D.E., Pleiss,J.A., Maeder,C., *et al.* (2010) Conformational dynamics of single pre-mRNA molecules during in vitro splicing. *Nat. Struct. Mol. Biol.*, **17**, 504–512.

17. Abelson,J., Hadjivassiliou,H. and Guthrie,C. (2010) Preparation of fluorescent pre-mRNA substrates for an smFRET study of pre-mRNA splicing in yeast. *Methods Enzymol.*, **472**, 31–40.

18. Hoskins,A.A., Friedman,L.J., Gallagher,S.S., Crawford,D.J., Anderson,E.G., Wombacher,R., Ramirez,N., Cornish,V.W., Gelles,J. and Moore,M.J. (2011) Ordered and dynamic assembly of single spliceosomes. *Science*, **331**, 1289–1295.

Chapter Four

The Spliceosome and Pre-mRNA Splicing

This chapter is published in the following form:

Carrocci TJ and Hoskins AA. 2016. The Spliceosome and pre-mRNA Splicing. *Encyclopedia of Cell Biology.* 1:495-502.

ABSTRACT

Eukaryotic mRNAs are often first transcribed as precursor mRNAs (pre-mRNAs) that contain introns that must be removed. These introns are spliced out by a large molecular machine called the spliceosome. The major components of the spliceosome are small nuclear ribonucleproteins (snRNPs) made up of small nuclear RNAs (snRNAs) and dozens of proteins. Spliceosomes form on RNA transcripts, recognize specific nucleotide sequences, and use basepairing interactions to bring catalytic groups together in order to remove introns. In humans, splicing is a central step in gene expression and alternative splicing greatly expands the coding capacity of the genome. Defects in splicing can lead to a number of diseases including cancers, blindness, or muscular atrophies.

INTRODUCTION

In the mid-20th century, eukaryotic biologists were presented with a dilemma. It had been well-established in prokaryotes that messenger RNAs (mRNAs) were direct copies of genes found within DNA. It was widely predicted that the same would hold true for eukaryotes; however, it was clear that RNA metabolism was occurring differently in organisms with a nucleus. Eukaryotic cells were found to transcribe a huge amount of nuclear RNA (called heterogeneous nuclear RNA or hnRNA) that never appeared to be exported to the cytoplasm (Sharp 1993; Soeiro et al. 1966). Furthermore, both hnRNAs and mRNAs were found to contain 5' cap structures and 3' polyadenylation tracts. If hnRNAs were the precursor molecules (pre-mRNAs) to cytosolic mRNAs, how could the larger hnRNA be shortened in the middle but keep both the 5' cap and 3' polyadenylation signals at the ends intact in the mRNA? To answer this question, the laboratories of Richard Roberts and Phillip Sharp independently carried out an elegant experiment: adenovirus RNA was hybridized to its coding DNA and directly visualized by electron microscopy (Berget et al. 1977; Chow et al. 1977). The RNA/DNA hybrid structures revealed loops—regions of DNA not present in the mRNA. The answer was then apparent: eukaryotic mRNAs are made by cutting and ligating together different regions of a pre-mRNA. This process is called RNA splicing, and Roberts and Sharp were awarded the Nobel Prize in Medicine in 1993 for their discovery.

In the years since the discovery of RNA splicing, much has been learned about how splicing is catalyzed and how cells integrate splicing with gene regulation. The intragenic regions removed from pre-mRNAs are called introns and the retained segments are called exons. Introns are ubiquitous in eukaryotic genomes, with most human transcripts having an average of 8 introns, and ultimately several hundred thousand introns are present in the human genome (Scherer 2008). Introns must be removed precisely: a splicing error of just one nucleotide will disrupt the correct reading frame of the mRNA. The splicing reaction is carried out by large cellular machine made up of both RNA and proteins, dubbed the spliceosome (Brody & Abelson

1985). The spliceosome minimally performs two functions during splicing: it defines the boundaries of the intron and then catalyzes its removal (**Figure 4.4.1**).

The Chemistry of Splicing and the Architecture of the Intron

The chemistry of splicing is relatively simple, comprising only two consecutive S_N2-type transesterification reactions (Moore & Sharp 1993). During the first catalytic step, the phosphodiester bond between the 5' exon and the intron is cleaved by nucleophilic attack from the 2' hydroxyl of the branchpoint adenosine (**Figure 4.1B**). The first step of catalysis results in the formation of a 2', 5' branched RNA (known as a lariat intermediate) and a free 5' exon. Rearrangements within the active site of the spliceosome then occur that permit the free 3' hydroxyl of the 5' exon to attack the intron-3' exon junction and result in the simultaneous ligation of the exons and excision of the lariat intron.

Human introns vary greatly in size from less than a hundred nucleotides to several hundred thousand (Scherer 2008). Despite the range of intron sizes, most exons are much shorter, ranging just a few hundred nucleotides. Because of this disparity in length, selecting the proper position of the pre-mRNA for excision represents a formidable task. This problem is further complicated by the relative lack of sequence conservation between introns. Nonetheless, the spliceosome manages to recognize introns though direct interaction with consensus sequences known as the 5' splice site, the branchsite sequence, and the 3' splice site. The 5' splice site is found at the junction between the 5' exon and the intron. Downstream of the 5' splice site is the branchsite sequence containing the adenosine that serves as the branching point in the lariat intron product. The final components of the intron necessary for splicing are the polypyrimidine tract and the 3' splice site, with the latter marking the junction between the intron and 3' exon. The consensus splice site and branchsite sequences in eukaryotes such as the yeast *Saccharomyces cerevisiae* are well conserved whereas these sequences are more variable in metazoans. The human consensus sequences are shown in **Figure 4.1A**.

Despite the existence of consensus sequences, splice site recognition is complicated by the presence of both non-consensus splice sites that must be used by the spliceosome and pseudo-splice sites that must be avoided. The spliceosome can accurately discriminate *bona fide* splice sites from false signals and does so with input from other factors in nuclear RNA processing. Amazingly, the human spliceosome is proposed to properly recognize >2500 different 5' splice site sequences (Roca & Krainer 2009). Identification of the correct splice sites is accomplished through multiple types of non-covalent interactions occurring during spliceosome assembly and activation. The spliceosome is an incredibly sensitive and flexible enzyme that can discriminate between subtle differences in a large number of RNA sequences.

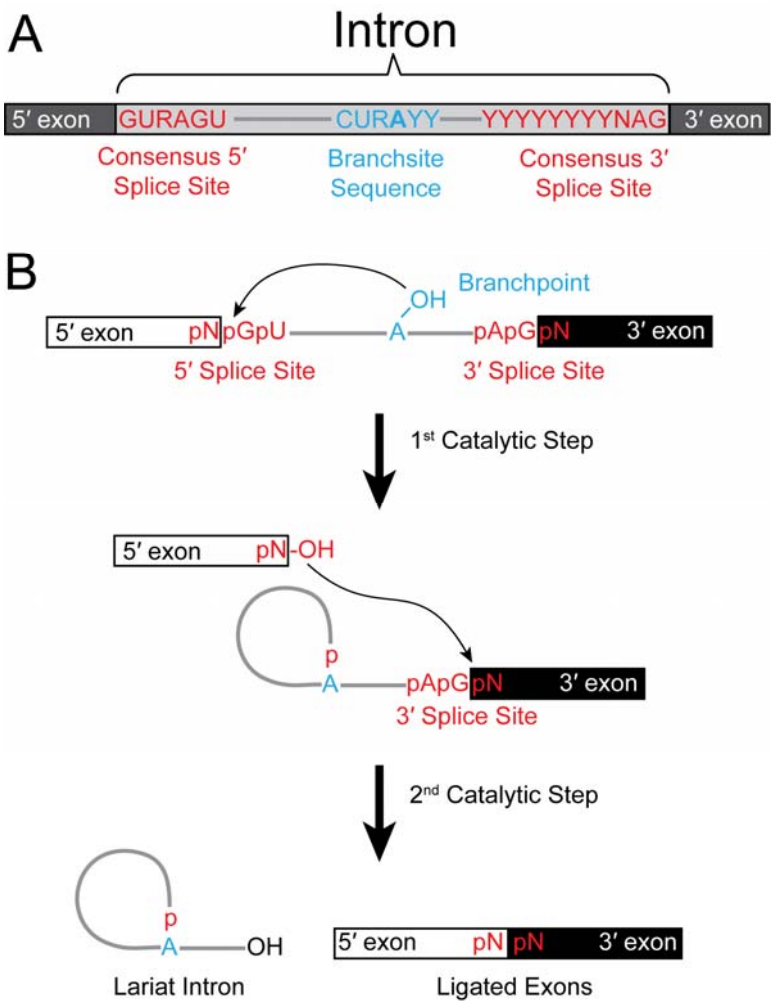


Figure 4.1. (A) The architecture of the intron. The boundaries of the intron are known as the 5' and 3' splice sites (shown in red). In addition to the splice sites, the pre-mRNA contains a branchsite region (blue) that contains the reactive adenosine (bold) necessary for catalysis. Polypyrimidine tracts are often found between the branchsite sequence and the 3' splice site. The consensus sequences as shown with *R* being any purine, *Y* being any pyrimidine, and *N* being any base (B) The chemical steps of splicing. During the first step, the 2' hydroxyl of the branchsite adenosine (blue), attacks the 5' splice site to liberate the 5' exon and generate a lariat intermediate. In the second step, the 3' hydroxyl of the free 5' exon attacks the 3' splice site to generate the spliced mRNA and lariat intron products.

The Spliceosome is Composed of snRNAs and Proteins

Along with the discovery of introns and splicing by Roberts and Sharp, a second major breakthrough in splicing occurred when Joan Steitz proposed that small nuclear ribonucleoproteins (snRNPs, pronounced “snurps”) were the complexes that carried out the splicing reaction (Lerner et al. 1980). Spliceosomal snRNPs are large complexes containing both small nuclear, uridine-rich RNAs (the “U” snRNAs) and proteins. Sequencing of the U1 snRNA revealed that its 5' end was complementary to many 5' splice site sequences, suggesting that the U1 snRNP played a role in the splicing reaction. It was later determined that U1 along with other snRNPs assembled on pre-mRNAs into spliceosomal complexes (Wahl et al. 2009).

The core of the spliceosome is composed of four major, highly conserved subcomplexes: the U1 and U2 snRNPs, the U4/U6.U5 tri-snRNP (composed of three individual snRNPs), and the protein-only nineteen complex (NTC) (Will & Lührmann 2011). These subcomplexes work together with a host of additional protein factors to process the RNA transcript. Each snRNP contains one or more snRNA and a variable number of snRNP-specific proteins (**Table 4.1**). The U1, U2, U4, and U5 snRNPs all contain a heptameric Sm protein ring bound to a U-rich sequence near their 3' end, while the U6 snRNA is bound by the LSm2-8 heptamer (Matera & Wang 2014). Each snRNP serves a distinct purpose in splicing. The U1 and U2 snRNPs are responsible for recognizing the 5' splice site and branchsite sequences, respectively, through snRNA:pre-mRNA interactions. Basepairing between the U2 snRNA and the pre-mRNA branchsite sequence is responsible for positioning the reactive branchpoint adenosine by bulging it from a U2 snRNA/pre-mRNA duplex to allow for catalysis (Smith et al. 2009).

The U4/U6.U5 tri-snRNP is pre-assembled prior to spliceosome formation from individual U4/U6 di-snRNPs and U5 snRNPs. U4/U6 contains extensive basepairing between the U4 and U6 snRNAs. The U6 snRNA is the mostly highly conserved of the spliceosomal snRNAs, and it plays a critical in catalysis by the spliceosome. The U6 snRNA contains at least three distinct features that are absolutely required for catalysis: the ACAGA-box (pronounced “ah-cah-ga”),

the U6 internal stem loop (ISL), and the AGC catalytic triad (Brow 2002). The U6 ISL and AGC triad are thought to play important roles in catalysis through the coordination of catalytic metal ions, while the ACAGA-box positions the 5' splice site for exon ligation. Proper formation and alignment of these features is aided, in part, by duplex formation between the U2 and U6 snRNAs in the spliceosome. Possibly to prevent premature or aberrant splicing, the U6 ISL is disrupted in the tri-snRNP by annealing of the U6 snRNA to the U4 snRNA, which acts as a chaperone for U6. Therefore, the U4/U6 di-snRNA must be unwound prior to catalysis (Brow 2002; Wahl et al. 2009).

The U5 snRNA is responsible for contacting the exons of the pre-mRNA to ensure proper splice site alignment (Newman 1997). Furthermore, the U5 snRNP contains three important protein components: Prp8, Snu114 and Brr2. Prp8 is a highly conserved 220 kDa protein that is thought to help structure and modulate the active site of the spliceosome as evidenced by a large number of chemical crosslinks between the protein and critical RNA components (Teigelkamp et al. 1995). Prp8 has been shown to interact with the U5 and U6 snRNAs, the 5' and 3' splice sites, and the branchpoint adenosine of the pre-mRNA substrate. Prp8 contains multiple protein domains including those resembling endonucleases, reverse transcriptases, and RNase H (Galej et al. 2013). The endonuclease and reverse transcriptase domains are particularly noteworthy since these resemble proteins encoded by group II self-splicing introns that facilitate DNA homing and intron insertion. This similarity suggests that at least some protein components of the spliceosome, like the snRNAs as described below, evolved from a common ancestor shared by the group II intron machinery. In addition to scaffolding the active site, Prp8 also regulates the activity of Brr2, which is responsible for unwinding the U4/U6 snRNA duplex to allow catalytic structures within U6 to form. The last unique protein component of U5, Snu114, is homologous to the translation elongation factor EF-G and uses GTP hydrolysis to further coordinate Brr2 activity (Small et al. 2006).

The final major subcomplex of the spliceosome is the NTC (Hogg et al. 2010). The NTC was named for a core spliceosomal protein, Prp19, and contains approximately seven other proteins. The NTC is required for stabilizing interactions between the U5 and U6 snRNAs with the pre-mRNA. Additionally, the NTC protein Cwc2 is essential for splicing and contacts the U6 ISL and the duplex formed between the ACAGA-box and 5' splice site (Rasche et al. 2012). This suggests that one function of the NTC may be to bring distal regions of the spliceosome into close proximity for catalysis.

The spliceosome requires many other accessory proteins not stably associated with the U snRNPs or the NTC complex to properly excise introns. For example, two important components for spliceosome assembly are splicing factor 1 (SF1), which recognizes the pre-mRNA branchsite prior to U2 association, and the U2 auxillary factor heterodimer (U2AF35/65, also known as U2AF1/2) that binds the polypyrimidine tract and 3' splice site AG dinucleotide (Wahl et al. 2009). SF1 is ultimately replaced at the branchsite by U2. Another major class of proteins that associate with the spliceosome, are the DExH/D box proteins. These proteins contain a signature sequence motif of aspartate (D), glutamate (E), usually a hydrophobic amino acid (x), and either a second aspartate or histidine (H). DExH/D box proteins are RNA-dependent ATPases that are extensively involved in RNA metabolism often through destabilization or unwinding of RNA duplexes (Jankowsky 2011). Finally, metazoan spliceosomes are frequently associated with splicing regulatory (SR) proteins. SR proteins usually contain arginine-serine (RS) repeat motifs and help to guide spliceosome formation at the proper locations in a pre-mRNA (Long & Caceres 2009).

In sum, a spliceosome contains 5 snRNAs and a core set of 75 proteins that have been conserved between yeast and humans. The number of accessory proteins that have been isolated with spliceosomes in humans numbers over 300 (Jurica & Moore 2003). The snRNPs and spliceosomes are consequently some of the largest macromolecular machines inside the cell. Due to their size and compositional complexity, only limited high-resolution crystallographic

structural information is available for the individual snRNPs or spliceosomal proteins. However, lower resolution cryo-electron microscopy structures are available for several snRNPs and spliceosomal complexes. Obtaining high-resolution structural information of spliceosomes and spliceosomal components is currently being pursued by many laboratories.

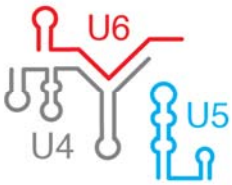
Complex	*S. cerevisiae*		*H. sapiens*		snRNA Structure
	snRNA length	# proteins	snRNA length	# proteins	
U1 snRNP	568nt	17	164nt	10	
U2 snRNP	1,175nt	19	187nt	19	
U4/U6.U5 tri-snRNP	112nt 160nt 179-214nt	34	107nt 145nt 116nt	36	
Nineteen Complex (NTC)	———	8	———	7	———

Table 4.1. Comparison of the snRNP components for *S. cerevisiae* and *H. sapiens* spliceosomes. Cartoon structures of the human snRNAs are shown.

Spliceosomes are Assembled from snRNPs on Introns

The spliceosome is not a static complex, but rather a dynamic machine that assembles stepwise from preformed catalytically inactive subcomplexes for each round of splicing (Hoskins et al. 2011). In the canonical stepwise assembly pathway, the U1 snRNP first recognizes the 5' splice site through base pairing of the 5' end of the U1 snRNA to the pre-mRNA (**Figure 4.2**). Elsewhere on the intron, the branchsite, polypyrimidine tract, and 3' splice sites are recognized by SF1, U2AF65, and U2AF35, respectively. Together, this forms the first intermediate in spliceosome assembly, the E complex. E complex is then converted to A complex upon ATP-dependent exchange of SF1 for U2 at the branchsite. On long introns, U1 and U2 may first interact across exons to form an exon definition complex prior to interactions that occur across the intron to be spliced (Wahl et al. 2009). Regardless, A complex formation is often followed by the recruitment of the U4/U6.U5 tri-snRNP to the pre-mRNA substrate to form the catalytically inactive B complex spliceosome that contains all five of the U snRNPs. Assembly of B complex is followed by Brr2-dependent activation and the formation of B^{ACT}, wherein the NTC has joined the complex, U1 and U4 snRNPs have been lost, and the active site of the spliceosome has begun to form. Activation involves multiple rearrangements that result in U6 replacement of U1 at the 5' splice site, formation of the U6 ISL, and duplex formation between the U2 and U6 snRNAs. The action of the DExD/H box protein Prp2 helps to promote the first step of catalysis (Kim & Lin 1993), and spliceosomes capable of carrying out the chemistry of splicing are called C complexes. The first step of catalysis occurs, followed by several rearrangements within the active site facilitated by the DExD/H box protein Prp16 that reposition the components for the second step of splicing (Schwer & Guthrie 1991). After the second step, the spliced mRNA is released from the intron lariat spliceosome (ILS), and the post-catalytic complex is disassembled and the components are recycled for another round of splicing (Fourmann et al. 2013). The spliceosome is a single turnover enzyme, meaning that the assembly and activation must occur anew on each intron to be spliced.

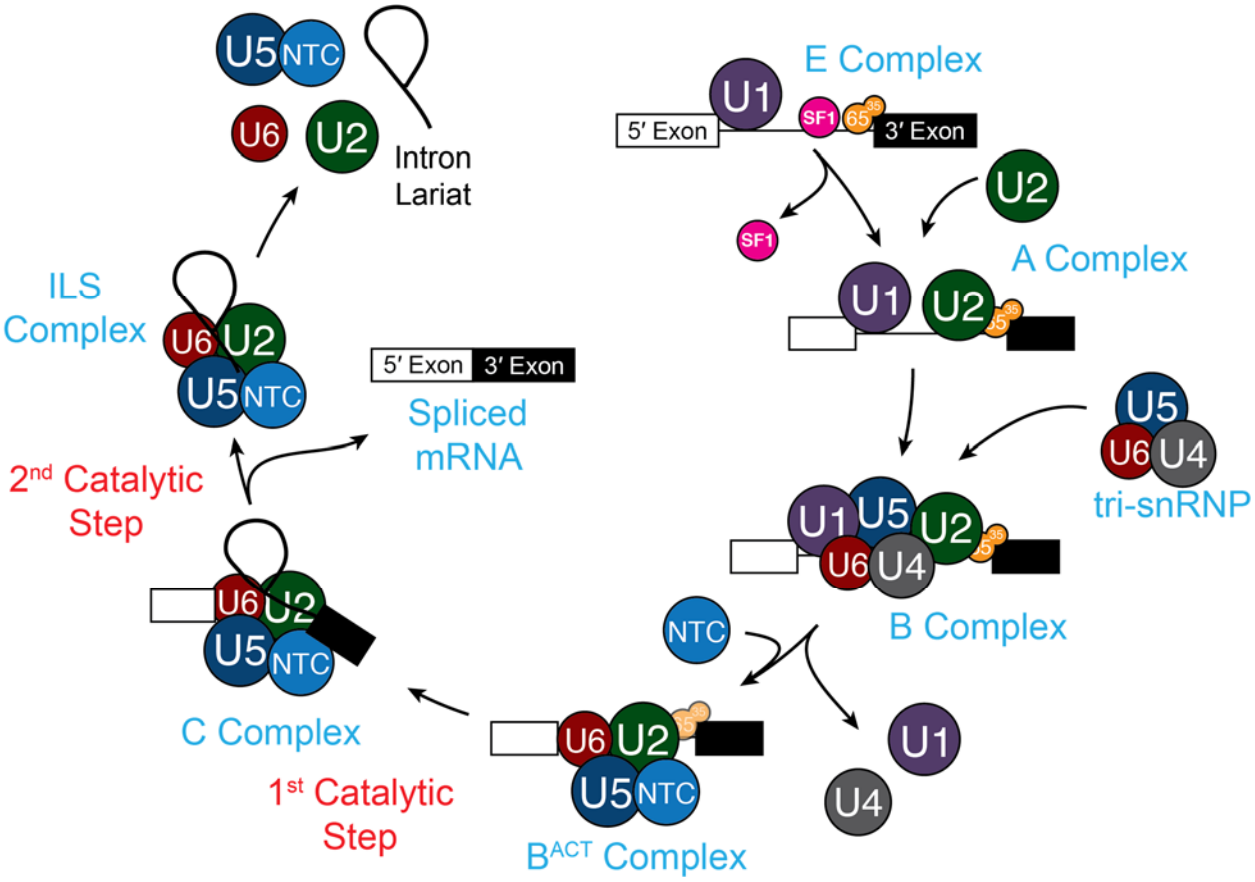


Figure 4.2. Stepwise assembly mechanism of the spliceosome. Spliceosome assembly begins at E complex, with U1 bound at the 5' splice site, SF1 at the branchsite sequence and U2AF65/35 at the 3' splice site. Exchange of SF1 for U2 generates A complex, which is converted to B complex upon addition of the U4/U6.U5 tri-snRNP. Activation of the spliceosome releases the U1 and U4 snRNPs from the spliceosome and generates the catalytically ready B^{ACT} complex. It is unknown if U2AF65/35 are present in the B^{ACT} spliceosome and are thus shown in lighter colors. The first step of catalysis occurs and C complex forms. After the second step of catalysis, the mRNA is released and the intron lariat spliceosome (ILS) is disassembled for further rounds of splicing.

The Active Site of the Spliceosome.

Spliceosomes must bring together distant regions of the pre-mRNA along with spliceosomal snRNAs and proteins that enable catalysis. Alignment of the reactive groups occurs in part through scaffolding of the snRNA and pre-mRNA through the basepairing interactions between the intron and the U2 and U6 snRNAs (**Figure 4.3A**). These reactive groups must also be aligned with the U6 ISL, and it is thought that spliceosomal proteins such as Prp8 and the NTC play key roles in juxtaposing all of these functional groups. A number of experiments have found evidence that the U6 ISL plays a key role in coordinating essential magnesium ions for catalysis (Fica et al. 2013). Despite the necessity for proper alignment, the spliceosome active site must also be flexible since it is remodeled between the first and second steps of splicing to permit juxtaposition of the 5' exon and 3' splice site for exon ligation.

One striking feature of the spliceosomal active site is its similarity to that found in group II introns. Group II introns are catalytic pieces of “self-splicing” RNA that carry out identical chemical steps (5' splice site cleavage/lariat formation and exon ligation) to those of the spliceosome. For both spliceosomes and group II introns, the chemical steps have been shown to be reversible and metal-ion dependent (Chin & Pyle 1995) (Tseng & Cheng 2008). Remarkably, domain V of group II introns possesses many of the same catalytic features as the U6 snRNA (**Figure 4.3B**). Crystal structures of group II introns have been used to model the spliceosomal active site and enabled experiments that found direct evidence for the presence of two catalytic magnesium ions (Toor et al. 2008; Fica et al. 2013). It is likely that the spliceosome and group II intron use a “two-metal mechanism” (Steitz & Steitz 1993) for catalyzing phosphodiester bond cleavage and formation with each magnesium ion playing a role in stabilizing the nucleophile or leaving group. In addition to the similarity between Prp8 domains and group II intron encoded proteins, this conservation of active site structure and mechanism between the spliceosome and group II introns provide strong evidence for evolution of the spliceosome from a group II intron-like ancestor.

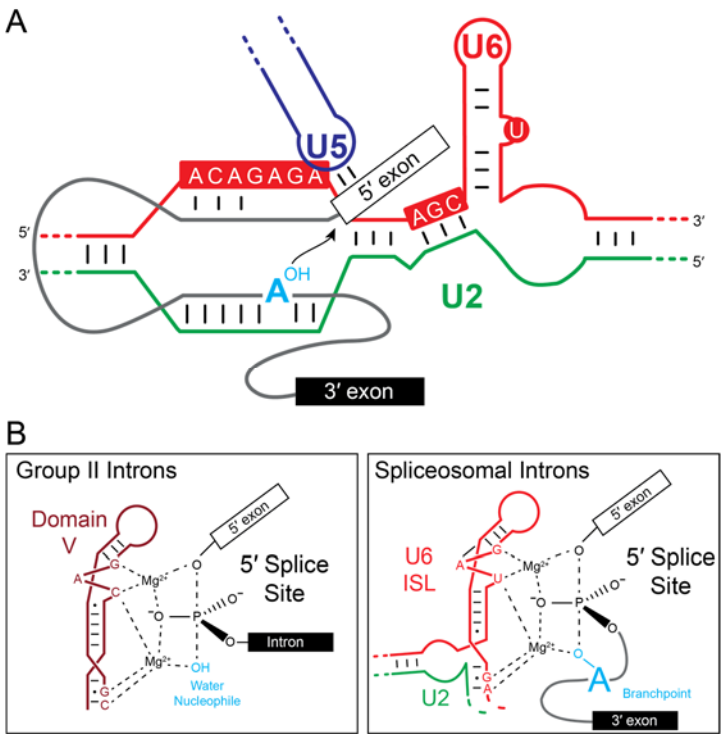


Figure 4.3. The active site of the spliceosome. (A) The catalytically active spliceosome contains a complex containing the U2, U6 and U5 snRNPs. U2 and U6 snRNAs are extensively basepaired with each other and the substrate. U2 makes contact with the branchsite region and results in the formation of the bulged adenosine residue that provides the nucleophile for the first step. U6 contacts the 5' splice site with the conserved ACAGAGA sequence and helps to coordinate catalytic metal ions with the AGC triad, the U6 ISL, and a bulged uridine. The U5 snRNA contacts the 5' exon to help position the substrate for catalysis. (B) Comparison of the group II intron and spliceosomal active sites. The U6 ISL is homologous the domain V of the group II intron. Both coordinate catalytic metal ions (probably Mg^{2+}) and contain similar geometries to promote catalysis. The first step of catalysis is shown for both enzymes and some group II introns can use water nucleophiles to promote 5' exon cleavage. Figures in (A) and (B) were adapted from Konarska *et al.* (2006) and Fica *et al.*, *Nature* (2013), respectively.

DExD/H Box Proteins are Essential Cofactors of the Spliceosome

DExD/H-box proteins are ubiquitously associated with processes that involve RNA, ranging from pre-mRNA splicing to mRNA translation and degradation. These enzymes couple ATP hydrolysis to RNA binding or duplex unwinding (helicase) activity in order to facilitate structural or compositional rearrangements in different RNA-protein complexes. Eight different DExD/H-box ATPases are required for splicing: the DEAD-box proteins UAP56, Prp5, and Prp28; the DEIH-box U4/U6 unwindase Brr2; and the DEAH-box proteins Prp2, Prp16, Prp22, and Prp43 (Chang et al. 2013). Each plays a distinct role in splicing. UAP56 and Prp5 are involved in spliceosome assembly; Prp28, Brr2, and Prp2 are involved in activating the spliceosome for catalysis; Prp16 and Prp22 are responsible for remodeling the catalytic spliceosome; and Prp43 is necessary for spliceosome disassembly. Consequently, while the chemical steps of splicing do not require ATP, ATP hydrolysis by DExD/H box proteins is necessary for splicing.

In addition to their roles in regulating spliceosomal conformational changes, several of the ATPases also have roles in splicing fidelity. Examples of this can be found in studies of the yeast homologs of Prp16 and Prp22 (Mayas et al. 2006) (Koodathingal et al. 2010). The ATPase activity of Prp16 is involved in transitioning the spliceosome from the first to second step of catalysis. Prior to exon cleavage, Prp16 is also involved in “kinetic proofreading” of suboptimal substrates (**Figure 4.4**). Spliceosomes that have assembled on pre-mRNAs with mutations in the branchsite region proceed through the catalytic steps more slowly than RNAs that contain the consensus sequences. In these suboptimal spliceosomes, hydrolysis of ATP diverts the complex from the splicing pathway to a spliceosome discard pathway. In this way, Prp16 is able to increase the fidelity of pre-mRNA splicing by rejecting suboptimal substrates or poorly assembled spliceosomes prior to the first step. This kinetic proofreading step is also seen with Prp22, which monitors the spliceosome after exon cleavage but prior to exon ligation. The DEAH-box protein Prp43 provides a critical, irreversible step during discard by disassembling these spliceosomes.

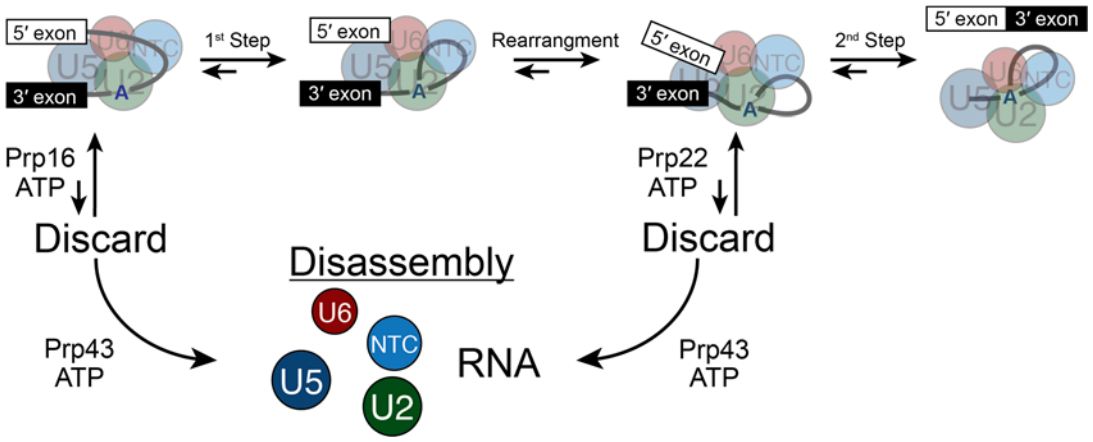


Figure 4.4. Kinetic proofreading during splicing. The forward reactions (toward spliced mRNA) are favored but suboptimal substrates can enter into the pre-mRNA discard pathway. Entry to these pathways are controlled by the action of the DEAH-box ATPases Prp16 and Prp22. While entry into the discard pathway is reversible, discard becomes irreversible if the spliceosomes are subsequently disassembled by the DEAH-box ATPase Prp43.

Alternative Splicing Creates Multiple Products From a Single Gene

While many introns present in eukaryotic genes are spliced from the pre-mRNA transcript constitutively, the removal of other introns can be subject to additional layers of regulation. By altering the splice site decisions during intron removal, multi-exonic transcripts can be spliced together to produce many isoforms of the same gene (**Figure 4.5**). This process is known as alternative splicing. In mammalian cells, 95% of gene transcripts are thought to undergo alternative splicing (Nilsen & Graveley 2010). The generation of multiple mRNA isoforms from a single pre-mRNA is a powerful method to increase the protein-encoding capacity of the genome without increasing the size of the genome. In this way, alternative splicing represents an important source of proteomic diversity. The most common form of alternative splicing is exon skipping in which specific exons are expressed in only a fraction of the total transcripts. Exon skipping can commonly be found in different tissue isoforms of the same gene. The first example of alternative splicing in humans was for gene encoding the peptide hormone calcitonin (Leff & Rosenfeld 1986). Two distinct isoforms of the calcitonin mRNA exist, one that contains only exons 1-4 and a second that contains exons 1-3, 5, and 6. The first mRNA transcript is found predominately in the brain and encodes for calcitonin and the second encodes another peptide called calcitonin-gene-related peptide, which predominates in the thyroid.

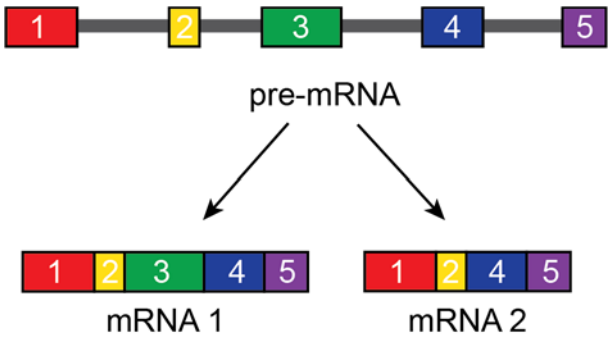


Figure 4.5. Alternative splicing of a pre-mRNA. Single pre-mRNAs can generate multiple splicing products through the use of alternative splice sites. mRNA 1 contains exon 3 (green) which is not found in mRNA 2.

Splicing is Coupled to Other Cellular Processes

Splicing does not occur in isolation but rather is connected to other cellular processes such as transcription, RNA capping, polyadenylation, and chromatin remodeling. Work in several organisms suggests that splicing can occur during or after the synthesis of pre-mRNA. Indeed, interactions between RNA polymerase II and the spliceosomal snRNPs are thought to recruit them to the nascent transcript for processing (Bentley 2014). These interactions may ultimately influence the alternative splicing fate of the pre-mRNA. In humans and other metazoans, nuclear splicing and cytosolic translation and RNA decay can even be linked (Nott et al. 2004). This occurs through deposition of a set of proteins called the exon junction complex (EJC) by the spliceosome just upstream of ligated exons. The presence of EJCs on an mRNA can have profound consequences for the fate of that mRNA once it enters the cytosol. EJCs can influence the number of proteins that can be translated from a mRNA, target the mRNA for rapid degradation, or impact other aspects of mRNA biology (Moore & Proudfoot 2009).

Defects in Splicing Cause Disease

More that 15% of human genetic diseases arise from single point mutations that cause defects related to pre-mRNA splicing (Faustino & Cooper 2003). These mutations can be classified as either *cis*- or *trans*-acting. *Cis*-acting mutations affect the pre-mRNA substrate, usually at constitutive or alternative splice sites but other regions of the RNA can be affected as well. These mutations change how the pre-mRNA is processed by the spliceosome and can result in, for example, failure to remove introns, choice of incorrect splices sites, or deletions of entire exons in the mRNA. Ultimately this may change the protein that is encoded by the mRNA; sometimes by just a few amino acids as is the case in Frasier syndrome. In this disease a mutant 5' splice site causes deletion of just three amino acids from the Wilms Tumor suppressor protein, WT1, and results in severe defects in urogenital development.

Trans-acting mutations affecting the basal splicing machinery can also cause severe human diseases such as retinitis pigmentosa (RP), cancer, or spinal muscular atrophy (SMA). RP

is the leading cause of inherited blindness in the world and is characterized by progressive retinal degeneration. This eventually results in total blindness from loss of rod photoreceptor cells. Several genetic causes of RP exist, with some resulting from defects in components of the U4/U6.U5 tri-snRNP, such as Prp8 or Brr2. Some cancers, such as the blood cell cancers caused by myelodysplastic syndromes, have also been found to be specifically associated with single point mutations in components of the U2 snRNP, SF1, or U2AF35/65. Degenerative neuromuscular disorders such as SMA can arise from defects in production and assembly of the snRNPs themselves. SMA is attributed to mutations in the survival of motor neuron (SMN) complex, which plays a major role in the assembly of newly produced U1, U2, U4, and U5 snRNPs. Patients with SMA exhibit a progressive loss of spinal cord motor neurons leading to total paralysis of the voluntary muscles and early death.

Conclusion

The discovery of RNA splicing represented one of the great surprises in biology during the last century. Equally surprising was the complexity of the spliceosomal machinery that carries out this reaction: a multi-megadalton complex of snRNAs and proteins that briefly associate together on a transcript to remove an intron before being dismantled. In addition to this complexity and transience, splicing is absolutely essential for processing of the vast majority of human mRNAs. It is likely that the complexity of the spliceosome is precisely what allows it to function on a vast array of substrates and integrate with other cellular processes to produce correct mRNAs. In the coming years advances in structural biology, biophysical methods, and genetic engineering will lead to new insights into the structures of the spliceosome and the underlying biochemical mechanisms of RNA splicing.

REFERENCES

Bentley, D.L., 2014. Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*. 15, 163-175.

Berget, S.M., Moore, C. & Sharp, P.A., 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences*, 74(8), pp.3171–3175.

Brody, E. & Abelson, J., 1985. The “spliceosome”: yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science*, 228(4702), pp.963–967.

Brow, D.A., 2002. Allosteric cascade of spliceosome activation. *Genetics*, 36, pp.333–360.

Chang, T-H. et al., 2013. Functions of the DExD/H-box proteins in nuclear pre-mRNA splicing. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(8), pp.764-774.

Chin, K. & Pyle, A.M., 1995. Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proofreading mechanism for 5'-splice site selection. *RNA*, 1(4), pp.391–406.

Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.J., 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell*, 12(1), pp.1–8.

Faustino, N.A. & Cooper, T.A., 2003. Pre-mRNA splicing and human disease. *Genes & Development*, 17(4), pp.419–437.

Fica, S.M., Mefford, M.A., Piccirilli, J.A., Staley, J.P., 2014. Evidence for a group II intron–like catalytic triplex in the spliceosome. *Nature Structural & Molecular Biology*, 21(5), pp.464–471.

Fica, S.M. et al., 2013. RNA catalyses nuclear pre-mRNA splicing. *Nature*, 503(7475), pp.229–234.

Fourmann, J.B. et al., 2013. Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes & Development*, 27(4), pp.413–428.

Galej, W.P., Oubridge, C., Newman, A.J., Nagai, K., 2013. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature*, 493(7437), pp.1–7.

Hogg, R., McGrail, J. & O'Keefe, R., 2010. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochemical Society Transactions*, 38(4), p.1110.

Hoskins, A.A. et al., 2011. Ordered and Dynamic Assembly of Single Spliceosomes. *Science*, 331(6022), pp.1289–1295.

Jankowsky, E., 2011. RNA helicases at work: binding and rearranging. *Trends in Biochemical Sciences*, 36(1), pp.19–29.

Jurica, M.S. & Moore, M.J., 2003. Pre-mRNA splicing: awash in a sea of proteins. *Molecular Cell*, 12(1), pp.5–14.

Kim, S-H. & Lin, R-J., 1993. Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. *Proceedings of the National Academy of Sciences*, 90(3), pp.888–892.

Konarska, M.M., Vilardell, J. & Query, C.C., 2006. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Molecular Cell*, 21(4), pp.543–553.

Koodathingal, P., Novak, T., Piccirilli, J.A., Staley, J.P., 2010. The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5′ splice site cleavage during pre-mRNA splicing. *Molecular Cell*, 39(3), pp.385–395.

Leff, S.E. & Rosenfeld, M.G., 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annual Review of Biochemistry*, 55(1), pp.1091–1117.

Lerner, M.R. et al., 1980. Are snRNPs involved in splicing?. *Nature*, 283(5743), pp. 220-224.

Long, J.C. & Caceres, J.F., 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, 417(1), p.15.

Matera, A.G. & Wang, Z., 2014. A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*, 15(2), pp.108–121.

Mayas, R.M., Maita, H. & Staley, J.P., 2006. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nature Structural & Molecular Biology*, 13(6), pp.482–490.

Moore, M.J. & Sharp, P.A., 1993. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*, 365(6444), pp.364–368.

Moore, M.J. & Proudfoot, N.J., 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*, 136(4), pp.688–700.

Newman, A.J., 1997. The role of U5 snRNP in pre-mRNA splicing. *The EMBO Journal*, 16(19), pp.5797–5800.

Nilsen, T.W. & Graveley, B.R., 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463(7280), pp.457–463.

Nott, A., Le Hir, H. & Moore, M.J., 2004. Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes & Development*, 18(2), pp.210–222.

Rasche, N. et al., 2012. Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. *The EMBO Journal*, 31(6), pp.1–14.

Roca, X. & Krainer, A.R., 2009. Recognition of atypical 5′ splice sites by shifted base-pairing to U1 snRNA. *Nature Structural & Molecular Biology*, 16(2), pp.176–182.

Scherer, S., 2008. *A Short Guide to the Human Genome*, Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY USA.

Schwer, B.B. & Guthrie, C.C., 1991. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature*, 349(6309), pp.494–499.

Sharp, P.A., 1993. Split genes and RNA splicing. *Nobel Lectures, Physiology or Medicine*, 1995, pp.145–174.

Small, C.E., Leggett, S.R., Winans, A.A., Staley, J.P., 2006. The EF-G-like GTPase Snu114p Regulates Spliceosome Dynamics Mediated by Brr2p, a DExD/H Box ATPase. *Molecular Cell*, 23(3), pp.389–399.

Smith, D.J., Konarska, M.M. & Query, C.C., 2009. Insights into branch nucleophile positioning and activation from an orthogonal pre-mRNA splicing system in yeast. *Molecular Cell*, 34(3), pp.333–343.

Soeiro, E., Birnboim, H.C. & Darnell, J.E., 1966. Rapidly labeled HeLa cell nuclear RNA: II. Base composition and cellular localization of a heterogeneous RNA fraction. *Journal of Molecular Biology*, 19(2), pp.362–372.

Steitz, T.A. & Steitz, J.A., 1993. A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences*, 90(14), pp.6498–6502.

Teigelkamp, S., Newman, A.J. & Beggs, J.D., 1995. Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *The EMBO Journal*, 14(11), p.2602-2612.

Toor, N., Keating K., Taylor, S., Pyle, A.M., 2008. Crystal Structure of a Self-Spliced Group II Intron. *Science*, 320(5872), pp.77–82.

Tseng, C.K. & Cheng, S.C., 2008. Both Catalytic Steps of Nuclear Pre-mRNA Splicing Are Reversible. *Science*, 320(5884), pp.1782–1784.

Wahl, M.C., Will, C.L. & Lührmann, R., 2009. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, 136(4), pp.701–718.

Will, C.L. & Lührmann, R., 2011. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, 3(7), p.a003707.

Chapter Five

SF3b1 Mutations Associated with Myelodysplastic Syndromes Alter the Fidelity of Branchsite Selection in Yeast

This chapter is published in the following form:

Carrocci, T.J. et al. (2017) SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. Nucleic Acids Res. Published online January 6. 2017. http://dx.doi.org/10.1093/nar/gkw1349

ABSTRACT

RNA and protein components of the spliceosome work together to identify the 5' splice site, the 3' splice site, and the branchsite (BS) of nascent pre-mRNA. SF3b1 plays a key role in recruiting the U2 snRNP to the BS. Mutations in human SF3b1 have been linked to many diseases such as myelodysplasia (MDS) and cancer. We have used SF3b1 mutations linked to MDS to interrogate the role of the yeast ortholog, Hsh155, in BS selection and splicing. These alleles change how the spliceosome recognizes the BS and alter splicing when nonconsensus nucleotides are present at the -2, -1, and +1 positions relative to the branchpoint adenosine. This indicates that changes in BS usage observed in humans with SF3b1 mutations may result from perturbation of a conserved mechanism of BS recognition. Notably, different *HSH155* alleles elicit disparate effects on splicing: some increase the fidelity of BS selection while others decrease fidelity. Our data support a model wherein conformational changes in SF3b1 promote U2 association with the BS independently of the action of the DEAD-box ATPase Prp5. We propose that SF3b1 functions to stabilize weak U2/BS duplexes to drive spliceosome assembly and splicing.

INTRODUCTION

The spliceosome is emerging as a potential therapeutic target and a potent driver of human disease (1, 2). While defects in the splicing machinery have previously been implicated in spinal muscular atrophies (3) and some forms of retinitis pigmentosa (4-6), recent evidence suggests strong links between the splicing machinery and cancer (7). The spliceosome is an intricate molecular machine composed of 5 U-rich small nuclear ribonucleoproteins (the U1, U2, U4, U5, U6 snRNPs) that function in concert with numerous other splicing factors to excise introns from nascent pre-mRNA (8, 9). Mutations in several snRNP proteins are implicated in a variety of cancers, while the splicing machinery in general appears to be critical for proliferation of c-MYC associated cancers (10) as well as DNA repair through the ATM signaling pathway (11).

Among splicing factors implicated in disease, the U2 snRNP protein SF3b1 is of particular interest since SF3b1 mutation is strongly correlated with cancers such as uveal melanoma, chronic lymphocytic leukemia (CLL) and myelodysplastic syndromes (MDS) (12-14). Many of the same mutations are associated with different diseases arising from distinct cell lineages (15). Bioinformatic analysis has shown that SF3b1 mutations are correlated with changes in alternative splicing, often due to the selection of cryptic, upstream 3' SS (16). Recent experiments have pointed to alternative BS usage by the spliceosome instigating cryptic 3' SS activation (17-19); however, the mechanisms by which SF3b1 mutations can influence usage of one BS or 3' SS over another are unclear.

SF3b1 is the largest protein of the SF3 complex, which itself is a component of the U2 snRNP. U2 is recruited to introns early in spliceosome assembly and subsequent ATP-dependent transitions result in basepairing of the U2 snRNA to the branchsite (BS) in the pre-spliceosome or spliceosome A complex (**Figure 5.1A**) (20). These transitions require the DEAD-box helicase Prp5/DDX46 (21). U2 then undergoes dramatic conformational changes during splicing resulting in basepairing between the U2 and U6 snRNAs to form the catalytic core of the spliceosome (9).

SF3b1 crosslinks both up- and downstream of the BS in the spliceosome A complex , underlying a role in stabilizing the U2 snRNA/BS duplex and positioning protein factors within the spliceosome that interact with this duplex (22, 23). Recent structures of the catalytically activated (B^{act}) yeast spliceosome (24, 25) and the isolated SF3b complex (26) have revealed the molecular architecture of both human and yeast SF3b1/Hsh155 and other components of the SF3b complex. Hsh155 directly contacts the U2 snRNA/BS duplex and may help stabilize the bulged branchpoint adenosine. Missense mutations found in MDS map to the surface of the HEAT-repeat domain of SF3b1 in the region that interacts with the intron between the BS and 3' SS and nearby the DEAH-box helicase Prp2. This region of SF3b1 is highly conserved among eukaryotes, suggesting its function within the spliceosome is also conserved (**Figure 5.1B**).

SF3b1 is also the target of several antitumor compounds, such as spliceostatin A (27), pladienolide B (28), and herboxidiene (29). The antitumor compound E7107 targets SF3b1 to block ATP-dependent A complex formation as well as a conformational change in U2 that exposes the snRNA region responsible for basepairing to the BS (30). SF3b1 must undergo additional conformational changes during splicing in order to release the U2/BS duplex. Prior to 5' splice site (SS) cleavage, Prp2 remodels the spliceosomal active site, resulting in juxtaposition of the 5' SS and BS as well as a decrease in affinity between the entire SF3 complex, including SF3b1, and the catalytic spliceosome (31-34). Despite this reduced affinity, SF3b1 still influences splicing chemistry, as pladienolide B binds to SF3b1 to both prevent spliceosome assembly and inhibit exon ligation (35). Together, these data from the E7107 and pladienolide B splicing inhibitors suggest that U2 and SF3b1 may undergo similar conformational changes during assembly of the spliceosome and catalysis.

To investigate the impact of SF3b1 on the molecular mechanisms of splicing, we have incorporated naturally occurring human MDS alleles into the yeast SF3b1 ortholog and studied their impact on the well-characterized yeast spliceosome. *In vivo* splicing assays in combination

with an MDS allele-centered yeast two-hybrid (Y2H) screen have allowed us to define the consequences of mutation of a core U2 snRNP protein on both splicing and the association of essential splicing factors. SF3b1 mutations alter usage of nonconsensus BS containing substitutions at the same positions impacted by mutation of the DEAD-box ATPase Prp5; however, the mechanisms by which mutation of these two splicing factors influence BS usage are distinct. Moreover, the Y2H screen also suggests that SF3b1 is a central hub for recruitment of splicing factors to the spliceosome active site, and we show that MDS mutations can interact genetically with Prp2 mutants. Combined, these results suggest that branchsite selection arises from balancing the opposing activities of SF3b1 and Prp5 during spliceosome assembly.

METHODS AND MATERIALS

Saccharomyces cerevisiae strains used in these studies were derived from 46α (kind gift of David Brow), BJ2168, or ySSC026 (kind gift of Soo-Chen Cheng) (36, 37). **Tables 1 and 2** contain detailed lists of strains and plasmids. Yeast transformation and growth was carried out using standard techniques and media (38).

Site-Directed Mutagenesis

Point mutants were generated using inverse polymerase chain reaction (PCR) with Phusion DNA polymerase (New England Biolabs; Ipswich, MA). The PCR was performed for 16 cycles using primers with the desired mutations incorporated at the 5' ends. PCR products were treated with DpnI (New England Biolabs; Ipswich, MA) to remove template, 5' phosphorylated and self-ligated using T4 polynucleotide kinase (New England Biolabs) and T4 DNA ligase (New England Biolabs; Ipswich, MA) and transformed into Top10 competent cells. Individual colonies were screened by sequencing to identify the desired mutation.

Temperature Growth Assays

Yeast strains expressing WT or mutant *HSH155* alleles were grown to mid-log phase in YPD, the OD was adjusted to OD_{600} = 0.5 and equal volumes were spotted onto YPD plates. Plates were incubated at the indicated temperature and scored after 3 days growth at 23°C, 30°C, 37°C and 10 days at 16°C.

ACT1-CUP1 Copper Assays

ACT1-CUP1 reporters and growth assays have been described previously (36). Briefly, yeast strains expressing WT or mutant proteins and *ACT1-CUP1* reporters were grown to mid-log phase in the appropriate media to maintain selection for the plasmids, adjusted to OD_{600} = 0.5 and equal volumes were spotted onto plates containing 0, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4,

0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.25, or 2.5 mM $CuSO_4$. Plates were scored after 3 days growth at 30°C.

RNA Analysis

Yeast were grown in selective liquid media until OD_{600} reached 0.5–0.8. Cells (10 OD_{600} units) were harvested by centrifugation, and total cellular RNA was isolated using a MasterPure Yeast RNA Purification Kit (Epicentre BioTechnologies; Madison, WI) according to the vendor's instructions. Primer extensions reactions were performed using SuperScriptIII reverse transcriptase (ThermoFisher Scientific; Waltham, MA) and the primers YAC6 (5'-GGCACTCATGACCTTC-3') and yU6 (5'-GAACTGCTGATCATCTCTG-3') (39). Primer extension reactions contained 10 µg total cellular RNA, 10 U Superscript III, and 400 nM each [^{32}P]-end-labeled primer. Assembled extension reactions were incubated at 55°C for 1 h and extension products were analyzed using denaturing PAGE (7% acrylamide:bisacrylamide (19:1), 8M urea, 1x TBE). Gels were then transferred to BioRad filter paper, dried, and exposed to a PhosphorImager screen. The PhosphorImager screen was imaged using a Typhoon FLA 9000 biomolecular imager (GE Healthcare Life Sciences; Chicago, IL) and band intensities were quantified using ImageJ software.

Yeast-Two Hybrid Assays

The Hsh155 open reading frame (ORF) was cloned into pGADT7 (Clontech) generating a GAL4-activation domain-Hsh155 fusion. The ORFs of Bud13, Clf1, Cus1, Cus2, Hsh49, Mud2, Prp2, Prp5, Prp11, Prp22, Prp28, Prp43, and Ysf3 were fused to the C-terminus of the Gal4-DNA binding domain in plasmid pGBKT7. Each pair of plasmids was transformed into the S. *cerevisiae* strain Y2H GOLD, which has the Gal4 UAS upstream of the *HIS3* and *ADE2* loci. Expression of fusion proteins was confirmed by western blotting and plasmids were assayed for autoactivation in combination with empty vectors (i.e. pGADT7 without anything cloned into the vector). Interactions were examined by growth on media lacking histidine, leucine and tryptophan. Briefly,

Y2H expressing both fusions were grown in media lacking leucine and tryptophan to maintain selection for the plasmids. Ten-fold serial dilutions beginning with OD_{600} = 0.5 was plated onto solid media and plates were scored after 3 days incubation at 30°C.

TCA Precipitation and Western Blotting

Total protein was isolated by tricholoracetic acid (TCA) precipitation (40). Yeast were grown in selective media until reaching OD_{600} = 0.5-1.0. Ten OD_{600} units were harvested by centrifugation, and cell pellets were washed with 20% TCA. After washing, pellets were suspended in 20% TCA and subjected to mechanical lysis using glass beads. Glass beads were removed and 5% TCA was added to achieve a final concentration of ~10% TCA and precipitated proteins were collected by centrifugation. Pellets were washed with 70% ethanol, followed by solublization in 1M Tris pH 8.0 and subsequent SDS-PAGE.

For western blotting, one volume of 2x Laemmli Buffer was added to TCA-precipitated total protein or soluble yeast whole cell extract and the sample was denatured by incubation at 95°C for 5 min. Centrifugation was used to remove insoluble material and the resulting supernatant was resolved on 4-20% Criterion TGX midi protein gel (200V for 1 hr; Bio-rad; Hercules, CA). Proteins were subsequently transferred to a nitrocellulose membrane for blotting (30 min, 100V, 4°C). The membrane was blocked using 5% (w/v) non-fat dry milk dissolved in 1X TBST and probed using the appropriate antibodies. Blots were developed using Clarity Western ECL substrate (Biorad; Hercules, CA) and imaged using an Imagequant LAS 4000 Imager (GE Healthcare Life Sciences; Chicago, IL). The Prp8 antibody was a kind gift of Soo-Chen Cheng. Antibodies against the HA and c-myc tags were conjugated to horseradish peroxidase (HRP) and obtained from Sigma Aldrich and ThermoFisher Scienitific, respectively. V5 antibody was purchased from Bio-rad AbD Serotech (Hercules, CA). Goat α-rabbit-HRP and goat α-mouse-HRP secondary antibodies Bio-rad AbD Serotech (Hercules, CA).

Table 5.1. Yeast strains used in this study.

Name	Genotype	Description
46α	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2*	Used to generate Hsh155 shuffle strain. Also used to test Hsh155 strains as merodiploids
yAAH0465	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS416-HSH155*	Derived from 46α (kind gift of Dave Brow). Copper sensitive strain used to generate MDS mutant strains for ACT1-CUP1 by plasmid shuffle. Hsh155 was deleted using a KanMX cassette.
yAAH0648	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT}*	Copper sensitive strain expressing Hsh155^{WT}. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH0656	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{E291D}*	Copper sensitive strain expressing Hsh155^{E291D}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0659	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{Y292C}*	Copper sensitive strain expressing Hsh155^{Y292C}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0660	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{R294C}*	Copper sensitive strain expressing Hsh155^{R294C}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0655	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{R294L}*	Copper sensitive strain expressing Hsh155^{R294L}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0661	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{N295D}*	Copper sensitive strain expressing Hsh155^{N295D}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0650	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{H331D}*	Copper sensitive strain expressing Hsh155^{H331D}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0657	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{H331Q}*	Copper sensitive strain expressing Hsh155^{H331Q}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0654	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E}*	Copper sensitive strain expressing Hsh155^{K335E}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0651	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335N}*	Copper sensitive strain expressing Hsh155^{K335N}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0658	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335R}*	Copper sensitive strain expressing Hsh155^{K335R}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0653	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335T}*	Copper sensitive strain expressing Hsh155^{K335T}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0652	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{P369E}*	Copper sensitive strain expressing Hsh155^{P369E}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH1147	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{P369K}*	Copper sensitive strain expressing Hsh155^{P369K}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0649	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G}*	Copper sensitive strain expressing Hsh155^{D450G}. Used for ACT1-CUP1 assays after transformation with reporter plasmids

yAAH1151	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{G409E}*	Copper sensitive strain expressing Hsh155^{G409E}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH1152	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K410N}*	Copper sensitive strain expressing Hsh155^{K410N}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yTJC0161	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155$^{R294L/H331D}$*	Copper sensitive strain expressing Hsh155$^{R294L/H331D}$. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yTJC0162	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155$^{R294L/K335E}$*	Copper sensitive strain expressing Hsh155$^{R294L/K335E}$. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yTJC0163	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155$^{R294L/D450G}$*	Copper sensitive strain expressing Hsh155$^{R294L/D450G}$. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yTJC0158	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155$^{N295D/H331D}$*	Copper sensitive strain expressing Hsh155$^{N295D/H331D}$. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yTJC0159	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155$^{N295D/K335E}$*	Copper sensitive strain expressing Hsh155$^{N295D/K335E}$. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yTJC0160	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155$^{N295D/D450G}$*	Copper sensitive strain expressing Hsh155$^{N295D/D450G}$. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH1569	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} prp5::hphMX pRS416-PRP5*	Copper sensitive strain used to generate Prp5 mutants with Hsh155^{WT} expressed from a plasmid. Prp5 mutations were incorporated by plasmid shuffle. Used for ACT1-CUP1 assays
yAAH1570	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} prp5::hphMX pRS416-PRP5*	Copper sensitive strain used to generate Prp5 mutants with Hsh155^{K335E} expressed from a plasmid. Prp5 mutations were incorporated by plasmid shuffle. Used for ACT1-CUP1 assays
yAAH1571	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} prp5::hphMX pRS416-PRP5*	Copper sensitive strain used to generate Prp5 mutants with Hsh155^{D450G} expressed from a plasmid. Prp5 mutations were incorporated by plasmid shuffle. Used for ACT1-CUP1 assays
yAAH1572	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} prp5::hphMX pRS413-PRP5^{WT}*	Copper sensitive strain expressing Hsh155^{WT} and Prp5^{WT} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1573	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} prp5::hphMX pRS413-PRP5^{AAAA}*	Copper sensitive strain expressing Hsh155^{WT} and Prp5^{AAAA} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1574	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} prp5::hphMX pRS413-PRP5^{E235A}*	Copper sensitive strain expressing Hsh155^{WT} and Prp5^{E235A} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1575	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} prp5::hphMX pRS413-PRP5^{N399D}*	Copper sensitive strain expressing Hsh155^{WT} and Prp5^{N399D} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1576	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} prp5::hphMX pRS413-PRP5^{TAG}*	Copper sensitive strain expressing Hsh155^{WT} and Prp5^{TAG} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1579	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} prp5::hphMX pRS413-PRP5^{WT}*	Copper sensitive strain expressing Hsh155^{K335E} and Prp5^{WT} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.

yAAH1580	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} prp5::hphMX pRS413-PRP5^{AAAA}*	Copper sensitive strain expressing Hsh155^{K335E} and Prp5^{AAAA} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1581	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} prp5::hphMX pRS413-PRP5^{E235A}*	Copper sensitive strain expressing Hsh155^{K335E} and Prp5^{E235A} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1582	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} prp5::hphMX pRS413-PRP5^{N399D}*	Copper sensitive strain expressing Hsh155^{K335E} and Prp5^{N399D} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1583	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} prp5::hphMX pRS413-PRP5^{TAG}*	Copper sensitive strain expressing Hsh155^{K335E} and Prp5^{TAG} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1586	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} prp5::hphMX pRS413-PRP5^{WT}*	Copper sensitive strain expressing HSH155^{D450G} and Prp5^{WT} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1587	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} prp5::hphMX pRS413-PRP5^{AAAA}*	Copper sensitive strain expressing HSH155^{D450G} and Prp5^{AAAA} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1588	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} prp5::hphMX pRS413-PRP5^{E235A}*	Copper sensitive strain expressing HSH155^{D450G} and Prp5^{E235A} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1589	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} prp5::hphMX pRS413-PRP5^{N399D}*	Copper sensitive strain expressing HSH155^{D450G} and Prp5^{N399D} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yAAH1590	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} prp5::hphMX pRS413-PRP5^{TAG}*	Copper sensitive strain expressing HSH155^{D450G} and Prp5^{TAG} from plasmids. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
Y2HGOLD	*MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1UAS–Gal1TATA–His3, GAL2UAS–Gal2TATA–Ade2 URA3::MEL1UAS–Mel1TATA AUR1-C MEL1*	Strain used to test interactions by Y2H (Clontech).
yAAH1568	*MAT a prc1 prb1 pep4 leu2 trp1 ura3 PRP5V5 hsh155::KANMX pRS416-HSH155*	Derived from ySSC026 (kind gift of Soo-Chen Cheng). Protease deficient strain with Prp5 V5-tagged at the C-terminus. The genomic copy of Hsh155 was deleted using a KanMX cassette and Hsh155 was expressed from pRS416. This strain was used to generate MDS mutant strains with Prp5-tagged by plasmid shuffle.
yAAH1599	*MAT a prc1 prb1 pep4 leu2 trp1 ura3 PRP5V5 hsh155::KANMX pRS414-HSH155^{WT}*	Protease deficient strain with Prp5 V5-tagged at the C-terminus. Expresses Hsh155^{WT} from pRS414. This strain was used to generate MDS mutant strains by plasmid shuffle.
yAAH1600	*MAT a prc1 prb1 pep4 leu2 trp1 ura3 PRP5V5 hsh155::KANMX pRS414-HSH155^{R294L}*	Protease deficient strain with Prp5 V5-tagged at the C-terminus. Expresses Hsh155^{R294L} from pRS414. This strain was used to generate MDS mutant strains by plasmid shuffle.
yAAH1603	*MAT a prc1 prb1 pep4 leu2 trp1 ura3 PRP5V5 hsh155::KANMX pRS414-HSH155^{K335E}*	Protease deficient strain with Prp5 V5-tagged at the C-terminus. Expresses Hsh155^{K335E} from pRS414. This strain was used to generate MDS mutant strains by plasmid shuffle.
yAAH1604	*MAT a prc1 prb1 pep4 leu2 trp1 ura3 PRP5V5 hsh155::KANMX pRS414-HSH155^{D450G}*	Protease deficient strain with Prp5 V5-tagged at the C-terminus. Expresses Hsh155^{D450G} from pRS414. This strain was used to generate MDS mutant strains by plasmid shuffle.

yAAH0982	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS416-HSH155 cus2::hphMX*	Copper sensitive strain used to generate MDS mutant strains for ACT1-CUP1 by plasmid shuffle. Cus2 is deleted using an hphMX deletion cassette
yAAH0985	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{WT}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0993	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{E291D} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{E291D}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0996	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{Y292C} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{Y292C}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0997	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{R294C} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{R294C}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0992	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{R294L} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{R294L}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0998	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{N295D} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{N295D}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0987	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{H331D} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{H331D}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0994	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{H331Q} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{H331Q}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0991	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{K335E}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0988	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335N} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{K335N}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0995	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335R} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{K335R}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0990	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335T} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{K335T}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0989	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{P369E}*	Copper sensitive strain expressing Hsh155^{P369E}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yAAH0986	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} cus2::hphMX*	Copper sensitive strain expressing Hsh155^{D450G}. Used for ACT1-CUP1 assays after transformation with reporter plasmids
yTJC0267	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} prp2::hphMX pRS415-PRP2^{WT}*	Copper sensitive strain expressing Hsh155^{WT} and Prp2^{WT} from plasmids. Used for cs assays.
yTJC0269	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} prp2::hphMX pRS415-PRP2^{WT}*	Copper sensitive strain expressing Hsh155^{K335E} and Prp2^{WT} from plasmids. Used for cs assays.
yTJC0271	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} prp2::hphMX pRS415-PRP2^{WT}*	Copper sensitive strain expressing HSH155^{D450G} and Prp2^{WT} from plasmids. Used for cs assays.
yTJC0278	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT} prp2::hphMX pRS415-PRP2^{Q548N}*	Copper sensitive strain expressing Hsh155^{WT} and Prp2^{Q548N} from plasmids. Used for cs assays.

yTJC0270	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K335E} prp2::hphMX pRS415-PRP2^{Q548N}*	Copper sensitive strain expressing Hsh155^{K335E} and Prp2^{Q548N} from plasmids. Used for cs assays.
yTJC0272	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{D450G} prp2::hphMX pRS415-PRP2^{Q548N}*	Copper sensitive strain expressing HSH155^{D450G} and Prp2^{Q548N} from plasmids. Used for cs assays.

Table 5.2. Plasmids used in this study.

#	Plasmid Name	Plasmid Description
pAAH0380	pRS416-HSH155^{WT}	Plasmid used to maintain viability after deletion of genomic copy of *HSH155*. Hsh155 coding region +/- 250bp was cloned into pRS416 using SacI/XhoI sites.
pAAH0403	pRS414-HSH155^{WT}	Plasmid used to generate Hsh155^{WT} strain. Hsh155 coding region +/- 250bp was cloned into pRS414 using SacI/XhoI sites.
pAAH0437	pRS414-HSH155^{E291D}	Plasmid used to generate Hsh155^{E291D} strain. Cloned using SDM of pAAH0403
pAAH0466	pRS414-HSH155^{Y292C}	Plasmid used to generate Hsh155^{Y292C} strain. Cloned using SDM of pAAH0403
pAAH0467	pRS414-HSH155^{R294C}	Plasmid used to generate Hsh155^{R294C} strain. Cloned using SDM of pAAH0403
pAAH0436	pRS414-HSH155^{R294L}	Plasmid used to generate Hsh155^{R294L} strain. Cloned using SDM of pAAH0403
pAAH0468	pRS414-HSH155^{N295D}	Plasmid used to generate Hsh155^{N295D} strain. Cloned using SDM of pAAH0403
pAAH0405	pRS414-HSH155^{H331D}	Plasmid used to generate Hsh155^{H331D} strain. Cloned using Gibson assembly on pAAH0403 digested with SphI/AfeI (SDM) and a gBlock (IDT) spanning the same region and containing the H331D
pAAH0425	pRS414-HSH155^{H331Q}	Plasmid used to generate Hsh155^{H331Q} strain. Cloned using Gibson assembly on pAAH0403 digested with SphI/AfeI (SDM) and a gBlock (IDT) spanning the same region and containing the H331Q
pAAH0435	pRS414-HSH155^{K335E}	Plasmid used to generate Hsh155^{K335E} strain. Cloned using Gibson assembly on pAAH0403 digested with SphI/AfeI (SDM) and a gBlock (IDT) spanning the same region and containing the K335E
pAAH0406	pRS414-HSH155^{K335N}	Plasmid used to generate Hsh155^{K335N} strain. Cloned using Gibson assembly on pAAH0403 digested with SphI/AfeI (SDM) and a gBlock (IDT) spanning the same region and containing the K335N
pAAH0438	pRS414-HSH155^{K335R}	Plasmid used to generate Hsh155^{K335R} strain. Cloned using Gibson assembly on pAAH0403 digested with SphI/AfeI (SDM) and a gBlock (IDT) spanning the same region and containing the K335R
pAAH0424	pRS414-HSH155^{K335T}	Plasmid used to generate Hsh155^{K335T} strain. Cloned using Gibson assembly on pAAH0403 digested with SphI/AfeI (SDM) and a gBlock (IDT) spanning the same region and containing K335T
pAAH0407	pRS414-HSH155^{P369E}	Plasmid used to generate Hsh155^{P369E} strain. Cloned using Gibson assembly on pAAH0403 digested with SphI/AfeI (SDM) and a gBlock (IDT) spanning the same region and containing P369E
pAAH0595	pRS414-HSH155^{P369K}	Plasmid used to generate Hsh155^{P369K} strain. Cloned using SDM of pAAH0403
pAAH0599	pRS414-HSH155^{G409E}	Plasmid used to generate Hsh155^{G409E} strain. Cloned using SDM of pAAH0403
pAAH0600	pRS414-HSH155^{K410N}	Plasmid used to generate Hsh155^{K410N} strain. Cloned using SDM of pAAH0403
pAAH0404	pRS414-HSH155^{D450G}	Plasmid used to generate Hsh155^{D450G} strain. Cloned using Gibson assembly on pAAH0403 digested with

		SphI/AfeI (SDM) and a gBlock (IDT) spanning the same region and containing D450G
pAAH0731	pRS414-HSH155$^{R294L/H331D}$	Plasmid used to generate Hsh155$^{R294L/H331D}$ strain. Plasmid pAAH0436 was digested with SphI/SacI to generate the backbone and then ligated with the SphI/SacI insert from pAAH0405 (H331D)
pAAH0732	pRS414-HSH155$^{R294L/K335E}$	Plasmid used to generate Hsh155$^{R294L/K335E}$ strain. Plasmid pAAH0436 was digested with SphI/SacI to generate the backbone and then ligated with the SphI/SacI insert from pAAH0435 (K335E)
pAAH0733	pRS414-HSH155$^{R294L/D450G}$	Plasmid used to generate Hsh155$^{R294L/D450G}$ strain. Plasmid pAAH0436 was digested with SphI/SacI to generate the backbone and then ligated with the SphI/SacI insert from pAAH0404 (D450G).
pAAH0728	pRS414-HSH155$^{N295D/H331D}$	Plasmid used to generate Hsh155$^{N295D/H331D}$ strain. Plasmid pAAH0468 was digested with SphI/SacI to generate the backbone and then ligated with the SphI/SacI insert from pAAH0405 (H331D)
pAAH0729	pRS414-HSH155$^{N295D/K335E}$	Plasmid used to generate Hsh155$^{N295D/K335E}$ strain. Plasmid pAAH0468 was digested with SphI/SacI to generate the backbone and then ligated with the SphI/SacI insert from pAAH0435 (K335E)
pAAH0730	pRS414-HSH155$^{N295D/D450G}$	Plasmid used to generate Hsh155$^{N295D/D450G}$ strain. Plasmid pAAH0468 was digested with SphI/SacI to generate the backbone and then ligated with the SphI/SacI insert from pAAH0404 (D450G)
-	pGADT7	Obtained from Clontech
pAAH0499	pGADT7-HSH155^{WT}	Plasmid used for Y2H. A PCR product was generated from pAAH0403 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0553	pGADT7-HSH155^{E291D}	Plasmid used for Y2H. A PCR product was generated from pAAH0437 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0554	pGADT7-HSH155^{Y292C}	Plasmid used for Y2H. A PCR product was generated from pAAH0466 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0555	pGADT7-HSH155^{R294C}	Plasmid used for Y2H. A PCR product was generated from pAAH0467 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0556	pGADT7-HSH155^{R294L}	Plasmid used for Y2H. A PCR product was generated from pAAH0436 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0557	pGADT7-HSH155^{N295D}	Plasmid used for Y2H. A PCR product was generated from pAAH0468 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0501	pGADT7-HSH155^{H331D}	Plasmid used for Y2H. A PCR product was generated from pAAH0405 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0558	pGADT7-HSH155^{H331Q}	Plasmid used for Y2H. A PCR product was generated from pAAH0425 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0559	pGADT7-HSH155^{K335E}	Plasmid used for Y2H. A PCR product was generated from pAAH0435 and then ligated into pGADT7 at the XmaI and BamHI sites.

pAAH0560	pGADT7-HSH155^{K335N}	Plasmid used for Y2H. A PCR product was generated from pAAH0406 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0561	pGADT7-HSH155^{K335R}	Plasmid used for Y2H. A PCR product was generated from pAAH0438 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0562	pGADT7-HSH155^{K335T}	Plasmid used for Y2H. A PCR product was generated from pAAH0424 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0500	pGADT7-HSH155^{P369E}	Plasmid used for Y2H. A PCR product was generated from pAAH0407 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0634	pGADT7-HSH155^{P369K}	Plasmid used for Y2H. A PCR product was generated from pAAH0595 and then ligated into pGADT7 at the XmaI and BamHI sites.
pAAH0502	pGADT7-HSH155^{D450G}	Plasmid used for Y2H. A PCR product was generated from pAAH0404 and then ligated into pGADT7 at the XmaI and BamHI sites.
-	pGBKT7	Obtained from Clontech
pAAH0494	pGBKT7-BUD13	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0495	pGBKT7-CLF1	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0511	pGBKT7-CUS1	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0496	pGBKT7-CUS2	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0644	pGBKT7-HSH49	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0497	pGBKT7-MUD2	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0645	pGBKT7-PRP2	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0498	pGBKT7-PRP5	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0690	pGBKT7-PRP9	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the XmaI and SalI sites.
pAAH0646	pGBKT7-PRP11	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0715	pGBKT7-PRP22	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and SalI sites.
pAAH0716	pGBKT7-PRP28	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and SalI sites.

pAAH0647	pGBKT7-PRP43	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and SalI sites.
pAAH0722	pGBKT7-SLU7	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0720	pGBKT7-YSF3	Plasmid used for Y2H. A PCR product was generated from *S. cerevisiae* genomic DNA and then ligated into pGBKT7 at the NdeI and BamHI sites.
pAAH0691	pRS416-PRP5^{WT}	Plasmid used to maintain viability after deletion of genomic copy of *PRP5.* Prp5 coding region +/- 500bp was cloned into pRS416 at the BamHI and NotI sites.
pAAH0692	pRS413-PRP5^{WT}	Plasmid used to generate Prp5^{WT} strains.
pAAH0708	pRS413-PRP5^{AAAA}	Plasmid used to generate Prp5^{AAAA} strains. Cloned using SDM of pAAH0692
pAAH0709	pRS413-PRP5^{E235A}	Plasmid used to generate Prp5^{E235A} strains. Cloned using SDM of pAAH0692
pAAH0710	pRS413-PRP5^{N399D}	Plasmid used to generate Prp5^{N399D} strains. Cloned using SDM of pAAH0692
pAAH0711	pRS413-PRP5^{TAG}	Plasmid used to generate Prp5^{TAG} strains. Cloned using SDM of pAAH0692
-	pSP6ACT1_WT	Plasmid used for the transcription of the ACT1 pre-mRNA. Gift of Soo-Chen Cheng
-	pSP6ACT1_U257C	Plasmid used for the transcription of the ACT1 U257C pre-mRNA. Gift of Soo-Chen Cheng
-	pACT1CUP1 WT	WT reporter used for ACT1-CUP1 assays. Gift of Dave Brow
-	pACT1CUP1 G1A	5' SS G1A mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U2A	5' SS U2A mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A3U	5' SS A3U mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U4A	5' SS U4A mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U4G	5' SS U4G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 G5A	5' SS G5A mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U254G	BS U254G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A255G	BS A255G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 C256A	BS C256A mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U257A	BS U257A mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U257C	BS U257C mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U257G	BS U257G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A258C	BS A258C mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A258G	BS A258G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query

-	pACT1CUP1 A258U	BS A258U mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 BS C (A259C)	BS A259C mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 C260G	BS C260G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U301G	3' SS gAG mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A302U	3' SS UuG mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	yCP50 Prp2	Plasmid used to maintain viability after deletion of genomic copy of *PRP2.* Gift of RJ Lin
-	pRS415 Prp2	Plasmid used to generate $Prp2^{WT}$ strains. Gift of Dave Brow
pAAH0790	pRS415 $Prp2^{Q548N}$	Plasmid used to generate $Prp2^{Q548N}$ strains. Cloned by SDM of pRS415 Prp2

RESULTS

Given that cancer-causing mutations in human SF3b1 have been implicated in altering BS selection by the spliceosome (18), we reasoned that a library of SF3b1 mutations could be used to produce a set of alleles in yeast that would allow us to dissect the role of the protein. The majority of SF3b1 mutations associated with MDS and other diseases cluster within a region corresponding to the C-terminal HEAT repeats of the protein, specifically repeats four through nine (**Figure 5.1B**). This region is highly conserved (>50% identical) between humans and the yeast SF3b1 ortholog, Hsh155. We deleted the chromosomal *HSH155* gene and maintained yeast viability by expression of wild type (WT) Hsh155 from a low-copy *URA3/CEN6*-containing plasmid. We then generated yeast strains expressing only the MDS alleles by transformation of the *WT/URA3* yeast with *TRP1/CEN6*-containing plasmids with MDS mutant alleles and subsequent 5-FOA selection of the resulting transformants. Because the most frequently mutated position in human disease, K700, corresponds to P369 in yeast, we generated both P369K and P369E alleles. Additionally, we also incorporated two disease alleles (corresponding to G409E and K410N in Hsh155) that have so far only been observed in patients diagnosed with CLL but not MDS (41). All transformants were viable when grown on 5-FOA-containing media and the genotypes were confirmed by plasmid rescue and DNA sequencing (**Figure 5.1C and Figure 5.2B**). In total, we generated a library of 17 isogenic strains containing either WT or one of 16 different missense mutations corresponding to MDS and CLL disease alleles (collectively labeled Hsh155^{MDS}; **Figure 5.1B, Figure 5.2A,** and **Table 5.1**).

Disease Alleles Do Not Affect Cellular Proliferation in Yeast

We initially screened the mutant yeast strains for defects in proliferation or temperature sensitivity, which has often been observed upon mutation of the splicing machinery. All of the mutant yeast strains were viable when expressing only mutant Hsh155. Measurement of doubling

times in liquid culture at 30°C also showed no significant differences between the mutant and WT strains (**Figure 5.2C**). When the growth of each strain was assayed at different temperatures ranging from 16 to 37°C, we detected no discernable difference between any of the mutants and the WT control (**Figure 5.1D and Figure 5.2D**). These data suggest that *HSH155*^{MDS} alleles do not result in general defects in proliferation. As a consequence, MDS mutant Hsh155 proteins are functional and mutations likely do not cause general disruption of pre-mRNA splicing in yeast.

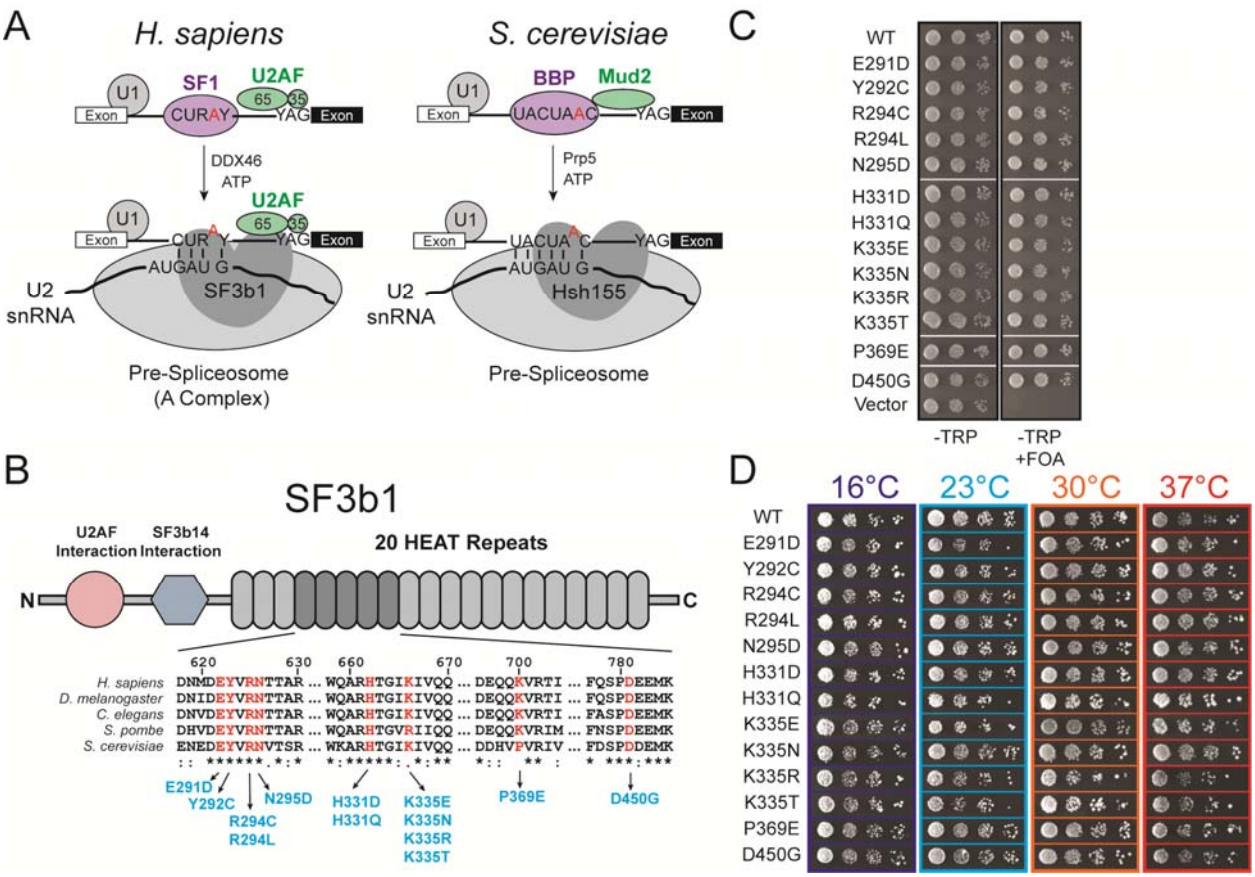


Figure 5.1. MDS alleles of Hsh155 do not affect proliferation in yeast

(A) Schematic comparison of pre-spliceosome formation in *S. cerevisiae* and *H. sapiens*. Hsh155/SF3b1 function as part of the U2 snRNP, interacting with the BS/U2 snRNA duplex and downstream intronic RNA. (B) (Top) Schematic primary structure of SF3b1, with regions known to interact with other splicing factors indicated. (Bottom) Alignment of sequences from *H. sapiens, D. melanogaster, C. elegans, S. pombe* and *S. cerevisae*. Positions found to be frequently mutated in MDS and CLL are shown in red and the amino acid numbering corresponds to *H. sapiens* SF3b1. The most frequently occurring mutations at those positions are shown in blue with the numbering for *S. cerevisiae* Hsh155. (C) Haploid yeast expressing only *HSH155*^{MDS} alleles are viable when plated on FOA. (D) Representative temperature sensitivity growth assays of Hsh155^{MDS} strains plated on YPD. No growth defects are observed in haploid strains

expressing only Hsh155^{MDS} plated on YPD at 16, 25, 30, or 37°C. Successive 10-fold dilutions of a OD_{600} = 0.5 culture are shown.

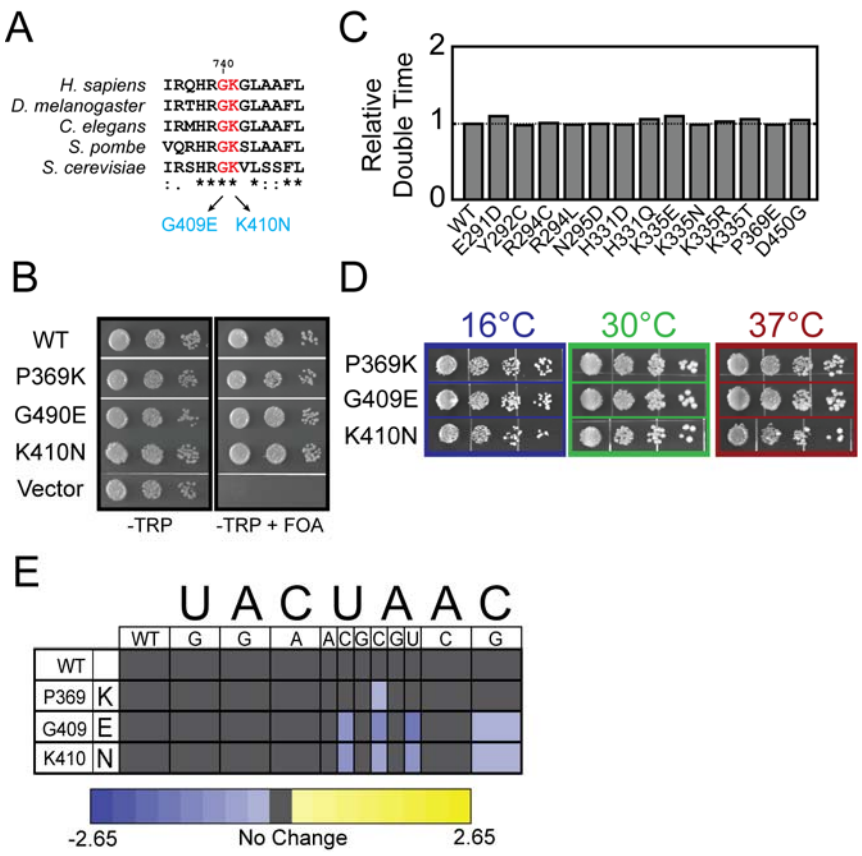


Figure 5.2. Additional MDS and CLL-related mutations do not impair yeast growth

(A) Alignment of SF3b1 sequences from *H. sapiens, D. melanogaster, C. elegans, S. pombe* and *S. cerevisae* highlighting the conservation of amino acid identity at positions corresponding to G740 and K741 in *H. sapiens*. Mutations associated with CLL are shown in blue with *S. cerevisiae* numbering. (B) Haploid yeast strains expressing the shown Hsh155 alleles are viable when plated on FOA. Hsh155^{P369K} corresponds to mutation of *S. cerevisiae* proline 369 to the lysine normally found in *H. sapiens* SF3b1 at that position. (C) Relative doubling times of the given Hsh155 alleles in haploid yeast when grown in solution culture in YPD at 30°C. (D) Growth of haploid yeast strains expressing the given Hsh155 alleles is not impaired when plated on YPD at 16, 30, or 37°C. (**E**) Heatmap summarizing the impact of the given Hsh155 mutations on splicing of ACT1-CUP1 reporter pre-mRNAs containing BS substitutions. The heatmap was generated as in **Figure 5.3F**.

MDS Mutations Alter the Splicing of Pre-mRNAs with Nonconsensus Branchsites

We next assayed our Hsh155^{MDS} mutant library using the *ACT1-CUP1* splicing reporter to evaluate the capacity of each mutant to splice pre-mRNA. This assay utilizes a reporter plasmid expressing the *CUP1* copper resistance gene fused to an intron-containing portion of the actin (*ACT1*) pre-mRNA (**Figure 5.3A**) (36). Expression and proper splicing of this reporter gene confers growth in the presence of Cu^{2+}, with the maximum concentration of Cu^{2+} upon which the yeast grow proportional to the extent of *ACT1-CUP1* pre-mRNA splicing.

Consistent with the proliferation data in **Figure 5.1**, all of the Hsh155^{MDS} strains grew equally well in the presence of Cu^{2+} while expressing an *ACT1-CUP1* reporter with consensus splice sites (**Figure 5.3B** and **Figure 5.2E**). To probe *ACT1-CUP1* pre-mRNA and mRNA levels directly, total cellular RNA was isolated from each strain and primer extension reactions were performed. In all cases we observed the spliced *ACT1-CUP1* mRNA as the predominant species and only small amounts of unspliced pre-mRNA (**Figure 5.3C)**. Taken together these data indicate that the splicing of introns containing consensus splice sites is not affected by these mutations of Hsh155.

To investigate if MDS alleles would alter the splicing of nonconsensus introns, we combined our mutant library with an *ACT1-CUP1* reporter with a single substitution in the BS sequence (*i.e.*, A258U: UACUuAC, substitution in lowercase; **Figure 5.3A)**. In contrast to our results with the consensus *ACT1-CUP1* reporter, yeast strains transformed with the A258U reporter no longer grew equally well in the presence of Cu^{2+} (**Figure 5.3D**). Most strains (*e.g.*, Hsh155^{K335E}) could only support growth at lower levels of Cu^{2+} than Hsh155^{WT}. However, some mutants grew more robustly than Hsh155^{WT} and supported growth at high Cu^{2+} levels (the E291D, R294L and D450G mutants). To validate that the changes in growth are correlated with changes in pre-mRNA splicing, we isolated total RNA from each strain and characterized the relative

amounts of spliced and unspliced reporter by primer extension. The general trends observed in the Cu^{2+} growth assay with the A258U reporter are recapitulated with the primer extension assay with the strains showing the greatest growth inhibition also showing the smallest accumulation of spliced mRNA (**Figure 5.3E**). Thus, MDS variants of Hsh155 alter splicing of introns containing the nonconsensus BS substitution A258U but not the consensus BS.

To assess whether or not the splicing of introns with BS substitutions other than A258U is impacted by MDS mutations, we singly transformed each member of our missense library with ten additional *ACT1-CUP1* reporters encoding at least one substitution at each position within the BS. We then tested each strain to determine the extent of growth on Cu^{2+}-containing media. Given the size of the resultant data set, we created a heatmap showing the growth of each strain with each reporter as the $\log_2$ transform of the ratio of the maximum [Cu^{2+}] tolerated by the Hsh155 mutant strain relative to the maximum [Cu^{2+}] tolerated by the WT strain (**Figure 5.3F**).

The data show a striking and highly specific impact of MDS alleles on the splicing of introns containing substitutions at positions -2, -1, and +1 relative to the branchpoint adenosine (*i.e.,* substitutions at U257, A258, and C260). Every MDS allele tested in our library altered the splicing of at least one of the *ACT1-CUP1* reporters with substitutions at these positions. As with the A258U reporter, most of the MDS alleles tested showed impaired growth on Cu^{2+} relative to WT for other BS reporters and a corresponding decrease in mRNA by primer extension (blue boxes, **Figure 5.3F** and **Supplemental Figure 5.4A-C**). Splicing of reporters with substitutions immediately 5' of the branchpoint (-1, A258) was strongly affected by MDS alleles, with A258U showing effects with every missense Hsh155 mutant tested. However, not all substitutions at A258 impacted splicing equally: the A258G substitution showed no change between the WT and MDS alleles while the A258C mutation was nearly as impactful as A258U. Many but not all MDS alleles that showed decreased growth relative to WT with the A258U reporter also showed decreased growth with substitutions at the -2 and +1 positions (U257C and C260G, respectively).

The Hsh155^{P369E} mutation corresponding to the frequently observed K700E MDS allele was more disruptive than incorporation of the lysine found in SF3b1 at that position (Hsh155^{P369K}) (**Figure 5.3F** and **5.2D**). However, less frequently observed MDS alleles affect yeast growth more significantly than either P369 mutation (*c.f.*, Hsh155^{P369E} vs. Hsh155^{K335E}). Together these results show that *Hsh155^{MDS}* alleles impact the splicing of introns containing nonconsensus nucleotides at the -2, -1, and +1 BS positions, these alleles are most sensitive to transversions at the -1 position, and the most common result is a decrease in splicing of introns with these nonconsensus BS.

The majority of the SF3b1 mutations tested in our *ACT1-CUP1* assay have been implicated in both CLL and MDS. Despite the fact that many of the same mutations are found in both diseases, the prognostic outcome for an MDS patient differs greatly from a CLL patient, with SF3b1 mutation being favorable in MDS and unfavorable in CLL (15, 42). We sought to further understand this disparity by investigating the mutations G409E and K410N, which have thus far only been linked to CLL. Like mutations associated with both diseases, combination of the CLL-specific mutations with *ACT1-CUP1* reporters bearing nonconsensus BS revealed that Hsh155^{G409E} and Hsh155^{K410N} are only affected by substitutions at the -2, -1 and +1 position of the BS (**Figure 5.2E**). These results suggest that while different mutations in SF3b1 in humans are correlated with distinct cancers, the mechanism of action in yeast for the mutations in the HEAT repeat is likely the same.

While most of the MDS mutants grew less well than Hsh155^{WT} with BS-substituted *ACT1-CUP1* reporters, a few alleles exhibited the opposite effect and showed increased growth on Cu^{2+} relative to Hsh155^{WT}. The strains Hsh155^{E291D}, Hsh155^{R294L}, and Hsh155^{D450G} all showed increased growth in the presence of Cu^{2+} compared to Hsh155^{WT} (**Figure 5.3F**; yellow boxes). While Hsh155^{E291D} only displayed this phenotype with the A258U reporter, both Hsh155^{R294L} and Hsh155^{D450G} showed increased growth with multiple *ACT1-CUP1* reporters and were sensitive to

both the A258U and A258C substitutions. Hsh155^{D450G} displayed the broadest impact on splicing, affecting growth in yeast with reporters containing substitutions at U257 and A258 (**Figure 5.3A, 5.3F**). Strikingly, a single position mutated to different amino acids yielded opposite phenotypes. While Hsh155^{R294L} showed increased growth with the A258U and A258C reporters, Hsh155^{R294C} showed a decrease in growth using these same reporters. Combined with the results described above, these experiments demonstrate that MDS alleles can increase or decrease splicing of an intron containing BS substitutions at the -2, -1, or +1 positions and that different missense mutations of the same amino acid can have opposite effects.

It is possible that mutations in *HSH155* are destabilizing and lead to changes in nonconsensus intron splicing by reducing the concentration of the protein in cells. To test this, we generated strains with three copies of the HA epitope at the C-terminus of Hsh155^{WT} as well as two of the Hsh155^{MDS} mutants showing the strongest phenotypes in our Cu^{2+} growth assay (Hsh155^{K335E} and Hsh155^{D450G}) and assayed protein levels by western blot (**Figure 5.5**). All mutants showed similar levels of Hsh155 relative to both Prp8 and Prp5, suggesting that the mutations do not affect Hsh155 expression. Additionally, we generated merodiploid strains expressing both mutated and wild-type Hsh155 to determine whether the effect of MDS mutants on splicing the U257C and A258U reporters is dominant or recessive. In all cases tested, the effect of expressing Hsh155 with MDS mutations alone is recapitulated in the merodiploid strains, including the small effect of the R294L mutation on splicing the U257C reporter (**Figure 5.3G**). However, the magnitudes of the changes in splicing are less than what is observed when expressing only the mutant copy, which is consistent with the incorporation of both isoforms into functional spliceosomes and indicates that Hsh155^{MDS} mutations are semi-dominant. These data show that Hsh155 plays an active role in BS selection and mutations associated with MDS compromise the ability of Hsh155 to act during splicing.

The Effects of Hsh155^{MDS} Mutations are Additive

We further explored the effect of these mutations by generating additional strains bearing multiple Hsh155^{MDS} mutations and assaying them for altered reporter splicing. For this, we chose the mutations R294L and N295D (which decrease and increase growth in *ACT1-CUP1* reporter assays, respectively) and individually combined them with the mutations H331D, K335E, and D450G to generate six additional strains. When tested using *ACT1-CUP1* reporters with substitutions in the BS at the -2 or -1 position, Hsh155 double mutants displayed additive effects (**Figure 5.3H**). For example, the Hsh155$^{R294L/H331D}$ double mutant strain was less tolerant of Cu^{2+} than Hsh155^{R294L} alone and the Hsh155$^{R294L/D450G}$ double mutant was more tolerant than Hsh155^{R294L} alone. The same additive trend was also observed for the N295D mutation when combined with H331D or D450G. The Hsh155$^{N295D/K335E}$ double mutant strain was the only variant to deviate from this trend, but this may be the result of being unable to further reduce splicing and Cu^{2+} tolerance in a strain already severely impaired by the K335E mutation. Interestingly, these double mutant strains still showed no changes in splicing consensus intron reporters, further supporting the notion that MDS mutations give rise to change by altering the splicing of specific nonconsensus introns rather than by causing a general pre-mRNA splicing defect.

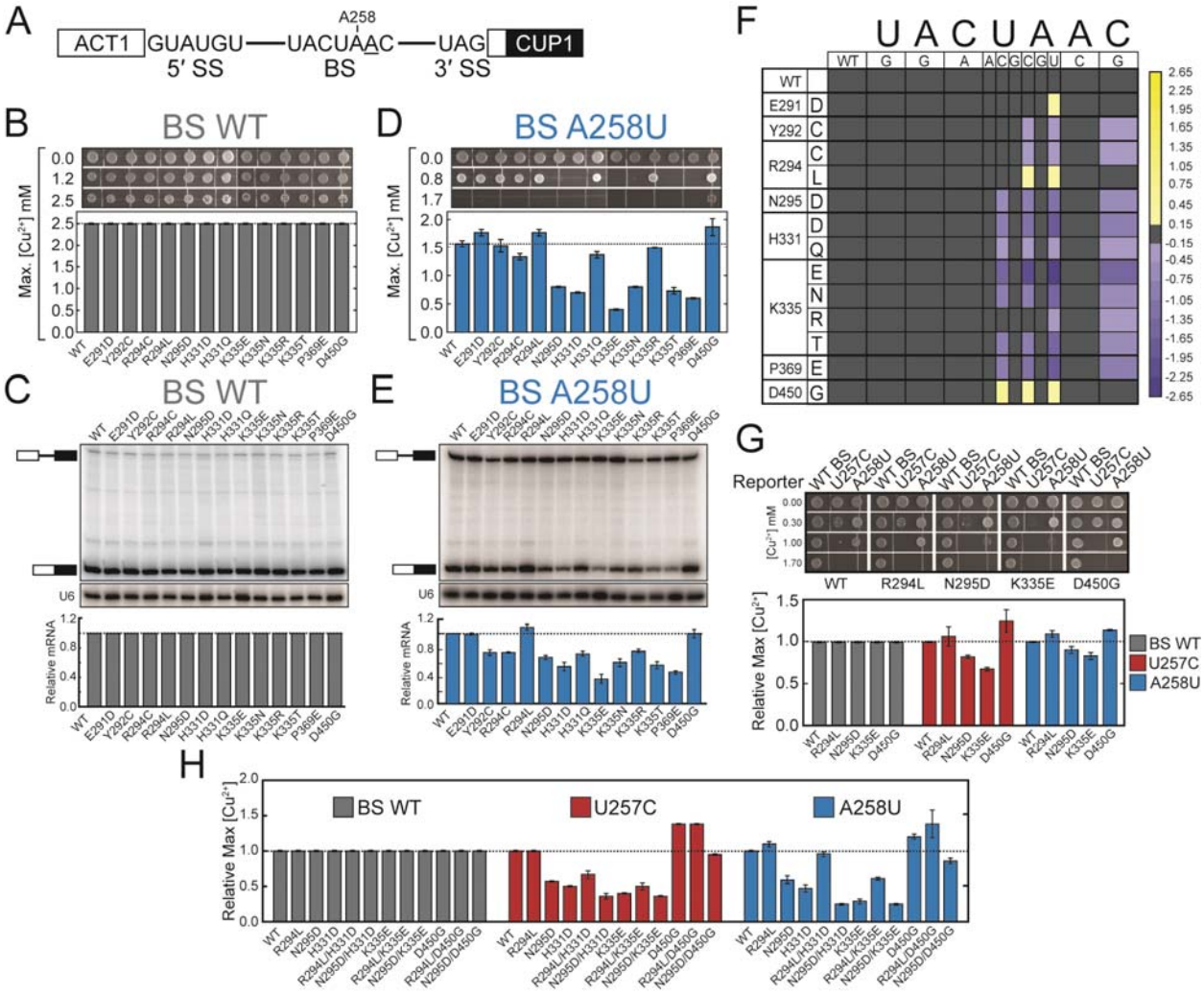


Figure 5.3. MDS mutations alter the splicing of introns with nonconsensus BS sequences. (**A**) Schematic representation of the *ACT1-CUP1* reporter pre-mRNA. The consensus sequences of the yeast 5' SS, BS, and 3' SS are shown. The position of A258 is noted and the branchpoint adenosine is underlined. (**B**) Cu^{2+} growth assay of strains carrying an *ACT1-CUP1* reporter plasmid with a consensus intron. Representative images are shown at the top and the maximum [Cu^{2+}] at which growth was observed is plotted below. (**C**) Determination of *ACT1-CUP1* reporter RNA levels by primer extension from isolated total yeast RNA. (Top) Positions of the pre-mRNA and mRNA are noted in the primer extension polyacrylamide gel. (Middle) Primer extension analysis of the U6 snRNA was used as an internal control and analyzed on the same gel as shown in the top panel. (Bottom) Quantification of the amount of *ACT1-CUP1* mRNA after normalization

to U6 for each strain. U6 bands are taken from the same gel and contrast has been adjusted. (**D**) Cu^{2+} growth assay of strains carrying an *ACT1-CUP1* reporter plasmid with a A258U nonconsensus BS. (**E**) Determination of A258U *ACT1-CUP1* reporter RNA levels by primer extension from isolated total yeast RNA. (**F**) Heatmap summarizing mutant *ACT1-CUP1* reporter data for all BS reporters tested. Plotted data represent the $\log_2$ transform of the ratio of the maximum [Cu^{2+}] at which growth was observed for the indicated $Hsh155^{MDS}$ mutant to the maximum [Cu^{2+}] at which growth was observed for $Hsh155^{WT}$. Purple colors indicate decreased growth relative to $Hsh155^{WT}$, and yellow colors indicate improved growth. (**G**) Cu^{2+} growth assay of merodiploid strains expressing the indicated $HSH155^{MDS}$ allele from a plasmid in addition to the chromosomal copy of $Hsh155^{WT}$ for the WT, U257C, and A258U *ACT1-CUP1* splicing reporters. (**H**) Cu^{2+} growth assay of strains expressing Hsh155 proteins harboring multiple MDS mutations for the WT, U257C, and A258U *ACT1-CUP1* splicing reporters. In panels B, D-H, each bar represents the average of three independent experiments, and error bars represent the standard deviation.

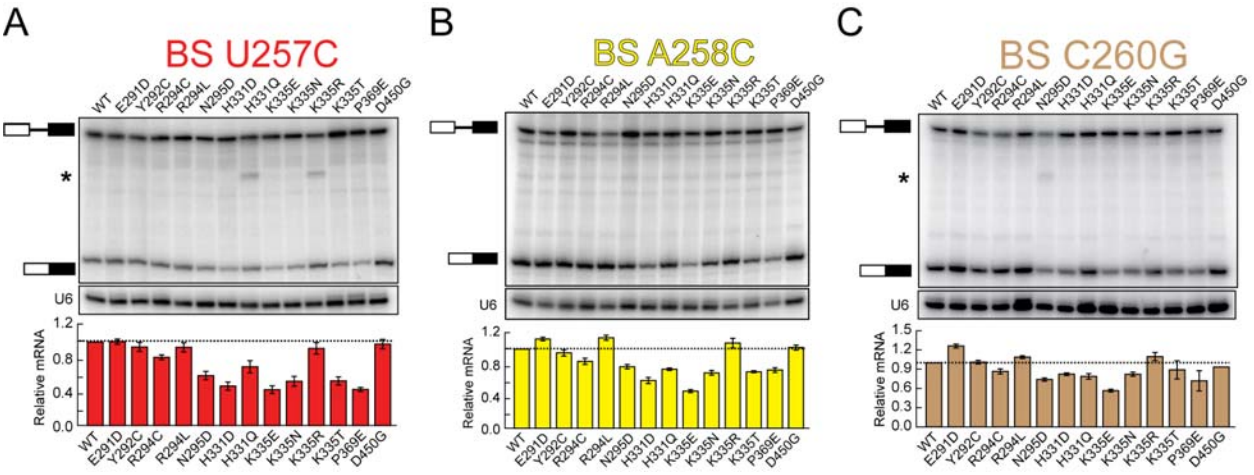


Figure 5.4. Primer extension analysis of ACT1-CUP1 spliced products from additional reporter pre-mRNAs harboring nonconsensus BS substitutions.

(**A**) Determination of U257C BS substitution ACT1-CUP1 reporter RNA levels by primer extension from isolated total yeast RNA. (Top) Positions of the pre-mRNA and mRNA are noted in the primer extension polyacrylamide gel. (Middle) Primer extension analysis of the U6 snRNA was used as an internal control and analyzed on the same gel as shown in the top panel. (Bottom) Quantification of the amount of ACT1-CUP1 mRNA relative to the total amount of mRNA and pre-mRNA after normalization to U6 for each strain. (**B**) Determination of A258C ACT1-CUP1 reporter RNA levels by primer extension from isolated total yeast RNA. (**C**) Determination of C260G ACT1-CUP1 reporter RNA levels by primer extension from isolated total yeast RNA. In panels A-C, bars represent the average of three independent experiments, and error bars represent the standard deviation. Asterisks (*) indicate unknown bands that were also irreproducible.

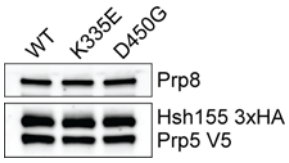


Figure 5.5. MDS mutations in Hsh155 that significantly impact BS usage do not affect Hsh155 protein expression

SDS-PAGE and western blotting analysis of protein expression of Prp8, Prp5, and the noted Hsh155 alleles in yeast. Prp8 was detected using a primary antibody to the protein while Prp5 and Hsh155 proteins were detected by genetically encoding V5 and 3xHA epitope tags, respectively. No changes in relative expression of Prp8, Prp5, and the Hsh155 mutants were apparent.

Hsh155^{MDS} Mutations Do Not Alter Splicing of Nonconsensus 5' or 3' Splice Sites and

To investigate if the impact of MDS alleles is limited to BS substitutions, we tested eight additional *ACT1-CUP1* reporters with single nucleotide substitutions in the consensus 5' splice site (5' SS) or 3' splice site (3' SS). In all cases, yeast strains with MDS alleles grew to levels equivalent to Hsh155^{WT} in the presence of Cu^{2+} (**Figure 5.6A, B**), supporting the notion that splicing of reporters with mutations at these sites are not affected by MDS alleles. This is consistent with SF3b1/Hsh155 primarily functioning near the BS and nearby, downstream sequences.

Hsh155^{MDS} Mutations Do Not Affect Cryptic 3' SS Discrimination

To evaluate directly whether Hsh155^{MDS} mutants are intrinsically impaired at discriminating against cryptic 3' SS, we employed an *ACT1-CUP1* reporter mutated to include a second consensus 3' SS 10 nucleotides (nt) downstream of the branchpoint adenosine and 34 nt upstream of the canonical 3' SS (**Figure 5.6C**) (34, 43). We tested for use of the proximal and distal 3' SS in Hsh155^{WT}, Hsh155^{R294L}, Hsh155^{K335E}, and Hsh155^{D450G} mutant strains by primer extension (**Figure 5.6D, left panel**). We observed very little change in 3' SS discrimination. Further testing of these strains with a reporter bearing both the A258U BS substitution and a cryptic 3' SS also showed similar ratios of 3' SS usage between *HSH155* alleles (**Figure 5.6D, right panel**). Together, our *ACT1-CUP1* reporter data support the idea that MDS alleles likely do not affect 5' or 3' SS usage or discrimination between cryptic and bona fide 3' SS. Changes in cryptic 3' SS usage observed in humans with MDS may instead arise from a defect in the ability of the spliceosome to utilize weak BS, leading to alternative positioning of U2 on the intron and selection of a different BS.

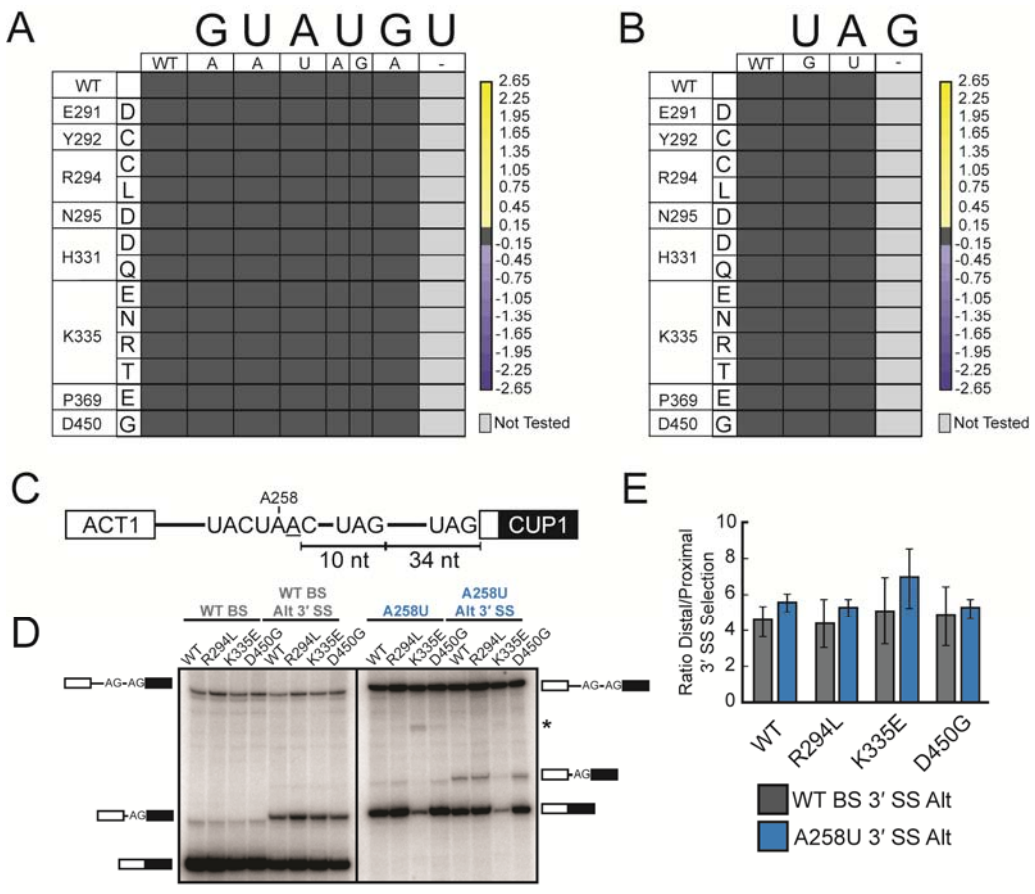


Figure 5.6. MDS mutations do not affect the splicing of introns containing nonconsensus 5' SS and 3' SS or 3' SS selection.

(**A**) Heatmap summarizing mutant *ACT1-CUP1* reporter data for all 5' SS substitution reporters tested. Data were normalized and the heatmap generated as in Figure 2F. No changes in 5' SS usage were observed. (**B**) Heatmap summarizing mutant *ACT1-CUP1* reporter data for all 3' SS substitution reporters tested. Data were normalized and the heatmap generated as in Figure 2F. No changes in 3' SS usage were observed. (**C**) Schematic representation of the *ACT1-CUP1* reporters used to evaluate cryptic 3' SS selection. The cryptic 3' SS is located 10 nt downstream of the branchpoint adenosine and 34 nt upstream of the canonical 3' SS. Reporters containing both a consensus BS and an A258U substitution were used. (**D**) Primer extension and PAGE analysis of spliced products of the *ACT1-CUP1* reporters shown in (C) from total RNA isolated from the given yeast strains. Positions of the pre-mRNA and mRNA products are noted. The

reporter containing the A258U nonconsensus BS also contains a larger 3' exon leading to shift in electrophoretic mobility between the consensus and nonconsensus reporter RNAs. The asterisk (*) indicates an unknown band that was not reproducible. (**E**) Quantification of the data shown in (D) for 3' SS usage by the $Hsh155^{WT}$ and given $Hsh155^{MDS}$ strains. Bars represent the average of three independent experiments, and error bars represent the standard deviation.

Hsh155 Interacts with Multiple Components of the Splicing Machinery

Our experiments using the *ACT1-CUP1* reporter reveal that SF3b1 mutations alter usage of nonconsensus BS. Recent structures have implicated the mutated HEAT repeats in direct binding of RNA downstream of the U2/BS duplex (24, 25). It is possible that mutation of these HEAT repeats either directly or indirectly distort the conformation of Hsh155/SF3b1 thereby altering contacts with other components of the spliceosome and leading to the observed pre-mRNA splicing changes. To test this idea, we used a yeast two-hybrid assay to screen for altered interactions upon mutation of Hsh155. A number of proteins that interact with Hsh155 have previously been identified by Y2H (44), and we assayed these identified interactions in combination with MDS mutations (**Representative Images in Supplemental Figure 5.7; Figure 5.8A**). Since SF3b1 has recently been implicated in influencing steps after pre-spliceosome formation (35), we also included a number of other factors that interact with the spliceosome during splicing.

Hsh155 was fused to the GAL4 activation domain (AD) while each potential interacting protein was fused to the GAL4 DNA binding domain (BD). We confirmed expression of each AD-Hsh155 mutant by western blotting, and all mutants expressed equally well in the Y2H strain (**Figure 5.8B**). Similarly, we confirmed expression of potential interaction partners and only the fusions that were shown to express by western blotting were included in the assay. We screened 15 alleles of Hsh155 against 15 components of the splicing machinery, for a total of 225 potential interactions.

The Y2H screen using an AD-Hsh155^{WT} fusion confirmed previously known interactions with Bud13, Clf1, Cus2, Mud2, and Prp5 as well as identified new potential binding partners. Novel Y2H interactions were detected between Hsh155 and the SF3b components Cus1 and Ysf3. We did not observe any Y2H interaction between AD-Hsh155^{WT} and either the SF3a protein Prp11 or

SF3b protein Hsh49. These results suggest that the AD-Hsh155 Y2H assay is reporting on a subset of protein-protein interactions occurring within U2 or the spliceosome.

The Y2H screen also identified previously unknown interactions between Hsh155 and Prp2, Prp43, and Slu7. Prp2 and Prp43 are both spliceosomal DEAH-box ATPases (32), while Slu7 is a second step factor important for selection of 3' SS (43). Our observed interaction between Hsh155 and Prp2 agrees with the role of Prp2 in activation and remodeling of the U2/U6 active site (which involves destabilization of SF3) as well as recent cryo-electron microscopy (cryo-EM) structures of spliceosomes (32-34, 45). To our knowledge, a Y2H interaction between Hsh155 and Prp43 has not previously been reported. Prp43 has multiple roles in the splicing cycle and is responsible for disassembly of lariat-intron product complexes as well as spliceosomes rejected by proofreading mechanisms (46, 47). Prp43 may interact with Hsh155 to gain access to the U2/U6 active site during disassembly (48). We observed no interaction between AD-Hsh155^{WT} and the DEAD-box ATPase Prp28 or DEAH-box ATPase Prp22. This is consistent with Prp28 and Prp22 acting on the spliceosome at regions other than the BS: Prp28 isomerizes interactions between the 5' SS and U1 and U6, while Prp22 promotes mRNA release and crosslinks to the 3' exon (32). Together these results suggest that SF3b1 may interact with a subset of spliceosomal ATPases that need to function at or near the U2/BS pairing region.

Interactions with Hsh155 Remain Intact upon Inclusion of SF3b1 Disease Alleles with the Exception of Prp5

HSH155^{MDS} alleles altered a small subset of the Y2H interactions while leaving most others unaffected (**Figure 5.8A).** None of the MDS mutations changed interactions between Hsh155 and Bud13, Cus2, or Clf1. The R294C and R294L mutations disrupted interactions among the greatest number of splicing factors, including components of U2 snRNP (Cus1, Ysf3), factors involved in early spliceosome assembly (Mud2 and Prp5) and factors involved in spliceosome activation, catalysis, or disassembly (Prp2, Slu7, and Prp43, respectively) (**Figure**

5.8A). The disruptions caused by missense mutations of R294 could be due to changes in Hsh155 structure that influence several binding sites or interactions, a result possibly amplified in the context of the Y2H assay. In support of this idea, transformation and subsequent 5-FOA selection of the *HSH155* shuffle strain with the AD-Hsh155^{R294L} plasmid resulted in viable yeast, showing that AD-Hsh155^{R294L} is active for splicing notwithstanding these altered Y2H interactions (**Figure 5.9**). Surprisingly, both R294L and R294C disrupted identical sets of interactions despite these alleles showing opposite phenotypes in our *ACT1-CUP1* reporter assay (**Figure 5.3F**). This suggests that while R294L and R294C disturb binding of many of the same splicing factors, the mutations likely alter Hsh155 structure in unique ways.

Aside from the R294C and R294L mutations, interactions between the other *HSH155^{MDS}* alleles and the 3' SS selection factor Slu7 remained intact (**Figure 5.8A**). This indicates that while a molecular signature of MDS in humans is selection of cryptic 3' SS, disruption of the interaction between Hsh155 and Slu7 is not likely to be a major driver of the process in yeast. Supporting this conclusion is our observation that 3' SS choice in the *ACT1-CUP1* assay is unaffected even by the Hsh155^{R294L} mutation (**Figure 5.6C, D**).

The majority of *HSH155* mutant alleles (10 of 14) altered Y2H interactions to Prp5, implying that many MDS mutations either directly or indirectly influence interactions between these two proteins during spliceosome assembly. Interestingly, previous work has shown that Prp5 mutations also alter BS fidelity at the same positions flanking the branchpoint adenosine that we observe to be altered by the MDS alleles (**Figure 5.3F**) (49). Taken together, the Y2H data support the notion that most protein-protein interactions between Hsh155 and other splicing factors are unaffected by Hsh155^{MDS}. The major exception is the interaction between Hsh155 and Prp5.

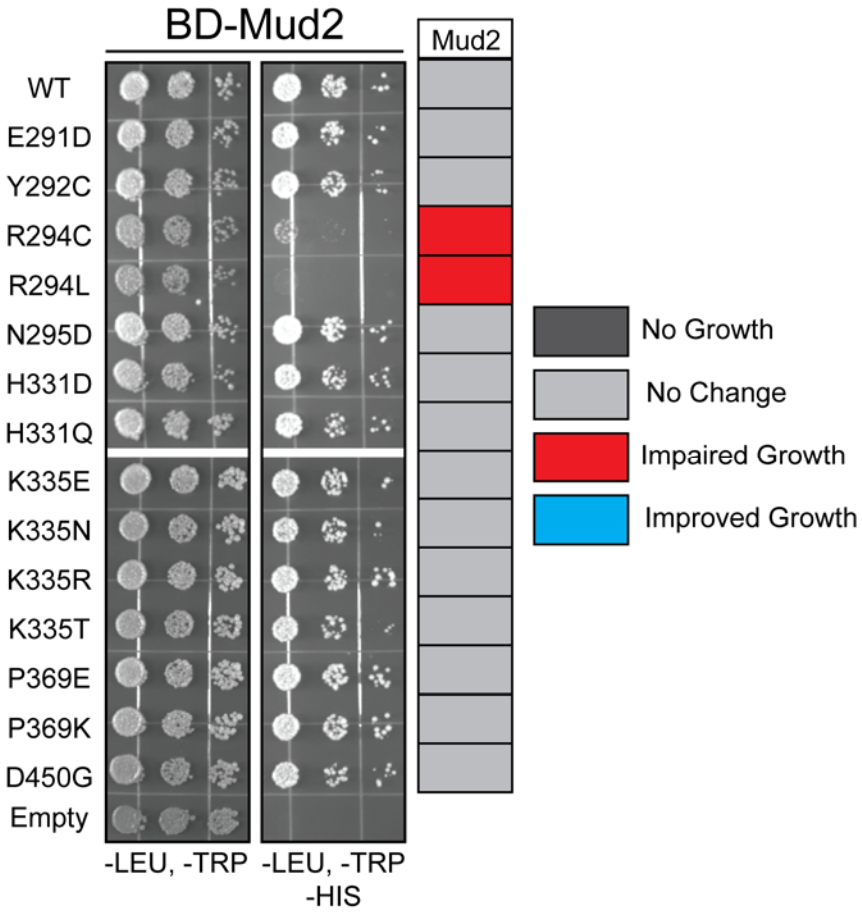


Figure 5.7. Representative Y2H Images.

Yeast expressing AD-Hsh155 mutants and BD-Mud2 were plated on selective media to investigate interaction between the components. Results are depicted as a pseudo-heatmap where red shows impaired interactions.

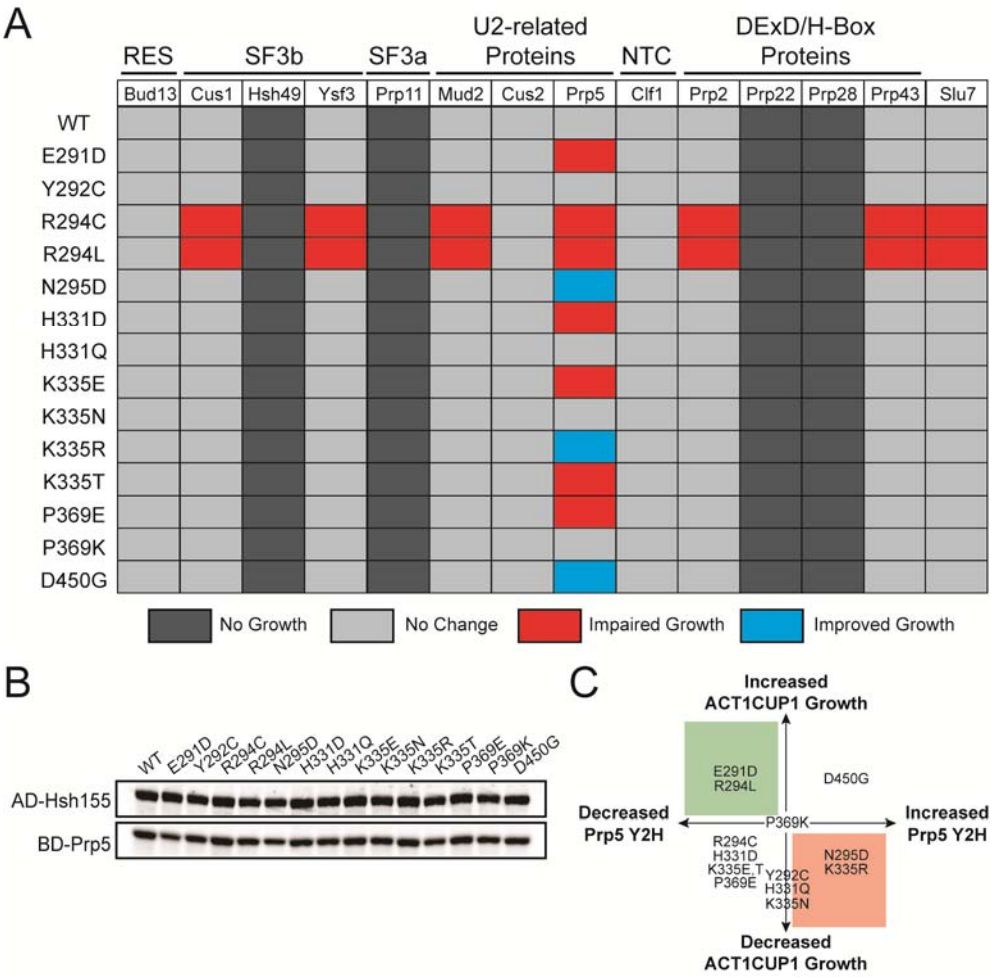


Figure 5.8. MDS mutations perturb interactions between Hsh155 and Prp5 but leave most other interactions intact.

(**A**) Pseudo-heatmap showing the observed Y2H interactions of Hsh155 upon incorporation of MDS mutations with the given splicing factors. Red indicates an impaired growth relative to $Hsh155^{WT}$ when plated on media that is selective for the Y2H interaction (-His dropout media), blue indicates improved growth, and light grey indicates no change. Dark grey indicates no observable Y2H interaction. (**B**) Representative western blot confirming expression of the fusion proteins to $HSH155^{MDS}$ used in the Y2H assay. Expression of each potential interacting partner was also confirmed by western blotting and expression of Prp5 is also shown as a representative example. (**C**) Graphical representation of the relationship between changes in yeast growth

observed with the BS A258U *ACT1-CUP1* splicing reporter (Figure 2D) and altered interactions observed by Y2H (Figure 4A). Shaded areas represent predictions made from a previously described model for Prp5-based BS fidelity in which retention of Prp5 leads to increased fidelity (red) and weakening of the Prp5 interaction leads to relaxed fidelity (green) (37).

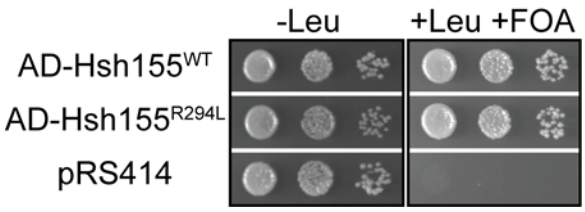


Figure 5.9. The AD-Hsh155^{R294L} is viable as the sole copy of Hsh155 in yeast.

Yeast expressing AD-Hsh155^{WT}, AD-Hsh155^{R294L,} and pRS414 were plated on FOA-containing media to demonstrate that the R294L mutation is functional.

Hsh155^{MDS} Mutations Affect Splicing in a Manner Distinct from Prp5 Proofreading

Since Prp5 plays essential roles during pre-spliceosome assembly and U2/BS pairing, we tested whether the altered interactions between Hsh155 and Prp5 were the cause of the observed changes in BS usage. We noticed that results from the Y2H screen with Prp5 do not directly correlate with those from the *ACT1-CUP1* assay (**Figure 5.8C**). For example, *HSH155^{MDS}* alleles that decrease growth in the *ACT1-CUP1* assay with the nonconsensus A258U BS reporter show a variety of effects in Hsh155/Prp5 in the Y2H assay. This suggests that changes in BS usage arise from more complicated mechanisms than simply disrupting or strengthening the interactions between Hsh155 and Prp5. To resolve this, we investigated the known roles of Prp5 to look for impacts on these functions by MDS mutations.

In an ATP-dependent role, Prp5 has been proposed to displace Cus2 from the U2 snRNP to permit U2 association with the pre-mRNA (**Figure 5.10A**) (50, 51). We hypothesized that mutation of Hsh155 could be impacting Cus2 displacement by Prp5, and that this could lead to defects in spliceosome assembly and the observed changes in *ACT1-CUP1* splicing. To investigate this, we generated MDS strains with *CUS2* deleted and assayed them in an *ACT1-CUP1* reporter assay using the A258U BS mutant. All *CUS2Δ* strains grew equally well as strains with intact *CUS2* (**Figure 5.10B**). This shows that the observed changes in pre-mRNA splicing do not result from changes in Cus2 displacement by Prp5 during pre-spliceosome formation.

In addition to Cus2 displacement, Prp5 has also been implicated in proofreading at the BS during spliceosome assembly although the mechanism remains unclear (49, 52). A number of mutations in Prp5 change the splicing of introns with BS substitutions and previous work has suggested that these may function in part by altering interactions between Prp5 and other splicing factors or by modulating Prp5 transitions between open and closed conformations (**Figure 5.10C**) (37, 53). For example, alanine mutation of the N-terminal DPLD motif of Prp5 (AAAA) disrupts the interaction with U2/SF3b and causes greatly improved splicing of nonconsensus reporter substrates *in vivo*

(54). The Prp5 mutation E235A disrupts the open conformation of the protein and diminishes splicing of nonconsensus reporters, while mutation of the Prp5 DEAD-box SAT motif to TAG may disrupt the closed conformation and improve splicing of nonconsensus reporters (**Figure 5.10C**) (53). The Prp5 mutation N399D also increases splicing of nonconsensus reporters; however, its mechanism is unclear (49). It has recently been proposed that all of these Prp5 mutations ultimately impact splicing by influencing how Prp5 is retained on the pre-spliceosome (37). In this model, Prp5 ensures BS fidelity by recognizing mispairing between the U2 snRNA and the intron BS and preventing tri-snRNP recruitment in the presence of a mismatch. Prp5 mutants with higher affinity for the pre-spliceosome (*e.g.* $Prp5^{E235A}$) impair nonconsensus BS usage by retention of Prp5 in the pre-spliceosome and preventing tri-snRNP addition. Opposing mutants (*e.g.* $Prp5^{AAAA}$, $Prp5^{TAG}$, and $Prp5^{N399D}$) promote Prp5 release and progression of spliceosome assembly.

We next investigated the outcome of combining Prp5 mutations with MDS alleles during splicing. For this, we employed the MDS alleles $HSH155^{K335E}$ and $HSH155^{D450G}$ because these alleles show opposing effects in BS usage and interaction with Prp5 (**Figures 5.3C and 5.3F; Figure 5.8A**). We generated strains expressing each combination of Prp5 and Hsh155 mutations and tested them in *ACT1-CUP1* reporter assays using a consensus intron (**Figure 5.10D**). No differences were observed for any combination of Hsh155/Prp5 mutations for the WT *ACT1-CUP1* reporter. When Hsh155/Prp5 mutant strains were tested in combination with the U257C and A258U BS substitution reporters, we observed that the Prp5 mutations AAAA, N399D, and TAG improved growth on Cu^{2+} while E235A diminished growth regardless of the Hsh155 background (**Figure 5.10E and 5.10F**). However, the MDS alleles of *HSH155* still affected growth, as strains with $Hsh155^{K335E}$ showed generally diminished growth relative to $Hsh155^{WT}$ and $Hsh155^{D450G}$ strains showed improved growth irrespective of the *PRP5* allele. This suggests that the mechanism of action of the Hsh155 mutations is independent from the mechanism of Prp5

mutation: in our assays, Hsh155 establishes a baseline level of BS usage that Prp5 mutations either can raise or lower.

To further evaluate the effects of Prp5 mutations on interactions between Prp5 and Hsh155, we expanded our Y2H assay to include the Prp5^{AAAA}, Prp5^{E235A}, Prp5^{N399D}, and Prp5^{TAG} mutants. We confirmed expression of all BD-Prp5 variants by western blot (**Figure 5.11**). BD-Prp5^{AAAA} shows a complete loss of interaction with all AD-Hsh155 variants by Y2H (**Figure 5.11**). This result is consistent with earlier reports that showed that this region of Prp5 is important for the interaction of Prp5 with the SF3b complex (54). The BD-Prp5^{TAG} mutant also decreased the interaction with Hsh155. These data support the model that the Prp5^{AAAA} and Prp5^{TAG} mutations improve nonconsensus BS usage by weakening the interaction between Prp5 and other splicing factors. Interestingly, BD-Prp5^{E235A} and BD-Prp5^{N399D} mutants showed only minor changes in growth relative to BD-Prp5^{WT} in Y2H assays despite the strong influence these mutations have on BS usage in *ACT1-CUP1* reporter assays. The Prp5^{E235A} mutant modestly improved growth relative to Prp5^{WT} for a number of Hsh155 mutations (*e.g.* WT, H331D, K335E, etc.) while Prp5^{N399D} showed slightly impaired growth (*e.g.* WT, H331D, K335N, etc.). The directions of these changes are consistent with Prp5^{E235A} and Prp5^{N399D} interacting with the pre-spliceosome with different affinities to impact BS usage (37), and our data support Prp5^{E235A} having higher affinity than Prp5^{N399D}.

Importantly, the growth pattern of the *HSH155^{MDS}* alleles relative to one another was maintained independent of the Prp5 mutation. For example, AD-Hsh155^{N295D} grew better than AD-Hsh155^{WT} and AD-Hsh155^{H331D} grew worse than AD-Hsh155^{WT} in all instances. While mutation of BD-Prp5 changed the Y2H interaction with AD-Hsh155^{WT} and all Hsh155 alleles equivalently, the AD-Hsh155^{MDS} variants showed distinct changes in Y2H interactions with BD-Prp5. Our results from the *ACT1-CUP1* splicing reporter and Y2H assays argue that MDS alleles influence BS usage at a step distinct from that influenced by Prp5 mutations.

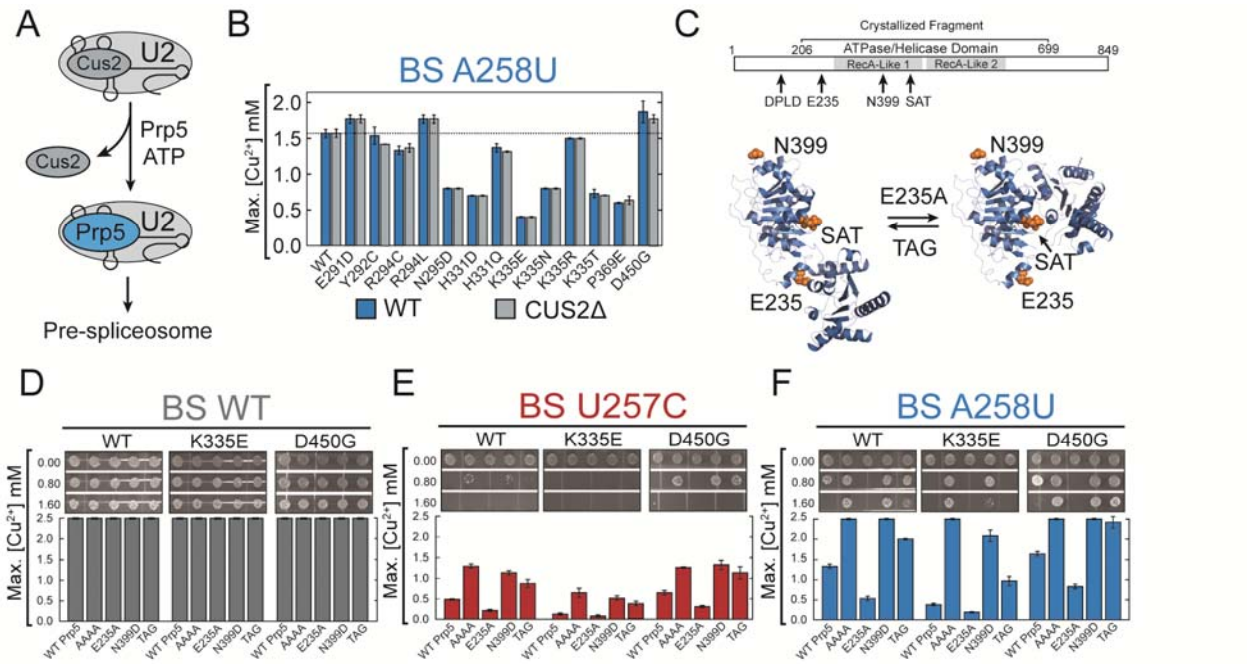


Figure 5.10. Hsh155 MDS mutants affect BS fidelity at a different step than Prp5 proofreading

(**A**) Cartoon depicting the proposed ATP-dependent function of Prp5 in displacement of Cus2 from the U2 snRNP during spliceosome assembly. (**B**) Comparison of Cu^{2+} growth assays using the nonconsensus A258U BS substitution *ACT1-CUP1* reporter pre-mRNA between strains containing and lacking the Cus2 protein. No significant differences were observed between the two strains. (**C**) (Top) Schematic of Prp5 structure indicating the positions of Prp5 mutations used in this assay relative to the two DEAD-box RecA-like domains and the fragment of Prp5 whose structure has been determined by X-ray crystallography. (Bottom) Prp5 has been proposed to undergo a conformational change to promote splicing. The open structure (left) represents the structure determined by X-ray crystallography (pdb 4LJY) while the closed structure (right) is believed to be necessary for ATP hydrolysis and was modeled based on structures of other DEAD-box proteins (coordinates for the closed structure were obtained from Yong-Zhen Xu and Charles Query) (53). Positions of the Prp5 mutations used in this study are noted. The E235A mutation is believed to favor the closed conformation while the TAG mutation of the SAT-motif is

believed to favor the open conformation. It is unclear if the N399D mutation used here would impact conformational switching. (**D**) Cu^{2+} growth assay for strains containing the Hsh155 WT, K335E, or D450G alleles in combination with the given Prp5 mutations (see text for additional explanation of each Prp5 mutation) with the *ACT1-CUP1* reporter containing a consensus BS. (**E**) Cu^{2+} growth assay for combinations of Hsh155 and Prp5 as in part (D) except the U257C nonconsensus BS *ACT1-CUP1* reporter was used. (**F**) Cu^{2+} growth assay for combinations of Hsh155 and Prp5 as in part (D) except the A258U nonconsensus BS *ACT1-CUP1* reporter was used. In panels B and D-F, bars represent the average of three independent experiments, and error bars represent the standard deviation.

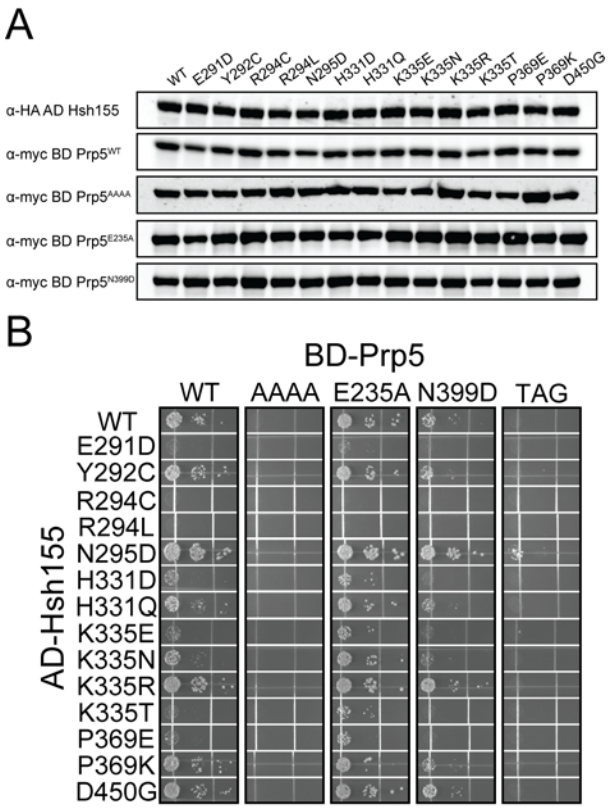


Figure 5.11. Prp5 mutations affect Y2H interactions with Hsh155, but MDS-allele specific changes in the Prp5/Hsh155 Y2H interaction are maintained.

(**A**) Confirmation by western blotting of Hsh155-allele and Prp5-allele expression in yeast strains used for Y2H analysis. (**B**) Representative images depicting yeast growth in a Y2H assay for strains containing combinations of Hsh155 and Prp5 mutations.

MDS Mutations Show Genetic Interactions with a Prp2 ATPase Mutant

To investigate whether MDS mutations can impact splicing at steps subsequent to assembly, we looked for genetic interactions with Prp2. Prp2 is responsible for destabilizing the SF3b complex from the U2/BS duplex to allow further steps in splicing to occur (**Figure 5.12A**), likely resulting in release of the U2 snRNA/BS duplex so that it may enter the spliceosome active site (45). We generated strains with the Prp2 ATPase mutation Q548N in backgrounds with two different Hsh155^{MDS} mutations (K335E and D450G). Prp2^{Q548N} confers cold sensitivity (*cs*) to yeast, likely due to poor ATPase and/or helicase activity (34)**.** Like Prp5, Prp2 is a fidelity factor and its function is correlated with proper basepairing between U2 and U6 snRNAs. Prp2^{Q548N} also suppresses lethal mutations in U2 that perturb U2/U6 helix Ia basepairing, suggesting that efficient ATPase activity of Prp2 prevents catalytic activation of spliceosomes with improperly formed active sites (34).

Strains expressing Prp2^{Q548N} grew better at 16°C in the presence of Hsh155^{K335E} than with Hsh155^{WT}, while Prp2^{Q548N} in the presence of Hsh155^{D450G} grew slightly worse. (**Figure 5.12B**). These data indicate that MDS mutations such as Hsh155^{K335E} that impair the splicing of reporter pre-mRNAs containing BS substitutions also partially suppress Prp2^{Q548N} cold sensitivity. This suggests that Hsh155^{K335E}-containing SF3 complexes may be partially destabilized and thus aid Prp2^{Q548N} activity, possibly by altering the structure of the SF3 complex during a Prp2-dependent step in spliceosome activation. Mutations that improve splicing (Hsh155^{D450G}) may do the opposite. Significantly, since BS sequences sensitive to Hsh155^{K335E} and Hsh155^{D450G} mutations (**Figure 5.3**) are extremely rare in yeast introns (if they are present at all) (55-57), it is likely that the difference in cold sensitivity for Prp2^{Q548N} strains is due to changes in the splicing of introns containing consensus BS (or nonconcensus BS other than those affected by Hsh155^{MDS} in our *ACT1-CUP1* assay). MDS mutations may alter the stability of yeast U2 proteins at the BS in general and this altered stability of the B^{act} spliceosome can in turn modulate the requirement for

Prp2. *HSH155*^{MDS} alleles might function at multiple steps in splicing and in response to different sequence elements within the BS during each step.

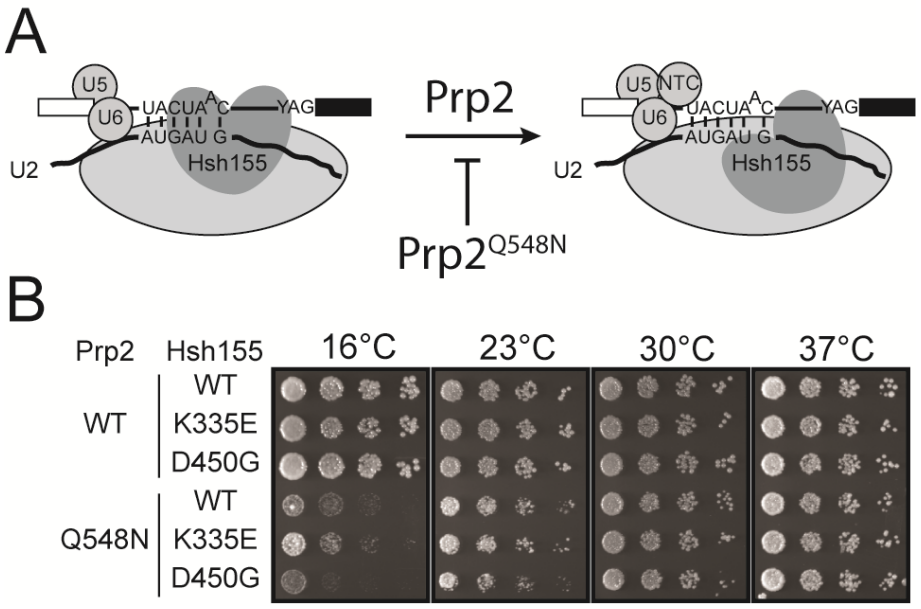


Figure 5.12. MDS mutations interact genetically with Prp2 mutations.

(**A**) Cartoon schematic of Prp2-dependent activation of the spliceosome. Prp2 is believed to destabilize Hsh155 as well as the rest of the SF3b complex from interacting with the BS. The *PRP2*^{Q548N} allele stalls this process at low temperatures (34). (**B**) Representative temperature sensitivity growth assays of the given Hsh155 variants in combination with Prp2^{WT} or Prp2^{Q548N} when plated on YPD at the given temperatures. Hsh155^{K335E} partially suppresses Prp2^{Q548N} and Hsh155^{D450G} enhances cold sensitivity.

Discussion

Mounting evidence has implicated mutations in the splicing machinery as potent drivers of human disease (7). Among splicing factors that have been linked to disease, the essential and conserved U2 component SF3b1 has been found to be frequently mutated (24, 41). We sought to understand the role of SF3b1 during splicing and the impact that MDS mutations can have on the function of the protein. We find that mutations associated with MDS and CLL alter the usage of certain nonconsensus BS without affecting the splicing of introns with consensus BS. Different mutations result in disparate changes in BS usage, and these mutations by themselves do not change 3' SS selection. Furthermore, these mutations appear to modify BS usage by a novel mechanism, potentially by disrupting Hsh155/SF3b1 conformations that stabilize weak U2:BS RNA duplexes formed on nonconsensus introns (**Figure 5.13B**). We show that the interaction network of Hsh155 is largely unperturbed by MDS mutations with the exception of Prp5; however, Prp5-dependent proofreading is not driving changes in BS usage. Finally, we show that mutations in Hsh155 have a genetic interaction with *cs* Prp2, suggesting a role for the protein in stabilizing the SF3b complex at the U2/BS duplex until activation. Together, these data suggest that SF3b1 mutations may cause disease through disruption of a BS recognition step conserved between yeast and humans that impacts how nonconsensus splice sites are utilized by the spliceosome.

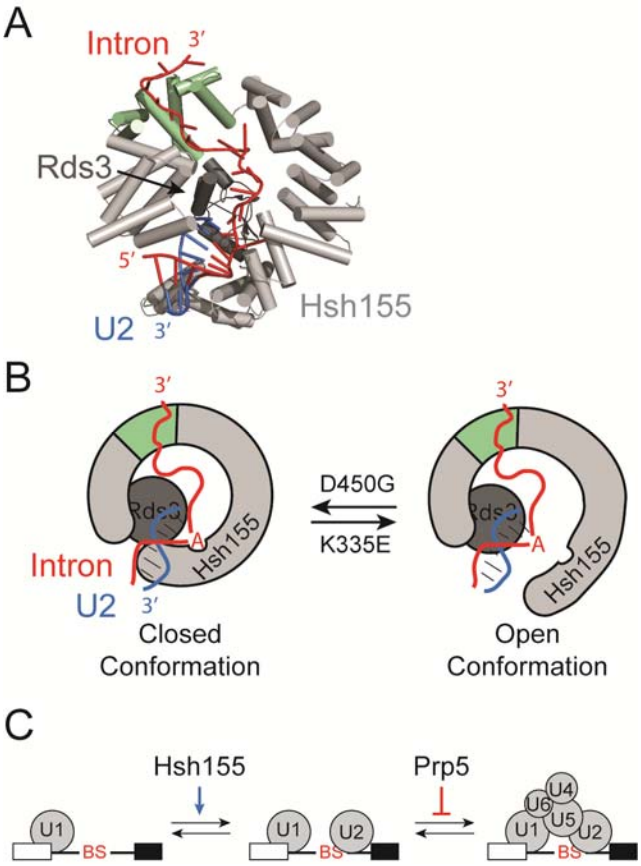


Figure 5.13. Models for SF3b1/Hsh155 function in BS duplex stabilization during splicing. (**A**) Cartoon representation of Hsh155 (light grey and green) and Rds3 (dark grey) bound to the U2 snRNA/BS duplex from the cryo-EM structure of the yeast B^{act} spliceosome (25). The region of Hsh155 containing the MDS mutations studied here is shown in green. (**B**) Model for SF3b1/Hsh155 action at the BS. In addition to the structure shown in (A), Hsh155 must also exist in a conformation that permits release of the U2 snRNA/BS RNA duplex and splicing. MDS mutations impact Hsh155 conformation and lead to changes that affect recognition and stabilization of the BS duplex. Mutations that increase splicing of nonconsensus BS (*e.g.,* $Hsh155^{D450G}$) could stabilize the "closed" or BS duplex bound form whereas mutations that inhibit splicing (*e.g.,* $Hsh155^{K335E}$) could favor an open form that is necessary for splicing catalysis but does not help stabilize a mismatched duplex during spliceosome assembly. (**C**) Model for opposing activities of SF3b1/Hsh155 and Prp5 during splicing. SF3b1 functions to stabilize U2

snRNA/BS duplex formation, particularly at nonconsensus or weak BS. Prp5's proofreading ability opposes this function to enforce BS fidelity by blocking tri-snRNP association. The relative activities of SF3b1/Hsh155 and Prp5 at particular BS may be used to promote or inhibit spliceosome assembly.

SF3b1/Hsh155 Interacts with Numerous Splicing Factors During Splicing

In both yeast and humans, Hsh155/SF3b1 directly contacts the pre-mRNA substrate in the region near the BS and is probably present throughout splicing (8, 23). This places Hsh155/SF3b1 in a position to influence how BS are selected during assembly as well as later steps in catalysis. One mechanism by which Hsh155/SF3b1 could influence splicing is by regulating the recruitment and retention of spliceosome proteins. Consistent with this role is prior Y2H data showing interactions between Hsh155 and numerous splicing factors, and our data identifying novel Y2H interactions between Hsh155 and Prp2, Prp43, and Slu7 (**Figure 5.8A**) (44). Our Y2H results confirm recently observed crosslinks between Hsh155 and a Prp43 variant in $B^{act\Delta prp2}$ spliceosomes (48). Our Y2H data additionally suggest that destabilization of the SF3 complex by Prp2 may occur in part through direct contact between Prp2 and Hsh155, in agreement with recent cryo-EM structures (24, 25). We speculate that changes in Slu7 function due to altered interaction with Hsh155/SF3b1 may in turn explain how small molecules that bind SF3b1 also impact exon ligation in human spliceosomes, since Slu7 has previously been implicated in 3' SS selection in both yeast and humans (43, 58, 59). Thus, by modulating interactions between the spliceosome and transiently associated splicing factors, Hsh155/SF3b1 could potentially regulate spliceosome assembly via Prp5, spliceosome activation via Prp2, 3' SS selection via Slu7, and spliceosome disassembly or discard via Prp43. Hsh155/SF3b1 may act as a general hub on the spliceosome for proteins needing BS access throughout splicing. This hypothesis is intriguing because the N-terminus of SF3b1 in humans contains numerous ULM regions that interact with additional partners not found in yeast (60). These additional factors could modulate constitutive or alternative splicing by binding to and acting through SF3b1.

Hsh155/SF3b1 Functions to Stabilize the U2/BS Duplex

Accurate recognition of splice sites is essential for maintaining the integrity of a spliced mRNA, and the spliceosome has evolved numerous mechanisms to ensure high fidelity at nearly

every stage of splicing. Many spliceosome proofreading mechanisms rely on coupling the activity of DExH ATPases with the stalling or discard of spliceosomes (32). For example, one mechanism proposed for proofreading BS selection involves recognition of mispairing between the BS and U2 snRNA by Prp5 (49). Mispairing triggers Prp5 retention on the spliceosome, thereby blocking subsequent assembly steps (37). Our data suggest the functions of Prp5 are unaffected by MDS mutations. First, deletion of Cus2 (which is believed to be removed from U2 by ATP-dependent Prp5 activity) showed no changes in reporter RNA splicing, suggesting that MDS mutations do not act through retention of Cus2. Additionally, Prp5 mutations and MDS alleles of Hsh155 are not epistatic. Prp5 mutations known to affect fidelity still impact splicing when used in combination with MDS alleles, suggesting these mutants act at different times during splicing. Based on these data and our Y2H results, we propose that Hsh155/SF3b1 modulates BS usage in a manner distinct from Prp5. Additionally, we propose a function of Hsh155 is to confer stability to weak duplexes, thereby improving spliceosome assembly and splicing on introns containing weak BS.

Hsh155 may help to stabilize structures within the spliceosome that are ultimately necessary for catalysis. Specifically, we propose that SF3b1/Hsh155 can help bolster U2/BS duplexes with mismatches near the branchpoint adenosine at the -2, -1, and +1 positions early during spliceosome assembly and that this stabilization allows for progression to subsequent steps in splicing (**Figure 5.13C**). In this model, mutations that impact splicing, such as those found in MDS, are those that affect the ability of SF3b1/Hsh155 to stabilize the U2/BS duplex, with some mutations conferring greater stability (*e.g*. D450G) than WT and others conferring less (*e.g.* K335E). Consistent with this hypothesis are our observations that transversions occurring at A258, which immediately flanks the branchpoint at the -1 position, impact splicing in these mutants. These transversions introduce C/U and U/U mismatches within the snRNA/BS duplex. The A258G transition, which can likely form a stable G/U wobble pair with the snRNA, shows no splicing defects. Also consistent with this hypothesis is that SF3b1/Hsh155 mutations do not

change the splicing of an intron containing a consensus BS sequence or a sequence with substitution of the branchpoint adenosine with cytidine. This position is not paired with the snRNA and therefore may contribute less to the overall stability of the helix (61). Recognition and proofreading of the branchpoint nucleotide is performed by other splicing factors (branchpoint bridging protein/SF1 during assembly and Prp16 prior to 5' SS cleavage) (62-64) while Hsh155 is critical for formation the U2 snRNA/BS duplex. Finally, our results agree with recent structures of the yeast B^{act} spliceosome. In those structures, the nucleotides of the U2/BS duplex immediately flanking the branchpoint adenosine also make extensive contacts with Hsh155 (24); these are the same nucleotide positions shown by our *ACT1-CUP1* to be impacted by $Hsh155^{MDS}$. Thus, these MDS alleles of Hsh155 may change how Hsh155 interacts with the U2/BS duplex in the branchpoint region and ultimately lead to stabilization or destabilization of duplexes containing nearby mismatches.

Our data also show that MDS mutations that impair BS usage can affect mRNA levels to a greater extent than those that improve usage (**Figure 5.3E**). The biological function of Hsh155 may be to relax the specificity of the spliceosome and allow it to splice introns with BS that deviate from the consensus sequence and form metastable U2 snRNA/BS duplexes. We propose that this role is of greater necessity in organisms like humans that have introns with poorly conserved splice sites. Indeed, the lack of observable growth phenotypes in yeast expressing MDS alleles is likely the result of the scarcity of nonconsensus BS sequences (**Figure 5.3F**) in native yeast introns (55-57). Intriguingly, this mechanism also suggests a finely tuned balance between Hsh155 and Prp5 function to modulate splicing. In this model, Prp5 would counteract Hsh155 activity and reduce nonconsensus BS usage at a step subsequent to Hsh155 binding of the U2 snRNA/BS duplex (**Figure 5.13C**). This is consistent with our data that Prp5 and Hsh155 mutations both impact splicing of reporter RNAs when combined (**Figure 5.10D-F**). Balancing the

competing activities of Hsh155 and Prp5 could be used to regulate splicing of transcripts containing particular BS sequences.

Conformational Flexibility in SF3b1 Leads to Changes in BS Usage

Proposed structural changes in SF3b1 could play a role in coupling snRNA/BS duplex recognition and stabilization with conformations that either promote or impede splicing. Indeed, conformational changes in SF3b1 have been previously proposed based on low resolution EM reconstructions and the crystal structure of the human SF3b complex (26, 65, 66). A number of positions mutated in MDS are involved in intramolecular interactions between helices in the HEAT repeats (26). Directly affecting the conformational equilibrium between multiple states of Hsh155 is one way to achieve disparate effects on splicing. In our model, binding of intronic RNA downstream of the BS to HEAT repeats mutated in MDS helps to select a closed conformation of Hsh155 that will bind and stabilize the U2/BS duplex (**Figure 5.13B**). The K335E mutation and others like it impede these processes and favor an open conformation, potentially by interfering with RNA binding or an allosteric switch. Conversely, mutations that improve nonconsensus intron splicing (*e.g.,* D450G) shift the equilibrium towards the closed, duplex-bound state. This model is supported by our observation that MDS mutations have additive effects on splicing, as the shift in equilibrium can be restored by including a second antagonist allele (**Figure 5.3H**). We cannot exclude an opposing model in which MDS alleles primarily function by disrupting intronic RNA binding near the site of the mutations. In this case, a stable RNA duplex formed between the U2 snRNA and a consensus BS would overcome the effect of mutating this RNA binding site on Hsh155. However, we do not favor this model, as a functional purpose of RNA binding to this region of Hsh155 is unclear. Moreover, a recent mutational analysis of human SF3b1 revealed no detectable changes in RNA binding by SF3b1 after introduction of a MDS mutation (26). Finally, we note that Hsh155 structural equilibrium may also impact the competition between U2 branchsite stem loop formation and snRNA/BS basepairing to facilitate coupled snRNA and

protein conformational changes during splicing (67). Structures of B^{act} spliceosomes and U2 snRNPs containing MDS mutations would provide interesting insight to the effects these mutations have on the overall structure.

Further evidence supporting the model that conformational change in Hsh155 is an important driver of BS usage comes from our experiments in which MDS alleles were combined with a mutant of Prp2 (**Figure 5.12B**). These results are consistent with our model that the K335E mutation favors the open, unbound conformation of Hsh155 and D450G favors the closed, bound conformation Hsh155. The former conformation facilitates Prp2-dependent spliceosome activation, while the latter impedes this step. It is likely that spliceosome activation involves release of contacts between Hsh155 and the U2/BS duplex to allow this RNA to enter the active site and splicing to proceed. As suggested by our data, RNA release and structural transitions in Hsh155 are likely coupled to one another as well as to Prp2 activity. ATP hydrolysis by Prp2 may help to trigger Hsh155 conformational change during activation of the spliceosome.

How usage of a different intronic BS leads to alternative 3′ SS selection in MDS is not immediately obvious based on sequence predictions or structural models. Consistent with our observation that MDS mutants do not have defects in cryptic 3′ SS discrimination (**Figure 5.6D, E**), recent work has identified that most splice site changes arise from switching of the BS from a "weak" BS to "strong" BS located nearby and upstream of the canonical BS (17, 18). Changes in how MDS mutant SF3b1 stabilizes weak U2/BS duplexes could lead to repositioning of the spliceosome to regions of the intron with differing complementarity to the U2 snRNA. Whether BS repositioning occurs during assembly or in spliceosomes through the action of a DEAH-box helicase (*e.g.*, Prp2 or Prp16) is not known. These helicases facilitate sampling of multiple potential BS by the spliceosome (68), and how these BS are sampled and their competitiveness with one another may be influenced by MDS mutations in SF3b1. Altered BS sampling in MDS potentially rationalizes the observation that a weak polypyrmidine (Py) tract is necessary for BS

switching in humans. The binding of the splicing factors U2AF65/35 to strong Py tracts could help limit BS sampling of neighboring sequences by the spliceosome during assembly.

The work presented here supports a novel mechanism wherein SF3b1 helps to define the BS during pre-mRNA splicing. Furthermore, we have provided insight into how mutations in a splicing factor can change fundamental functions of the spliceosome. The specific changes in alternative splicing that predispose individuals to MDS is currently unclear. Recent work has shown that the MDS-linked U2AF35 S34F mutation predisposes the cell to transformation through aberrant processing of the *ATG7* transcript (69). A similar mechanism may be occurring in MDS patients with mutant SF3b1, wherein only a fraction of the misprocessed transcripts lead to disease. These misprocessed transcripts may be produced by subtle alteration of how BS compete with one another during splicing and/or by how human-specific splicing regulatory proteins interact with SF3b1 to stabilize BS duplexes containing mismatches. It has been speculated that BS switching due to MDS alleles arises from selection of sequences with increased pairing potential to the U2 snRNA (17, 18), consistent with our results showing that some of the homologous Hsh155 mutations impair splicing when mismatches between the BS and snRNA are present. This suggests that principles that emerge from understanding how these disease alleles alter splicing in yeast will be informative for studies of human splicing in cancer. Understanding how SF3b1 functions in molecular detail is crucial to remedying defects associated with these processes and for designing novel SF3b1-targeted therapeutics for patients suffering from these malignancies.

COMMENTARY

At the time of publication of my paper addressing the impact of MDS mutation in yeast, another paper studying the same topic was published from Charles Query and Yongzhen Xu's labs (70). They chose to focus on a limited number of MDS mutations and sought to biochemically dissect the impact of these mutations. They were able to express HEAT domain fragments of Hsh155 and show interaction with Prp5 via the N-terminus of Prp5. Prp5 interacts with two different HEAT regions of Hsh155 (H1-6 and H9-12). MDS mutations altered the amount of Prp5 pulled down using purified components, but not when immunoprecipitation was performed from extracts. Furthermore, mutation of the Prp5 DPLD motif to AAAA resulted in greater pulldown of Hsh155. This was particularly unexpected, as *in vitro* and *in vivo* evidence from multiple sources has shown that mutation of DPLD results in a weakened interaction with SF3b. The above results are very similar to our observation that mutations in Hsh155 impact the relationship with Prp5, but the importance of this change is not predictive of splicing changes (**Figure 5.8**). It should also be noted that mutations in Prp5 can readily change BS usage and the interaction with the SF3b complex. However, to date no patients with mutations in Prp5 have been associated with any cancers. This is not what would be predicted if the interaction with Prp5 was the primary mechanism of action for MDS.

Tang *et al.* also performed two genetic screens to identify mutations that alter BS usage. The first was a genome-wide screen performed using UV mutagenesis. From this screen they identified mutation in (exclusively) U2-related factors that can modulate BS usage. However, not all mutations were found in Hsh155, but rather mutations could be found in Cus1, Rse1, Prp9 and Prp5. These mutations are mapped onto the structure of SF3b in **Figure 5.14A** and are close to the U2/BS duplex. Tang *et al.* also performed a directed screen to identify mutations in Hsh155 that change BS usage using error-prone PCR of HEAT repeats 1-10. Mutations were identified in repeats 2-8 and many of these mutations are involved in intramolecular interactions in Hsh155 (**Figure 5.14B**). Given the location and function of these positions, it is likely they contribute to

the structure of Hsh155. Therefore, these mutations also support the model that we proposed (**Figure 5.13**). It is possible that these mutations impact BS usage by changing SF3 stability in the spliceosome.

At its heart, the model proposed for BS changes by Tang *et al.* is essentially the same as ours. We agree that Hsh155 has least two conformations necessary for splicing. The first is involved in BS recognition and the second is triggered to allow for U2/BS duplex release and entrance into the spliceosomal active site. The major difference is that the authors propose that altered interaction with Prp5 is a driving force behind the observed splicing changes, whereas we propose that the conformational equilibrium of Hsh155 itself is altered by mutations. Structural data and genetic evidence supports our model. However, further work to understand the action of Prp5 during BS choice and its relationship with Hsh155 is warranted.

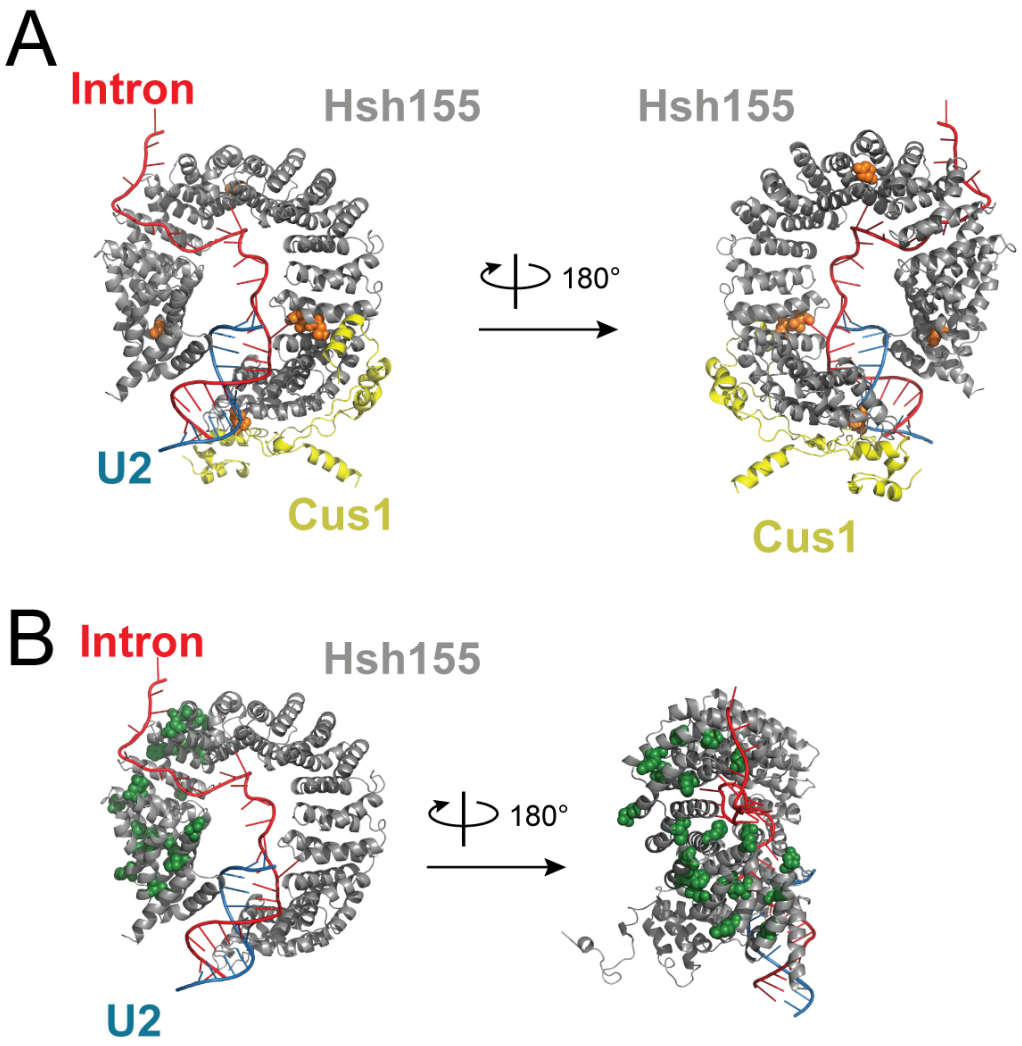


Figure 5.14. Location of mutations that alter BS usage identified in Tang *et al.* (70). (A) The location of mutations identified by UV mutagenesis in Cus1 (Cus1) and Hsh155 (gray; mutations shown in orange). Prp9 and the region in mutated in Rse1 are not observed in the Bact spliceosome structure. (B) Locations mutated in the directed screen (green).

ACKNOWLEDGEMENTS

We thank Charles Query, Soo-Chen Cheng, Jill Wildonger, and Dave Brow for strains, plasmids, and antibodies and Sandy Tretbar and George Luo for technical assistance. We also thank Sam Butcher, Dave Brow, Allison Didychuk, Jon Staley, and Betty Craig for careful reading of the manuscript, comments, and advice.

REFERENCES

1. Anczuków,O. and Krainer,A.R. (2016) Splicing-factor alterations in cancers. *RNA*, **22**, 1285–1301.

2. Dvinge,H., Kim,E., Abdel-Wahab,O. and Bradley,R.K. (2016) RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer*, **16**, 413–430.

3. Coovert,D.D., Le,T.T., McAndrew,P.E., Strasswimmer,J., Crawford,T.O., Mendell,J.R., Coulson,S.E., Androphy,E.J., Prior,T.W. and Burghes,A.H. (1997) The survival motor neuron protein in spinal muscular atrophy. *Hum. Mol. Genet.*, **6**, 1205–1214.

4. McKie,A.B., McHale,J.C., Keen,T.J., Tarttelin,E.E., Goliath,R., van Lith-Verhoeven,J.J., Greenberg,J., Ramesar,R.S., Hoyng,C.B., Cremers,F.P., *et al.* (2001) Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet.*, **10**, 1555–1562.

5. Vithana,E.N., Abu-Safieh,L., Allen,M.J., Carey,A., Papaioannou,M., Chakarova,C., Al-Maghtheh,M., Ebenezer,N.D., Willis,C., Moore,A.T., *et al.* (2001) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell*, **8**, 375–381.

6. Chakarova,C.F., Hims,M.M., Bolz,H., Abu-Safieh,L., Patel,R.J., Papaioannou,M.G., Inglehearn,C.F., Keen,T.J., Willis,C., Moore,A.T., *et al.* (2002) Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.*, **11**, 87–92.

7. Visconte,V., Makishima,H., Maciejewski,J.P. and Tiu,R.V. (2012) Emerging roles of the spliceosomal machinery in myelodysplastic syndromes and other hematological disorders. *Leukemia*, **26**, 2447–2454.

8. Will,C.L. and Lührmann,R. (2011) Spliceosome Structure and Function. *Cold Spring Harb. Perspect. Biol.*, **3**, a003707.

9. Wahl,M.C., Will,C.L. and Lührmann,R. (2009) The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, **136**, 701–718.

10. Hsu,T.Y.T., Simon,L.M., Neill,N.J., Marcotte,R., Sayad,A., Bland,C.S., Echeverria,G.V., Sun,T., Kurley,S.J., Tyagi,S., *et al.* (2015) The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature*, **525**, 384–388.

11. Tresini,M., Warmerdam,D.O., Kolovos,P., Snijder,L., Vrouwe,M.G., Demmers,J.A.A., van IJcken,W.F.J., Grosveld,F.G., Medema,R.H., Hoeijmakers,J.H.J., *et al.* (2015) The core spliceosome as target and effector of non-canonical ATM signalling. *Nature*, **523**, 53–58.

12. Martin,M., Maßhöfer,L., Temming,P., Rahmann,S., Metz,C., Bornfeld,N., van de Nes,J., Klein-Hitpass,L., Hinnebusch,A.G., Horsthemke,B., *et al.* (2013) Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat. Genet.*, **45**, 933–936.

13. Papaemmanuil,E., Cazzola,M., Boultwood,J., Malcovati,L., Vyas,P., Bowen,D., Pellagatti,A.,

Wainscoat,J.S., Hellstrom-Lindberg,E., Gambacorti-Passerini,C., *et al.* (2011) Somatic SF3B1 Mutation in Myelodysplasia with Ring Sideroblasts. *N. Engl. J. Med.*, **365**, 1384–1395.

14. Quesada,V., Conde,L., Villamor,N., Ordóñez,G.R., Jares,P., Bassaganyas,L., Ramsay,A.J., Beà,S., Pinyol,M., Martínez-Trillos,A., *et al.* (2011) Exome sequencing identifies recurrent mutations of the splicing factor. *Nat. Genet.*, **44**, 47–52.

15. Gentien,D., Kosmider,O., Nguyen-Khac,F., Albaud,B., Rapinat,A., Dumont,A.G., Damm,F., Popova,T., Marais,R., Fontenay,M., *et al.* (2014) A common alternative splicing signature is associated with SF3B1 mutations in malignancies from different cell lineages. *Leukemia*, **28**, 1355–1357.

16. DeBoever,C., Ghia,E.M., Shepard,P.J., Rassenti,L., Barrett,C.L., Jepsen,K., Jamieson,C.H.M., Carson,D., Kipps,T.J. and Frazer,K.A. (2015) Transcriptome Sequencing Reveals Potential Mechanism of Cryptic 3’ Splice Site Selection in SF3B1-mutated Cancers. *PLoS Comp. Biol.*, **11**, e1004105.

17. Alsafadi,S., Houy,A., Battistella,A., Popova,T., Wassef,M., Henry,E., Tirode,F., Constantinou,A., Piperno-Neumann,S., Roman-Roman,S., *et al.* (2016) Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.*, **7**, 10615–10615.

18. Darman,R.B., Seiler,M., Agrawal,A.A., Lim,K.H., Peng,S., Aird,D., Bailey,S.L., Bhavsar,E.B., Chan,B., Colla,S., *et al.* (2015) Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep.*, **13**, 1033–1045.

19. Mupo,A., Seiler,M., Sathiaseelan,V., Pance,A. and Yang,Y. (2016) Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia*, 10.1038/leu.2016.251.

20. Behrens,S.E., Tyc,K., Kastner,B., Reichelt,J. and Lührmann,R. (1993) Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Mol. Cell Biol.*, **13**, 307–319.

21. Ruby, S.W., Chang, T.H. and Abelson, J. (1993) Four yeast spliceosomal proteins (PRP5, PRP9, PRP11, and PRP21) interact to promote U2 snRNP binding to pre-mRNA. *Genes Dev.*, **7**, 1909–1925.

22. McPheeters, D.S. and Muhlenkamp, P. (2003) Spatial organization of protein-RNA interactions in the branch site-3′ splice site region during pre-mRNA splicing in yeast. *Mol. Cell. Biol.*, **23**, 4174-4186.

23. Schneider,C., Agafonov,D.E., Schmitzová,J., Hartmuth,K., Fabrizio,P. and Lührmann,R. (2015) Dynamic Contacts of U2, RES, Cwc25, Prp8 and Prp45 Proteins with the Pre-mRNA Branch-Site and 3' Splice Site during Catalytic Activation and Step 1 Catalysis in Yeast Spliceosomes. *PLoS Genet.*, **11**, e1005539.

24. Rauhut,R., Fabrizio,P., Dybkov,O. and Hartmuth,K. (2016) Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. *Science*, **353**, 1399-1405.

25. Yan,C., Wan,R., Bai,R., Huang,G. and Shi,Y. (2016) Structure of a yeast activated spliceosome at 3.5 A resolution. *Science*, **353**, 904–911.

26. Cretu,C., Schmitzova,J., Ponce-Salvatierra,A., Dybkov,O., DeLaurentiis,E.I., Sharma,K., Will,C.L., Urlaub,H., Lührmann,R. and Pena,V. (2016) Molecular architecture of SF3b and structural consequences of its cancer-related mutations. Mol. Cell., **64**, 307–319.

27. Kaida,D., Motoyoshi,H., Tashiro,E., Nojima,T., Hagiwara,M., Ishigami,K., Watanabe,H., Kitahara,T., Yoshida,T., Nakajima,H., *et al.* (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.*, **3**, 576–583.

28. Kotake,Y., Sagane,K., Owa,T., Mimori-Kiyosue,Y., Shimizu,H., Uesugi,M., Ishihama,Y., Iwata,M. and Mizui,Y. (2007) Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat. Chem. Biol.*, **3**, 570–575.

29. Hasegawa,M., Miura,T., Kuzuya,K., Inoue,A., Won Ki,S., Horinouchi,S., Yoshida,T., Kunoh,T., Koseki,K., Mino,K., *et al.* (2011) Identification of SAP155 as the Target of GEX1A (Herboxidiene), an Antitumor Natural Product. *ACS Chem. Biol.*, **6**, 229–233.

30. Folco,E.G., Coil,K.E. and Reed,R. (2011) The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. *Genes Dev.*, **25**, 440–444.

31. Kim,S.-H. and Lin,R.-J. (1993) Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. *Proc. Natl. Acad. Sci. U S A*, **90**, 888–892.

32. Liu,Y.-C. and Cheng,S.-C. (2015) Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing. *J. Biomed. Sci.*, **22**, 54–54.

33. Ohrt,T., Prior,M., Dannenberg,J., Odenwalder,P., Dybkov,O., Rasche,N., Schmitzova,J., Gregor,I., Fabrizio,P., Enderlein,J., *et al.* (2012) Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA*, **18**, 1244-1256.

34. Wlodaver,A.M. and Staley,J.P. (2014) The DExD/H-box ATPase Prp2p destabilizes and proofreads the catalytic RNA core of the spliceosome. *RNA*, **20**, 282–294.

35. Effenberger,K.A., Urabe,V.K., Prichard,B.E., Ghosh,A.K. and Jurica,M.S. (2016) Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. *RNA*, **22**, 350–359.

36. Lesser,C.F. and Guthrie,C. (1993) Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics*, **133**, 851–863.

37. Liang,W.-W. and Cheng,S.-C. (2015) A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev.*, **29**, 81–93.

38. Amberg,D.C., Burke,D. and Strathern,J.N. (2005) Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

39. Siatecka,M., Reyes,J.L. and Konarska,M.M. (1999) Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.*, **13**, 1983–1993.

40. Keogh,M.-C., Kim,J.-A., Downey,M., Fillingham,J., Chowdhury,D., Harrison,J.C., Onishi,M., Datta,N., Galicia,S., Emili,A., *et al.* (2005) A phosphatase complex that dephosphorylates γH2AX regulates DNA damage checkpoint recovery. *Nature*, **439**, 497–501.

41. Wan,Y. and Wu,C.J. (2013) SF3B1 mutations in chronic lymphocytic leukemia. *Blood*, **121**, 4627–4634.

42. Rossi,D., Spina,V., Bomben,R., Rasi,S., Dal-Bo,M., Bruscaggin,A., Rossi,F.M., Monti,S., Degan,M., Ciardullo,C., *et al.* (2013) Association between molecular lesions and specific B-cell receptor subsets in chronic lymphocytic leukemia. *Blood*, **121**, 4902–4905.

43. Frank,D. and Guthrie,C. (1992) An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes Dev.*, **6**, 2112–2124.

44. Wang,Q., He,J., Lynn,B. and Rymond,B.C. (2005) Interactions of the Yeast SF3b Splicing Factor. *Mol. Cell. Biol.*, **25**, 10745–10754.

45. Lardelli,R.M., Thompson,J.X., Yates,J.R. and Stevens,S.W. (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA*, **16**, 516–528.

46. Martin,A., Schneider,S. and Schwer,B. (2002) Prp43 Is an Essential RNA-dependent ATPase Required for Release of Lariat-Intron from the Spliceosome. *J. Biol. Chem.*, **277**, 17743–17750.

47. Mayas,R.M., Maita,H. and Staley,J.P. (2006) Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat. Struct. Mol. Biol.*, **13**, 482–490.

48. Fourmann,J.B., Dybkov,O., Agafonov,D.E. and Tauchert,M.J. (2016) The target of the DEAH-box NTP triphosphatase Prp43 in Saccharomyces cerevisiae spliceosomes is the U2 snRNP-intron interaction. *eLife*. **5**, e15564.

49. Xu,Y.-Z. and Query,C.C. (2007) Competition between the ATPase Prp5 and Branch Region-U2 snRNA Pairing Modulates the Fidelity of Spliceosome Assembly. *Mol. Cell*, **28**, 838–849.

50. Perriman,R. and Ares,M. (2000) ATP can be dispensable for prespliceosome formation in yeast. *Genes Dev.*, **14**, 97–107.

51. Perriman,R., Barta,I. and Voeltz,G.K. (2003) ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. Proc. Natl. Acad. Sci. U S A, **24**, 13857-13862.

52. Dalbadie-McFarland,G. and Abelson,J. (1990) PRP5: a helicase-like protein required for mRNA splicing in yeast. *Proc. Natl. Acad. Sci. U S A*, **87**, 4236–4240.

53. Zhang,Z.-M., Yang,F., Zhang,J., Tang,Q., Li,J., Gu,J., Zhou,J. and Xu,Y.-Z. (2013) Crystal Structure of Prp5p Reveals Interdomain Interactions that Impact Spliceosome Assembly. *Cell Rep.*, **5**, 1269–1278.

54. Shao,W., Kim,H.S., Cao,Y., Xu,Y.Z. and Query,C.C. (2011) A U1-U2 snRNP Interaction Network during Intron Definition. *Mol. Cell. Biol.*, **32**, 470–478.

55. Gould,G.M., Paggi,J.M., Guo,Y., Phizicky,D.V. and Zinshteyn,B. (2016) Identification of new branch points and unconventional introns in Saccharomyces cerevisiae. *RNA*, **20**, 1522-1534.

56. Qin,D., Huang,L., Wlodaver,A., Andrade,J. and Staley,J.P. (2016) Sequencing of lariat termini in S. cerevisiae reveals 5′ splice sites, branch points, and novel splicing events. *RNA*, **22**, 237–253.

57. Grate,L. and Ares,M. (2002) Searching yeast intron data at Ares lab web site. *Methods Enzymol.*, **350**, 380–392.

58. Chua,K. and Reed,R. (1999) The RNA splicing factor hSlu7 is required for correct 3′ splice-site choice. *Nature*, **402**, 207-210.

59. Zhang,X. and Schwer,B. (1997) Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Res.*, **25**, 2146–2152.

60. Thickman,K.R., Swenson,M.C., Kabogo,J.M., Gryczynski,Z. and Kielkopf,C.L. (2006) Multiple U2AF65 Binding Sites within SF3b155: Thermodynamic and Spectroscopic Characterization of Protein–Protein Interactions among pre-mRNA Splicing Factors. *J. Mol. Biol.*, **356**, 664–683.

61. Query,C.C., Moore,M.J. and Sharp,P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.*, **8**, 587–597.

62. Berglund,J.A., Chua,K., Abovich,N., Reed,R. and Rosbash,M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, **89**, 781–787.

63. Koodathingal,P., Novak,T., Piccirilli,J.A. and Staley,J.P. (2010) The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5′ splice site cleavage during pre-mRNA splicing. *Mol. Cell*, **39**, 385–395.

64. Burgess,S.M. and Guthrie,C. (1993) A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell*. **73**, 1377-1391.

65. Golas,M.M., Sander,B., Will,C.L., Lührmann,R. and Stark,H. (2005) Major Conformational Change in the Complex SF3b upon Integration into the Spliceosomal U11/U12 di-snRNP as Revealed by Electron Cryomicroscopy. *Mol. Cell*, **17**, 869–883.

66. Rakesh,R., Joseph,A.P., Bhaskara,R.M. and Srinivasan,N. (2016) Structural and mechanistic insights into human splicing factor SF3b complex derived using an integrated approach guided by the cryo-EM density maps. *RNA Biol.*, **13**, 1025-1040.

67. Perriman,R. and Ares,M. (2010) Invariant U2 snRNA Nucleotides Form a Stem Loop to Recognize the Intron Early in Splicing. *Mol. Cell*, **38**, 416–427.

68. Semlow,D.R., Blanco,M.R., Walter,N.G. and Staley,J.P. (2016) Spliceosomal DEAH-Box ATPases Remodel Pre- mRNA to Activate Alternative Splice Sites. *Cell*, **164**, 985–998.

69. Park,S.M., Ou,J., Chamberlain,L., Simone,T.M. Yang,H., Virbasius C-M., Ali A.M., Zhu L.J., Mukherjee S., Raza A. and Green M.R. (2016) U2AF35 (S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3′ End Formation. *Mol Cell*. **62**, 479-490.

70. Park,S.M., Ou,J., Chamberlain,L., Simone,T.M. Yang,H., Virbasius C-M., Ali A.M., Zhu L.J., Mukherjee S., Raza A. and Green M.R. (2016) U2AF35 (S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3′ End Formation. *Mol Cell*. **62**, 479-490.

71. Tang,Q., Rodriguez-Santiago,S. and Wang,J. *et al.* (2016) SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. Genes. Dev. **30**, 2710–2723.

Chapter Six

Functional Characterization of SF3b Contacts to the U2/BS duplex

Joshua Paulson was a tremendous help with the cloning and reagent generation necessary for this work.

ABSTRACT

The SF3b complex plays a conserved but poorly understood role in pre-mRNA splicing. Here, we explored the function of SF3b components, Hsh155 and Rds3, by probing molecular contacts observed in recently released structures of the spliceosome. This data reveals the importance of individual contacts to the U2/BS duplex and defines their contribution to branchsite usage by the spliceosome. We have also studied the divergence of human and yeast SF3b1/Hsh155 using chimeric protein constructs and defined essential regions of Hsh155. In total, this work enhances the genetic framework for understanding the function of the SF3b complex in yeast.

INTRODUCTION

Recent advances in single particle cryo-electron microscopy (cryo-EM) have revealed the molecular architecture of several spliceosome intermediates, including an activated spliceosomal complex containing the SF3b subcomplex(1, 2). These structures represent a huge leap forward in our understanding of splicing. However, functional characterization of many of the protein:protein, protein:RNA, and RNA:RNA contacts observed in these structures is lacking. This is significant, as myself and others have shown that mutation of many of the SF3 components can modulate BS usage in yeast (**Chapter 5**)(3, 4). To address this deficiency, the work presented here has sought to understand the importance of individual contacts to the U2/BS duplex by SF3b components for pre-mRNA splicing. I focused my efforts on Hsh155 (human homolog SF3b1; for clarity the yeast protein will hereafter be referred to as Hsh155 and the human protein as SF3b1) and the protein Rds3 (human homolog PHF5a). These two proteins directly contact the U2/BS duplex at numerous positions, have been implicated in disease states in human cells, and have been shown to affect splicing in yeast(1-7).

In the structure of the activated spliceosome, Hsh155 contacts the intron and U2 snRNA in two distinct regions and makes numerous contacts with other components of the SF3b complex(1, 2). The overall architecture of Hsh155 is formed by the HEAT domain and reminiscent of an open washer in which the U2/BS duplex binds along a cleft formed between the N-terminal and C-terminal HEAT repeats (**Figure 6.1A AND 6.1B**)(1, 2, 8). Contacts in the region (HEATs 4-9) that bind RNA located between the branchsite (BS) and the 3′ splice site (3 SS) are implicated in myelodysplastic syndrome and are discussed extensively in **Chapter 5**. However, prior to the cryo-EM structure, binding of U2/BS duplex by the N- and C-terminal regions directly was unanticipated. The N-terminal region of the protein upstream of the HEAT repeats is substantially larger in SF3b1 than Hsh155 and has been implicated in interacting with other splicing factors through U2AF ligand motifs (ULMs)(9, 10). ULMs are protein:protein interaction motifs that tightly bind to U2AF homology motifs (UHMs). While the N-terminus of Hsh155 is smaller, it still contains

at least one *bona fide* ULM and several other putative ULMs that are divergent from the consensus. Therefore, it is likely that binding of RNA or protein in this region may have functional significance at different stages of the splicing cycle. However, the importance of this region is currently unclear.

SF3b1 is the target of many small molecule inhibitors that are active in both human cells and human splicing extracts(11-14). SF3b1 is extremely well conserved across organisms, however, the efficacy of splicing inhibitors in yeast has not been explored. Many of these inhibitors have found use as drugs treating various forms of cancer. Mutations in SF3b1 such as R1074H can also grant resistance to drugs such as pladienolide B(15).

Nestled in the interior of the open washer structure of Hsh155 is another SF3b component called Rds3 (**Figure 6.1A**). Rds3 is a small, 12.2 kDa protein that folds into a motif belonging to the plant homeodomain (PHD) family(5, 16, 17). Heat inactivation of temperature sensitive alleles of Rds3 result in disassembly of the SF3b complex and a block of splicing in yeast(5). Mutations in the human homolog of Rds3, PHF5a, have also been identified in a number of solid-state cancers(6, 7). Rds3 directly contacts both Hsh155 and the U2/BS duplex in multiple regions but the importance of these contacts and the role of Rds3 in splicing have not been well studied.

Here, we have used a combination of genetic and biochemical tools to dissect the role of contacts between components of the SF3b complex. We have also explored the ability of the human protein to complement loss of Hsh155 and studied the impact of a splicing inhibitor on yeast *in vitro* splicing. This work further defines the role of Hsh155 and Rds3 during splicing.

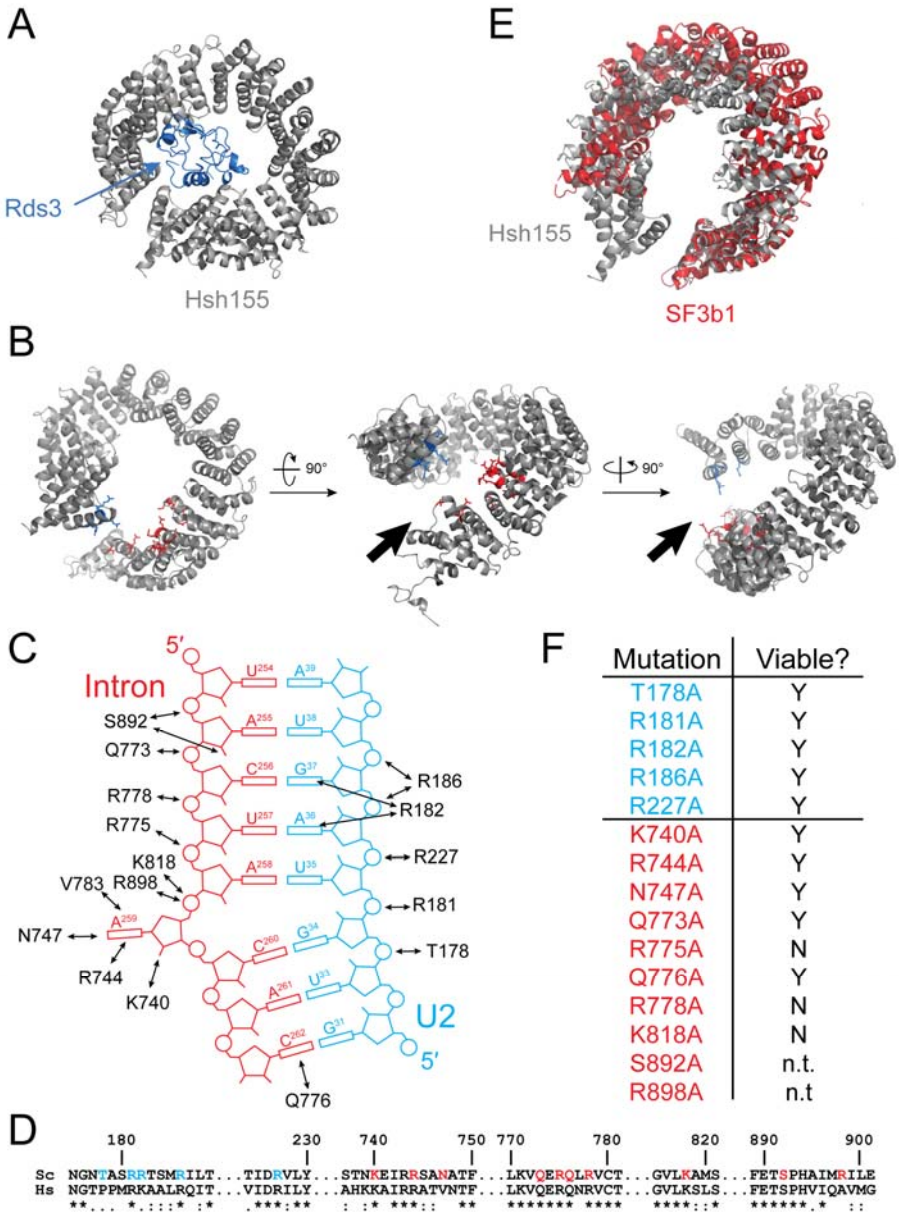


Figure 6.1: Overview of amino acids within 4 Å of the U2/BS duplex. (A) Structure of Hsh155 from the B^{act} spliceosome (PDB: 5GM6). (B) Residues in the N-terminal HEAT repeats contacting the U2/BS duplex are shown in blue. C-terminal intron-contacting residues are colored in red. Arrow indicates the location of the cleft in which the U2/BS duplex resides. (C) Cartoon scheme showing contacts to the U2/BS duplex. Contacts are not base specific and RNA is colored matched with (B). (D) Alignment of human SF3b1 (bottom) and *S. cerevisae* (top) in regions containing residues that contact the U2/BS duplex. Most residues are conserved. (E) Overlay of human SF3b1 and

Hsh155 shows the different curvature of the superhelix (F) Table summarizing results of alanine scanning both the N-terminal (blue) and C-terminal (red) regions. Residues R775, R778 and K818 are lethal when mutated.

MATERIALS AND METHODS

Saccharomyces cerevisiae strains used in these studies were derived from 46α (kind gift of David Brow) or BJ2168. **Tables 6.1 and 6.2** contain detailed lists of strains and plasmids. Yeast transformation and growth was carried out using standard techniques and media.

Site-Directed Mutagenesis

Point mutants were generated using inverse polymerase chain reaction (PCR) with Phusion DNA polymerase (New England Biolabs; Ipswich, MA). Chimeric Hsh155/SF3b1 plasmids were made by restriction enzyme cloning using a yeast codon optimized sequence for SF3b1 purchased from IDT. SF3b1 cDNA from humans inserted into a plasmid cannot be replicated in *E. coli*(18).

In Vitro Splicing Assays

Splicing extracts were prepared from yeast strains derived from BJ2168 and prepared as previously described(19). Briefly, yeast were grown in YPD to an OD_{600} = 1.8, collected by centrifugation and washed with AGK buffer (10 mM HEPES●KOH pH 7.9, 10% glycerol, 200 mM KCl, 1.5 mM MgCl2, 1mM DTT**)**. Cells were collected again by centrifugation, suspended in 7 mL fresh AGK buffer and flash frozen in liquid nitrogen. Yeast were disrupted by cryogenic mechanical lysis using a Retsch Ball Mill. Lysate was thawed, centrifuged at 18000 rpm in a JA-20 rotor for 30 min and again at 36000 rpm for 1 hr in a SW 55 rotor at 4°C. The middle layer was extracted, dialyzed at 4°C against 2 L buffer (25 mM HEPES•KOH pH 7.9, 20% (v/v) glycerol, 50 mM KCl, 1 mM DTT) followed by flash freezing in liquid nitrogen and storage at -80°C until use. Capped RP51a pre-mRNA substrate were transcribed using T7 RNA polymerase. Splicing reactions consisted of 100 mM KPi pH 7.3, 3% (w/v) PEG 8000, 2.5 mM MgCl2, 1 mM DTT, 2 mM ATP, 40 U Murine RNase Inhibitor, 300 pM RNA, 40% (v/v) whole cell extract from yeast. Splicing reactions were performed at room temperature for 45 min. Samples were phenol:chloroform:isoamyl alcohol (25:24:1) extracted and ethanol precipitated to isolate nucleic

acid. Samples were analyzed using a 7% denaturing gel (19:1 acrylamide:bisacrylamide) and imaged using autoradiography. Pladienolide B was purchased from Santa Cruz Biotechnology. It was dissolved in DMSO and used at concentrations ranging from 0.1-20 μM. Extract was incubated with PB for 10 min at room temperature prior to the addition of RNA where indicated.

ACT1-CUP1 Copper Assays

ACT1-CUP1 reporters and growth assays have been described previously(3, 20). Briefly, yeast strains expressing WT or mutant proteins and *ACT1-CUP1* reporters were grown to mid-log phase in the appropriate media to maintain selection for the plasmids, adjusted to OD_{600} = 0.5 and equal volumes were spotted onto plates containing 0, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.25, or 2.5 mM $CuSO_4$. Plates were scored after 3 days growth at 30°C.

Table 6.1. Yeast strains used in this study.

Name	Genotype	Description
46α	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2*	Used to generate Hsh155 shuffle strain.
yAAH0465	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS416-HSH155*	Derived from 46α (kind gift of Dave Brow). Copper sensitive strain used to generate MDS mutant strains for ACT1-CUP1 by plasmid shuffle. Hsh155 was deleted using a KanMX cassette.
yAAH0648	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{WT}*	Copper sensitive strain expressing Hsh155^{WT}. Used for ACT1-CUP1 assays after transformation with reporter plasmids.
yTJC0472	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{N4A}*	Copper sensitive strain expressing Hsh155 T178A R181A R182A R186A.
yTJC0298	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K740A}*	Copper sensitive strain expressing Hsh155 K740A. Transformed with ACT1CUP1 reporters.
yTJC0483	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K740R}*	Copper sensitive strain expressing Hsh155 K740R. Transformed with ACT1CUP1 reporters.
yTJC0300	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{N747A}*	Copper sensitive strain expressing Hsh155 N747A. Transformed with ACT1CUP1 reporters.
yTJC0484	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{N747D}*	Copper sensitive strain expressing Hsh155 N747D. Transformed with ACT1CUP1 reporters.
yTJC0487	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{K818R}*	Copper sensitive strain expressing Hsh155 K818R. Transformed with ACT1CUP1 reporters.
yTJC0301	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{V783A}*	Copper sensitive strain expressing Hsh155 V783A. Transformed with ACT1CUP1 reporters.
yTJC0442	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hsh155::KanMX pRS414-HSH155^{R743H}*	Copper sensitive strain expressing Hsh155 R743H. Transformed with ACT1CUP1 reporters.
yTJC0467	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 rds3::KanMX pRS416-RDS3*	Derived from 46α (kind gift of Dave Brow). Copper sensitive strain used to generate Rds3 mutant strains for ACT1-CUP1 by plasmid shuffle. Rds3 was deleted using a hphMX cassette.
yTJC0674	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 rds3::KanMX pRS413-RDS3^{WT}*	Copper sensitive strain expressing Rds3 WT. Transformed with ACT1CUP1 reporters.
yTJC0675	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 rds3::KanMX pRS416-RDS3^{R38A}*	Copper sensitive strain expressing Rds3 R38A. Transformed with ACT1CUP1 reporters.
yTJC0676	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 rds3::KanMX pRS416-RDS3^{R38C}*	Copper sensitive strain expressing Rds3 R38C. Transformed with ACT1CUP1 reporters.
yTJC0677	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 rds3::KanMX pRS416-RDS3^{N95A}*	Copper sensitive strain expressing Rds3 N95A. Transformed with ACT1CUP1 reporters.
yTJC0678	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 rds3::KanMX pRS416-RDS3^{R99A}*	Copper sensitive strain expressing Rds3 R99A. Transformed with ACT1CUP1 reporters.
yAAH1568	*MAT a prc1 prb1 pep4 leu2 trp1 ura3 hsh155::KANMX pRS414-HSH155^{wt}*	Protease deficient strain used for in vitro splicing.

Table 6.2. Plasmids

#	Plasmid Name	Plasmid Description
pAAH0403	pRS414-HSH155^{WT}	Plasmid used to generate Hsh155^{WT} strain. Hsh155 coding region +/- 250bp was cloned into pRS414
pAAH0854	Hsh155 T178A	Plasmid used to generate mutant Hsh155 strain
pAAH0855	Hsh155 R181A	Plasmid used to generate mutant Hsh155 strain
pAAH0856	Hsh155 R182A	Plasmid used to generate mutant Hsh155 strain
pAAH0857	Hsh155 R186A	Plasmid used to generate mutant Hsh155 strain
pAAH0858	Hsh155 T178A R181A R182A R186A (N4A)	Plasmid used to generate mutant Hsh155 strain
pAAH0860	Hsh155 K227A	Plasmid used to generate mutant Hsh155 strain
Paah0835	Hsh155 R743H	Plasmid used to generate mutant Hsh155 strain
pAAH0862	Hsh155 Q773A	Plasmid used to generate mutant Hsh155 strain
pAAH0863	Hsh155 R775A	Plasmid used to generate mutant Hsh155 strain
pAAH0864	Hsh155 Q776A	Plasmid used to generate mutant Hsh155 strain
pAAH0865	Hsh155 R778A	Plasmid used to generate mutant Hsh155 strain
pAAH0866	Hsh155 Q773A R775A R776A Q778A (C4A)	Plasmid used to generate mutant Hsh155 strain
pAAH0875	Hsh155 K740R	Plasmid used to generate mutant Hsh155 strain
pAAH0876	Hsh155 N747D	Plasmid used to generate mutant Hsh155 strain
pAAH0879	Hsh155 K818R	Plasmid used to generate mutant Hsh155 strain
pAAH0818	HSH155 K740A	Plasmid used to generate mutant Hsh155 strain
pAAH0819	HSH155 R744A	Plasmid used to generate mutant Hsh155 strain
pAAH0820	HSH155 N747A	Plasmid used to generate mutant Hsh155 strain
pAAH0821	HSH155 V783A	Plasmid used to generate mutant Hsh155 strain
pAAH0822	HSH155 K818A	Plasmid used to generate mutant Hsh155 strain
pAAH0818	HSH155 K740A	Plasmid used to generate mutant Hsh155 strain
pAAH0850	pRS413 RDS3 WT	Plasmid used to generate Rds3 WT strain
pAAH0929	pRS413 RDS3 R38A	Plasmid used to generate mutant Rds3 strain
pAAH0930	pRS413 RDS3 R38C	Plasmid used to generate mutant Rds3 strain
pAAH0931	pRS413 RDS3 N95A	Plasmid used to generate mutant Rds3 strain
pAAH0932	pRS413 RDS3 R99A	Plasmid used to generate mutant Rds3 strain
-	pACT1CUP1 WT	WT reporter used for ACT1-CUP1 assays. Gift of D. Brow
-	pACT1CUP1 U254G	BS U254G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A255G	BS A255G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 C256A	BS C256A mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U257A	BS U257A mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U257C	BS U257C mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U257G	BS U257G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A258C	BS A258C mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A258G	BS A258G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A258U	BS A258U mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 BS C (A259C)	BS A259C mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query

-	pACT1CUP1 BS G (A259G)	BS A259C mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 C260G	BS C260G mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 U301G	3' SS gAG mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
-	pACT1CUP1 A302U	3' SS UuG mutant reporter used for ACT1-CUP1 assays. Gift of Charles Query
pAAH0869	HSH155 3xFLAG	Same as pAAH0403 but with 3xFLAG at C-terminus
pAAH0945	HSH155 Hs1-5 3xFLAG	Human HEATs 1-5 have been inserted into pAAH0869.
pAAH0944	HSH155 Hs1-12 3xFLAG	Human HEATs 1-12 have been inserted into pAAH0869.
pAAH0946	HSH155 Hs1-16 3xFLAG	Human HEATs 1-16 have been inserted into pAAH0869.
pAAH0940	HSH155 Hs5-12 3xFLAG	Human HEATs 5-12 have been inserted into pAAH0869.
pAAH0943	HSH155 Hs5-16 3xFLAG	Human HEATs 5-16 have been inserted into pAAH0869.
pAAH0939	HSH155 Hs5-20 3xFLAG	Human HEATs 5-20 have been inserted into pAAH0869.
pAAH0941	HSH155 Hs12-20 3xFLAG	Human HEATs 12-20 have been inserted into pAAH0869.
pAAH0933	HSH155 NΔ63 TRP1	Deletions of Hsh155 at indicated position
pAAH0934	HSH155 NΔ95 TRP1	Deletions of Hsh155 at indicated position
pAAH0935	HSH155 NΔ156 TRP1	Deletions of Hsh155 at indicated position
pAAH0936	HSH155 NΔ236 TRP1	Deletions of Hsh155 at indicated position
pAAH0937	HSH155 CΔ26 TRP1	Deletions of Hsh155 at indicated position
pAAH0938	HSH155 CΔ61 TRP1	Deletions of Hsh155 at indicated position

RESULTS AND DISCUSSION

My work on Hsh155 mutations linked to MDS (**Chapter 5**) suggested a role for Hsh155 in stabilizing the U2/BS duplex and so I began investigating the role of residues that directly contact the duplex guided by the recent cryo-EM structures. Residues in Hsh155 in the structure of the B^{act} spliceosome within 4 Å of the U2/BS duplex were identified using PyMOL (**Figure 6.1**).

Hsh155 contacts the U2/BS duplex

Hsh155 interacts with the U2/BS duplex via the N-terminal and C-terminal HEAT repeats (**Figure 6.1B**). The N-terminal HEAT region interacts exclusively with the U2 snRNA and the C-terminal HEATs binds the intron. Almost all of the contacts are made with the phosphate backbone of the duplex (**Figure 6.1C**)(1, 2). This observation is consistent with utilization of nonconcensus BS in yeast and humans, as sequence-specific contacts might constrain the versatility of the spliceosome. The branchpoint adenosine is bulged from the duplex and inserted into a pocket formed on Hsh155(21). The adenosine base is surrounded by R744, N747, and V783, while the ribose 2′OH and the phosphate backbone interact with K740 and K818, respectively. A sequence alignment between Hsh155 and SF3b1 revealed that the majority of these residues are conserved, suggesting they play a similar role in the human spliceosome (**Figure 6.1D**). This is surprising, as the structure of the isolated human SF3b complex showed a different pitch of the HEAT domain of SF3b1 (**Figure 6.1E**)(1, 2, 8). In this structure, the cleft formed between the N- and C-terminus is much wider. The domain must move to close the gap if SF3b1 binds the U2/BS duplex in a manner similar to Hsh155. Furthermore, this supports the conformational change model proposed for SF3b1/Hsh155 in **Chapter 5**.

We performed alanine scanning of these residues to identify contacts that are required for viability in yeast. Surprisingly, mutation of all the N-terminal residues (T178, R181, R182, R186) had no observable effect in yeast when mutated alone or in combination (**Figure 6.1F**). This data argues that contacts with the U2/BS duplex in the cleft formed by Hsh155 are not critical for splicing. Surprisingly, mutation of many of the residues that contact the branchpoint adenosine

also had no effect on viability. However, when positively charged residues that interact with the BS at positions immediately upstream of the BP were mutated (R775, R778, K818), the yeast could not survive. These data suggest that the only critical contacts between Hsh155 and the BS occur in the C-terminus of the protein and many of the other contacts are unimportant or redundant.

Contacts between Hsh155 and the U2/BS duplex are important for splicing

My work in Chapter 5 demonstrated that changes in splicing caused by mutation of Hsh155 could occur without noticeable defects in yeast growth(3, 4). For this reason, we chose to examine many of the viable mutants discussed above for defects in splicing using the *ACT1-CUP1* reporter assay. We generated a quadruple alanine mutant of the four most N-terminal residues (T178, R181A, R182A, R186A) and found the yeast to be viable after selection on 5-FOA. This yeast strain had no defects in splicing ACT-CUP1 reporters with a consensus intron (**Figure 6.2A**). However, when this strain was assayed for the ability to splice the A258U reporter (BS UACUuAC), we found that the N4A mutant was more copper tolerant than the WT strain (**Figure 6.2A**). This suggests that ablation of N-terminal contacts improves nonconsensus intron splicing. A possible explanation for this improvement is that these contacts must be broken to allow the U2/BS to enter the spliceosome active for the first step of catalysis. Ablation of these contacts facilitates this process. The C4A mutant (Q773A, R775A, Q776, R778A) was lethal and could not be assayed.

The branchpoint adenosine is essentially the only nucleotide in the U2/BS duplex recognized in ways other than electrostatic interactions with the phosphate backbone. For this reason, I generated alanine mutants as well as mutants with more conservative mutations at these positions and assayed them for splicing defects. Specifically, we chose to make N747A/Q, K740A/R, K818A/R, and V783A. Consistent with all other data for Hsh155, the splicing of consensus reporters is not affected by these mutations (**Figure 6.2B**). K740 interacts with the 2′OH of the BP ribose base, which also acts as the nucleophile for the first step of splicing.

Mutation of K740 resulted in improved splicing when mutated to alanine and impaired splicing when mutated to arginine (**Figure 6.2B**). N747 interacts with the adenosine base and mutation of this residue to either A or Q resulted in improved splicing (**Figure 6.2B**). V783 helps to form a pocket the adenosine base is inserted into, and alanine mutation of this residue resulted in greatly improved splicing (**Figure 6.2B**). The K818A mutation was lethal and the K818R mutation was very poorly tolerated in yeast during the ACT1-CUP1 assay (**Figure 6.2B**). In total, this data validate contacts observed in the structure of the B^{act} spliceosome and reveals the importance of these contacts for nonconsensus intron splicing in yeast.

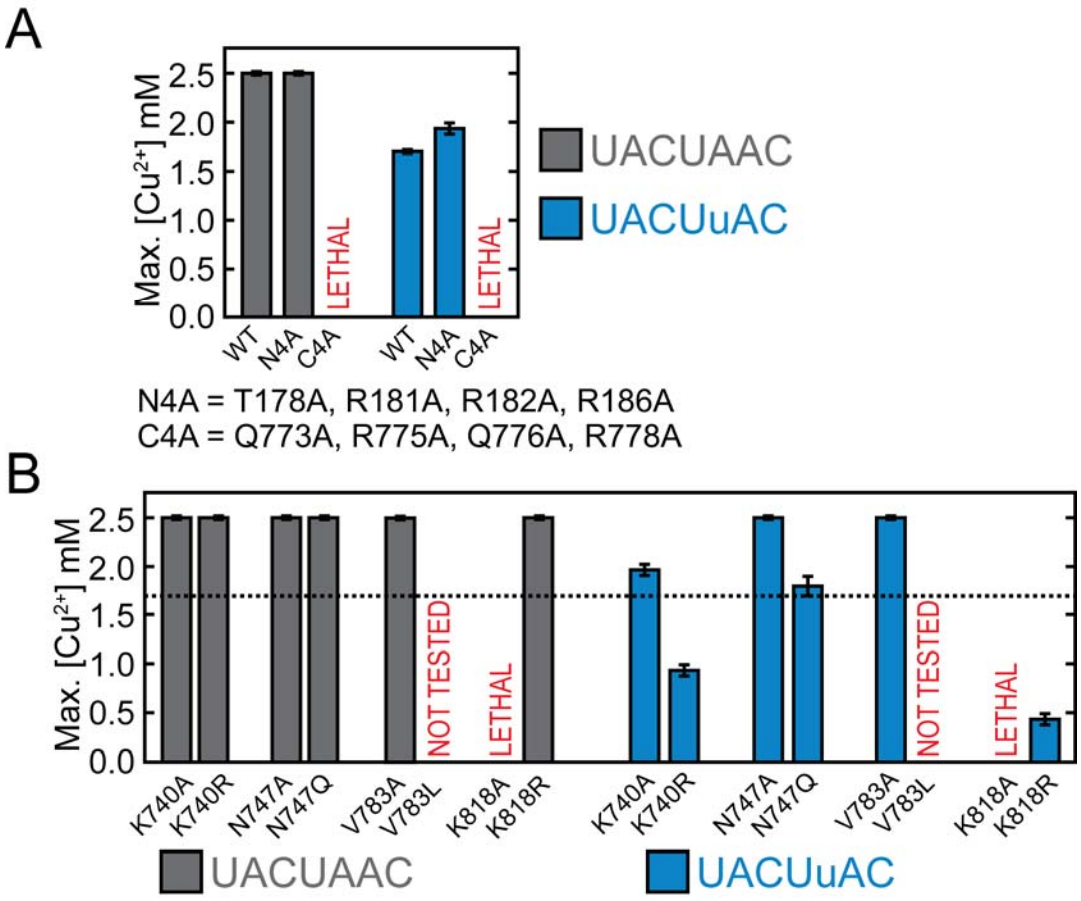


Figure 6.2: *ACT1-CUP1* reporter data for U2/BS duplex mutants. (A) Mutation of four N-terminal residues that contact the U2/BS duplex slightly improves nonconcensus intron splicing but does not affect consensus intron splicing. (B) Positions that contact the branchpoint adenosine were mutated to alanine or a similar amino acid. The impact on BS A258U and BS WT splicing is shown. Dotted line indicated maximum copper tolerance for $Hsh155^{WT}$.

Rds3 impacts BS usage in yeast

We next explored whether contacts observed for Rds3 are important for pre-mRNA splicing in yeast. In the B^{act} spliceosome, Rds3 contacts Hsh155 at both the N- and C-terminal regions of the protein and the U2/BS duplex(1, 2). We chose four mutations in Rds3, specifically R38A, R38C, N95A, and R99A, and tested their impact. N95 and R99 are both located in a C-terminal alpha helix that interacts with the U2/BS duplex whereas R38 contacts Hsh155 (**Figure 6.3A and B**). R38 is conserved in the human homolog PHF5a. Mutation of this position to cysteine is implicated in solid state cancers and so we tested R38C in addition to the alanine mutant in yeast(6, 7). All four mutants showed no effect with the WT, U254G, A255G, and A258G BS reporters (**Figure 6.3C**). Surprisingly, all effects observed for these Rds3 mutants were relatively mild and only R99A showed defects with all BS reporters tested. R38A, R38C and N95A did however impact splicing at A258C and A258U. Interestingly, we also observed a defect in splicing the gAG 3′SS reporter for the R38C and N95D mutants. This change was not observed with mutations in Hsh155. These data demonstrate that Rds3 mutations can impact splicing of nonconsensus reporters.

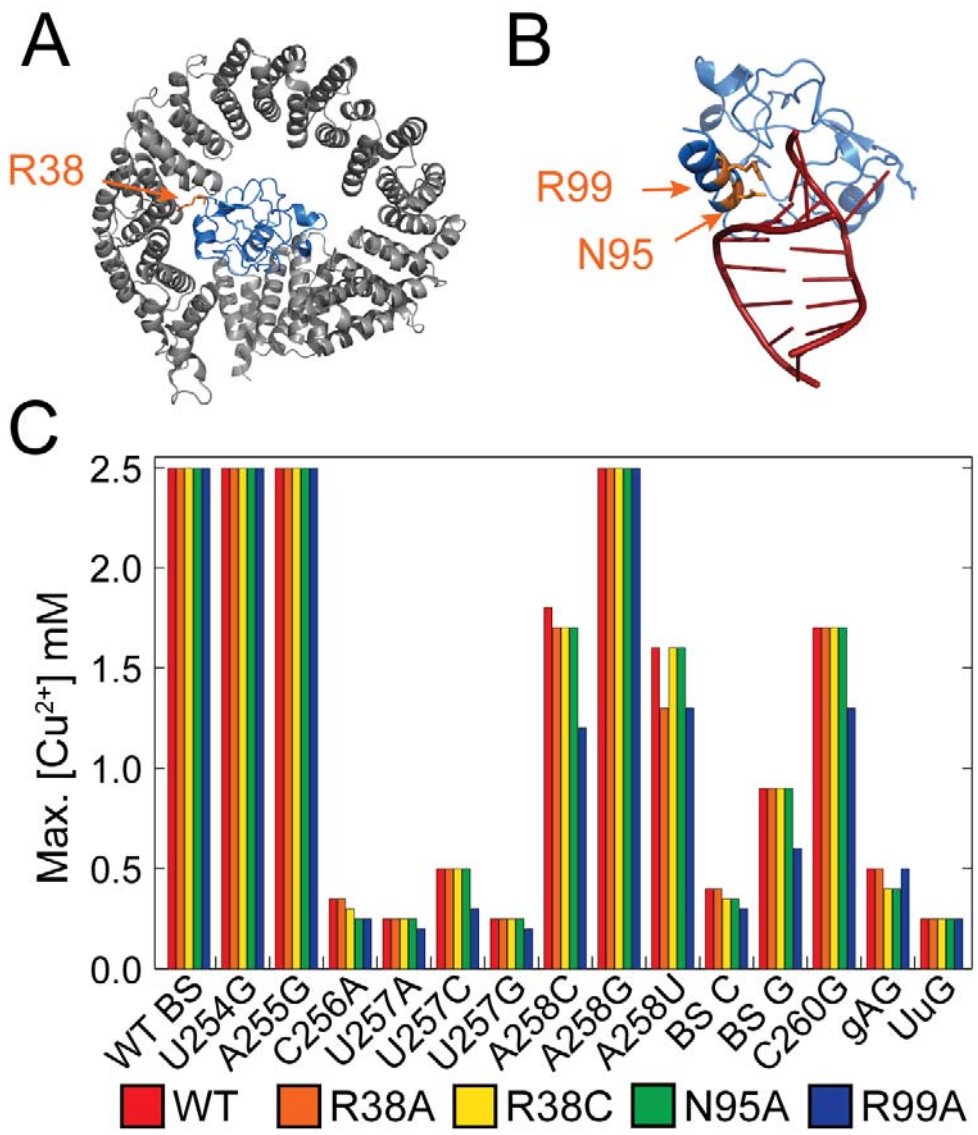


Figure 6.3: *ACT1-CUP1* data for Rds3 mutants. (A) Structure of Rds3 from the B^{act} spliceosome (PDB: 5GM6) with Hsh155 in grey and Rds3 in blue. R38 contacts Hsh155 and is highlighted in orange. (B) The U2/BS duplex is shown and N95 and R99 are colored orange. (C) Rds3 mutants display a phenotype by copper tolerance when assayed using BS and 3′ SS mutant reporters.

The HEAT domain of yeast and human Hsh155/SF3b1 are not identical

We were intrigued by the possibility of testing splicing inhibitors in yeast extracts prepared from MDS strains to determine whether MDS mutants were more susceptible to inhibition by pladienolide B (PB). To this end, we first asked whether PB would block splicing in yeast extracts. PB concentrations below 200 nM are sufficient to block splicing in vitro in human nuclear extract(11). PB concentrations ranging from 100 nM to 20 μM failed to inactivate splicing regardless of whether or not the compound was incubated with the extract prior to the addition of pre-mRNA (**Figure 6.4**). Consistent with this observation is data from a collaborator at H3 Biomedicine that shows that [^{3}H]-PB does not bind Hsh155 (data not shown). These data show that yeast Hsh155 is resistant to inhibition by PB. However, we also tested a mutation homologous to the R1074H mutation that grants resistance to PB in yeast (R743H). We find that mutation of this residue greatly improves nonconsensus intron splicing (**Figure 6.5**) suggesting that while the two positions are not identical, they may function similarly.

The HEAT domain of Hsh155 is extremely well conserved and some individual HEAT repeats are more than 70% identical between yeast and human (**Figure 6.6A**). The lack of efficacy for PB in yeast extracts is therefore surprising. Indeed, the R1074H mutation in SF3b1 that grants resistance to PB is found in repeat 15, which is one of the most conserved repeats(15). We sought to determine the point of divergence for the two organisms by generating chimeric proteins derived from the yeast and human sequences. We began by replacing HEAT repeats 1-5 (H1-5), repeats 1-12 (H1-12) and repeats 1-16 (H1-16) with the human sequence. The N-terminal extension of Hsh155 was not altered in these constructs. H1-5 and H1-12 grow equally well as WT, but further exchange of repeats in H1-16 is less well tolerated (**Figure 6.6B**). This demonstrates that the region implicated in MDS (H4-9) likely functions the same in yeast as in human.

We next exchanged sequence internal to the protein starting with HEAT repeat 5. Consistent with the N-terminal chimera data, H5-12 was well tolerated and, surprisingly, mutation

of region 5-16 was also tolerated (**Figure 6.6C**). Further exchange to repeat 20 (H5-20) was lethal. We also tested H12-20 in yeast because of the difference observed for H1-16 and H5-16 and found it to also be lethal. This data shows that exchange of the C-terminal HEAT repeats of the protein is not tolerated whereas the N-terminal region is functionally interchangeable. Interestingly, the C-terminal HEATs are extremely well conserved and are, on average, more conserved than the N-terminus HEATs (**Figure 6.6A**).

We evaluated the structure of the SF3b complex in the B^{act} spliceosome in an attempt to determine why exchange of the C-terminus is lethal. Hsh155 makes contact with all but one component of the SF3b complex (Hsh49, not shown) (**Figure 6.7A**). We further mapped contacts to individual HEAT repeats (**Figure 6.7B**). As one would expect, repeats that contact multiple other factors in the spliceosome are also the most well conserved. However, there is no obvious connection between the chimera data and the contacts made by Hsh155. Given that replacing HEAT repeats is only poorly tolerated if a large number are exchanged, it is possible that the deciding factor may be the angle of the superhelix formed by the HEAT domain(1, 8). This has been shown to differ between the human and yeast structures.

We also performed a series of truncations in yeast Hsh155 outside of the HEAT domain to determine the regions essential for viability (**Figure 6.8**). We truncated the N-terminus at four positions and the C-terminus at two positions. Loss of the first 63 or 95 amino acids has no effect on viability whereas Hsh155 without the first 156 amino acids cannot support growth. Surprisingly, further truncation of Hsh155 to remove the first two HEAT repeats of Hsh155 (position 236) restored growth on FOA. This suggests that the N-terminal repeats are not essential and may rationalize the lack of an observed effect with mutation of N-terminal residues that contact the U2/BS duplex (**Figure 6.2A**). Truncation of the C-terminal extension of Hsh155 is lethal, as is loss of HEAT 20 (**Figure 6.8**), further cementing the idea that the C-terminal region is essential for Hsh155 function.

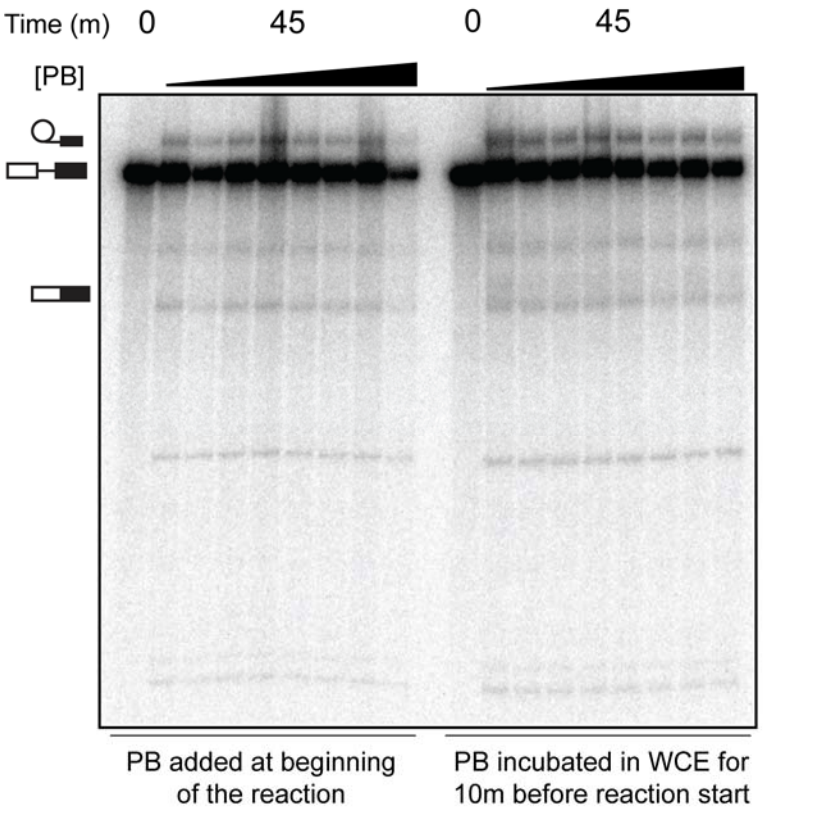


Figure 6.4: Pladienolide B (PB) has no effect on *in vitro* splicing in yeast extract. Prior incubation of PB with extract or simultaneous addition of PB with splicing substrate has no effect on the amount of mRNA formed.

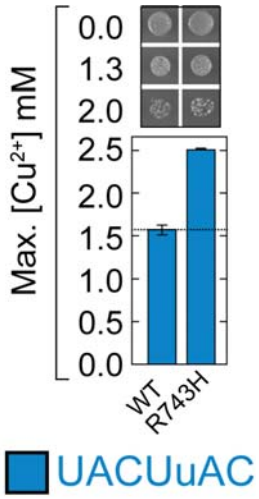


Figure 6.5: The Hsh155 mutation R743H improves nonconcensus BS usage in yeast.

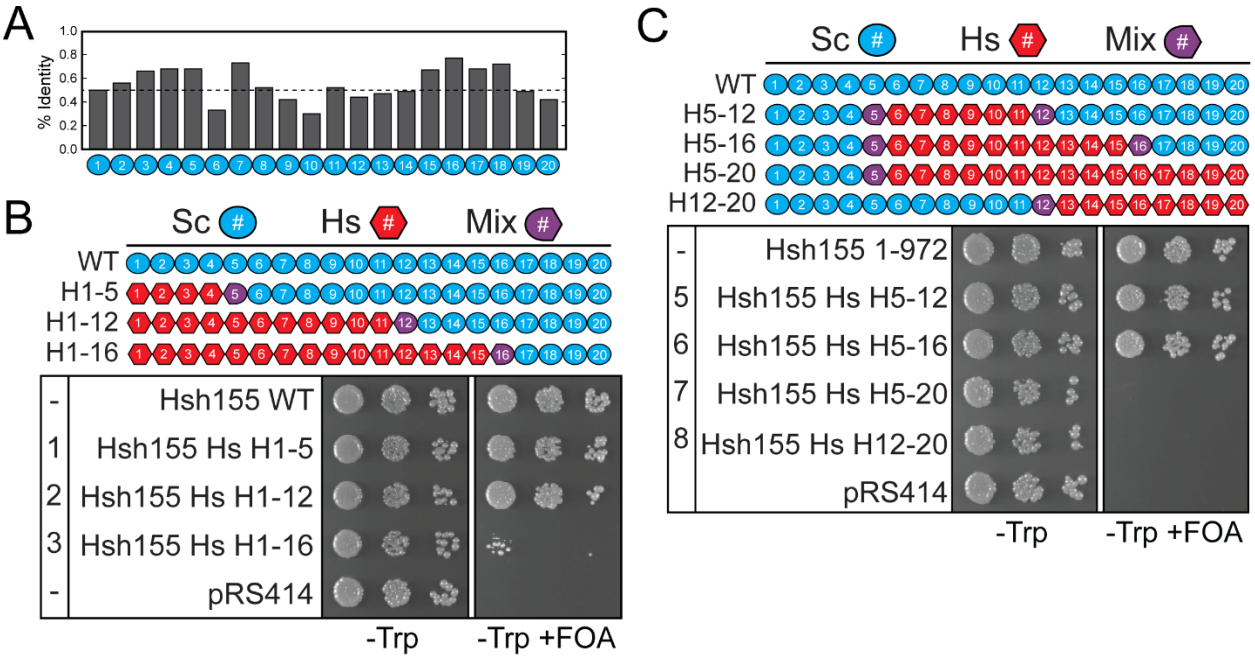


Figure 6.6: Chimeric Hsh155/SF3b1 tested in yeast. (A) Graph showing the conservation of individual HEAT repeats. The dotted line indicates 50% sequence identity. The boundaries of HEAT repeats were determined from the structure of the yeast B^{act} spliceosome and the human SF3b complex (pdb: 5GM6 and 5IFE). The region implicated in MDS (H4-9) and the regions that bind the U2/BS duplex are most well conserved (H15-20). (B) The sequence of Hsh155 was exchanged for SF3b1. HEAT repeats 1-12 can be exchanged without noticeable effects on viability. Exchange of an additional 4 repeats is not well tolerated. (C) Exchange of the middle and C-terminal HEAT repeats. Exchanges of repeats 5-12 and 5-16 are tolerated but not repeats 5-20 or 12-20.

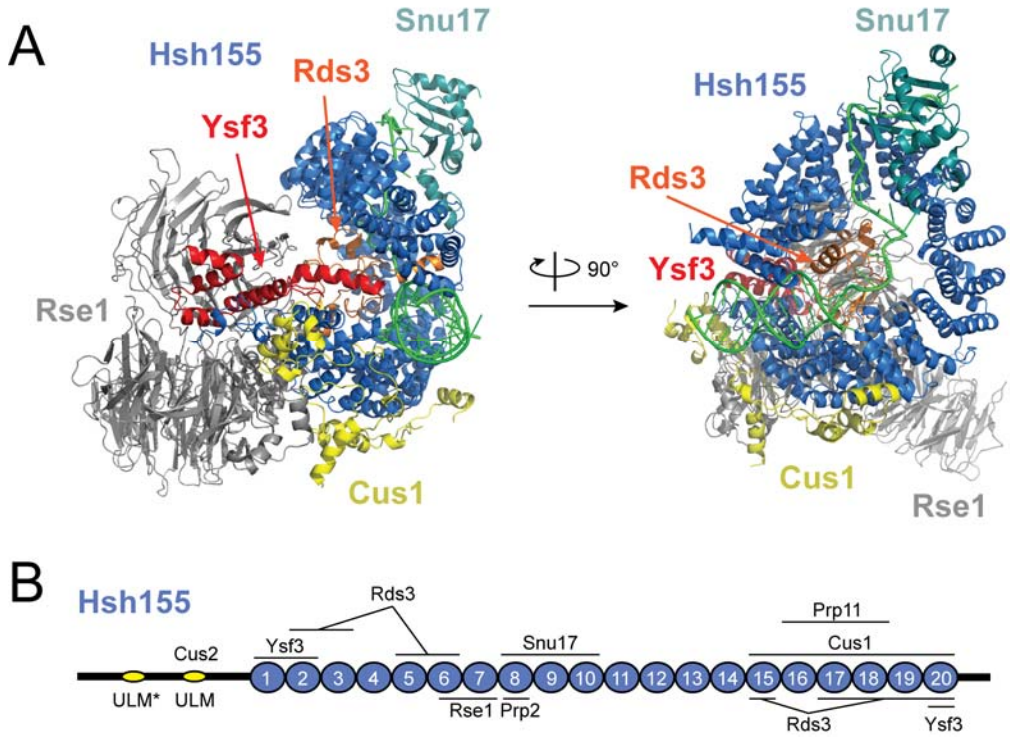


Figure 6.7: Overview of protein contacts with Hsh155 from the B^{act} spliceosome (PDB: 5GM6). (A) Structure of the SF3b complex containing Hsh155, Ysf3, Rds3, Snu17, Rse1 and Cus1 in two orientations. (B) Schematic showing individual contacts between Hsh155 and other components of the SF3b.

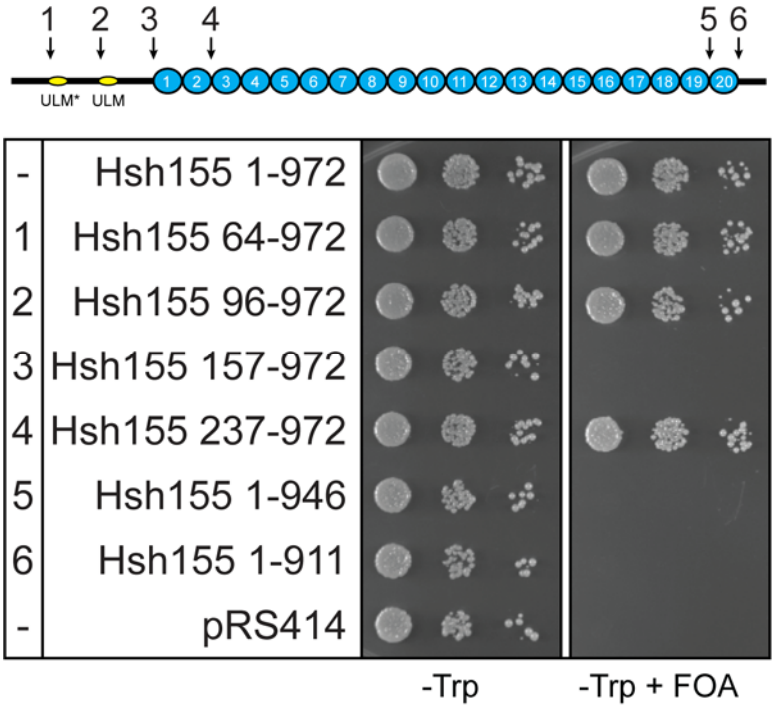


Figure 6.8: Truncation analysis reveals the essential regions of Hsh155. Numbers correspond with truncation locations in the schematic at top.

Conclusions and Perspective

The work presented here was a major effort to understand the function of Hsh155, its contacts with other components of the SF3b complex, and the differences that exist between the yeast and human proteins. Multiple avenues to continue research exist. For example, a complete examination of individual interactions with the U2/BS duplex could provide interesting insight to how specific regions contribute to BS usage. It would also be interesting to probe how other components of the SF3b complex contribute to splicing. Ysf3 would be particularly interesting, as it likely helps to close the Hsh155 ring. Release of the U2/BS duplex may require loss of Ysf3 to destabilize Hsh155. Testing whether extracts prepared from chimeric Hsh155 are susceptible to inhibition by PB would also be incredibly interesting and would be of great interest to the medical community. These chimeric constructs may reveal the location of the PB binding site on SF3b1/Hsh155. Finally, exploration of essential HEAT repeats would be intriguing. It would be interesting to see if the entire region implicated in binding downstream intronic RNA (HEATs 4-9) is dispensable for growth. This data would have provide significant insight into the function of Hsh155/SF3b1 and diseases linked to these proteins.

REFERENCES

1. Yan,C., Wan,R., Bai,R., Huang,G. and Shi,Y. (2016) Structure of a yeast activated spliceosome at 3.5 A resolution. *Science*, **353**, 904–911.

2. Rauhut,R., Fabrizio,P., Dybkov,O. and Hartmuth,K *et al*. (2016) Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. *Science*, **353**, 1399-1405.

3. Carrocci,T.J., Zoerner,D.M. and Paulson,J.C. *et al* (2017) SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. Nucleic Acids Res. Published online January 6. 2017. http://dx.doi.org/10.1093/nar/gkw1349.

4. Tang,Q., Rodriguez-Santiago,S. and Wang,J. *et al.* (2016) SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. Genes. Dev. **30**, 2710–2723.

5. Wang,Q. and Rymond,B.C. (2003) Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. *Mol. Cell. Biol.*, **23**, 7339–7349.

6. Falck,E. and Klinga-Levan,K. (2013) Expression patterns of Phf5a/PHF5A and Gja1/GJA1 in rat and human endometrial cancer. *Cancer Cell Int*. **13**, 13-43.

7. Hubert,C.G., Bradley,R.K., Ding,Y. and Toledo,C.M. et al.(2013) Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev*. **27**, 1032–1045.

8. Cretu,C., Schmitzova,J., Ponce-Salvatierra,A., Dybkov,O., DeLaurentiis,E.I., Sharma,K., Will,C.L., Urlaub,H., Lührmann,R. and Pena,V. (2016) Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell.*, **64**, 307–319.

9. Kielkopf,C.L. (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev.*, **18**, 1513–1526.

10. Thickman,K.R., Swenson,M.C., Kabogo,J.M., Gryczynski,Z. and Kielkopf,C.L. (2006) Multiple U2AF65 Binding Sites within SF3b155: Thermodynamic and Spectroscopic Characterization of Protein–Protein Interactions among pre-mRNA Splicing Factors. *J. Mol. Biol.*, **356**, 664–683.

11. Effenberger,K.A., Urabe,V.K., Prichard,B.E., Ghosh,A.K. and Jurica,M.S. (2016) Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. *RNA*, **22**, 350–359.

12. Hasegawa,M., Miura,T., Kuzuya,K., Inoue,A., Won Ki,S., Horinouchi,S., Yoshida,T., Kunoh,T., Koseki,K., Mino,K., *et al.* (2011) Identification of SAP155 as the Target of GEX1A (Herboxidiene), an Antitumor Natural Product. *ACS Chem. Biol.*, **6**, 229–233.

13. Kaida,D., Motoyoshi,H., Tashiro,E., Nojima,T., Hagiwara,M., Ishigami,K., Watanabe,H., Kitahara,T., Yoshida,T., Nakajima,H., *et al.* (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nature Chemical Biology*, **3**, 576–583.

14. Kotake,Y., Sagane,K., Owa,T., Mimori-Kiyosue,Y., Shimizu,H., Uesugi,M., Ishihama,Y.,

Iwata,M. and Mizui,Y. (2007) Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nature Chemical Biology*, **3**, 570–575.

15. Yokoi,A., Kotake,Y., Takahashi,K. and Kadowaki,T. (2011) Biological validation that SF3b is a target of the antitumor macrolide pladienolide. *FEBS J.*, **278**, 4870–4880.

16. Wang,Q., He,J., Lynn,B. and Rymond,B.C. (2005) Interactions of the Yeast SF3b Splicing Factor. *Molecular and Cellular Biology*, **25**, 10745–10754.

17. Vincent,K., Wang,Q., Jay,S., Hobbs,K. and Rymond,B.C. (2003) Genetic interactions with CLF1 identify additional pre-mRNA splicing factors and a link between activators of yeast vesicular transport and splicing. *Genetics*, **164**, 895–907.

18. Darman,R.B., Seiler,M., Agrawal,A.A., Lim,K.H., Peng,S., Aird,D., Bailey,S.L., Bhavsar,E.B., Chan,B., Colla,S., *et al.* (2015) Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep.*, **13**, 1033–1045.

19. Hoskins,A.A., Friedman,L.J., Gallagher,S.S., Crawford,D.J., Anderson,E.G., Wombacher,R., Ramirez,N., Cornish,V.W., Gelles,J. and Moore,M.J. (2011) Ordered and Dynamic Assembly of Single Spliceosomes. *Science*, **331**, 1289–1295.

20. Xu,Y.-Z. and Query,C.C. (2007) Competition between the ATPase Prp5 and Branch Region-U2 snRNA Pairing Modulates the Fidelity of Spliceosome Assembly. *Mol. Cell*, **28**, 838–849.

21. Query,C.C., Moore,M.J. and Sharp,P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev*, **8**, 587–597.

Chapter Seven

Single Molecule Investigation of Conformational Dynamics in the DEAD-box Helicase Prp5

Much of this work is done in collaboration with an undergraduate researcher, David H. Beier, and with the help of our lab manager, Joshua Paulson. I also thank Sandy Tretbar for preliminary work on the project and Margaret L. Rodgers for teaching me smFRET.

ABSTRACT

Prp5 plays a significant role in maintaining the fidelity of pre-mRNA splicing by influencing how the spliceosome utilizes nonconsensus branchsites. Multiple mechanisms for how Prp5 acts during splicing have been proposed, but knowledge of the ways in which Prp5 dynamics or conformational changes contribute to its function is unknown. We have used single molecule Förster resonance energy transfer to study the conformational dynamics of the Prp5 RecA domains in the isolated Prp5 helicase core and the full-length protein. In isolation, the helicase core behaves as a canonical DEAD-box protein and binds RNA in an ATP-dependent manner. However, the N- and C-terminal extensions of Prp5 abolish ATP- and RNA-dependent conformational changes observed by smFRET, suggesting these regions may help to organize the helicase core in an orientation competent for RNA binding without ATP. Organization of the RecA domains by the extensions may contribute to ATP-independent RNA unwinding observed for Prp5.

INTRODUCTION

Eukaryotic precursor messenger RNAs (pre-mRNAs) are modified in the nucleus by the spliceosome to remove non-coding intragenic sequences (introns) in a process known as splicing (1). The spliceosome assembles stepwise on each intron from pre-assembled particles (the U1, U2, U4, U5, and U6 small nuclear ribonucleoproteins or snRNPs) to recognize the branchsite (BS) sequence used in catalysis and define the boundaries of the intron (the 5' and 3' splice sites; SS)(1-3). Accurate recognition of these sites is key to ensuring the integrity of the transcript. Members of the DEAD-box helicase family, including Prp5, Prp28, and Sub2, help to regulate splice site recognition and usage during spliceosome assembly and activation(4-8).

DEAD-box helicases play central roles in RNA metabolism and are directly involved in transcription, translation, pre-mRNA splicing, mRNA export, and RNA decay(9). All DEAD-box helicases are comprised of a conserved helicase core composed of two RecA-like domains and usually contain N- or C-terminal extensions that help modulate the activity of the core(10, 11). DEAD-box helicases are not processive enzymes, but rather can act in an ATP-dependent manner to destabilize RNA duplexes locally(12). Current models for RNA unwinding by DEAD-box helicases suggest this is accomplished through an ATP-dependent conformational cycle and that ATP hydrolysis is important for release after unwinding(13-15). Indeed, factors can modulate the affinity of DEAD-box helicases for RNA by blocking ATP hydrolysis. In the exon junction complex, the MAGOH-Y14 heterodimer keeps eIF4AIII in an ATP-bound, closed state to clamp it to the RNA(16, 17). Structural studies of several DEAD-box helicase cores indicate that the RecA-like domains are splayed apart and do not contact each other in the absence of ATP (open state)(18). Structural data suggests that the arrangement of the RecA domains relative to one another in the open state can be extremely heterogeneous. Binding of ATP and RNA triggers a conformational change to a closed complex (closed state) that forces the RNA to adopt a

conformation incompatible with duplex formation(11, 12). ATP hydrolysis returns the helicase core to an open state(14).

Prp5 interacts with the U2 snRNP during spliceosome assembly and plays both ATP-dependent and independent roles during pre-mRNA splicing(5, 6, 19-21). Prp5 has been shown to directly crosslink with the U2 snRNA and ATP hydrolysis by Prp5 is essential for viability in yeast unless another U2-related splicing factor, Cus2, is deleted or the snRNA is mutated (**Figure 7.1**)(19, 20, 22). Additional work using branch point-interacting stem loop (BSL) mutations in the U2 snRNA implicated Prp5 ATPase activity in destabilization of the BSL to allow for base pairing of the U2 snRNA to the intron and formation of a stable pre-spliceosome(23). The function of the ATP-independent mechanism of Prp5 is less well understood, but Prp5 remains essential even after bypass of the ATP-dependent processes(22).

Prp5 has also been shown to function as a fidelity factor that ensures proper base pairing between the U2 snRNA and the intron BS (5′-UACUAAC)(6). The current model for Prp5 proofreading suggests that it monitors base pairing and, in the presence of a mismatch, is retained on the spliceosome, blocking further assembly steps (**Figure 7.1**)(20). Additionally, mutations that are predicted to affect the helicase conformational cycle have been shown to affect the fidelity of BS usage, suggesting a link between Prp5 conformational dynamics and Prp5 activity(6). These mutations are found in conserved helicase motifs in Prp5, such as the SAT motif (**Figure 7.2A**). Mutation outside of the helicase core in the N-terminal DPLD motif of Prp5 has also been shown to affect fidelity by diminishing the interaction of Prp5 with the U2 snRNP SF3b complex, specifically Hsh155(21, 24, 25).

Crystallographic evidence from apo and ADP-bound proteins suggests that Prp5 interdomain interactions stabilize the open conformation in the absence of ATP and RNA(26). By analogy with other DEAD-box helicases such as Vasa, binding of ATP and RNA would trigger formation of a closed complex that would create an extended RNA binding surface along both

RecA domains (**Figure 7.2B**). Interestingly and unlike some other DEAD-box helicases, recombinantly expressed and purified full-length Prp5 displays RNA binding and unwinding activity in the absence of ATP(27). Therefore, the role ATP binding in driving conformational changes during spliceosome assembly is unclear.

Here, we use single molecule Forster Resonance Energy Transfer (smFRET) to analyze conformational dynamics in Prp5. My work shows that the Prp5 helicase core adopts the open conformation and only transitions to the closed state upon binding of ATP and RNA. Interestingly, full-length Prp5 labeled in the same positions does not show a similar effect, suggesting that flanking domains of Prp5 may play a role in regulating Prp5 conformation. To further dissect this, we analyzed the impact truncation of the N- and C-terminal regions on viability in yeast.

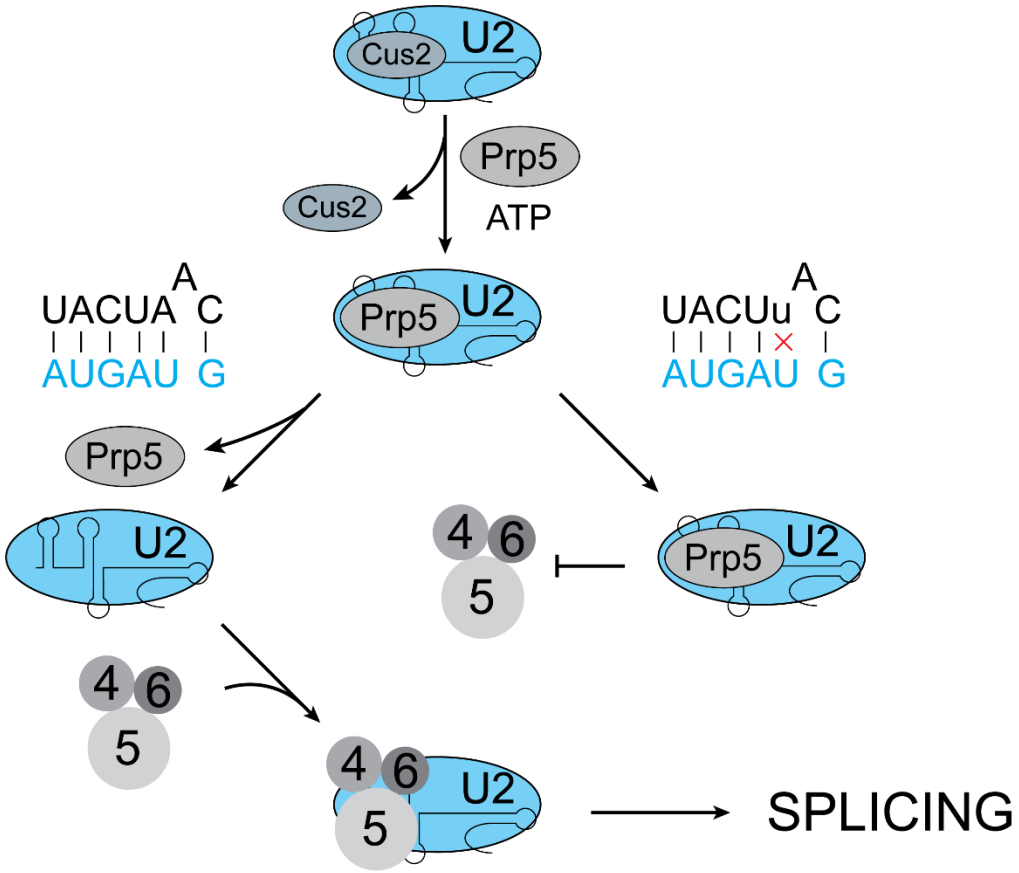


Figure 7.1: Schematic of Prp5 function during splicing. ATP hydrolysis by Prp5 is implicated in Cus2 displacement from the U2 snRNP. After displacement, Prp5 remains bound to U2, which can associate with the E complex spliceosome to form the pre-spliceosome (A complex). Consensus base pairing between the U2 snRNA and the intron BS results in ejection of Prp5 from the spliceosome. The U4/U6.U5 tri-snRNP then joins the pre-spliceosome to form B complex. Alternatively, if Prp5 senses mispairing between U2 and the intron, then it remains bound to the pre-spliceosome and blocks further assembly and splicing.

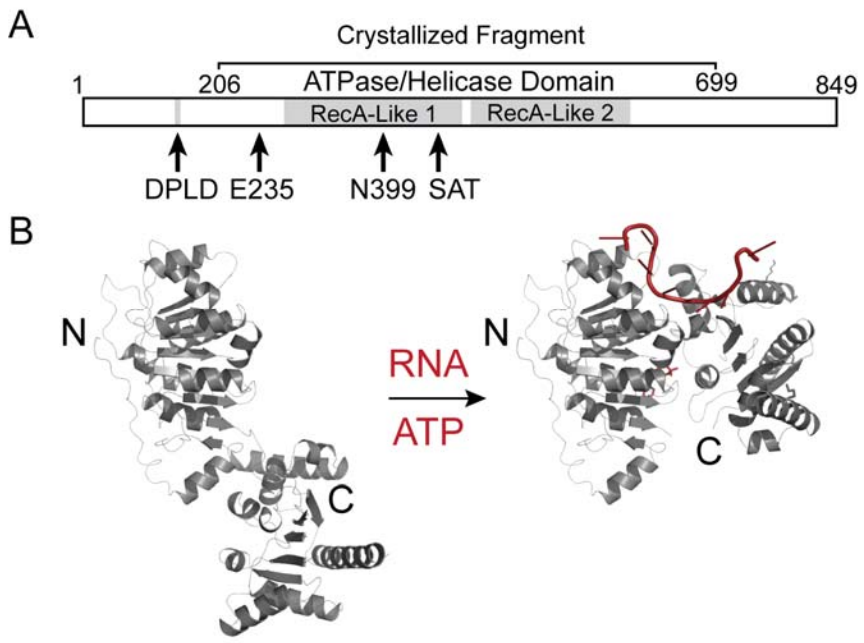


Figure 7.2: Schematic of Prp5 structure. (A) Prp5 contains two RecA domains and an N-terminal DPLD motif implicated in binding the SF3b components of the U2 snRNP. Positions in Prp5 that cause changes in splicing when mutated are indicated with arrows. E235A impairs splicing of nonconsensus introns, whereas N399D and SAT→TAG improve splicing of nonconsensus introns. (B) Prp5 is thought to undergo a conformational change upon binding of RNA and ATP. The apo complex is denoted “open” (left) and the RNA/ATP-bound complex is denoted “closed” (right). The open structure (left) represents the structure determined by X-ray crystallography (pdb 4LJY) while the closed structure (right) is believed to be necessary for ATP hydrolysis and was modeled based on structures of other DEAD-box proteins (coordinates for the closed structure were obtained from Yong-Zhen Xu and Charles Query)

MATERIALS AND METHODS

Cloning, Expression, Purification and Labeling of Prp5

Prp5$^{206-699}$ was expressed from a modified pET28b vector that includes an N-terminal biotin acceptor peptide (GLNDIFEAQKIEWHE) downstream of the TEV protease cleavage site(28). Endogenous cysteines were removed by site-directed mutagenesis using the inverse-PCR method to yield the following mutations: C250V, C364V, C365V, C442T, C486V, C538A, C553M, and C581A. Full-length cysteine-less Prp5 also contains the C785A mutation.

Protein was expressed in BL21(DE3) *E. coli* also expressing exogenous GST-BirA to ensure biotinylation of Prp5(29). After the cultures reached OD_{600} = 0.6-0.8, they were chilled on ice for 30 min and protein was subsequently induced by addition of 0.5 mM IPTG and 20 μM d-biotin. After growth overnight at 16°C, cells were harvested and the protein was purified as previously described with slight modifications(26). Briefly, cells were resuspended in lysis buffer (50 mM bis-tris pH 6.5, 500 mM NaCl, 10 mM imidazole, 0.2 mg/ml lysozyme), incubated at RT for 1 hr and then sonicated. The lysate was clarified by centrifugation, filtered through a 0.45 um PES filter and Prp5 was purified by nickel affinity chromatography using 500 mM imidazole for elution. The His tag was removed by addition of TEV protease while the protein was dialyzed overnight (20 mM Bis-Tris pH 6.5, 100 mM KCl, 10% glycerol, 1mM TCEP). The protein was further purified using a CM FF column (GE Healthcare) and a linear salt gradient to 1 M KCl. Ion exchange was performed without reducing agents to facilitate downstream maleimide labeling, as both DTT and TCEP have been shown to interfere with labeling. Peak fractions were pooled, concentrated and either immediately labeled or stored at -80°C until use. Full-length protein was expressed and purified in the same manner but with much more gradual elution gradients during ion exchange.

Fluorescent labeling of surface cysteines was achieved using 20-fold excess of each maleimide dye (Cy3 and Cy5; GE Healthcare) relative to Prp5 (10 μM) in labeling buffer (20 mM

Bis-Tris pH 6.5, 50 mM KCl, 10% glycerol)(30). After overnight incubation at 4°C, free dye was removed by ion exchange chromatography using a CM FF column. Labeled Prp5 was frozen and stored at -80C until use.

Single Molecule Assays

smFRET assays were performed at room temperature on a PEG/PEG-biotin passivated quartz slide(31). Prp5 was immobilized by biotin-streptavidin linkage and experiments were carried out in conditions that mimicked *in vitro* splicing (100 mM KPi pH 7.3, 3% PEG 8000, 2.5 mM $MgCl_2$, 10 mM HEPES pH 7.9, 20 mM KCl) and used the protocatechuic acid/protocatechaute dioxygenase oxygen scavenger system(32, 33). The triplet state quenchers trolox (1 mM), cyclooctatetraene (1 mM), 4-nitrobenzyl alcohol (1 mM) and N-propyl gallate (0.5 mM) were also added. Single-molecule data was collected on a home-built prism-based total internal reflection microscope using 532 and 640 nm for excitation of the donor and acceptor fluorophores, respectively. Emission was passed through a DV2 dual view apparatus (Photometrics) to spatially separate Cy3 and Cy5 signal and images were recorded using a cooled electron multiplying-charged coupled device camera (Andor). Data was collected continuously at a rate of 10 frames/s (100ms exposure) and images were recorded using Metamorph software (Molecular Devices).

Raw micrographs were analyzed using custom Matlab (Mathworks) software. Fluorescence intensities were background corrected for both the donor (I_{Cy3}) and acceptor (I_{Cy5}) fluorophores and apparent FRET efficiencies were calculated using: $E_{FRET} = I_{Cy5}/(I_{Cy3}+I_{Cy3})$(31). Hidden Markov modeling was performed using HaMMy software and data was also analyzed using vbFRET(34, 35).

Preparation of RNAs

An RNA consisting of U2 nucleotides 1-90 was purchased from IDT and used without further purification. U6 snRNA was a gift of S. Butcher and A. Didychuk.

Yeast Strains and Copper Assays

The copper sensitive Prp5 shuffle strain was a gift of Charles Query and has been previously described(6). The yeast plasmid was derived from pRS413 and contained the genomic copy of PRP5 +/- 500 base pairs. Copper assays were carried out as previously described and plates were scored after 3 days growth at 30°C(25). Truncation experiments were performed in the copper sensitive strain or a derivative wherein CUS2 had been replaced by homologous recombination with an hphMX cassette. Yeast were spotted on synthetic media lacking histidine and containing 5-FOA and grown for three days at 30°C before imaging.

Table 7.1. Yeast strains

Name	Genotype	Description
yAAH0054	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS416-PRP5*	Derived from 46α (kind gift of Charles Query). Copper sensitive strain used to generate Prp5 mutant strains for ACT1-CUP1 by plasmid shuffle.
yTJC0514	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS413 PRP5*	Copper sensitive strain expressing Prp5 WT. Transformed with ACT1CUP1 reporters.
yTJC0515	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS413 PRP5^{cys}*	Copper sensitive strain expressing mutant Prp5 lacking cysteines. Transformed with ACT1CUP1 reporters.
yTJC0526	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS413 PRP5^{cys} V268C K666C*	Copper sensitive strain expressing mutant Prp5 lacking cysteines. Transformed with ACT1CUP1 reporters.
yTJC0516	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS413 PRP5^{cys} T430C K544C*	Copper sensitive strain expressing mutant Prp5 lacking cysteines. Transformed with ACT1CUP1 reporters.
yTJC0523	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS413 PRP5^{cys} E235A*	Copper sensitive strain expressing mutant Prp5 lacking cysteines. Transformed with ACT1CUP1 reporters.
yTJC0525	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS413 PRP5^{cys} V268C K666C E235A*	Copper sensitive strain expressing mutant Prp5 lacking cysteines. Transformed with ACT1CUP1 reporters.
yTJC0526	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS413 PRP5^{cys} TAG*	Copper sensitive strain expressing mutant Prp5 lacking cysteines. Transformed with ACT1CUP1 reporters.
yTJC0528	*MAT α cup1Δ ura3 his3 trp1 lys2 ade2 leu2 prp5::loxP pRS413 PRP5^{cys} V268C K666C TAG*	Copper sensitive strain expressing mutant Prp5 lacking cysteines. Transformed with ACT1CUP1 reporters.

Table 7.2. Plasmids

#	Plasmid Name	Plasmid Description
pAAH0692	pRS413 Prp5 WT HIS3	Plasmid used to generate Prp5^{WT} strain. Hsh155 coding region +/- 250bp was cloned into pRS413 using BamHII/NotI sites.
pAAH0947	pRS413 Prp5 NΔ11 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0948	pRS413 Prp5 NΔ40 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0949	pRS413 Prp5 NΔ68 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0950	pRS413 Prp5 NΔ86 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0951	pRS413 Prp5 NΔ206 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0952	pRS413 Prp5 CΔ656 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0953	pRS413 Prp5 CΔ700 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0954	pRS413 Prp5 CΔ750 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0955	pRS413 Prp5 CΔ800 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0956	pRS413 Prp5 R615A HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0713	pRS413 Prp5-3 HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0711	pRS413 Prp5 TAG HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0709	pRS413 Prp5 E235A HIS3	Plasmid used to generate mutant Prp5 strain
pAAH0848	pRS413 Prp5 8x	Plasmid used to generate mutant Prp5 strain. Prp5 C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, C785A
pAAH0887	pRS413 Prp5 8x E235A	Plasmid used to generate mutant Prp5 strain. Prp5 C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, C785A, E235A
pAAH0888	pRS413 Prp5 8x TAG	Plasmid used to generate mutant Prp5 strain. Prp5 C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, C785A, SAT to TAG
pAAH0848	pRS413 Prp5 8x14	Plasmid used to generate mutant Prp5 strain. Prp5 C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, C785A, V268C, K666C
pAAH0889	pRS413 Prp5 8x14 E235A	Plasmid used to generate mutant Prp5 strain. Prp5 C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, C785A, E235A V268C, K666C
pAAH0890	pRS413 Prp5 8x14 TAG	Plasmid used to generate mutant Prp5 strain. Prp5 C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, C785A, SAT to TAG V268C, K666C
pAAH0884	pRS413 Prp5 8x23	Plasmid used to generate mutant Prp5 strain. Prp5 C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, C785A, E235A, T430C, K544C
pAAH0781	pET28b TEV Prp5 206-699	Expression plasmid for Prp5 helicase core.
pAAH0782	pET28b TEV Prp5 C44S	Expression plasmid for Prp5 helicase core.
pAAH0783	pET28b TEV Prp5 C158S C159S	Expression plasmid for Prp5 helicase core.
pAAH0784	pET28b TEV Prp5 C236S	Expression plasmid for Prp5 helicase core.
pAAH0785	pET28b TEV Prp5 C280S	Expression plasmid for Prp5 helicase core.
pAAH0786	pET28b TEV Prp5 C332S	Expression plasmid for Prp5 helicase core.

pAAH0787	pET28b TEV Prp5 C347S	Expression plasmid for Prp5 helicase core.
pAAH0788	pET28b TEV Prp5 C375S	Expression plasmid for Prp5 helicase core.
pAAH0810	pET28b TEV Prp5 C158S C159V	Expression plasmid for Prp5 helicase core.
pAAH0811	pET28b TEV Prp5 C158V C159V	Expression plasmid for Prp5 helicase core.
pAAH0846	pET28b TEV Prp5 AP 7x14	Expression plasmid for Prp5 helicase core. C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, V268C, K666C
pAAH0899	pET28b TEV Prp5 AP 8x14 FL	Expression plasmid for Prp5 helicase core. C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, C785A, V268C, K666C
pAAH0902	pET28b TEV Prp5 AP 7x23	Expression plasmid for Prp5 helicase core. C250V, C364V, C365V, C442T, C486V, C538A, C553M, C581A, T430C, K544C

RESULTS AND DISCUSSION

Rationale of Prp5 smFRET Constructs

A prerequisite for smFRET experiments is the installation of fluorophores in known positions on the protein of interest. I employed maleimide labeling methods, which form adducts to surface exposed thiols and are relatively unreactive with other functional groups at slightly acidic pH(30). However, this approach was potentially complicated by the nine naturally occurring cysteines in Prp5, eight of which are found within the crystallized fragment of the Prp5 helicase core and flanking N- and C-terminal extensions (positions C250, C364, C365, C442, C486, C538, C553, C581 in the helicase core and C785 in the C-terminal extension). However, none of the endogenous surface-exposed cysteines are ideally located for smFRET studies. To this end, I began this project determining if endogenous cysteines would complicate labeling and by identifying regions in Prp5 that would make excellent positions for introducing a cysteine.

The truncation of Prp5 used for crystallography comprising the helicase core and N- and C-terminal extensions expresses considerably better than full-length Prp5 in *E. coli*(26). For this reason, we used the crystal construct to determine whether endogenous cysteines in Prp5 could be labeled and whether individual mutations are deleterious to recombinant expression. We first generated seven mutant Prp5 proteins containing single or double cysteine to serine mutations, expressed and purified these proteins, and performed test labeling using maleimide-Cy3. Mutation of the cysteine doublet at positions C364/C365 to serine resulted in totally insoluble protein, likely because C365 is oriented into a hydrophobic pocket. For this reason, we substituted valine at positions C364/C365, which restored solubility. Individual serine point mutants could be expressed for all other positions. We found that multiple positions are reactive with maleimide fluorophores, as removal of individual positions was insufficient to abolish labeling (**Figure 7.3**). Analysis of the Prp5 structure using the DUET server suggested multiple cysteine positions were

surface exposed in the open complex(36). However, since the server cannot predict the accessibility in other conformations, we choose to remove all endogenous cysteines.

We performed an alignment of *S. cerevisiae* Prp5 with *S. pombe* and *H. sapiens* homologs (Prp10 and DDX46; **Figure 7.4**). These organisms are distantly related to *S. cerevisiae* and we hoped to identify amino acid substitutions for the endogenous cysteines that would be tolerated by Prp5. This alignment revealed that only two of the nine of the cysteines found in Prp5 are conserved, suggested that most cysteines could be mutated without affecting the function of the protein. We replaced each of the nonconserved positions guided by the alignment to generate a construct with the following mutations: C365V, C442T, C486V, C538A, C553M, and C581A. The conserved cysteines are positions C250 and C364 in yeast. We previously identified C364V as tolerated and therefore made serine, alanine, and valine mutants of C250. Of these, C250V was chosen for further use. Prp5 containing those mutations was well-expressed in *E. coli* and viable in yeast (discussed in detail below).

After removing all native cysteines in Prp5, we reintroduced two site-specific cysteines into Prp5 for FRET studies. Malleable positions that could be modified to cysteine were selected based on divergence between the *S. cerevisiae* and the closely related *K. lactis* protein (Uniprot: Q6CKI1) as well as relative solvent accessibility and location in the crystal structure. Alignment of the two yeast homologs shows that they are highly conserved, particularly in the helicase domain (**Figure 7.5**). Four positions were chosen based on lack of conservation: V268, T430, K544, and K666 (**Figure 7.5**). V268 and T430 are located on opposite sides of the RecA1 domains whereas K544 and K666 are located in RecA2 (**Figure 7.6A**). Mutants with cysteine at these positions can be expressed and labeled with maleimide fluorophores. Labeling with both Cy3 and Cy5 simultaneously creates a mixed population of Cy3/Cy3, Cy3/Cy5, Cy5/Cy3, and Cy5/Cy5 labeled proteins (**Figure 7.6B**). As FRET only occurs on molecules labeled with both Cy3 and

Cy5, the labeling mixture can be resolved by selecting specific molecules during smFRET analysis.

We generated two different FRET constructs for Prp5: Prp5 V268C/K666C and Prp5 T430C/K544C. These constructs were designed to behave oppositely in a smFRET assay. The V268C/K666C construct should exhibit a high FRET state in the absence of ATP and RNA indicative of the open conformation of Prp5 (**Figure 7.6C**). Conversely, for the same conformation, the T430C/K544C construct should yield a low FRET state in the absence of ATP and RNA (**Figure 7.6D**). We hypothesize that binding of ATP and RNA would trigger a transition to the opposite FRET state, as the domains of Prp5 move relative to one another (**Figure 7.6C and D**). Variants of Prp5 lacking native cysteines and containing these FRET pairs were expressed, purified and labeled using a combination of Cy3- and Cy5-maleimide fluorophores (**Figure 7.7**).

Finally, in order to perform smFRET studies with the doubly-labeled Prp5, we needed to develop a method for surface immobilization. Our lab relies on the biotin-streptavidin linkage to tether molecules near the surface of a slide in order to facilitate imaging via TIRF. To this end, I modified the N-terminus of Prp5 to include a biotin acceptor peptide(28). This peptide sequence is recognized by *E. coli* biotin ligase, BirA, and can be used to biotinylate the protein construct *in vivo* during protein expression. Protein biotinylation was confirmed using a supershift assay that relies on the biotin-streptavidin interaction(29). Streptavidin binds biotin with very high affinity and will remain bound during electrophoresis on a denaturing SDS-PAGE gel. Addition of streptavidin to the labeled Prp5 V268C/K666C construct shows a change in mobility, demonstrating that all of the protein has been efficiently biotinylated (**Figure 7.8**).

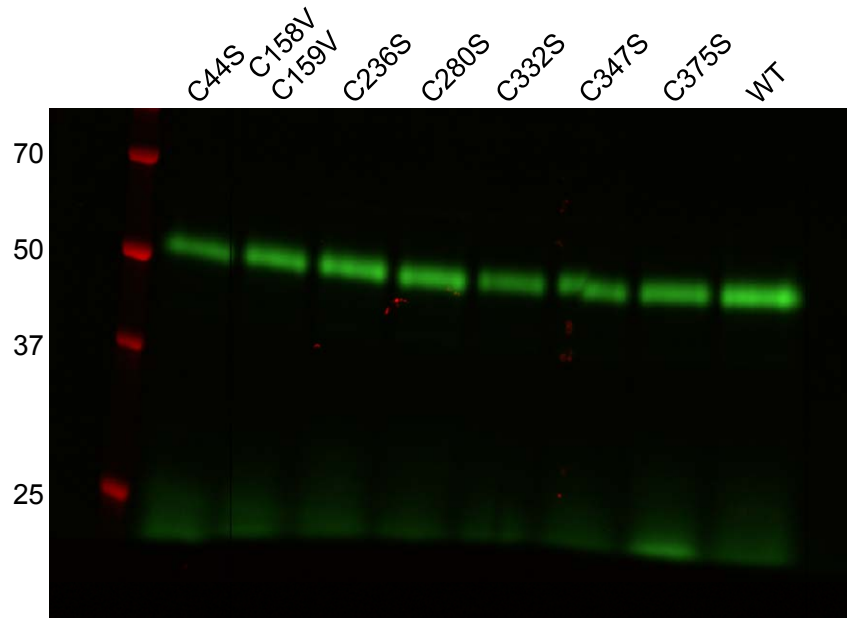


Figure 7.3: Test labeling of single cysteine point mutants in Prp5. The indicated Prp5 206-699 protein was incubated overnight with maleimide Cy3. The sample was then denatured and analyzed by SDS-PAGE. Cy3 fluorescence is observed from all mutant proteins (top green band), demonstrating that multiple cysteines are surface-exposed and can react with maleimide fluorophores. The lower green band is unincorporated maleimide Cy3.

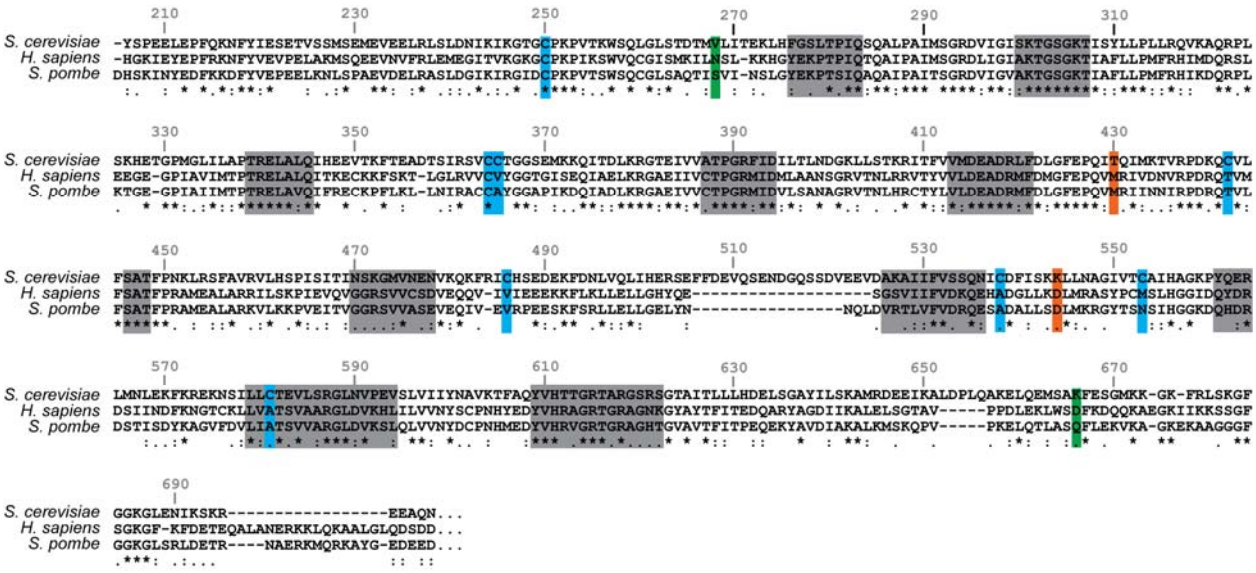


Figure 7.4: Alignment of *S. cerevisiae*, *S. pombe,* and *H. sapiens* Prp5 orthologs (Prp5, Prp11, and DDX46 respectively) generated using MUSCLE. Conserved motifs in DEAD-box helicases are highlighted in grey. Endogenous cysteines are highlighted in blue and nonconserved positions selected to become cysteine pairs are shown in orange and green. Mutations of endogenous cysteines were chosen based on the alignment (i.e., C583A). Conserved positions (C250, and C364) were mutated to valine, alanine and serine individually. Valine was chosen based on solubility of the construct when expressed in *E. coli*.

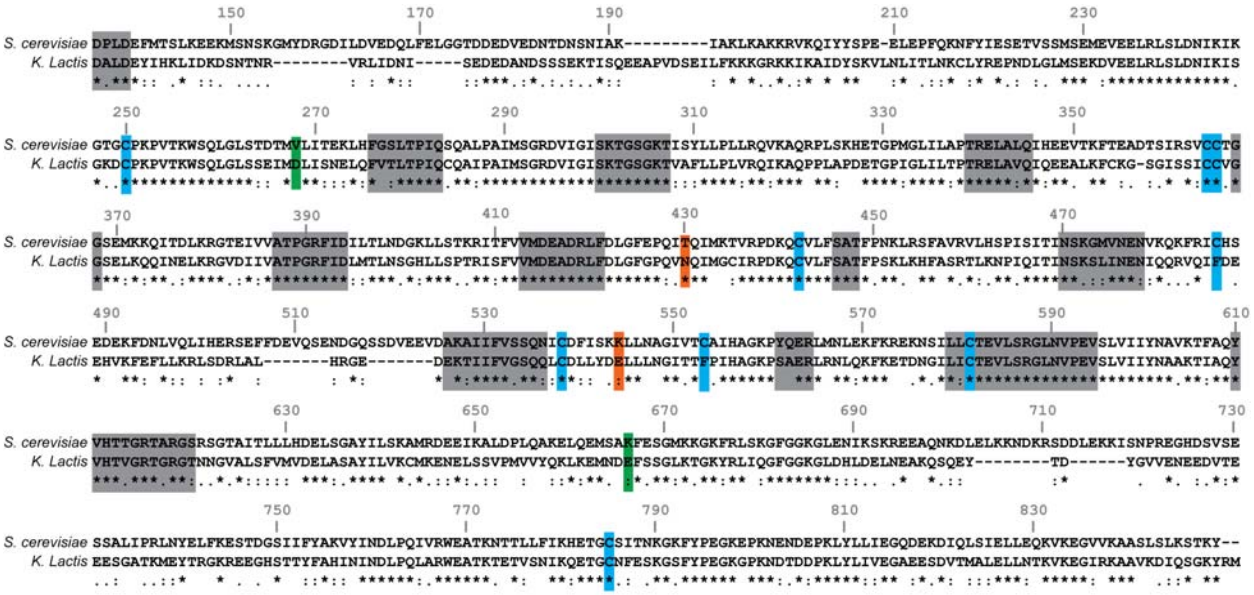


Figure 7.5: Alignment of *S. cerevisiae* and *K. lactis* Prp5. A protein sequence alignment of the *S. cerevisiae* and *K. lactis* orthologs was generated using Multiple Sequence Comparison by Log-Expectation (MUSCLE). The N-terminal region of Prp5 is variable in length and sequence and so the alignment starts from the conserved DPLD motif (amino acid 135 in *S. cerevisiae*). Conserved motifs in DEAD-box helicases and the DPLD motif are highlighted in grey. Endogenous cysteines are highlighted in blue and nonconserved positions selected to become cysteine pairs are shown in orange and green.

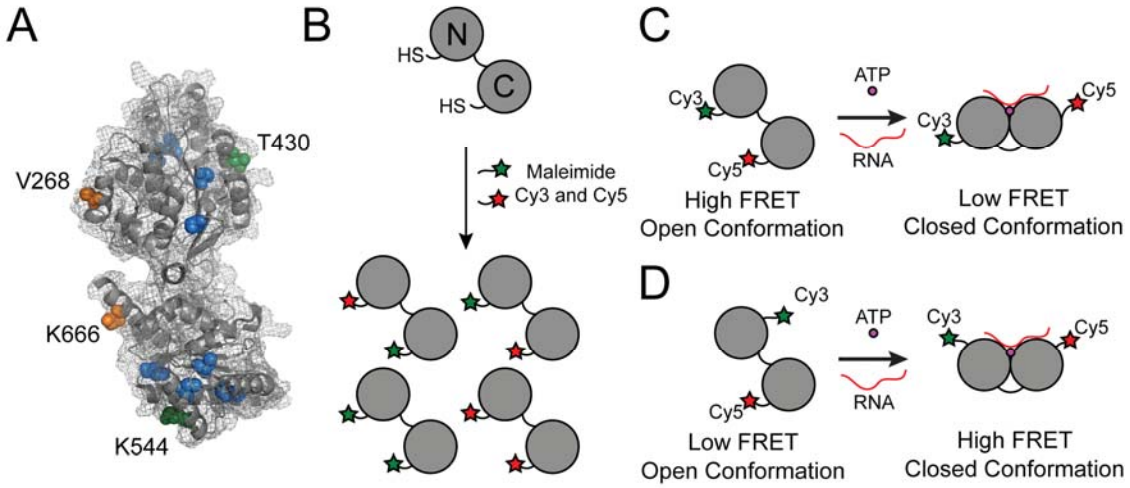


Figure 7.6: Schematic showing construct design for Prp5 smFRET experiments. (A) The crystal structure of $Prp5^{206-699}$ showing endogenous cysteine positions (blue), and positions chosen to become FRET pairs after labeling is depicted. Positions V268 and K666 (orange) give a high FRET state in the open conformation and positions T430 and K544 give a low FRET in the closed conformation based on distances measures in PyMol. (B) Maleimide fluorophores label surface-exposed cysteines in Prp5 to generate a mixture of labeled populations that can be resolved by single molecule methods. Molecules without both Cy3 and Cy5 can be identified and ignored during analysis. (C) Schematic showing the conformational and FRET efficiency changes for the labeled V268C/K666C double mutant. This construct is predicted to give a high FRET signal in the open conformation and a low FRET signal in the closed conformation. (D) Schematic showing the conformational and FRET efficiency changes for the labeled T430C/K544C double mutant. This construct is predicted to give a low FRET signal in the open confirmation and a high FRET signal in the closed conformation.

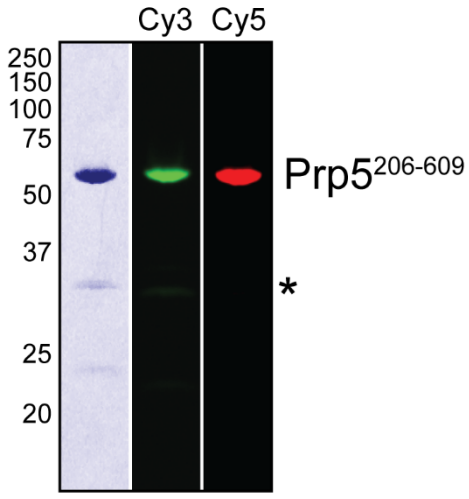


Figure 7.7: Purified Prp5 can be labeled with both Cy3 and Cy5 fluorophores. The asterisk (*) indicates an unknown minor contaminant.

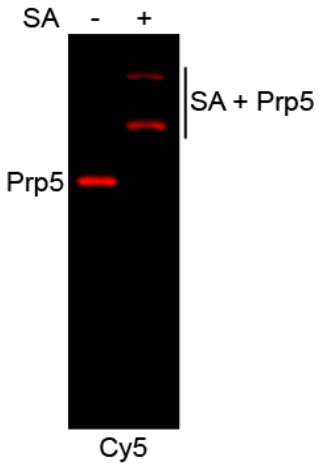


Figure 7.8: Gel shift assay demonstrating Prp5 biotinylation. Twenty-fold excess streptavidin was added to the labeled Prp5 construct and analyzed on a 9% SDS-PAGE gel without heat denaturation. Streptavidin binding to the biotinylated protein causes a shift in electrophoretic mobility indicating that the Prp5 construct is biotinylated.

Genetic Characterization of Prp5 smFRET Constructs

We tested the full-length Prp5 lacking cysteines for functionality *in vivo* using an *ACT1-CUP1* assay(6, 37). *ACT1-CUP1* reporter assays use copper tolerance as indirect measure of splicing. The *ACT1-CUP1* reporter pre-mRNA must be properly spliced in order to confer resistance or the yeast will die when plated on media containing copper (**Figure 7.9A**). Furthermore, the extent of mRNA accumulation correlates with copper tolerance. We generated strains in a copper sensitive background bearing mutations in Prp5 to remove all native cysteines as well as introduce the V268C/K666C and T430C/K544C pairs individually. We further mutated the Prp5 lacking cysteines (-Cys) and the V268C/K666C mutant to include the E235A and TAG mutations (**Figure 7.9A**)(6, 26). E235A abolishes interactions between the RecA domains in the open state, thereby impairing nonconsensus intron splicing, whereas TAG disrupts interactions in the closed state and improves splicing. Importantly, all variants of Prp5 complement loss of genomic *PRP5*. Additionally, mutant strains grew similarly well on copper-containing media when expressing reporters with consensus introns (**Figure 7.9B**). Mutations of the BS to include the U257C mutation (UACcAAC; mutation in lowercase) result in changes in yeast growth between Prp5 mutants (**Figure 7.9C**). Removal of cysteines results in improved growth, likely resulting from the C364 position. This region is near the DEAD-box GG motif and near the location of ATP binding. Despite this effect, mutation of the cysteine-less Prp5 to include E235A increases fidelity whereas TAG decreases fidelity. This is also true for the V268C/K666C and T430C/K544C mutants. Analysis of splicing of the A258U reporter (UACUuAC) by the cysteine-less mutants a showed similar outcome as U257C (**Figure 7.9D**). The impact of both E235A and TAG was diminished relative to wild-type protein with the same mutation. Together, these data show that Prp5 is still functioning in these backgrounds, albeit with slightly different activity.

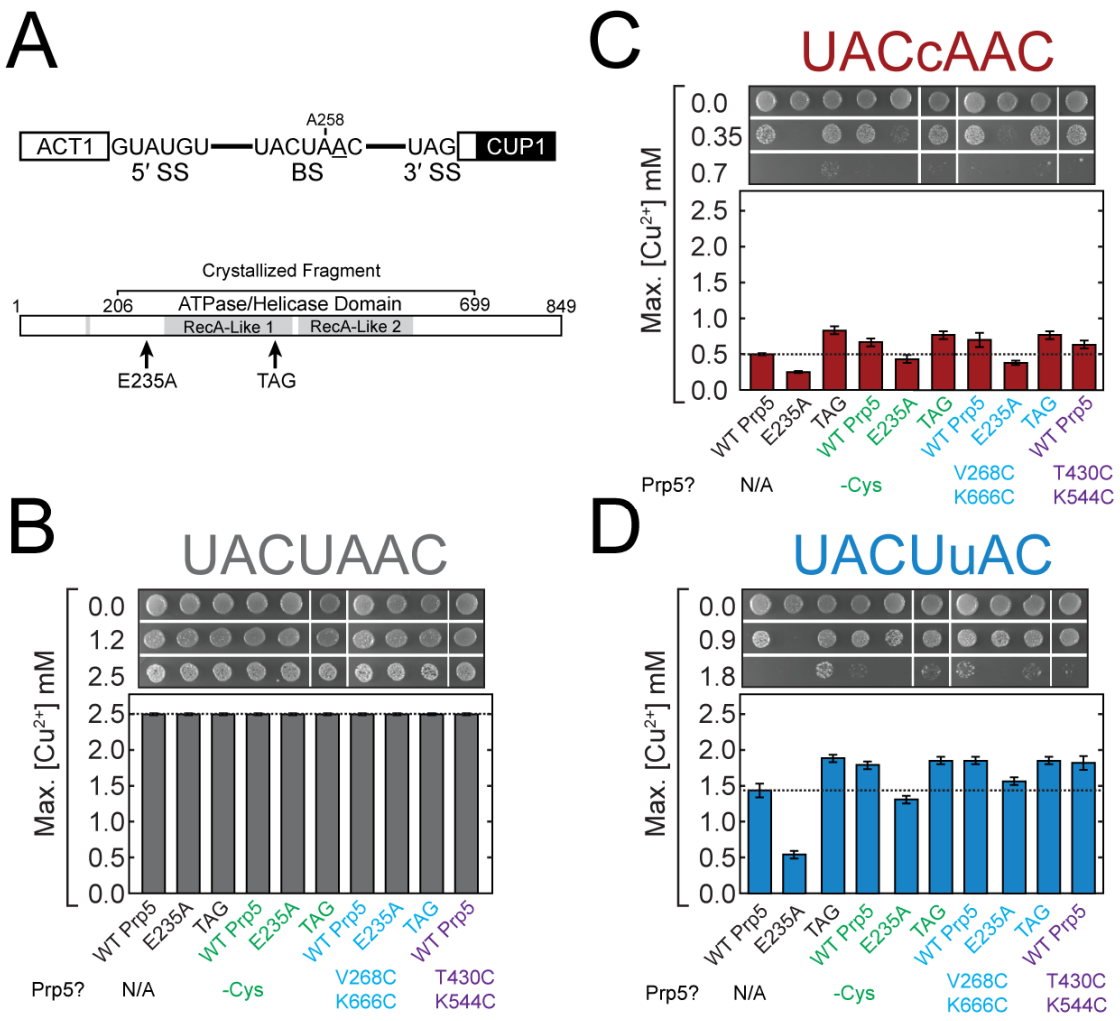


Figure 7.9: ACT1-CUP1 reporter assay using Prp5 mutants. (A) Cartoon schematic of ACT1-CUP1 (top) reporter and the Prp5 (bottom). Positions in Prp5 known to affect fidelity are indicated. E235A impairs nonconcensus branchsite usage and TAG improves usage. (B) Copper tolerance data for yeast expressing an ACT1-CUP1 reporter with a consensus intron. Prp5 background is shown by color: WT (black), -Cys (green), -Cys, V268C, K666C (light blue), -Cys T430C K544C (violet) (C) Copper tolerance data for the U257C reporter. BS sequence is shown at top. (D) Copper tolerance data for the A258U reporter. BS sequence is shown at the top.

Single Molecule Analysis of the Prp5 Helicase Core Conformational Dynamics

In my single molecule experiments, Prp5 is immobilized by the biotinylated acceptor peptide, and conformational changes between the RecA domains are detected as changes in FRET efficiency (**Figure 7.10A**). We first performed experiments using Prp5 V268C/K666C. Incubation of the labeled protein on the streptavidin-derivatized slide resulted in surface immobilized fluorescent particles (**Figure 7.10B**). Co-localization of a Cy5 signal with a Cy3 signal while both fluorophores were being directly excited revealed that ~10-20% of the surface tethered molecules were labeled with both fluorophores. Fluorescent particles were picked in the Cy5 (red) channel and translated to the Cy3 (green) channel using a mapping function. Integration of both fluorescent channels over the course of the movie (100s) revealed that in the absence of RNA and ATP, Prp5 is not dynamic within our time resolution and under conditions that mimic *in vitro* splicing (**Figure 7.10C**). Indeed, a FRET efficiency (E_{FRET}) histogram constructed from the calculated E_{FRET} of tens of thousands of observations shows a single peak centered at around 0.9 (**Figure 7.10D**). This result is consistent with the conformation of the Prp5 in the crystal structure in which the two positions labeled, V268C and K666C, are nearby each other and relatively fixed. This conformation is likely stabilized by interactions between the RecA domains.

We next determined the impact that ATP and RNA have on the conformation of Prp5. The addition of ATP alone did not have an effect on the high E_{FRET} state as can be seen in the single molecule time trajectories (**Figure 7.11A**). A histogram constructed from this data shows a similar distribution to what was observed for the protein alone (**Figure 7.10D vs. Figure 7.11B**). However, upon addition of both RNA and ATP, short-lived transitions to a low FRET state (0.6-0.7) were observed, suggesting that Prp5 becomes dynamic in the presence of ATP and RNA (**Figure 7.11C**). This observation is consistent with the formation of a transient closed complex wherein Prp5 has bound both ATP and RNA. Because of the short lifetime of the low FRET state, the E_{FRET} distribution appears similar to the distribution for protein alone (**Figure 7.11D**).

Monitoring conformational changes of the T430C/K544C construct in the presence of ATP and RNA showed a similar, but opposite, effect as the V268C/K666C construct. As predicted, the dominant state for the T430C/K544C construct is a low FRET signal centered around 0.35. Upon addition of RNA and ATP, it transitions to high FRET state around 0.8 (**Figure 7.12A and B**). Therefore, Prp5 appears to transition between open and closed states only in the presence of both ATP and RNA and under these conditions, the open state is preferred over the closed state.

We hypothesized that release of the ATP/RNA bound state (closed) back to the apo (open) state is connected to ATP hydrolysis. To confirm this, we monitored Prp5 V268C/K666C in the presence of RNA and a slowly hydrolyzed analog of ATP, ADPNP. In the presence of ADPNP and RNA, the low FRET state is observed with few dynamic transitions into the high FRET state (**Figure 7.13**). The corresponding E_{FRET} histogram exhibits a major peak centered at 0.65-0.7 indicative of a predominant low FRET state (closed conformation) in contrast to the predominant high FRET state observed in the presence of ATP. This confirms that ATP hydrolysis plays a role in the return to the open state. Taken together, these data support the hypothesis that the helicase domain of Prp5 functions as a somewhat typical DEAD-box helicase. Binding of ATP and RNA trigger the formation of a closed state wherein the orientation of the RecA domains change relative to one another.

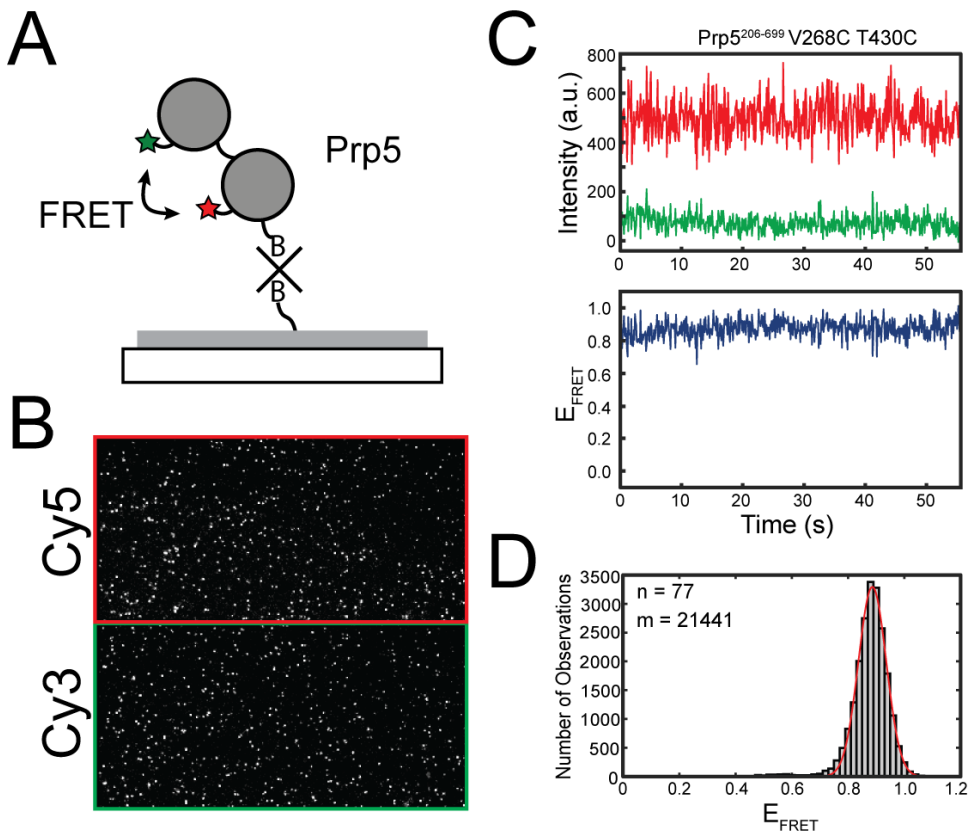


Figure 7.10: Single molecule FRET studies of immobilized Prp5. (A) Cartoon schematic showing Prp5 smFRET experiment. Prp5 is immobilized on a surface by a biotinylated N-terminal accepter peptide. (B) Micrograph showing single molecules of labeled Prp5 206-699 V268C/K666C (C) Cy3 and Cy5 trajectories for a single Prp5 in the absence of ATP and RNA (top) and calculated FRET efficiencies (bottom). (D) Histogram of FRET efficiencies of Prp5 in the absence of ATP and RNA. n represents the number of molecules used in the construction of the histogram and m is the number of observations.

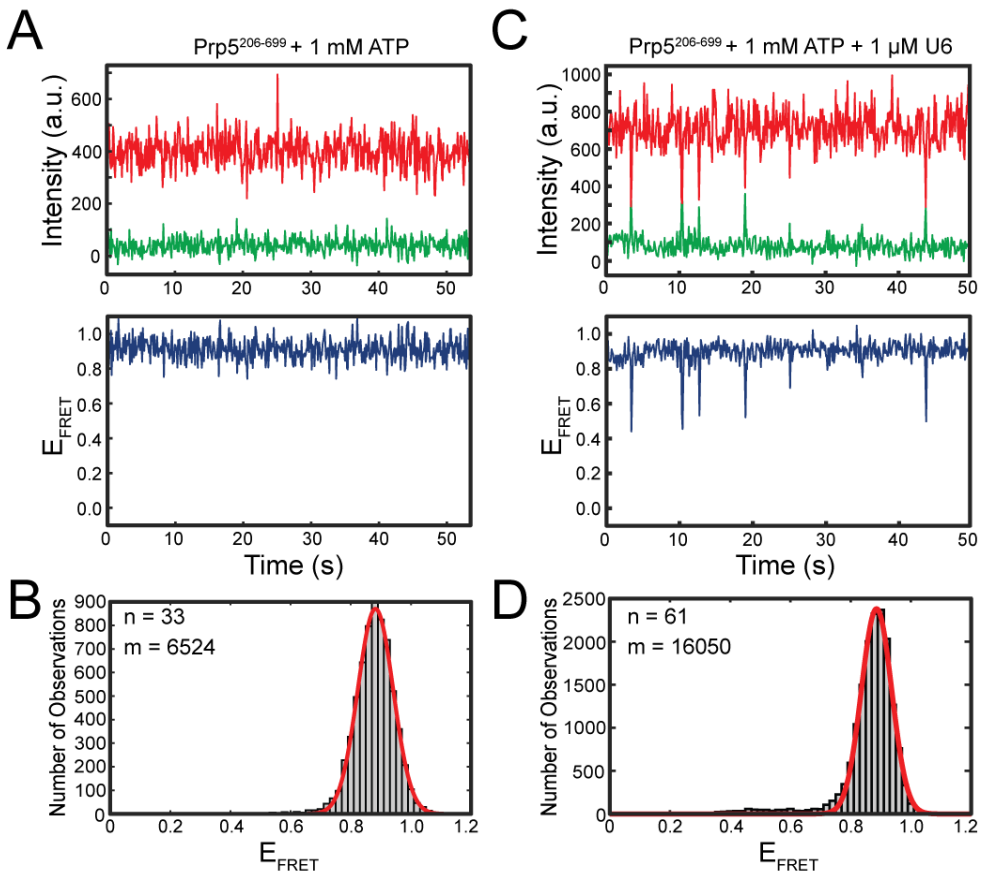


Figure 7.11: smFRET studies of Prp5 V268C/K666C in the presence of ATP and RNA. (A) Cy3 and Cy5 trajectories for a single Prp5 in the presence of ATP alone (top) and calculated FRET efficiencies (bottom). (B) Histogram of FRET efficiencies of Prp5 with ATP. (C) Cy3 and Cy5 trajectories for a single Prp5 in the presence of ATP and U6 snRNA (top) and calculated FRET efficiencies (bottom). (D) Histogram of FRET efficiencies of Prp5 with ATP and U6 snRNA.

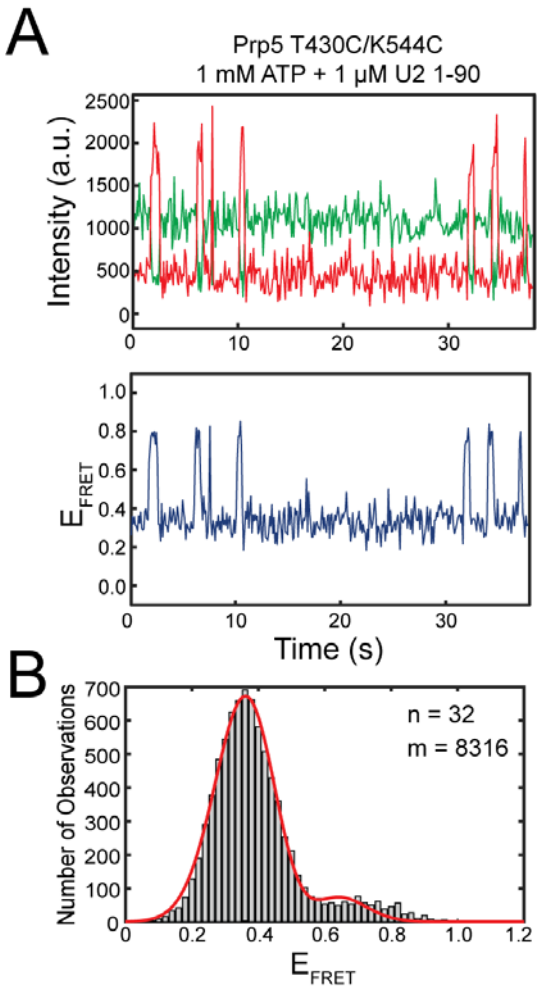


Figure 7.12: smFRET studies of Prp5 T430C/K544C. (A) Cy3 and Cy5 trajectories for a single Prp5 T430C K544C in the presence of ATP and U6 snRNA (top) and calculated FRET efficiencies (bottom). This construct gives a low FRET state in the open conformation and a high FRET state in the closed conformation (B) Histogram of FRET efficiencies of Prp5 with ATP and RNA.

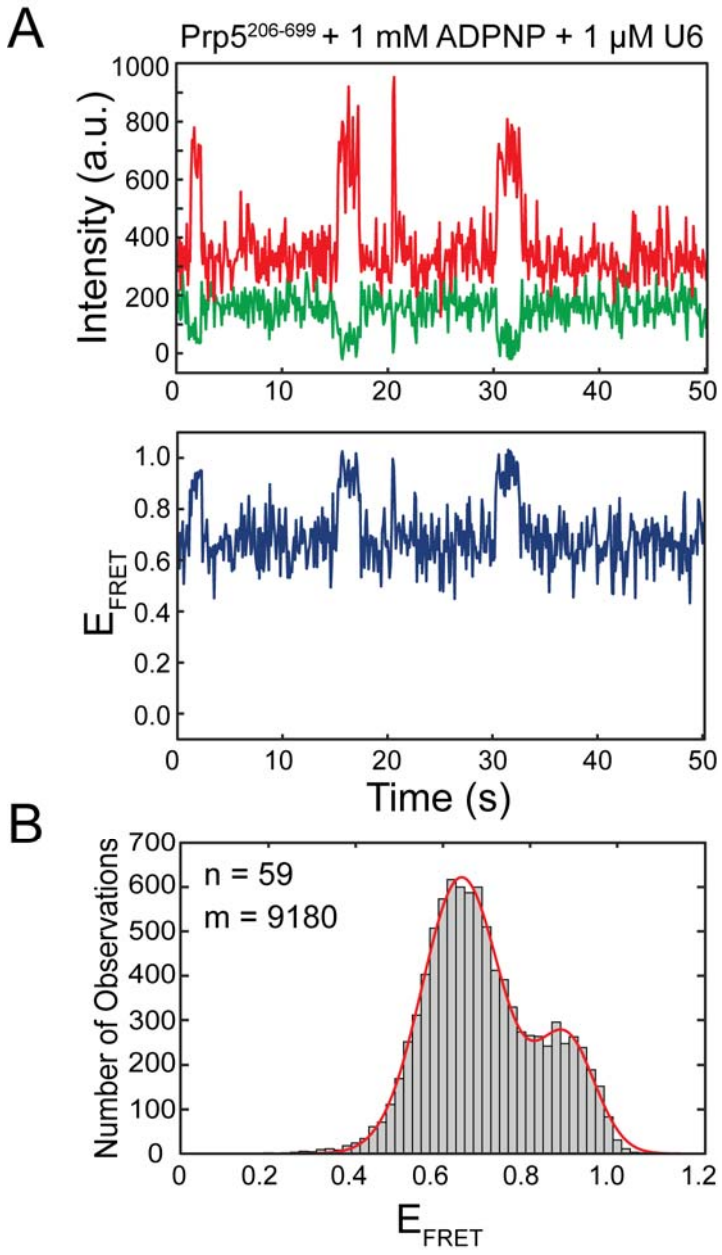


Figure 7.13: A non-hydrolysable analog of ATP (ADPNP) stabilizes the closed conformation of Prp5. (A) Cy3 and Cy5 trajectories for Prp5 206-699 V268C/K666C construct in the presence of ADPNP and RNA. (B) Histogram of FRET efficiencies constructed from thousands of observations of Prp5 with ADPNP and RNA.

Single Molecule Investigation of Full-length Prp5

The observation that the helicase domain acts as a typical DEAD-box protein is somewhat surprising, as Prp5 has been shown to have stable, ATP-independent RNA binding activity (unpublished work from S. Tretbar) (27, 38). The helicase domain alone has not been tested for RNA binding activity but most DEAD-box proteins only bind and unwind RNA in the presence of ATP or ATP analogs. While it is possible that other regions of the protein contribute to RNA binding, no identifiable domains directly implicated in RNA binding (i.e., RRMs, HEATS, etc.) have been discovered based on sequence homology. An alternative model is that regions that flank the helicase core can contribute to the formation of a closed state. Indeed, the C-terminal region of the DEAD-box protein Mss116 can impact the helicase core and modulate RNA binding and unwinding(39). To explore this possibility, we sought to study ATP- and RNA-induced conformational dynamics on full-length Prp5. Surprisingly, incubation of $Prp5^{FL}$ V268C/K666C in the absence of ATP and RNA resulted in smFRET trajectories that are considerably more heterogeneous than the helicase core (**Figure 7.14A**). For example, full length Prp5 showed multiple anti-correlated transitions to other states as opposed to the simple two-state system observed with the helicase core. Strikingly, the distribution is centered around 0.65-0.7, which resembles the low FRET state only observed upon binding of ATP and RNA to the core (**Figure 7.14B**). Addition of ATP and RNA did little to alter the dynamics observed in the smFRET trajectories or the distribution of the histogram (**Figure 7.14B vs. Figure 7.14D**). These data support that hypothesis that the flanking domains alter the conformation of the helicase core in the absence of ATP and RNA. It is intriguing that the FRET state corresponds with the form competent for RNA binding, as this may rationalize Prp5 binding to RNA in the absence of nucleotide cofactors. Investigation as to whether the helicase core alone can bind RNA would be necessary to confirm this potential activity of the flanking regions.

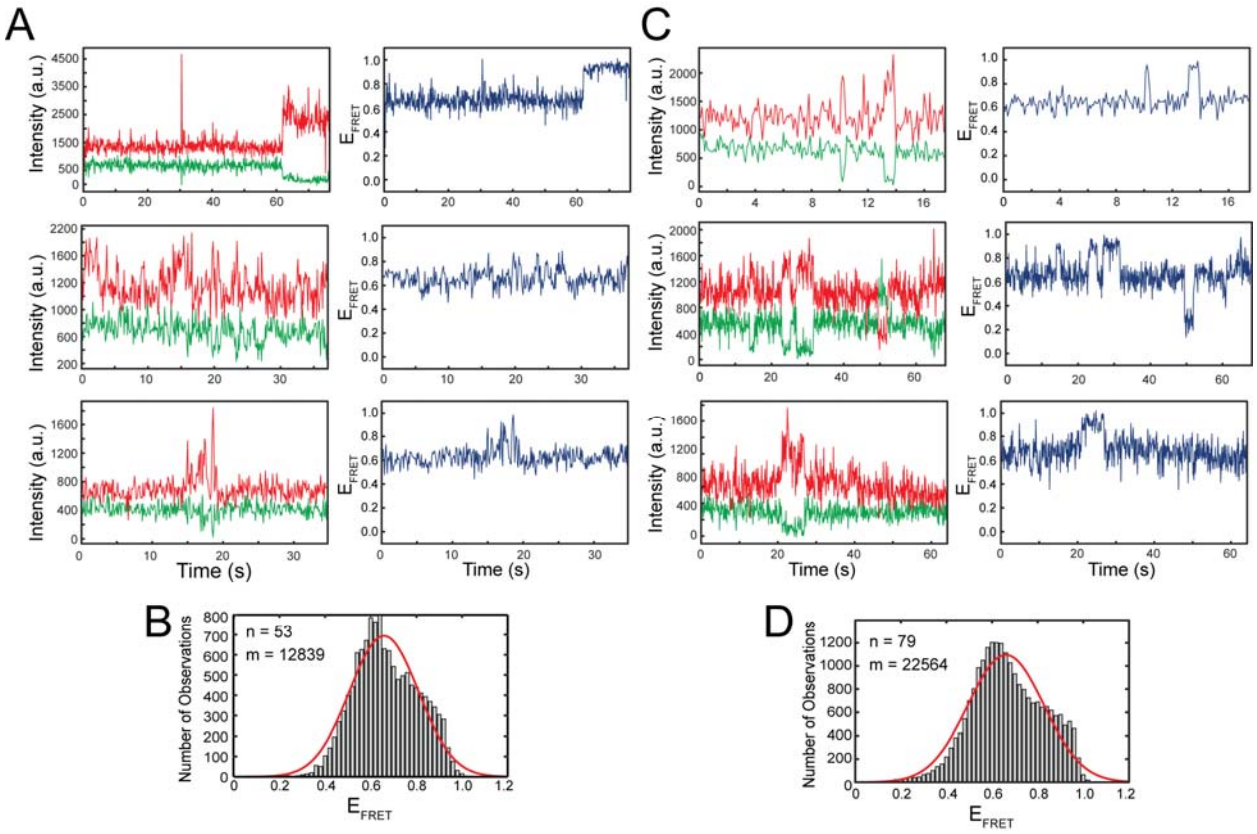


Figure 7.14: smFRET analysis of full-length Prp5 V268C/K666C. (A) Full length Prp5 was immobilized on a surface and observed in the absence of ATP and RNA. The trajectories are more dynamic and heterogeneous than the helicase core alone. Three representative traces are shown. (B) Histogram of data from A. (C) Full-length Prp5 was monitored in the presence of ATP and RNA. Representative traces are shown. (D) Histogram of data from C.

Genetic Analysis of the Essential Regions of Prp5

The single molecule data using full-length protein suggests that the helicase core behaves differently in the presence and absence of the flanking regions. We next explored the role of these regions on yeast growth with the goal of defining the minimum region necessary for viability. Numerous alleles of Prp5 resulting in truncation of the C-terminus of the protein have been shown to be viable *in vivo* (*prp5-3*, Prp5Δ494, etc.) as long as the ATPase requirement of Prp5 is bypassed(6, 23, 40). However, careful examination of the regions of Prp5 required for viability has not been performed, particularly with the N-terminal extension. We designed multiple truncations of the N-terminus of Prp5 upstream of the DPLD motif (amino acids 135-138) of Prp5 at positions 40, 68, 86 and downstream at position 206. Mutation of the DPLD motif to a quadruple alanine is viable in yeast but leads to defects in splicing(21, 41). Loss of the upstream region is viable, suggesting that this region does not have a secondary function (**Figure 7.15A**). However, truncation at position 206 is lethal, demonstrating that the N-terminal extension of Prp5 is important. It is possible that the DPLD motif plays a role in the requirement of residues 87-205. Bypass of the ATPase requirement of Prp5 has no effect on viability when combined with N-terminal truncations unlike the with the C-terminal truncations. Interestingly, truncation of Prp5 at positions 700, 750 and 800 are inviable whereas truncation at position 656 is not (**Figure 7.15B**). Position 656 is adjacent to the end of RecA2. This shows that partial deletion of the C-terminus is more deleterious than complete loss. Inclusion of these regions in the smFRET assay had a profound impact on Prp5 helicase dynamics, suggesting there may be an important connection between conformational dynamics mediated by the C-terminal extension and viability. Finally, mutation of arginine 615 to alanine is viable only when CUS2 is deleted. R615 is thought to stabilize the transition state during ATP hydrolysis by analogy to the DEAD-box protein DbpA(42). Loss of this interaction is consistent with the literature that demonstrates that Prp5 ATPase deficiency is only permissible in strains without Cus2 or with altered U2 snRNA.

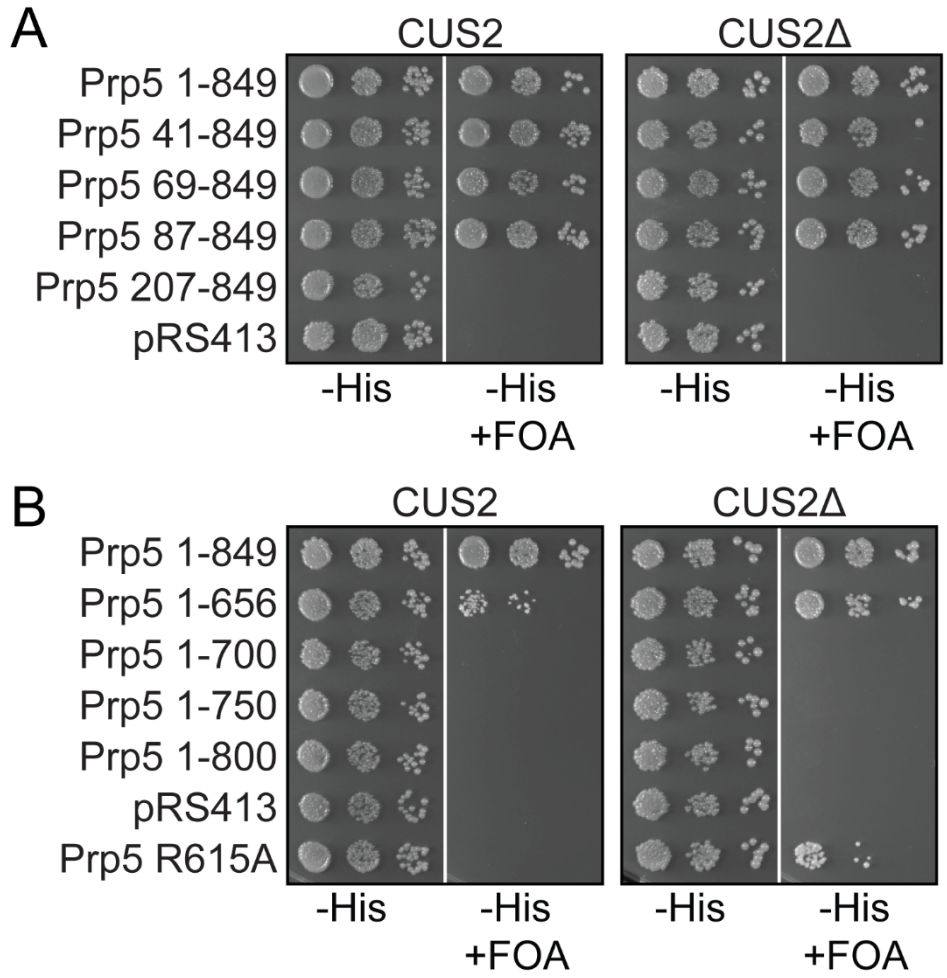


Figure 7.15: Truncation analysis of Prp5 in strains containing or lacking *CUS2*. (A) Liquid cultures were grown to mid-log phase, adjusted to the same OD_{600} and spotted onto plates lacking histidine and containing 5-FOA. Regions upstream of position 86 are non-essential for growth. Loss of *CUS2* does not improve growth (B) Truncation analysis of C-terminal truncations of Prp5. Truncation at position 656 deletes the entire C-terminal extension and is sick when expressed in yeast. Deletion of *CUS2* improves growth. Deletion of the C-terminus is lethal unless the entire sequence is removed. Loss of ATPase activity (R615A) is viable in strains lacking *CUS2.*

Conclusions and Perspectives

The work presented here has established methods to address many of the questions that exist concerning the multiple functions of Prp5 during splicing. It would be interesting to evaluate the impact of the E235A and TAG mutations on the lifetimes of the open and closed states of Prp5 using smFRET. If the helicase core of Prp5 is constrained into a single state by the N- and C-termini, then the mechanism for how these mutations impact splicing must be re-evaluated. This can be done using the Prp5 FL construct. It would also be interesting to investigate how other components of the U2 snRNP and the spliceosome in general impact dynamics in Prp5. Cus2 plays a complicated role in splicing and its function is linked with Prp5(19, 43). However, whether Cus2 directly impacts Prp5 dynamics is unknown. Prp5 also directly interacts with Hsh155/SF3b1, presumably through its N-terminal extension(24). Hsh155 is another factor that can be expressed recombinantly and added to smFRET experiments. An ambitious but worthwhile experiment would be to study Prp5 dynamics on the spliceosome during spliceosome assembly *in vitro*. This would require significant effort but would ultimately yield data in conditions that closely resemble the biological context. Another interesting avenue would be to biochemical characterize the Prp5 fragment in isolation for ATPase and RNA binding activity.

REFERENCES

1. Wahl,M.C., Will,C.L. and Lührmann,R. (2009) The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, **136**, 701–718.

2. Will,C.L. and Lührmann,R. (2011) Spliceosome Structure and Function. *Cold Spring Harb. Perspect. Biol.*, **3**, a003707.

3. Matera,A.G. and Wang,Z. (2014) A day in the life of the spliceosome. *Nat Rev Mol Cell Biol.*, **15**, 108–121.

4. Liu,Y.-C. and Cheng,S.-C. (2015) Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing. *J. Biomed. Sci.*, **22**, 54–54.

5. Dalbadie-McFarland,G. and Abelson,J. (1990) PRP5: a helicase-like protein required for mRNA splicing in yeast. *Proc. Natl. Acad. Sci. U S A*, **87**, 4236–4240.

6. Xu,Y.-Z. and Query,C.C. (2007) Competition between the ATPase Prp5 and Branch Region-U2 snRNA Pairing Modulates the Fidelity of Spliceosome Assembly. *Mol. Cell.*, **28**, 838–849.

7. Yang,F., Wang,X.Y., Zhang,Z.M., Pu,J., Fan,Y.J., Zhou,J., Query,C.C. and Xu,Y.Z. (2013) Splicing proofreading at 5' splice sites by ATPase Prp28p. *Nucleic Acids Res.*, **41**, 4660-70.

8. Saguez,C., Gonzales,F.A., Schmid,M. and Boggild,A. (2013) Mutational analysis of the yeast RNA helicase Sub2p reveals conserved domains required for growth, mRNA export, and genomic stability. *RNA.*, **19**, 1363-71.

9. Rocak,S. and Linder,P. (2004) DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol.*, **5**, 232-41.

10. Linder,P. and Jankowsky,E. (2011) From unwinding to clamping—the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol.*, **12**, 505-16.

11. Rudolph,M.G. and Klostermeier,D. (2015) When core competence is not enough: functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding. *Biol Chem.*, **396**, 849-65.

12. Del Campo,M. and Lambowitz,A.M. (2009) Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. *Mol. Cell.*, **35**, 598-609.

13. Mallam,A.L., Del Campo,M., Gilman,B., Sidote,D.J. and Lambowitz,A.M. (2012) Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature*, **490**, 121–125.

14. Liu,F., Putnam,A. and Jankowsky,E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. U S A*, **105**, 20209-14.

15. Liu,F., Putnam,A.A. and Jankowsky,E. (2014) DEAD-box helicases form nucleotide-dependent, long-lived complexes with RNA. *Biochemistry.*, **53**, 423-33.

16. Andersen,C., Ballut,L. and Johansen,J.S. *et al*. (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science*., **313**, 1968-72.

17. Bono,F., Ebert,J., Lorentzen,E. and Conti,E. (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell*., **126**, 713-25.

18. Mohlmann,S., Mathew,R. and Neumann,P. *et al*. (2014) Structural and functional analysis of the human spliceosomal DEAD-box helicase Prp28. *Acta Crystallogr D Biol Crystallogr*., **70**, 1622-30.

19. Perriman,R., Barta,I. and Voeltz,G.K., *et al*. (2003) ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. *Proc. Natl. Acad. Sci. U S A*, **24**, 13857-13862.

20. Liang,W.-W. and Cheng,S.-C. (2015) A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev.*, **29**, 81–93.

21. Xu,Y.Z., Newnham,C.M., Kameoka,S., Huang,T., Konarska,M.M. and Query,C.C. (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *The EMBO Journal*, **23**, 376–385.

22. Perriman,R. and Ares,M. (2000) ATP can be dispensable for prespliceosome formation in yeast. *Genes Dev.*, **14**, 97–107.

23. Perriman,R. and Ares,M. (2010) Invariant U2 snRNA Nucleotides Form a Stem Loop to Recognize the Intron Early in Splicing. *Mol Cell*, **38**, 416–427.

24. Tang,Q., Rodriguez-Santiago,S. and Wang,J *et al*. (2016) SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. *Genes Dev.,* **30**, 2710-2723.

25. Carrocci,T.J., Zoerner,D.M. and Paulson,J.C., *et al*. (2017) SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. *Nucleic Acids Res., pii:gkw1349.*

26. Zhang,Z.-M., Yang,F., Zhang,J., Tang,Q., Li,J., Gu,J., Zhou,J. and Xu,Y.-Z. (2013) Crystal Structure of Prp5p Reveals Interdomain Interactions that Impact Spliceosome Assembly. *Cell Rep.*, **5**, 1269–1278.

27. Banroques,J., Cordin,O., Doère,M. and Linder,P., Tanner,N.K., (2011) Analyses of the functional regions of DEAD-box RNA 'helicases' with deletion and chimera constructs tested *in vivo* and *in vitro*. *J Mol Biol*., **413**, 451-72.

28. Beckett,D., Kovaleva,E. and Schatz,P.J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase‐catalyzed biotinylation. *Protein Sci*. **8**, 921-9.

29. Kay,B.K., Thai,S. and Volgina,V.V. (2009) High-throughput biotinylation of proteins. *Methods Mol Biol*., **498**, 185-96.

30. Nanda,J.S. and Lorsch,J.R. (2014) Labeling of a Protein with Fluorophores Using Maleimide Derivitization. *Methods Enzymol*. **536**, 79-86.

31. Roy,R., Hohng,S. and Ha,T. (2008) A practical guide to single-molecule FRET. *Nat Methods*, **5**, 507–516.

32. Hoskins,A.A., Friedman,L.J., Gallagher,S.S., Crawford,D.J., Anderson,E.G., Wombacher,R., Ramirez,N., Cornish,V.W., Gelles,J. and Moore,M.J. (2011) Ordered and Dynamic Assembly of Single Spliceosomes. *Science*, **331**, 1289–1295.

33. Aitken,C.E., Marshall,R.A. and Puglisi,J.D. (2008) An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments. *Biophys J*, **94**, 1826–1835.

34. McKinney,S.A., Joo,C. and Ha,T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J*, **91**, 1941-51.

35. Bronson,J.E., Fei,J., Hofman,J.M. and Gonzalez,R.L., *et al*. (2009) Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys J.*, **97**, 3196-205.

36. Pires,D., Ascher,D.B. and Blundell,T.L. (2014) DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.,* **42**, W314-9.

37. Lesser,C.F. and Guthrie,C. (1993) Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics*. **133**, 851-63.

38. Wu,G., Adachi,H., Ge,J. and Stephenson,D., *et al*. (2016) Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. *EMBO J.,* **35***, 654-67.*

39. Mohr,G., Del Campo,M., Mohr,S., Yang,Q. and Jia,H., *et al*. (2008) Function of the C-terminal domain of the DEAD-box protein Mss116p analyzed in vivo and in vitro. *J Mol Biol., **375**, 1344-64.*

40. Ruby,S.W., Chang,T.H. and Abelson,J. (1993) Four yeast spliceosomal proteins (PRP5, PRP9, PRP11, and PRP21) interact to promote U2 snRNP binding to pre-mRNA. *Genes Dev*., **7**, 1909-25.

41. Shao,W., Kim,H.S., Cao,Y., Xu,Y.Z. and Query,C.C. (2011) A U1-U2 snRNP Interaction Network during Intron Definition. *Mol Cell Biol*, **32**, 470–478.

42. Elles,L. and Uhlenbeck,O.C. (2008) Mutation of the arginine finger in the active site of Escherichia coli DbpA abolishes ATPase and helicase activity and confers a dominant slow growth phenotype. *Nucleic Acids Res*., **36**, 41-50.

43. Yan,D., Perriman,R., Igel,H. and Howe,K.J., *et al*. (1998) CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. Mol Cell Biol., **18**, 5000-9.

Chapter Eight

Outlook

Many avenues exist to further the work presented in my thesis. Here, I present some ideas that could come from the work that I presented.

Improvement of the JX1 aptamer

In chapter two, I discussed work on the generation of the JX1 aptamer that binds benzylguanine. This aptamer is a potentially useful tool to purify RNAs from a heterogeneous mixture. However, JX1 could be improved through a more rigorous selection process (i.e. selection of aptamers that form covalent adducts upon binding of benzylguanine).This mechanism is analogous to the activity of the protein SNAP-tag. This improved aptamer could be coupled with quenched SNAP-dyes to generate a fluorogenic aptamer. I think this is an important step towards using the JX1 aptamer practically and makes it a competitive option relative the spinach, broccoli, or mango aptamers.

The selection of better deoxyribozymes

The second portion of this section of my thesis employed deoxyribozymes to prepare fluorescent RNAs. Deoxyribozymes are useful for a number of potential applications. In my thesis, I demonstrated that 10DM24 can be used to fluorescently label pre-mRNA substrates *in vitro* using fluorescent nucleotides. Despite this success, the conditions necessary to achieve efficient labeling are harsh and significantly degrade the RNA target. The selection of new and better deoxyribozymes that function in milder conditions is likely the best way to overcome this drawback. Partial randomization of the 10DM24 sequence and selection of active targets in the presence of decreased levels of Mg^{2+} and lanthanides is a potentially useful selection strategy. Ultimately, I believe labeling RNAs in this manner would be best served by the generation of novel catalysts specifically intended to modify long RNAs and this process could also be coupled with the selection of other important features. Since 10DM24 shows preference for incorporating guanine over other nucleotides, the selection of deoxyribozymes that rely on other nucleotides is warranted. This would allow for simultaneous labeling of two positions using orthogonal systems, which could be used to monitor intron removal or generate FRET constructs.

Dbr1 inhibitors to block debranching

The latter half of my chapter on deoxyribozymes discusses the development of linkages that cannot be cleaved by the debranching enzyme, Dbr1. We chose to pursue phosphorothioates as the primary means to overcome debranching. However, alternative approaches to this problem are also possible, including the development of Dbr1 inhibitors. One could imagine a screen to select inhibitors from a small molecule library using a debranching activity assay. Branched substrates that include a FRET pair or a fluorophore and a quencher pair could be used to monitor Dbr1 processing of substrates in a high-throughput fashion. We could include active inhibitors during single molecule splicing assays *in vitro*. Inhibitors identified in this way could also find use in other avenues of interest in our lab if they also function on metazoan Dbr1. Inhibitors could be useful tools for isolating metazoan lariats and mapping the locations of BS used.

Furthering our understanding of the role of Hsh155/SF3b1 during splicing

My work on MDS mutations in the splicing factor Hsh155 showed that they alter BS fidelity in yeast (Chapter 5). I further examined the contacts formed by Hsh155 in the B^{act} spliceosome to determine which contacts are significant for function in Chapter 6. The recent structures of the spliceosome provide important insight into how this molecular machine may function. However, structure without function is meaningless. Further characterization and validation of these newly revealed contacts of Hsh155/SF3b1 in addition to those made by other components of the SF3b complex, is critical to understanding the spliceosome in molecular detail. I began looking at the impact of mutations in Rds3 on splicing. Focusing on this factor is a logical next step, as the human homolog has been implicated in disease. An important initial step in treating disease is attempting to understand how the mutated factor natively functions to identify how it misfunctions in disease.

I proposed that a conformational change in Hsh155, along with the helicity of its HEAT repeats, play a crucial role in BS recognition. This model is supported by my observations that disrupting contacts observed in the B^{act} structure did not predictably affect viability in yeast.

Understanding how conformational changes in Hsh155 influence BS selection will be a challenging but necessary task to understand on a mechanistic level how the spliceosome transitions from definition of splice sites during assembly to a catalytic conformation. Furthermore, understanding how protein-protein contacts with other members of the SF3b complex and with Prp5 influence Hsh155 structure and function will be of key importance.

New inhibitors of pre-mRNA splicing

I demonstrated that spliceosome inhibitors that block splicing in humans do not function in yeast because they do not bind the target protein, SF3b1/Hsh155. This work emphasizes that differences between the homologs exist and also suggests that this difference could potentially be exploited. The identification of inhibitors that target yeast Hsh155 and not human SF3b1 could generate new antibiotic compounds. These could be used to treat pathogenic yeast infections.

One of the last projects that I initiated in the Hoskins' lab was to study Prp5 conformational dynamics. Understanding how the spliceosome is regulated by DEAD- and DEAH-box helicases is fundamental towards comprehending splice site choice and alternative splicing. My work on characterizing the dynamics of Prp5 was an important step towards this goal. My approach could easily be applied to other proteins that modulate different steps in splicing such as the DEAD-box Prp28, or the U4/U6 assembly chaperone Prp24. Prp28 may function similar to Prp5 in facilitating the exchange of the U1 snRNP for the U6 snRNP at the 5'SS. Additionally, smFRET could be a useful mechanism to understand how Prp24 unwinds U6 to anneal it to the U4 snRNA.

The mechanism of Prp5 in BS choice

Many questions still exist for Prp5. Mutations in Prp5 have been implicated in changing nonconsensus intron splicing, but the mechanism behind these changes remain controversial. The smFRET approach is powerful and allows for direct monitoring of the conformation of Prp5 as it functions. We can incorporate these mutations into the Prp5 smFRET assay to determine how they influence dynamics in the protein. Using this approach we can observe if these mutations bias Prp5 conformation and assign active and inactive conformations of the protein.

Beyond smFRET approaches, there are a number of other interesting ways to look at the function of Prp5. Prp5 has a known but poorly understood relationship with the splicing factor Cus2. One reason for a lack of understanding of this relationship is the unknown function of Cus2. Cus2 has no direct effect on splicing using reporter assays. Furthermore, Cus2 is non-essential in *S. cerevisiae,* but the *S. pombe* homolog, Uap2, is essential. The essential nature of Uap2 implies that these yeast are more reliant on the function of the protein. Therefore, *S. pombe* may be a better choice to dissect the relationship between Cus2 and Prp5 (Uap2 and Prp10 in *S. prombe*). Understanding this relationship could provide insight to important steps in snRNP recycling prior to incorporation into the prespliceosome.

My graduate work has contributed to scientific discourse and established interesting avenues to continue. The ideas that I present here are just some of the potentially interesting directions that can come from my work.

TUCKER CARROCCI

440 Henry Mall, Rm 2212 • Madison, WI 53706

ACADEMIC BACKGROUND

Doctor of Philosophy, Biochemistry, University of Wisconsin–Madison, Madison, WI, 2017
Bachelor of Science, Biochemistry and Molecular Biophysics & Molecular and Cellular Biology, University of Arizona, Tucson, AZ, 2011

RESEARCH EXPERIENCE

University of Wisconsin–Madison, Graduate Research Assistant 08/2011 – 05-2017
Principal Investigator: Prof. Aaron Hoskins
Department of Biochemistry, Madison, WI
Development of new methods for RNA visualization
Characterization of SF3b1/Hsh155 mutations associated with myelodysplastic syndromes

New England Biolabs, Summer Research Intern 06/2014 – 08/2014
Supervising Investigator: Dr. Ming-Qun Xu
Development of immobilized enzymes for various purposes

University of Arizona, Laboratory Assistant
Principal Investigator: Prof. Robin Polt 06/2009 – 08/2011
Department of Chemistry and Biochemistry, Tucson, AZ
Synthetic production of pure rhamnolipid congeners with the purpose of biosurfactant character analysis

Principal Investigator: Prof. Jeanne Pemberton 09/2010 – 06/2011
Department of Chemistry and Biochemistry, Tucson, AZ
Characterization of synthetic rhamnolipid precursors prepared in the Polt laboratory

Principal Investigator: Prof. W. Dan Stamer 08/2008 – 12/2009
Department of Ophthalmology and Vision Science, Tucson, AZ
Analysis of the effect of cortisol and cortisol derivatives on cellular outflow in trabecular meshwork cells

REFERENCES

Prof. Aaron A. Hoskins
University of Wisconsin-Madison
Phone: (608) 890-3101
Email: ahoskins@wisc.edu

Prof. Samuel E. Butcher
University of Wisconsin-Madison
Phone: (608) 263-3890
Email: sebutcher@wisc.edu

Prof. Charles Query
Albert Einstein College of Medicine
Phone: (718) 430-4174
Email: charles.query@einstein.yu.edu

TEACHING EXPERIENCE

- Graduate Teaching Assistant 09/2012 – 12/2013
  Supervisors: *Prof. Aaron Hoskins and Prof. Mike Cox*
  Department of Biochemistry, Madison, WI
  I led discussions, prepared additional practice problems and tutored students in an introductory biochemistry course.

- Undergraduate Teaching Assistant
  Supervisor: *Prof. Tsu-Shuen Tsao* 01/2011 – 05/2011
  Department of Chemistry and Biochemistry, Tucson, AZ
  I helped instruct an advanced biochemistry class with a focus on metabolism by leading class discussions, tutoring students and organizing concept reviews prior to exams.

  Supervisor: *Dr. Hamish Christie* 09/2010 – 12/2010
  Department of Chemistry, Tucson, AZ
  I helped instruct, led reviews, and tutored undergraduate students enrolled in organic chemistry

PUBLICATIONS

Didychuk AL, Montemayor EJ, Carrocci TJ, DeLaitsch AT, Lucarelli S, Brow DA, Hoskins AA, Butcher SE. Usb1 controls U6 snRNP assembly through evolutionarily divergent phosphodiesterase activities. (submitted)

Carrocci TJ, Zoerner ZM, Paulson JC, Hoskins AA. SF3b1 Mutations Associated with Myelodysplastic Syndromes Alter the Fidelity of Branchsite Selection in Yeast. *Nucleic Acid Res*. Published online Jan 6 2017. http://dx.doi.org/10.1093/nar/gkw1349

Xu, J, Carrocci TJ, Hoskins AA. Evolution and characterization of a benzylguanine-binding RNA aptamer. *Chem Commun (Camb).* 2016 Jan 11; 52(3):549-552.

Carrocci TJ, Hoskins AA. The Spliceosome and Pre-mRNA Splicing. Encyclopedia of Cell Biology. 2015. 1:495-502.

Carrocci TJ, Hoskins AA. Imaging of RNAs in live cells with spectrally diverse small molecule fluorophores. *Analyst.* 2014 Jan 7; 139(1):44-7.

Coss C, Carrocci T, Maier R, Pemberton J, and Polt R. Minimally Competent Lewis Acid Catalysts. In(III) and Bi(III) Salts Produce Rhamosides in High Yield and Purity. *Helv. Chim. Acta.* 2012; 95;2652-2659.

ORAL AND POSTER PRESENTATIONS

T. Carrocci, D. Zoerner, J. Paulson, A. Hoskins. SF3b1 Mutations Associated with Myelodysplastic Syndromes and Cancers Alter the Fidelity of Branchsite Selection. (2016). 21st Annual Meeting of the RNA Society, Tokyo, Japan, July 2016. (**Talk**)

T. Carrocci, D. Zoerner, J. Paulson, A. Hoskins. Spliceosomal Mutations Associated with Myelodysplastic Syndromes in Humans Change the Fidelity of Branchsite Selection in Yeast. (2015). CSHL Eukaryotic Pre-mRNA Processing. Cold Spring Harbor, NY, August 2016. (**Talk**)

T. Carrocci, D. Zoerner, A. Hoskins. Spliceosomal Mutations Associated with Myelodysplastic Syndromes in Humans Change the Fidelity of Branchsite Selection in Yeast. (2015). 20th Annual Meeting of the RNA Society, Madison, WI, June 2015. (**Poster**)

T. Carrocci, M. Ashton, L. Buettner, F. Javadi-Zarnaghi, C. Höbartner, A. Hoskins. Deoxyribozymes for Preparation of Fluorescent Pre-mRNAs without Splinted Ligation. (2014). 19th Annual Meeting of the RNA Society. Quebec City, Canada, June 2014. (**Poster**)

T. Carrocci, A. Hoskins. New methods for fluorescently labeling RNA in vitro and in vivo. (2013). ASBMB Special Symposia Series "Evolution and Core Processes in Gene Regulation". July 2013. (**Talk**)

T. Carrocci, J. Turri, A. Hoskins. Development of a New Fluorescent Toolbox for Imaging RNA in Live Cells (2012). 17th Annual Meeting of the RNA Society. Ann Arbor, MI, May 2012. (**Poster**)

FELLOWSHIPS

- *William H. Peterson Biochemistry Fellowship* 09/2014 – 09/2015
  Granting Agency: Department of Biochemistry
- *NIH Biotechnology Training Program Trainee (T32-GM08349)* 09/2011 – 08/2014
  Granting Agency: National Institute of General Medical Sciences (NIGMS)

HONORS AND AWARDS

- Denton Award for Undergraduate Mentoring (2015)
- RNA Biochemistry Poster Award (2014)
- National Science Foundation GRFP – Honorable Mention (2013)
- University of Arizona Honors Program Graduate (2011)
- Undergraduate Scholar in Integrated Science (2008-2011)
- Danny Brower Memorial Scholar Award (2009)

MENTORING AND TEACHING EXPERIENCE

- Teaching Assistant, Physics of Living Cells Summer School 2013, Center for the Physics of Living Cells, University of Illinois – Urbana-Champaign, July 2013

- Mentoring Undergraduate Students, University of Wisconsin-Madison

  David Beier, Biochemistry Student 06/2016 – Current
  Project Title: Single Molecule Studies Reveal the Impact of Prp5 Conformational Changes on Branchsite Fidelity

  Douglas Zoerner, Biochemistry Student 09/2014 – 08/2015
  Project Title: Characterizing the Impact of MDS Mutations on the Yeast Spliceosome
  Current Position: Medical Student at the University of Kentucky

  George Luo, Biochemistry Student 09/2014 – 08/2015
  Project Title: The Impact of Hsh155 and Prp21 MDS Mutations on Intron Splicing
  Current Position: Medical Student at the Case Western

  Matthew Ashton, Biochemistry Student 09/2012 – 09/2015
  Project Title: Deoxyribozymes as Tools for Fluorescent RNA Labeling
  Current Position: Medical Student at Tulane University

- Research Mentor Training Course, Delta Program, University of Wisconsin–Madison 09/2012

REVELANT OUTREACH

- *Adult Role Models in Science (ARMS) Club Leader (2011-2012)*
  I ran a science club with elementary school children (primarily second through fifth grade) with weekly demonstrations and activities meant to encourage their scientific interest and thought.

- *Department of Chemistry and Biochemistry Ambassador (2010-2011)*
  The club was coordinated by the department and help run multiple recruitment weekends for potential incoming high school students.

- *Chemistry & Biochemistry Club (2009-2011)*
  The club was run by the department of Chemistry and Biochemistry and aimed at recruiting future students as well as encouraging scientific understanding and education through community demonstrations.

SOCIETY MEMBERSHIPS

RNA Society Member