HIGH THROUGHPUT SINGLE-MOLECULE MICROSCOPY FOR SURVEYING RNA STRUCTURE

by

EMILY CARSTAIRS

A THESIS

Presented to the Department of Biology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

May 2024

An Abstract of the Thesis of

Emily Carstairs for the degree of Bachelor of Science
in the Department of Biology to be taken June 2024

Title: High Throughput Single-Molecule Microscopy for Surveying RNA Structure

Approved: *Dr. Julia Widom*
Primary Thesis Advisor

Ribonucleic acid, or RNA, is a vital biomolecule responsible for many functions in an organism. Messenger RNA (mRNA) is responsible for carrying the genetic instructions from DNA to protein synthesis, which is responsible for much of the cellular activity needed for life. mRNA is originally synthesized as "precursor" mRNA, called pre-mRNA, and undergoes a modification called splicing within the cell nucleus to develop into mature mRNA. Splicing ensures the mRNA contains the correct sequence of nucleotides for proper protein synthesis. RNA sequence informs its structure, which is intrinsically linked to its function and therefore has been the subject of much research. We have developed a method using single-molecule fluorescence resonance energy transfer (smFRET) that allows us to survey the structure of many RNA sequences in a high throughput manner. Using this method, we have varied an important sequence of a truncated yeast pre-mRNA, Ubc4, involved in splicing called the branchpoint sequence (BPS). We show that small changes in the sequence of the BPS of Ubc4 result in a variety of different structural behaviors. However, the overall behavior of the sampled pre-mRNA mimics the behavior exhibited by a wild-type construct, suggesting that secondary structure formation of pre-mRNAs is not influenced by changes to the BPS. This finding, along with further investigations into Ubc4 sequences and protein interactions, will illuminate pre-mRNA structural dynamics in splicing initiation.

Acknowledgements

I would like to extend a huge, huge thank you to everyone in the Widom lab, particularly Dr. Julia Widom and Dr. Satya Yadav. Without the opportunity I have had working in Dr. Widom's lab, I would not even be writing this thesis today. In addition, Dr. Widom has offered me much support throughout the writing process so I would like to give her a big, big thank you! Of course, I also could not have worked on this project without the help of Dr. Yadav. Dr. Yadav helped me worked through every single experiment in this thesis and much of the data analysis. Thank you so much! I would also to thank all the amazing staff and faculty here at UO. I have had so many amazing professors within the biology department and the honors college. Within the honors college I would specifically like to thank Professor Hinkle and Professor de Onis for help with this thesis. Without them, I would not have had the knowledge or skills to complete this thesis project.

Finally, I could not have made it to this point without the amazing support of my family and friends. Thank you, mom and dad, for constantly supporting me throughout the entirety of my education! Thank you to my wonderful roommates, Sarah, Sydney, and Sabrina, for listening to me endlessly explain the lasers and my thesis project!

Table of Contents

Introduction 6

Why study ribonucleic acid (RNA) structure? 6

What is single molecule fluorescence imaging? 9

Experimental Design 11

Results 12

RNA Ligation 12

smFRET 13

Discussion 17

Future Directions 18

Methods 20

Sample Preparation 20

Slide Preparation 20

smFRET 22

Data Analysis 23

Bibliography 24

List of Figures

Figure 1: Splicing and experimental schematic. 8
Figure 2: RNA ligation acrylamide gel visualized via fluorescent scan. 13
Figure 3: Comparison of density of raw data collection using Janelia Fluor and Alexa Fluor. 14
Figure 4: Example FRET traces of Ubc4 structural behaviors. 15
Figure 5: FRET Efficiency of Ubc4 pre-mRNA by frame occurrence. 16

Introduction

Why study ribonucleic acid (RNA) structure?

Ribonucleic acid, or RNA, is a vital biomolecule in all living organisms. It is similar to deoxyribonucleic acid, or DNA, with a few structural differences. Namely, RNA is generally single-stranded while DNA is generally double-stranded. Both RNA and DNA share the nucleotides adenosine, cytosine, and guanosine, but RNA contains uridine instead of the DNA analog thymidine. Functionally, RNA and DNA fulfill very different roles within the cell. DNA is the genetic information holder, or the "blueprint" for the cell. In contrast, RNA performs many different functions within the cell. Messenger RNA (mRNA) is the most well-known form of RNA and is responsible for carrying information from the DNA to the ribosome for protein synthesis. Other forms of RNA, like ribosomal RNA (rRNA) or transfer RNA (tRNA), also aid in protein synthesis. Some forms of RNA are involved in regulating gene expression like long-noncoding RNA (lncRNA) and micro RNA (miRNA)[1]. There are many other forms of RNA completing many other cellular functions, but this thesis will focus on the activities of mRNA.

RNA structure and function are intrinsically linked, and one informs the other. Therefore, to understand RNA function it is important to understand its structure as well. RNA structure can be thought of in four different ways: primary structure, secondary structure, tertiary structure, and quaternary structure. Primary structure describes the order of nucleotides within an RNA molecule. Secondary structure describes the base-pairing interactions between nucleotides, adenosine to uridine and cytosine to guanosine. Tertiary structure describes the three-dimensional form of the nucleotide chain. Quaternary structure describes interactions with other biomolecules. In this study, we have varied the primary structure, or sequence of a truncated

yeast RNA, Ubc4, and observed the effects on the higher order structures. We aim to discover how RNA sequence is related to its higher order structures specifically within the mRNA splicing process.

mRNA is originally synthesized as "precursor" mRNA, called pre-mRNA, in the cell nucleus and many chemical modifications must be made before it matures and is exported out of the nucleus. A major modification that pre-mRNA undergoes in eukaryotes is splicing. In eukaryotic genes, there are regions called introns that do not encode any information for protein synthesis. There can be many intronic regions surrounded by exons, the regions of eukaryotic genes that do encode for protein synthesis. Splicing joins together exons and excises introns (Figure 1A) of RNA which ensures the mature mRNA carries the correct sequence of nucleotides for protein synthesis. This process is facilitated by both regulatory sequences within the pre-mRNA and a large complex of a combination of RNA and proteins within the nucleus called the spliceosome.

The spliceosome is a large complex composed of five small nuclear ribonucleoproteins[2] (snRNPs) dubbed U1, U2, U4, U5, and U6, and over a hundred proteins.[2–4] This composition of molecules gives the spliceosome the ability to recognize and bind to specific pre-mRNA sequences to facilitate splicing. Spliceosome assembly occurs on each pre-mRNA strand as the snRNPs interact with the pre-mRNA[3] through base-pairing interactions. The splicing process is characterized by a series of fourteen conformational rearrangements[5] facilitated by different arrangements of the snRNP's. The commitment complex, or complex E, is formed when U1 binds the 5' splice site and SF1 binds the branchpoint sequence. Complex A is formed when SF1 is displaced by U2, which also binds the BPS.[2,3] Association of U4, U5, and U6 with U1 and U2 forms a loop structure in the RNA and brings the exons parallel to each other. This conformation

is known as complex B. Complex C is formed after the dissociation of U1 and U4, which brings the exons in even closer proximity.[6] The formation of Complex C is what allows the exons to be brought in close enough proximity for splicing to occur.

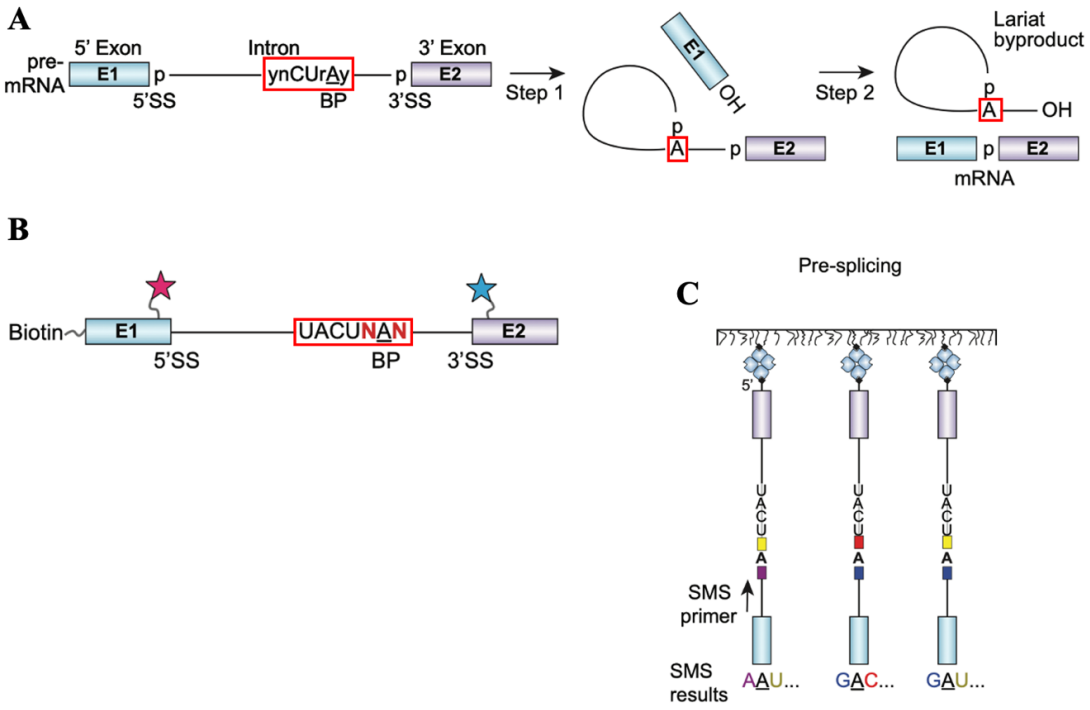

Figure 1: Splicing and experimental schematic.

A: Intramolecular chemistry that occurs during splicing. The branchpoint (BP) contains an adenosine residue that attacks the 5' splice site (SS). The 5' SS is then free to attack the 3' SS, thereby joining the exons and excising the intron as a lariat. **B:** Our designed Ubc4 pre-mRNA construct. The BPS follows the wild-type sequence except for random nucleotides, denoted by N, to either side of the BP adenosine. The construct is also modified with a 5' biotin for slide immobilization and 2 fluorophores, denoted by the stars, for single molecule microscopy. **C:** Experimental schematic for single molecule sequencing (SMS). When combined with structural data from smFRET movies, this will allow us to link pre-mRNA structure to sequence.

While the spliceosome is a very large complex, it is not pre-assembled in the nucleus and instead assembles directly on the pre-mRNA molecules to be spliced. There are many important sequences in the RNA to facilitate this process like the branchpoint sequence (BPS), 3' or 5' splice sites (SS), and polypyrimidine tract. The BPS is especially important for initial

spliceosome assembly.[7,8] The BPS varies from organism to organism and even gene to gene, but does contain one consistent adenosine residue across all pre-mRNAs. This adenosine residue is critical for chemistry that occurs within the splicing process (Figure 1A). The BPS is also located within all introns of pre-mRNA and is responsible for recruiting branchpoint binding proteins. In yeast, the BPS is fairly conserved across most pre-mRNAs, but in mammals the BPS is much more variable.[9] The BPS is thought to play an important role in binding of Msl5 in yeast[10] and SF1 in mammals.[2] Msl5 and SF1 are both branchpoint binding proteins that recognize a specific BPS within the pre-mRNA introns and form the initiation complex of the spliceosome.[6] Mutations of regulatory sequences such as the BPS have been implicated in many diseases, including spinal Duchenne muscular dystrophy and spinal muscular atrophy, which a leading genetic cause of death in children.[6] Disfunction in binding of SF1 to the BPS has been linked with cancer, especially hematopoietic malignancies.[11] For these reasons, the relationship between pre-mRNA primary structure and secondary or quaternary structures formed throughout splicing are of interest to our lab.

What is single molecule fluorescence imaging?

To understand the subtle structural variations within splicing due to variations within the BPS, we must be able to resolve structural heterogeneity between the RNA molecules. Previous studies have relied upon population surveys of RNA, where individual differences between molecules may be obscured.[12] In addition, it has been difficult to study RNA splicing previously due to an inability to observe conformational dynamics in real-time.[7] By using single-molecule microscopy techniques such as single molecule fluorescence resonance energy transfer (smFRET) and single molecule sequencing (SMS), we are able to observe the behavior of individual pre-mRNA molecule in real-time.

smFRET is a unique technique that reports on the distance between two fluorescent molecules, called fluorophores, and is resistant to instrumental noise as measurements are taken on the scale of individual molecules rather than the laboratory scale.[13] The technique relies upon energy transfer between an "acceptor" and "donor" molecule receptive to different wavelengths of light. The donor molecule is excited by a laser, and in turn emits its own energy which excites the acceptor molecule. The FRET efficiency (E) is a measure describing the distance between the fluorophores based on the energy transmitted between them. It is calculated as follows: $E = (1 + (R/R_0)^6)^{-1}$ where R is the distance between fluorophores and R_0 is the Forster radius at which E = 0.5.[13] High FRET efficiencies indicated the fluorophores are close together, and low FRET efficiencies indicate the fluorophores are further apart. When these fluorophores are attached to a single RNA molecule, we can visualize the changes in RNA structure over time using the FRET efficiency.[13] This process allows for continuous monitoring of the conformational changes that occur during the splicing process without the need for synchronization of splicing steps between different RNA molecules.

SMS is a single molecule technique that allows us to obtain the sequence of individual RNA molecules. This method relies upon a primer complimentary to the RNA of interest and the addition of modified nucleotides. Only one modified nucleotide can bind at a time until a blocking group is cleaved through a chemical reaction. The modified nucleotides fluoresce upon binding, allowing us to observe the sequence. Combining smFRET with SMS gives significant power to survey the conformational dynamics over many different RNA sequences at the same time and link specific structural dynamics to specific sequences. Using these techniques together we will be able to survey BPS the match the consensus sequences and ones that do not. We

hypothesize that RNA molecules with BPS that do not match the consensus will have trouble binding Msl5 or SF1, which will block splicing[10] and will retain the intronic sequence.

Experimental Design

Work began on this project with sample preparation. The Ubc4 pre-mRNA was ordered and synthesized in two pieces due to yield constraints. Our first step after receiving the RNA was to add the fluorescent dyes, followed by joining, or ligating, the two pieces together. The complete Ubc4 construct was then subject to smFRET experiments. Starting with the RNA only, we optimized several variables including the density of RNA imaged on a slide and the fluorophores used. Once satisfied with the data collection, the data was processed and analyzed through many codes in Matlab and Mathematica. This allowed me to infer what FRET states the Ubc4 pre-mRNA exists at with no protein present. Due to time constraints, SMS data could not be included in this thesis. Future steps in this project will include SMS data and smFRET data with branchpoint binding proteins.

Results

RNA Ligation

We desired to study an RNA construct of a hundred thirty-five nucleotides with multiple modifications including a 5'biotin to enable immobilization on a microscope slide, fluorophore labeling sites on modified uridines, and random nucleotides (Figure 1B). Due to length and complexity constraints on yield, we ordered the RNA construct in two separate pieces. The 5' piece contained a 5' biotin modification and a fluorophore labeling site at its fourteenth nucleotide. The 3' piece contained one random nucleotide on either side of the branchpoint adenosine and a fluorophore labeling site on its fifty-second nucleotide. Both the fluorophore labeling sites were just inside the exon. Before joining the pieces together through a process called ligation, each piece was fluorescently labeled. Fluorescent dyes like Cy3 and Cy5, used for donor and acceptor respectively, or their analogs are often used to monitor as recent advances have decreased fluorescent blinking and photobleaching, which can make data interpretation less clear.[14] In the first iteration, we labeled the 5' piece with Cy3 and the 3' piece with Cy5. However, we found these dyes were not as photostable as we would have liked, and so we replaced them with Janelia Fluor 549 and Alexa Fluor 647, respectively. The pieces were then ligated together to obtain our full construct (Figure 2). Ligated RNA was purified via acrylamide gel electrophoresis and electroelution.

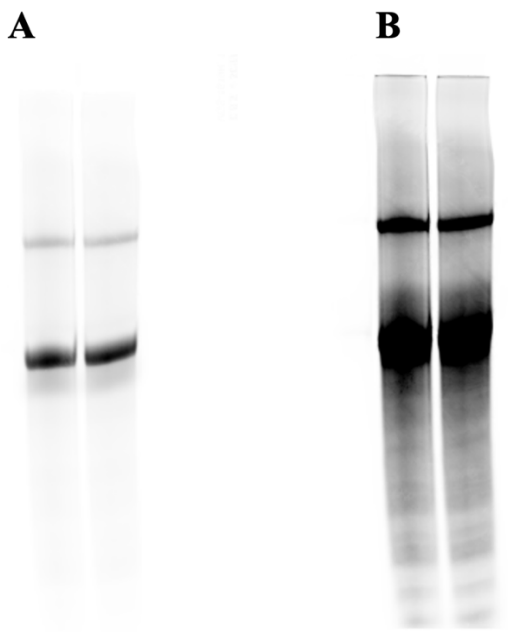

Figure 2: RNA ligation acrylamide gel visualized via fluorescent scan.

A: Cy3 scan of ligated Ubc4 pre-mRNA. Upper band represents the ligated RNA. The lower band represents unligated RNA from the reaction mixture. **B:** Cy5 scan of ligated Ubc4 pre-mRNA. Upper band represents the ligated RNA. The lower band represents unligated RNA from the reaction mixture.

smFRET

Data collection occurred over many iterations as there were many variables to adjust for optimization of the system. The raw data is a collection of fluorophore emissions takes the form of a video that ranges from two to five minutes (Figure 3). The acceptor and donor fluorophore emissions are split into two channels to independently measure the emissions of each. In the first iterations when using the Cy3 and Cy5 construct, we struggled with photobleaching of much of the RNA before two thousand frames, which prompted our switch to Janelia Fluor and Alexa Fluor. The new dyes were photostable beyond three thousand frames. The first iterations of our data collection were also collected at much too high density of RNA molecules (Figure 3A). By adjusting the concentration of the RNA added to our slide and the incubation time, we were able to adjust the density to an appropriate amount. When the RNA molecules are too dense on the

slide, they can be too close to each other so two molecules appear as one. When molecules are this close, the data is unreliable because you cannot be sure that FRET is not occurring between two molecules instead of within one molecule.

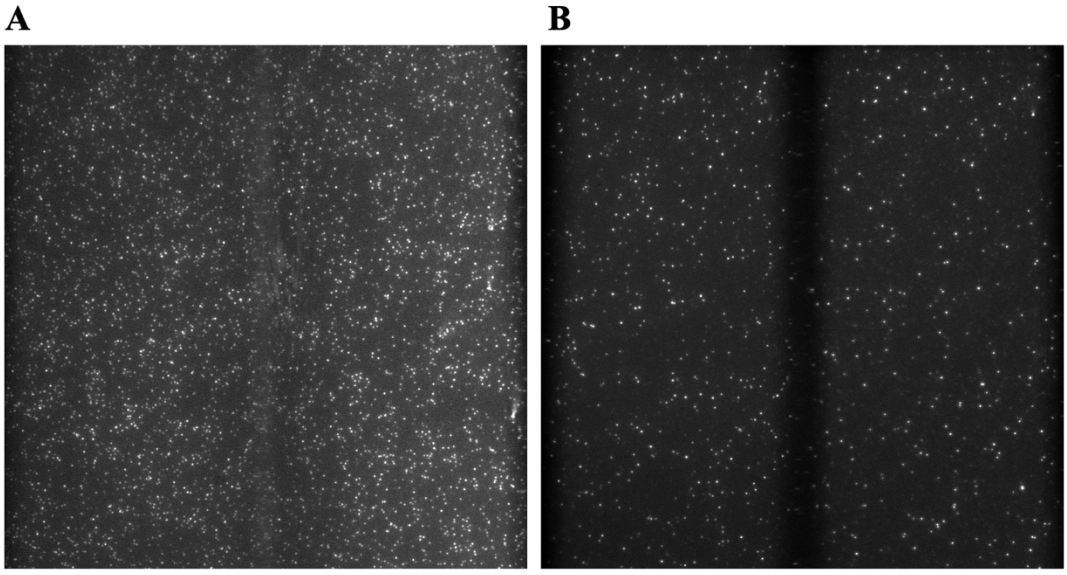

Figure 3: Comparison of density of raw data collection using Janelia Fluor and Alexa Fluor.

A: Example of high-density data collection iterations. The left-hand half of this image is the green channel; right is the red channel. **B:** Data collection used for data analysis present in this paper. The left-hand half is the green channel; right is the red channel.

Donor and acceptor fluorophore emission was measured, and the resulting images were mapped, so emissions from molecules in the green channel corresponded to emission at the same locations in the red channel. Fluorescence signals from individual RNA molecules, called "traces", were visualized via Matlab (Figure 4). Seventy-two traces were selected as usable and passed our criteria for trace selection. This criterion involves both set code parameters and judgment by visual analysis of traces. Most of the traces spent the majority of time in one state and would transiently switch to another state. This looked like a low FRET state transitioning into a higher FRET state or a high FRET state transitioning into a lower FRET state (Figure 4A). Other traces rapidly sampled several different FRET states and spent an around equal time in all

states (Figure 4B), while even others spent an equal amount of time in a low or high FRET state but transitioned much slower (Figure 4C-D). From the wide range of FRET behavior seen, we can assume that a wide variety of sequences sampled gives rise to a wide variety of structural behaviors and distance between the fluorophores over time. This logically makes sense and supports the existing literature that even a single base pair substitution can give rise to different conformational dynamics.[7]

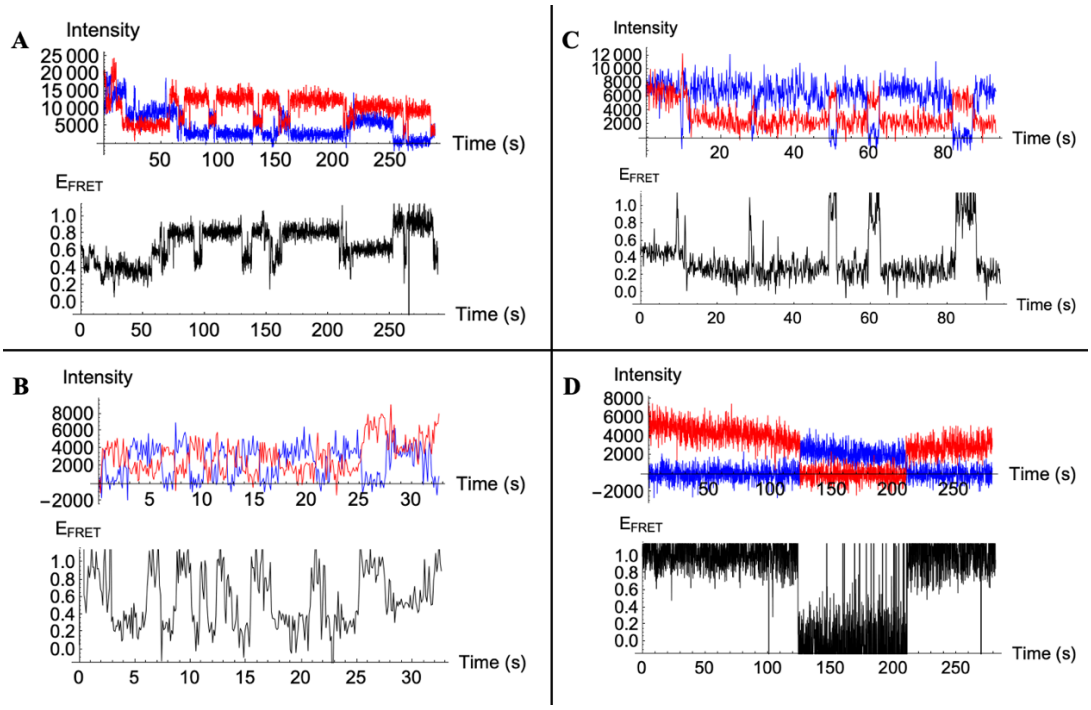

Figure 4: Example FRET traces of Ubc4 structural behaviors.

A: An RNA molecule occupying a high FRET state with brief transitions to a mid-FRET state. **B**: An RNA molecule that rapidly transitions from a low to high FRET state. **C**: An RNA molecule that spends roughly equal amounts of time in high and low FRET states. **D**: An RNA molecule that exists in a high or low FRET state for prolonged time periods.

From the seventy-two acceptable traces, we formed a histogram of the frequency of different FRET states over all their captured frames (Figure 5A). While maintaining structural heterogeneity in the data is important, displaying the data for all analyzed traces does allow us to see any larger patterns or trends within the sample. Fitting the histogram with Gaussian curves in

Mathematica (Figure 5B) can help distinguish the number and value of FRET states displayed. The Ubc4 pre-mRNA construct we surveyed mostly appears to range from a mid-to-high FRET state (Figure 5B). This seems to suggest that despite the randomness introduced within the BPS in our construct, the Ubc4 pre-mRNA naturally forms a folded secondary structure that brings the fluorophore, placed near the splice sites, in proximity.

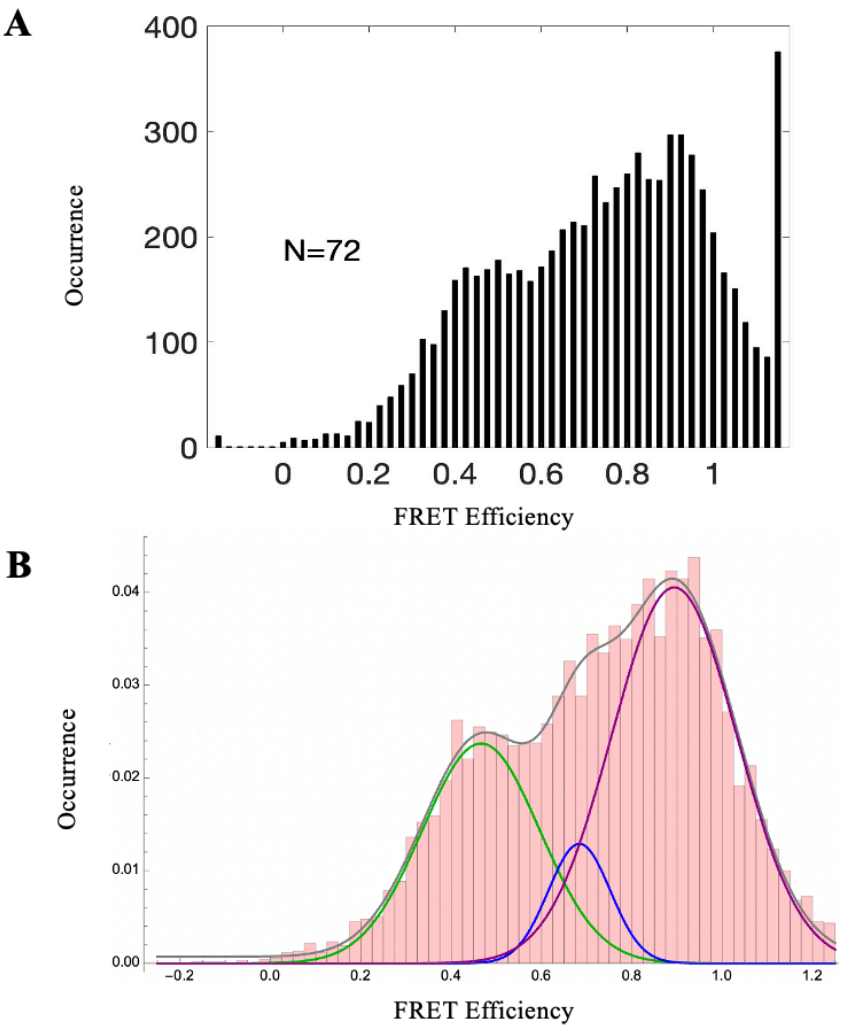

Figure 5: FRET Efficiency of Ubc4 pre-mRNA by frame occurrence.

A: Raw seventy-two traces that have passed our criteria for trace selection. The tall bar on the right represents all the occurrences of any FRET trace larger than one. **B:** Histogram fitted with a 3-Gaussian fit. The gray curve represents the sum of the individual green, blue, and purple curves.

Discussion

By using smFRET we were able to resolve structural heterogeneity of several different pre-mRNA sequences in a high throughput manner. We are using this method to investigate the structural dynamics of pre-mRNAs in splicing, but this powerful method can be applied to any number of constructs. When combined with SMS, it will reveal the structure of any number of different RNA sequences at the same time on the same slide. Defects in RNA expression are implicated in a number of debilitating diseases including: cancer, deafness, amyotrophic lateral sclerosis (ALS), autism, Huntington's disease, and others too numerous to list here.[15] Future research utilizing this method to address these diseases will be able to obtain data on many different RNA and potential mutants quickly and efficiently.

In previous experiments by Abelson et. al, it was revealed that a single base mutation in the 3' SS had larger effects on the secondary structure of the Ubc4 pre-mRNA than a single base mutation of the BPS adenosine. The Abelson et. al branchpoint mutant (BPM) construct and our BPS library (BPL) more closely mirror the wild-type pre-mRNA with FRET efficiency peaks at 0.4 and 1.0 than the 3' SS mutant (3' SSM).[7] However, there are still differences between the FRET efficiency histograms of the BPM, our BPL, and the wild-type pre-mRNA such as the increased amplitude of the peak at 1.0 FRET efficiency in the BPM and our BPL and decreased amplitude of lower FRET efficiencies in the BPM. This change may reflect our breadth of sequences sampled, as some of our constructs fit the wild-type BPS consensus while some differ in ways that may be similar to the BPM. It is intriguing that despite the important role the BPS plays during splicing, the secondary structure of the Ubc4 pre-mRNA is semi-resistant to changes made within the BPS compared to the 3' SS. The BPM exhibited no FRET efficiency changes between the pre-mRNA alone and with splicing extract whereas the wild-type and 3'

SSM both exhibited changes. This is likely because of the important role BPS plays not in majorly defining pre-mRNA secondary structure but in recruiting proteins to alter the pre-mRNA secondary structure which the BPM fails to accomplish.

Future Directions

Future work on this project will most immediately involve adding sequencing to the Ubc4 FRET data. This will allow us to correlate different FRET states with their matching sequences to determine which conformation different sequences spend their time in. The next step is to repeat this same experiment but introduce BPS binding proteins Mls5 and SF1. By introducing these proteins, we can investigate any structural differences that may help form the initiation complex of the spliceosome. Further work down the line will use the same experimental premise but throughout the splicing process. Different pre-mRNA molecules can be compared to link together sequence, structure, and splicing efficiency. Not only will it be interesting to compare the data from different pre-mRNA sequences, but also to compare to the Abelson et al. paper. Their sampled BPS mutant only varied in one nucleotide from the wild-type construct (the BPS adenosine was changed to a cytosine), which I predict will have more direct detrimental effects on splicing due to an inability to attack the splice sites than our constructs which do not vary this nucleotide. However, our varied sequence approach will allow us a broader and deeper understanding of how the BPS binds Msl5 and SF1 and how these interactions alter the pre-mRNA conformation to facilitate splicing.

Interestingly, it has been suggested that splicing of pre-mRNAs with non-consensus BPS is facilitated by phosphorylation of SF1 in humans.[16] Phosphorylation is a chemical modification made to proteins to affect their activity. This is hypothesized to occur to adjust gene expression in accordance with the environment. The occurrence of alternative splicing is also much higher

in mammals than compared to yeast.[9] Alternative splicing allows organisms to produce multiple different protein products from the same gene; this increases an organisms flexibility and responsiveness to the environment. Future projects or investigations could address this potential linkage between alternative splicing and SF1 phosphorylation.

Methods

Sample Preparation

First, the Ubc4 RNA was ordered in two pieces from Dharmacon with HPLC purification. This was done to maximize yield as each piece was moderately complex in length and modifications. Piece 1 contained a 5' biotin and fluorescent labeling site. Piece 2 contained two random nucleotides to either side of the branchpoint and a fluorescent labeling site. Each pieced was fluorescently labeled with Janelia Fluor and Alexa Fluor.

After fluorophore-labeling each piece of RNA, we ligated the pieces together using T4 RNA ligase 2. The DNA splint needed for enzymatic activity of T4 RNA ligase 2 was obtained from Integrated DNA Technologies. To ligate the pieces together, we combined both pieces and the splint in equal parts then heated the mixture to 90°C for 2 minutes to anneal. After annealing, we added buffer, water, ATP, and T4 RNA ligase 2 and incubated for 3 hours at 25° C. Ligated RNA was purified via 10% acrylamide gel electrophoresis followed by electroelution.

Slide Preparation

Microscope slides with holes drilled diagonally from each other in center were cleaned of adhesive by soaking in acetone 24-48 hours and scraping the debris off with a razor. Then, the slides were scrubbed with alconox paste for 20 seconds. Without rinsing, they were placed in a coplin jar with milliQ water and sonicated for 30 minutes. The slides were then rinsed with water, methanol, and water and sonicated in methanol. After 10 minutes, the slides were rinsed with water then sonicated in 1 M potassium hydroxide for 20 minutes. They were then rinsed with water again and then boiled for 20 minutes in a 20 mL solution of 29% NH_4OH and 30%

H_2O_2 with 100 mL of water. After rinsing the slides again with water, they were dried with nitrogen gas. Finally, rinse slides twice in the coplin jar with acetone.

To coat the slide in amine groups, the slides were immersed in a solution of 98 mL of acetone and 2 mL (3-aminopropyl)triethoxysilane for 10 minutes then sonicated for 1 minute. After sitting for an additional 10 minutes, the slides were rinsed with water and dried with nitrogen gas. Following the amine coating, the slide was then coated in PEG. First, the PEGylation reaction mixture was prepared by mixing 8 mg biotin-PEG, 80 mg mPEG and 500 uL bicarbonate PEGylation buffer (84 mg sodium bicarbonate and 10 mL water) and centrifuging to pellet sediment. 70 uL of this solution was placed onto the slide surface and covered with a coverslip. After a minimum 2-hour incubation in the dark at room temperature, the slides were rinsed with water and dried with nitrogen gas. The final solution, DST, is placed to neutralize remaining amine groups. The DST reaction mixture was prepared by mixing 10 mg DST and 500 uL bicarbonate buffer (840 mg sodium bicarbonate and 10 mL water) and centrifuged to pellet sediment. 60 uL of this solution was placed onto the slide surface and covered with a coverslip. After a 20-minute incubation in the dark at room temperature, the slides were rinsed with water and dried with nitrogen gas.

Sample chambers were assembled as follows. Double sided tape was placed diagonally on the slide on other side of the holes. A coverslip was placed on top of the tape and holes. The excess tape was removed with a razor blade. To seal the sample chamber, epoxy was mixed and carefully applied to areas not sealed by the tape. Flipping the slide over, pipette tips were trimmed and placed into the diagonal holes. This slide-pipette tip junction was also sealed with epoxy to prevent leaks. A thin tube was added to the top of 1 pipette tip and also sealed with epoxy. Prepared slides were stored in the dark at room temperature.

smFRET

When ready to perform experiments, the previously prepared slides were subjected to the following treatment. The slides were first checked for leaks by flowing 100 uL of T50 buffer (10 mM Tris HCl pH 8.0, 50 mM NaCl) into the pipette tip without the tube. A beaker should be placed under the tube to catch liquid waste. If no leaks were found, 100 uL of 0.2 mg/mL biotinylated-BSA in T50 was flowed in and incubated for 10 minutes. The channel was then flushed with T50. 100 uL of 0.2 mg/mL biotinylated-BSA in T50 was flowed in and incubated for 10 minutes. The channel was then flushed with T50. 100 µL of 0.2 mg/mL streptavidin was flowed in and incubated for 10 minutes. The channel was then flushed with T50. Finally, the Ubc4 RNA construct was flowed in and incubated for around 5 minutes until the desired density was achieved. For additional photostability and to prevent photoblinking, an oxygen scavenger system and Trolox were added after RNA immobilization.[7]

Slides were visualized using prism-based total internal reflection microscopy[13] using a 63x oil immersion objective. A 10-frame movie of a slide coated with fluorescent microspheres would first be taken every day of data collection. This procedure allows for minute adjustments of the laser path and prism placement as well as obtaining accurate mapping data. A green laser was focused on the slide at a power of 1 mW using the microscope's eyepiece. Once the focus was obtained, the camera's EM gain setting was adjusted to receive the microscope image. Minute changes to the focus may be made between the eyepiece and camera image. The bead slide did not require an EM gain, or one could use a small EM gain because of the high intensity of the fluorescence. Bead movies were captured with around 15 mW of green laser power. Sample movies were collected at 50 mW green laser power and 15 mW red laser power with an EM gain of 200. The sample movies generally contained around 3000 frames, each taken at 0.1

seconds and the red laser turned off after 50 frames and turned back on again 100 frames from the end of the movie.

Data Analysis

The collected movie was put through a series of Matlab codes to extract each RNA trace. First, the mapping file was created from the bead movie using the SM_MovieMapper code. This code generates files used for correlating the spots in the green channel to spots in the red channel and is applied to sample movies downstream in the process. Both the bead slide and sample movie were put into FIJI to find the maxima or locations of all the spots. The next code ran was SM_analysis which extracts the traces from the collected movie using the map generated from the bead movie. The final code in Matlab was Trace_analysis which matches mapped peaks to intensities, removes traces that don't fit specified conditions, and allows the user to sort through the remaining traces. Good traces exhibit anti-correlation between donor and acceptor molecules, 1-step photobleaching if bleaching is exhibited, and have low noise. Once traces are selected, this code will also output a histogram of occurrence of FRET efficiencies.

Histogram were fit with a Gaussian curve in Mathematica. Multiple Gaussians were tested to determine the best fit. Ultimately, a 3-Gaussian fit was selected due to the low Bayesian Information Criterion and high R-squared value. Visualizing the summed curve as well as the individual curves allows us to see the FRET states the RNA is occupying.

Bibliography

(1) Morris, K. V.; Mattick, J. S. The Rise of Regulatory RNA. Nat. Rev. Genet. 2014, 15 (6), 423–437. https://doi.org/10.1038/nrg3722.

(2) Crisci, A.; Raleff, F.; Bagdiul, I.; Raabe, M.; Urlaub, H.; Rain, J.-C.; Krämer, A. Mammalian Splicing Factor SF1 Interacts with SURP Domains of U2 snRNP-Associated Proteins. Nucleic Acids Res. 2015, 43 (21), 10456–10473. https://doi.org/10.1093/nar/gkv952.

(3) Papasaikas, P.; Valcárcel, J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. Trends Biochem. Sci. 2016, 41 (1), 33–45. https://doi.org/10.1016/j.tibs.2015.11.003.

(4) Jurica, M. S.; Moore, M. J. Pre-mRNA Splicing: Awash in a Sea of Proteins. Mol. Cell 2003, 12 (1), 5–14. https://doi.org/10.1016/S1097-2765(03)00270-3.

(5) Staley, J. P.; Guthrie, C. Mechanical Devices of the Spliceosome: Motors, Clocks, Springs, and Things. Cell 1998, 92 (3), 315–326. https://doi.org/10.1016/S0092-8674(00)80925-3.

(6) Suñé-Pou, M.; Prieto-Sánchez, S.; Boyero-Corral, S.; Moreno-Castro, C.; El Yousfi, Y.; Suñé-Negre, J. M.; Hernández-Munain, C.; Suñé, C. Targeting Splicing in the Treatment of Human Disease. Genes 2017, 8 (3), 87. https://doi.org/10.3390/genes8030087.

(7) Abelson, J.; Blanco, M.; Ditzler, M. A.; Fuller, F.; Aravamudhan, P.; Wood, M.; Villa, T.; Ryan, D. E.; Pleiss, J. A.; Maeder, C.; Guthrie, C.; Walter, N. G. Conformational Dynamics of Single Pre-mRNA Molecules during in Vitro Splicing. Nat. Struct. Mol. Biol. 2010, 17 (4), 504–512. https://doi.org/10.1038/nsmb.1767.

(8) Vijayraghavan, U.; Parker, R.; Tamm, J.; Iimura, Y.; Rossi, J.; Abelson, J.; Guthrie, C. Mutations in Conserved Intron Sequences Affect Multiple Steps in the Yeast Splicing Pathway, Particularly Assembly of the Spliceosome. EMBO J. 1986, 5 (7), 1683–1695.

(9) Nguyen, H.; Das, U.; Xie, J. Genome-Wide Evolution of Wobble Base-Pairing Nucleotides of Branchpoint Motifs with Increasing Organismal Complexity. RNA Biol. 2020, 17 (3), 311–324. https://doi.org/10.1080/15476286.2019.1697548.

(10) Jacewicz, A.; Chico, L.; Smith, P.; Schwer, B.; Shuman, S. Structural Basis for Recognition of Intron Branchpoint RNA by Yeast Msl5 and Selective Effects of Interfacial Mutations on Splicing of Yeast Pre-mRNAs. RNA 2015, 21 (3), 401–414. https://doi.org/10.1261/rna.048942.114.

(11) Love, S. L.; Emerson, J. D.; Koide, K.; Hoskins, A. A. Pre-mRNA Splicing-Associated Diseases and Therapies. RNA Biol. 2023, 20 (1), 525–538. https://doi.org/10.1080/15476286.2023.2239601.

(12) Walter, N. G.; Huang, C.-Y.; Manzo, A. J.; Sobhy, M. A. Do-It-Yourself Guide: How to Use the Modern Single-Molecule Toolkit. Nat. Methods 2008, 5 (6), 475–489. https://doi.org/10.1038/nmeth.1215.

(13) Roy, R.; Hohng, S.; Ha, T. A Practical Guide to Single-Molecule FRET. Nat. Methods 2008, 5 (6), 507–516. https://doi.org/10.1038/nmeth.1208.

(14) Rasnik, I.; McKinney, S. A.; Ha, T. Nonblinking and Long-Lasting Single-Molecule Fluorescence Imaging. Nat. Methods 2006, 3 (11), 891–893. https://doi.org/10.1038/nmeth934.

(15) Cooper, T. A.; Wan, L.; Dreyfuss, G. RNA and Disease. Cell 2009, 136 (4), 777–793. https://doi.org/10.1016/j.cell.2009.02.011.

(16) Lipp, J. J.; Marvin, M. C.; Shokat, K. M.; Guthrie, C. SR Protein Kinases Promote Splicing of Nonconsensus Introns. Nat. Struct. Mol. Biol. 2015, 22 (8), 611–617. https://doi.org/10.1038/nsmb.3057.