SPLICING AT SINGLE MOLECULE RESOLUTION: PRE-MRNA DYNAMICS THROUGHOUT SPLICEOSOME ASSEMBLY AND CATALYSIS

by

Mario Reynaldo Blanco

A dissertation submitted in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2013

Docotoral Committee:

Professor Nils G. Walter, Chair
Professor David R. Engelke
Assistant Professor Aaron C. Goldstrohm
Professor Edgar Meyhofer

Acknowledgements

I would like to thank my family, advisor, and Walter lab members for their support and encouragement. Nils has given me the opportunity to share in the exciting development of a new field of research. None of the work presented in this thesis would be possible without the excellent collaborations Nils has fostered and developed. In particular, I would like to thank our collaborators John Abelson and Christine Guthrie. Their expertise and insight has proven critical throughout the development of the techniques presented here. As members of the splicing group in the Walter lab, Ramya Krishnan and Matthew Kahlscheuer have contributed significantly to the experiments presented here. Working together we have been able to endure and overcome the challenges of developing single molecule assays for splicing.

Table of Contents

List of Figures:

List of Tables:

List of Appendices:

Abstract

The spliceosome is a multi-megadalton RNA-protein complex that catalyzes the removal of introns and the ligation of exons during pre-mRNA splicing. In humans, approximately 95% of all pre-mRNAs undergo alternative splicing, which allows for the dynamic expression of various protein isoforms from a single gene through cell- and tissue-specific networks of regulated splicing events. Catalytic activation of the spliceosome involves an intricate set of RNA:RNA and RNA:protein interactions which must be carefully coordinated at various points of assembly. The pre-mRNA substrate is an integral component of the catalytically competent spliceosome. It serves as both a scaffold for assembly and provides the reactive sites for chemistry. Despite 25 years of study, questions about the specific conformational rearrangements and particularly their kinetics leading to catalysis remain unanswered. We have developed single molecule fluorescence resonance energy transfer (smFRET) assays that have begun to dissect pre-mRNA conformational changes during splicing. These assays allow us to track the relative position of conserved splice sites in real-time throughout spliceosome assembly and catalysis. We observed a series of reversible conformational states that were dependent on *bona fide* splicing signals and the presence of ATP. Furthermore, we dissected the sequence of conformational changes that lead to the first step of catalysis. Assigning pre-mRNA conformations to specific assembly events required the development of novel analysis methods that have broad applicability to the rest of the single molecule field. Finally, we utilized recently developed single molecule

immunopurification techniques to help isolate a pre-catalytic spliceosomal complex. We then tracked the dynamics of this complex as it was chased through the first step of catalysis. We determined that, much like the ribosome, trans-acting factors influence catalysis by acting on the intrinsic dynamics of the spliceosome. By establishing smFRET as a viable tool for the investigation of splicing we have opened the field to new experimental possibilities. The results from these studies are already challenging some canonical assumptions, providing evidence for others, as well as providing new avenues of investigation.

Chapter I: Introduction

I.1 RNA in the driver's seat

Nothing in cell biology makes sense except in the light of the RNA world. Our increasing understanding of RNA's role at every stage of the gene expression pathway highlights the importance of this molecule in the life of a cell. RNA is no longer relegated to a simple intermediate between DNA and proteins. Rather, we understand that RNA has a role that spans from deep within the nucleus where it interacts with chromatin all the way out to the extracellular environment where it can act as a signaling molecule.[1,2] The chemical properties of RNA give it the unique ability to be such a versatile biomolecule. It can interact in a sequence specific manner with other DNA and RNA molecules to target nucleic acid modifying complexes like the spliceosome and RISC complex.[3,4] In addition, the ability of RNA to fold into stable structures allows it to serve as a scaffold for protein assembly such as telomerase and the polycomb repressive complex.[5,6] Furthermore, it can serve as the information carrier that it was first identified as, by transporting the genetic information from the nucleus to the cytoplasm for translation by the ribosome, the protein synthesizing macromolecular complex whose activity is dependent on the catalytic RNA within its proteinaceous shell.

There are many examples of complexes utilizing more than one of RNA's potential functions. The spliceosome, a macromolecular complex responsible for the bulk of all RNA

splicing, serves as perhaps the best example of a complex that utilizes nearly all RNA functions in a cell to complete its task. The spliceosome's RNA elements serve as a scaffold for assembly, utilizing sequence specificity to recognize and modify a pre-messenger RNA (pre-mRNA) molecule through what may be an RNA catalyzed mechanism. A thorough understanding of the spliceosome's function is key to unraveling its role in the gene expression pathway, disease, and newly discovered roles in gene regulation. Furthermore, it can provide us with information about how other complexes in the cell utilize the basic functions of RNA to enact control over nearly all cellular processes.

Introns: the heart of our genetic dark matter

Solving the structure of DNA helped establish the mechanism by which it could serve as our genetic material. The base pairs held information in the form of nucleotides (A, T, G, and C) that could be interpreted as code for protein production. The transfer of information from DNA to protein required an intermediate molecule, RNA, that is then used for protein synthesis.[7] This sequence of events eventually became known as the central dogma of molecular biology commonly paraphrased as: DNA makes RNA (through transcription), RNA makes proteins (through translation).[8] Together with previously established genetics, the molecular picture of heredity and gene expression were at last coming together. The assumption was that DNA was the stable molecule that contained the genes, which explained genetic inheritance, and that RNA, an intrinsically unstable molecule, would be a direct copy of the genetic material that could be used as a temporary messenger for protein synthesis.

The simplified view that the messenger RNA molecule was a direct copy of the DNA segment from which it derived was challenged by the discovery of processing steps that joined

distal parts of the gene into a single product. In 1977, Phillip Sharp and Richard Roberts independently and simultaneously discovered the presence of introns, or intragenic regions, within genes.[9,10] By utilizing electron microscopy they observed what Richard Roberts referred to as an "amazing rearrangement" of the 5' end an adenovirus messenger RNA. Hybridization of the messenger RNA product with its gene of origin made it evident that genes were not continuous bits of information but rather segments of exons, expressed sequences, interrupted by introns. Initially it was unknown if this was a general processing step or specific to this RNA. We now know that the removal of introns from pre-mRNAs to create a mature messenger RNA (mRNA) via splicing is a key distinguishing feature of higher organisms and plays a key role in the expression, regulation, and diversification of our genetic code.

Why genes in pieces?

A discontinuous product from the genetic code makes one wonder, like Walter Gilbert did in 1978: "why genes in pieces?"[11] The amount of genetic material removed during the splicing process is significant. In humans, introns comprise 90% of the RNA sequence from protein coding genes with an average of approximately 8 introns per protein coding gene.[12] Why would evolutionary pressures not have selected for the loss of introns if they are simply discarded portions of genes? Evidence suggests that introns are important for the eukaryotic way of life; in fact, there has not been a sequenced genome of a eukaryotic organism that has no introns.[13] From their prevalence in eukaryotic genomes and their maintenance throughout evolution it is clear that there is a significant role for the splicing process in complex organisms.

One immediate consequence of segmented genes is a challenge to the one-gene one-protein hypothesis.[14] The one-gene one-protein hypothesis holds that the unit of inheritance is the gene, and this gene is responsible for a single protein that contributes to some cellular function.

With the introduction of introns into genes, the ability for combinatorial expression of different segments of genes to create multiple products from a single gene became evident. Interestingly, Gilbert correctly hinted at this possibility and the role it may have in the immune system, a hypothesis later confirmed with experimental evidence.[15] The synthesis of multiple products from a single gene through the combination of different exons is known as alternative splicing. Alternative splicing allows for genomes to remain compact within increasingly complex organisms as well as provides a possible point of regulation by introducing an additional processing step that can be temporally and spatially controlled. Nearly 95% of mammalian genes are affected by alternative splicing, making it a general mechanism of regulation and genomic diversity in higher organisms.

Additional roles for introns in cells have become more prevalent as high throughput sequencing technology has allowed us to detect more non-protein coding RNAs (npcRNAs). It is estimated that ~98% of the transcriptional output in humans is npcRNAs.[16] Some of these RNAs are transcribed from well-defined transcriptional units but many are derived from introns. A special subclass of these RNAs are intron derived microRNAs or mirtrons. [17] microRNAs are a class of npcRNAs that act to regulate gene expression through translational repression and targeted mRNA decay. The canonical miRNA processing pathway involves the transcription of a primary miRNA by RNA polymerase II and cleavage by a series of enzymes that assemble the final ~22nt long RNA into a repressive complex. A large number of mirtrons have been discovered in the human genome through the careful analysis of sequencing data.[18] This new role for introns, as harbors of npcRNAs, provides additional justification for the retention of introns in our genomes.

Due to splicing's central role in the regulation and expression of genes any mutation that disrupts normal function can have dramatic effects. In humans, it is estimated that 60% of disease causing mutations affect splicing.[19] Thorough investigations elucidating the mechanism underlying RNA splicing will allow us to understand the effect these mutations have on the splicing process and to rectify these mutations through targeted therapies.[20]

I.2 The spliceosome catalyzes pre-mRNA splicing in eukaryotes

Eukaryotic pre-mRNA splicing is catalyzed by a multi-megadalton ribonucleoprotein (RNP) complex known as the splicesome.[21] The spliceosome's function is conserved throughout eukarya. Some of the most conserved cellular proteins and RNAs are components of the spliceosome.[22] Mechanistically, splicing catalysis is the same in single-celled eukaryotes like yeast to multicellular organisms like humans. The high degree of conservation has allowed for the use of a genetically tractable model organism like budding (Baker's) yeast to determine the pre-mRNA signals, protein components, and assembly processes required during splicing.[21, 23, 24]

In the budding yeast, *Saccharomyces cerevisiae,* the spliceosome is composed of about 80 proteins and 5 small nuclear RNAs (snRNAs) namely U1, U2, U4, U5 and U6. The pre-mRNA is recognized by the presence of three conserved sequences particularly the 5' splice site (5'SS), branch site (BS) and the 3' splice site (3'SS) (**Fig. I.1**). The first step of chemistry occurs when the 2'-hydroxyl of the conserved adenosine in the BS attacks the 5'SS resulting in a free 5' exon and an intron- 'lariat' intermediate with a 2'-5' phosphodiester bond at the branchsite adenosine (**Fig I.2**). During the second step of catalysis, the 5' exon and 3' exon are ligated with concomitant cleavage of the intron, yielding spliced mRNA.

Figure 1. 1: Conserved pre–mRNA splice sites.

The 5' splice site (5'SS), branchsite (BS) , and 3' splice site (3'SS) are all required for splicing. These sequences are nearly invariant in budding yeast but are more degenerate in metazoans. The branchsite adenosine responsible for the first step of chemistry is underlined.

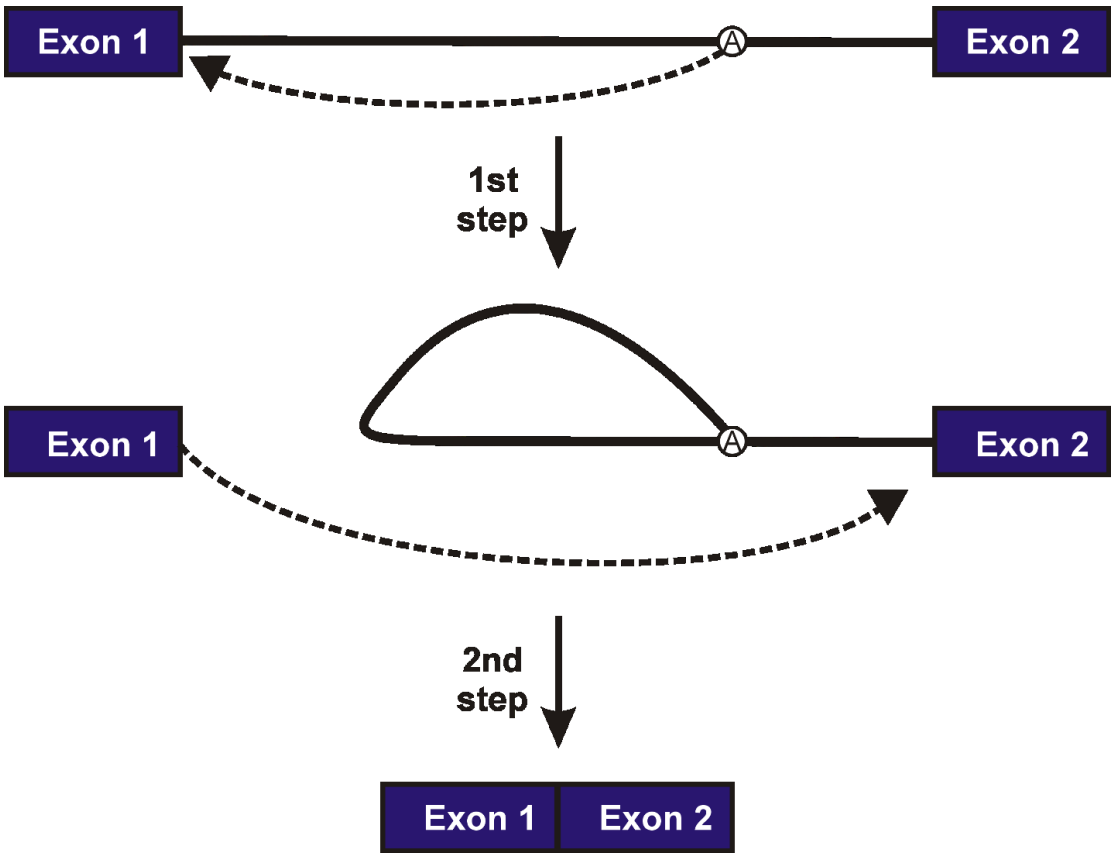


Figure 1. 2: The two-step splicing mechanism.

The first step of splicing is carried out by the attachment of the 5'SS to the BS adenosine, encircled for clarity. This yields a free 5'-exon and lariat intermediate. The first step of splicing utilizes the 2' hydroxyl of the branchsite adenosine for nucleophilic attack and does not require an intact 3'SS. In the second step of splicing, the 3' hydroxyl of the 5'SS attacks the 3'SS to achieve exon ligation and intron excision.

The spliceosome differs from many other complexes in that it does not exist as a preformed complex; rather it forms anew on pre-mRNA substrates where it assembles in a stepwise fashion. Spliceosome assembly is an ATP driven process that can be divided into four major steps: assembly, activation, catalysis, and disassembly. During assembly, the 5'SS is recognized by the U1 snRNP through base-pairing interactions between the U1 snRNA and 5'SS (**Fig. I.3)**. In the earliest ATP dependent event the U2 snRNP is assembled on the BS through the action of Sub2 and Prp5 RNA-dependent ATPases. Together the U1 and U2 snRNP form a stable complex that is termed the pre-spliceosome (Complex A). The tri-snRNP comprising of U4-U5-U6 snRNP is then recruited to the spliceosome, resulting in the formation of the pre-activated spliceosome (B complex). The activation of the spliceosome proceeds with the addition of an snRNA free complex known as the NineTeen Complex (NTC), which joins the spliceosome and through the action of Prp28 and Brr2 enact the removal of U1 and U4 snRNP, resulting in what is termed as the activated spliceosome (B^{act}). Although fully assembled, the spliceosome is still not poised for catalytic activity. Catalysis requires a significant reorganization of the spliceosomal core that is brought about through the action of Prp2. Prp2 consumes ATP resulting in the B* complex where the spliceosome is now poised for catalysis. The first step of splicing is then promoted by the activity of yet another RNA dependent ATPase Prp16. This yields the free 5'-exon and lariat intermediate that are indicative of the first step of splicing and formation of the C complex. The C complex can now finalize the ligation of the free 5' exon and 3' exon,the second step of splicing, which is dependent on the activity of Prp22. The spliceosome is disassembled and recycled for the next round of splicing, activities that are dependent on the action of Prp22 and Prp43.

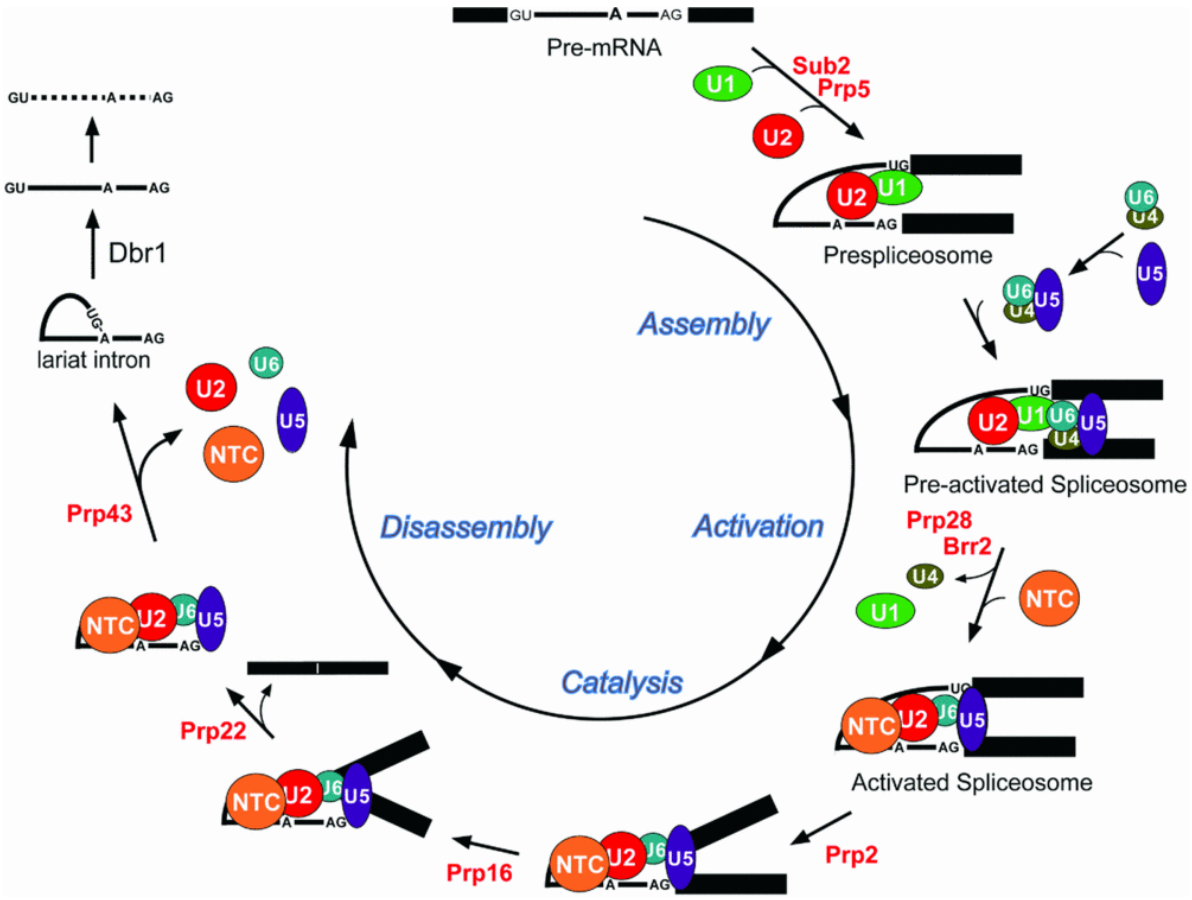


Figure 1. 3: The canonical spliceosome assembly pathway in yeast.

Figure reproduced with permission from Chen et al.[25] The canonical spliceosome assembly cycle in budding yeast proceeds through the stepwise ordered assembly of the subcomplexes U1, U2, U4:U5:U6, and NTC. Activation of the spliceosome involves a conformational rearrangement and loss of U1 and U4 snRNPs. Catalytic activation proceeds after the activity of DEXD-box helicases Prp2, Prp16, and Prp22.

The role of ATP in pre-mRNA splicing

Although the chemical steps of splicing are isoenergetic, ATP is consumed during the reaction. The role of ATP is to enact a series of conformational changes whose timing and coordination is controlled by a series of at least eight RNA dependent ATPases that are most similar to the DExD/H box family of helicases. The specific targets for these ATPases are still not known, yet well-defined roles for many of these proteins have emerged from biochemical and genetic studies in budding yeast.

Sub2 is frequently cited as the first ATPase to act in the splicing cycle. As is becoming more apparent for other ATPases in splicing, this protein acts in both an ATP independent and ATP dependent manner. The first role it serves is to stabilize the interaction of the Mud2 and Branchpoint Binding Protein (BBP) proteins with the BS during intron recognition in an ATP independent fasion.[26] In an ATP dependent fashion it then helps destabilize the Mud2-BBP heterodimer to allow for U2 snRNP binding at the BS.[27] Experimental evidence for an ATP dependent shift in pre-mRNA conformation that occurs before U2 snRNP binding is presented in **Chapter III** of this thesis. This shift in conformation may be reflective of Sub2's activity clearing the BS, making it accessible to the U2 snRNP. Binding of the U2 snRNP requires the action of yet another ATPase, Prp5. Prp5 is physically associated with the U2 snRNP and its ATPase activity is required for stable pairing between the U2 snRNA and the BS.[28] Once the U2 snRNP is stably associated, the U4/U5/U6 tri-snRNP is recruited to the spliceosome and the activity of yet another ATPase is required for conversion of the spliceosome to its activated form. The activated form of the spliceosome requires a large rearrangement of the RNA-RNA interaction network at the core of the spliceosome. The 5'SS:U1 snRNA interaction is replaced by the U6 snRNA (**Fig. I.4**). This shift in interactions requires the action of Prp28 and Brr2. Brr2

is responsible for the unwinding of the U6:U4 helix, which then frees the U6 to hybridize with the 5'SS and U2 snRNA. The exchange at the 5'SS is mediated by Prp28. Together, the action of these two ATPases allows the formation of the catalytic core of the spliceosome, a core that resembles the architecture that of the self-splicing group II introns.[29] The formation of the catalytic core leaves the spliceosome nearly competent for the first step of splicing. An additional conformational rearrangement carried out by Prp2 is required before the BS adenosine can attack the 5'SS. Prp2 was originally identified through a genetic screen that identified temperature sensitive mutations that affect pre-mRNA splicing.[24] The temperature sensitive allele Prp2-1 was identified that accumulated precursors. It was later determined that, although this mutant did not allow for catalysis, it did allow for assembly and activation of the spliceosome. Spliceosomes assembled and purified from heat inactivated Prp2-1 splicing extracts contain activated spliceosomes, but are not yet competent for catalysis, and are called B^{act}. The activity of these spliceosomes can be rescued through the addition of three factors, Prp2,Spp2, and Cwc25.[30] ATP hydrolysis by Prp2 causes a conformational change in the spliceosome that results in a shift in the sedimentation coefficient of the complex from 45S to 40S. This new conformation, termed B*, allows for inefficient first step catalysis that can be restored to wildtype levels by Cwc25. The mechanism by which Prp2/Spp2 and Cwc25 allow for first step catalysis is addressed in **Chapter IV**. Briefly, Prp2/Spp2 hydrolyze ATP to convert the B^{act} complex to B* which leads to a conformational rearrangement that now allows the 5'SS and BS to transiently interact. The presence of Cwc25 in the complex stabilizes the interaction between 5'SS and BS, allowing the reaction to proceed towards catalysis.

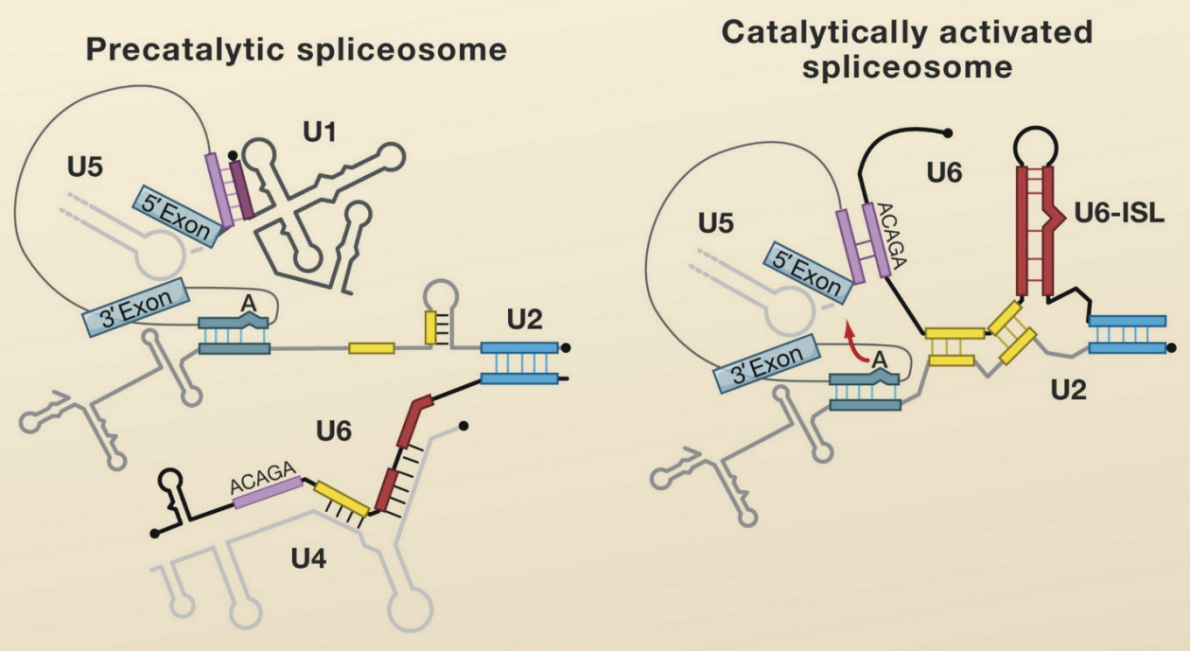


Figure 1.4: RNA-RNA rearrangements during catalytic activation.

The pre-mRNA is extensively base paired with the spliceosomal snRNAs before and after catalytic activation. In the pre-catalytic spliceosome (B complex) the 5'SS is occupied by the U1 snRNA and the BS by the U2 snRNA. Activation is achieved by the unwinding of the U6:U4 helix by Brr2 allowing U6 to base pair with the U2 snRNA. Additionally, the U1 snRNA is displaced from the 5'SS and replaced by the U6 snRNA by Prp28 (B^{act} complex). Figure reproduced with permission from Wahl *et al*.[3]

After the first step of catalysis, the splicesome enters what is call the C complex. This complex is responsible for the joining of the two exons and spliceosome disassembly. Prp16 hydrolyzes ATP to promote the transition towards the second step of catalysis. The main role of ATP hydrolysis is to eject some first step factors such as Cwc25 and allow for the binding of second step factors. The importance of Prp16 in this transition is emphasized by the discovery that substrates with mutations near the branchsite that are normally blocked from splicing are spliced efficiently in the presence of ATPase deficient Prp16 mutants.[31,32] This led to the hypothesis that Prp16's ATPase activity not only serves as a way to promote the second step of catalysis, but also to discourage aberrant splicing by proofreading the pre-mRNA substrate. The manner in which Prp16 accomplishes this is still not known, but the prevailing theory is that the ATPase activity of Prp16 acts as a timing mechanism that does not allow substrates to dwell in the first step conformation promoted by Prp2/Spp2/Cwc25. Suboptimal substrates are rejected because they do not splice fast enough to escape the conformational change cause by Prp16. Slow substrates are rejected and the DExD/H box ATPase Prp43 is recruited to disassemble the stalled complex.[31] This method of rejecting suboptimal substrates based on their reactivity has been termed the kinetic proofreading model that speculates that the spliceosomes are in dynamic states between conformations that either promote or disfavor catalysis. The ATPases act as timing mechanisms by enforcing kinetic constraints on splicing by changing the equilibrium among the available conformations.[33] Interestingly, Prp22, the ATPase responsible for the second step of splicing, seems to function in a similar fashion to Prp16. It promotes a rearrangement that favors exon ligation and then disassembly of the spliceosome. Substrates mutated at the 3'SS are blocked from splicing by the ATPase activity of Prp22.[34] Mutations that affect the ATPase activity of Prp22 now allow these mutated substrates to escape proofreading

leading to splicing. It is becoming evident that the spliceosome is utilizing the energy from ATP hydrolysis to control the coordination and timing of the rearrangements at its core to promote catalysis in a manner that prevents the deleterious effects aberrant splicing can have on a cell.

II.3 Single molecule FRET reveals the dynamics that are central to biological processes

Traditional biochemical and genetic tools utilized to identify and characterize the spliceosome ignore the inherently dynamic and heterogeneous nature of the process. The dynamic nature of the spliceosome is perhaps best highlighted by the discovery that both steps of splicing are chemically reversible.[35] This indicates that even after catalysis the spliceosome can toggle between conformations that change the chemical environment within its catalytic core. The inability to detect the dynamics of molecules in real-time leaves investigators with untestable hypothesis such as the kinetic proofreading model, whose sole premise is based on the dynamic sampling of conformational states. To help address this gap in knowledge we developed a single molecule Fluorescence Resonance Energy Transfer (smFRET) assay that allow us to peer into the core of the spliceosome by tracking the conformational dynamics of the pre-mRNA substrate. A prerequisite to chemical reversibility and the kinetic proofreading model is conformational reversibility. Utilizing our smFRET approach we provided the first evidence of the dynamic flexibility within the spliceosome by directly observing stable yet reversible conformations adopted by the pre-mRNA throughout splicing (**Chapter II**).

FRET: a sensitive tool for the study of structural dynamics

FRET is the non-irradiative energy transfer between two fluorescent molecules. In FRET a donor molecule enters its excited state through the absorption of a photon it then decays back to its ground state through a dipole-dipole interaction with an acceptor molecule. This interaction propels the acceptor molecule to its excited state, which then decays through the emission of a

photon. Because of the nature of the dipole-dipole interaction, the efficiency of the energy transfer is dependent on the inverse sixth distance between the donor and acceptor molecules. One can quickly calculate the relative efficiency of energy transfer by determining the fluorescence output of the donor and acceptor molecules. FRET is often summarized as a single value (E) that is calculated:

$$E=\frac{I_A}{I_A+I_D}$$

where I_A and I_D are the intensity of the acceptor and donor, respectively. The sensitivity of FRET to distance changes is dependent on the R_0 value, the distance at which energy transfer is 50% efficient (**Fig. I.5a**). Commonly used organic dyes, Cy3 (donor) and Cy5 (acceptor), have an R_0 value of ~54 Å, giving a dynamic range that spans from 20 Å to 100 Å. The conjugation of FRET pairs on biomolecules has allowed for the careful monitoring of relative distance changes in real-time. This technique was further extended to the study of single molecules through the advent of illumination technologies such as Total Internal Reflection (TIR) microscopy and exquisitely sensitive Charge Coupled Device (CCD) cameras (**Fig. 1.5b**).[36] In order to track a single molecule over time it is often immobilized on a solid surface suitable for microscopy such as a quartz slide. The immobilization gives us the ability to track the intensity changes in donor and acceptor intensities of individual molecules over time. (**Fig. 1.5c**) The changes in FRET over time reflect changes in the distance between the donor and acceptor molecules. Tracking these fluctuations provides insight into the role dynamics play in the function of biomolecules such as the catalysis of small ribozymes.[37,38] The ability to extract real-time dynamic information on structural changes has made smFRET a popular technique for the study of complex systems such as the ribosome, telomerase, and now the spliceosome.[39–41] The ribosome is perhaps the only macromolecular complex in the cell that rivals the spliceosome in complexity. Both

complexes represent a collision of the RNA and protein worlds. They are highly ordered RNA-protein complexes whose RNA element is buried deep within the core and in the case of the ribosome is responsible for chemistry.[42] Single molecule experiments on the ribosome have revealed the critical role dynamics play in protein synthesis. Thermal fluctuations intrinsic to the ribosome have been implicated in tRNA selection as well as translocation.[43,44] Clearly, a full understanding of the spliceosome's activity requires careful interrogation via single molecule techniques.

Only recently have the first single molecule experiments been conducted on the spliceosome. The first study from Melissa Moore's lab involved a simple proof of principle that showed that yeast pre-mRNA splicing activity could be observed under single molecule conditions.[45] Our study on pre-mRNA dynamics followed shortly thereafter, demonstrating that

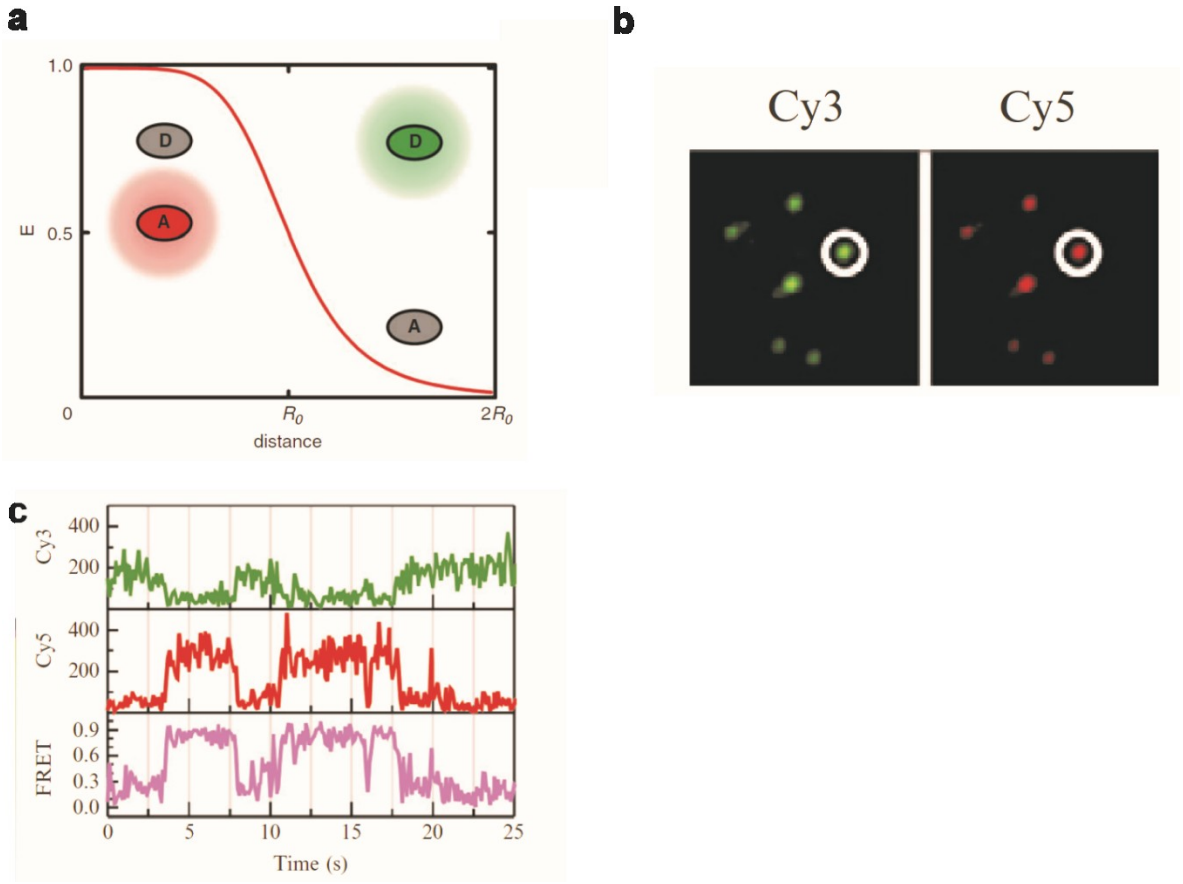


Figure 1.5: FRET from single molecules.

(a) The relationship of FRET efficiency (E) with R_0 distance. E decreases as Donor (D) fluorescence increases and increases as Acceptor (A) fluorescence increases.**(b)** An image captured from a CCD camera of Cy3 (donor) and Cy5 (acceptor) dually labeled single molecule. The fluorescent counts from each dye can be collected independently and used to calculate the FRET ratio. **(c)** Immobilizing molecules on a surface allows us to watch the dynamic fluctuations of that molecule. Here a single molecule's Cy3 (green) and Cy5 (red) intensity are tracked over a period of 25 seconds. The anti-correlated behavior of the signals are indicative of conformational changes in the molecule. Reproduced with permission from Johnson-Buck *et al*[46] and Blanco, M *et al*[47].

FRET could be utilized to track conformational changes during spliceosome assembly and catalysis.[41] We observed reversible conformational states that were dependent on *bona fide* splicing signals and the presence of ATP (**Chapter II)**. The following publication by the Moore lab, Hoskins *et al*, utilized labeled spliceosomal proteins in extract to track the assembly of the spliceosome in real-time. By tracking spliceosome assembly with single molecule resolution they were able to observe reversible binding and unbinding of individual components a result that

challenges the canonical assumption of a unidirectional assembly pathway.[48] In soon to be published work based on this thesis, we dissect the sequence of conformational changes that leads to the first step of catalysis. The lack of adequate single molecule analysis methods necessitated the development of a novel analysis technique that we expect will be adopted by other investigators utilizing single molecule tools to study systems with complex behaviors (**Chapter III**). Finally, we utilized recently developed single molecule immunopurification techniques to help isolate a pre-catalytic spliceosomal complex. We then tracked the dynamics of this complex as it was chased through the first step of catalysis. We determined that, much like the ribosome, trans-acting factors influence catalysis by acting on the intrinsic dynamics of the spliceosome (**Chapter IV**).

Within just a few years, the value of single molecule experiments on the spliceosome has been demonstrated. The results are challenging some canonical assumptions, providing evidence for others, as well as raising new questions. This should keep the splicing field busy for years to come.

Chapter II: Conformational Dynamics of Single pre-mRNA molecules during *in vitro* splicing[1]

II.1 Introduction

Introns in eukaryotic pre–mRNAs are removed by RNA splicing. This process is carried out by the spliceosome, a large supra–molecular assembly consisting of five small nuclear RNAs (snRNAs) and more than 100 proteins[3,49]. The spliceosome lacks a pre–formed catalytic core. Rather, spliceosomes are assembled on intron–containing substrates in a stepwise manner that requires both binding and release of small nuclear ribonucleoproteins (snRNPs; reviewed in Ref. 2). Notably, splicing entails two successive phosphoryl transfer reactions that are isoenergetic. Nonetheless, spliceosome assembly requires at least eight RNA–dependent ATPases of the so–called DEAD–box (DExD/H) family.[50] These enzymes, which catalyze successive rearrangements of RNA and protein pairing partners, are believed to enhance the fidelity of splicing by acting as proofreading clocks to allow the discard of mutant substrates at multiple steps in the assembly and catalytic pathway[34,51–53].

In the budding yeast *Saccharomyces cerevisiae*, assembly (reviewed in Ref. 2) is initiated by the ATP–independent formation of a commitment complex in which the U1 snRNP interacts

[1] Adapted from *Abelson, John, *Mario Blanco, Mark A. Ditzler, Franklin Fuller, Pavithra Aravamudhan, Mona Wood, Tommaso Villa, et al. "Conformational Dynamics of Single pre-mRNA Molecules During in Vitro Splicing." *Nature Structural & Molecular Biology* 17, no. 4 (April 2010): 504–512. doi:10.1038/nsmb.1767. *Contributed equally to this work

with the 5′ splice site (5′SS) while BBP and Mud2 interact with the branchpoint sequence (**Fig. 2.1**). If the branchpoint is mutated (BP mutant) this complex cannot form and spliceosome assembly is blocked. With a wild–type precursor (WT), the commitment complex is converted to the pre–spliceosome by the ATP–dependent binding of U2 to the branchpoint, followed by association of the U4/U6.U5 triple snRNP. Before the first transesterification can occur, U1 and U4 snRNPs must be released from the assembled spliceosome, again in ATP–dependent reactions. Importantly, the first chemical step, which results in formation of the lariat–intermediate and free 5′ exon, can occur even in the presence of a mutation at the 3′ splice site (3′SS). This mutation specifically blocks the second transesterification. Finally, the spliceosome undergoes ATP–dependent disassembly from the lariat intron and mature mRNA (**Fig. 2.1**).

Despite this level of understanding of the pathway, major questions about the specific conformational rearrangements and particularly their kinetics (and splicing signal dependence) remain unanswered. This gap in our knowledge is in large part due to the fact that to date splicing must be studied in a crude extract *in vitro*, where multi–step reversible processes are obscured by asynchronous progression along the pathway.[36] In addition, splicing in a yeast extract is generally inefficient, leading to only a fraction of molecules undergoing one or both steps of splicing. To overcome these severe limitations and monitor in real–time the conformational states of the pre–mRNA in spliceosome assembly, we have developed an *in vitro* assay based on single molecule fluorescence resonance energy transfer (smFRET) that obviates the need to isolate or synchronize reaction intermediates.

We first identified a natural yeast pre–mRNA with a small intron, Ubc4, which is efficiently spliced *in vitro*. This made possible the chemical synthesis of the pre–mRNA with donor (Cy 3) and acceptor (Cy 5) fluorophores in the exons adjacent to the 5′– and 3′– splice

sites, respectively. The Förster radius of the Cy3–Cy5 fluorophore pair is ~54 Å so that smFRET can effectively monitor changes in inter–fluorophore distance between ~20–100 Å. Cryo–EM micrographs of snRNPs and spliceosomal complexes range in size from 200–300 Å[54,55], yet cross–linking and footprinting indicate proximity of the two exons throughout splicing[56,57]. Using total internal reflection fluorescence microscopy (TIRFM) of immobilized Ubc4 we show here that the pre–mRNA reversibly folds into conformations with proximal exons even in the absence of splicing extract. Spliceosome assembly further accelerates and diversifies these conformational dynamics. We establish four criteria that suggest that we are indeed monitoring spliceosome assembly at the level of single molecules; an *in situ* assay verifies that the surface immobilized pre–mRNA is efficiently spliced; the observed complex smFRET transitions are time– and ATP–dependent; and they are dependent on appropriate intron signals. Our results suggest that the spliceosome is a cellular machine that operates close to equilibrium like the ribosome.

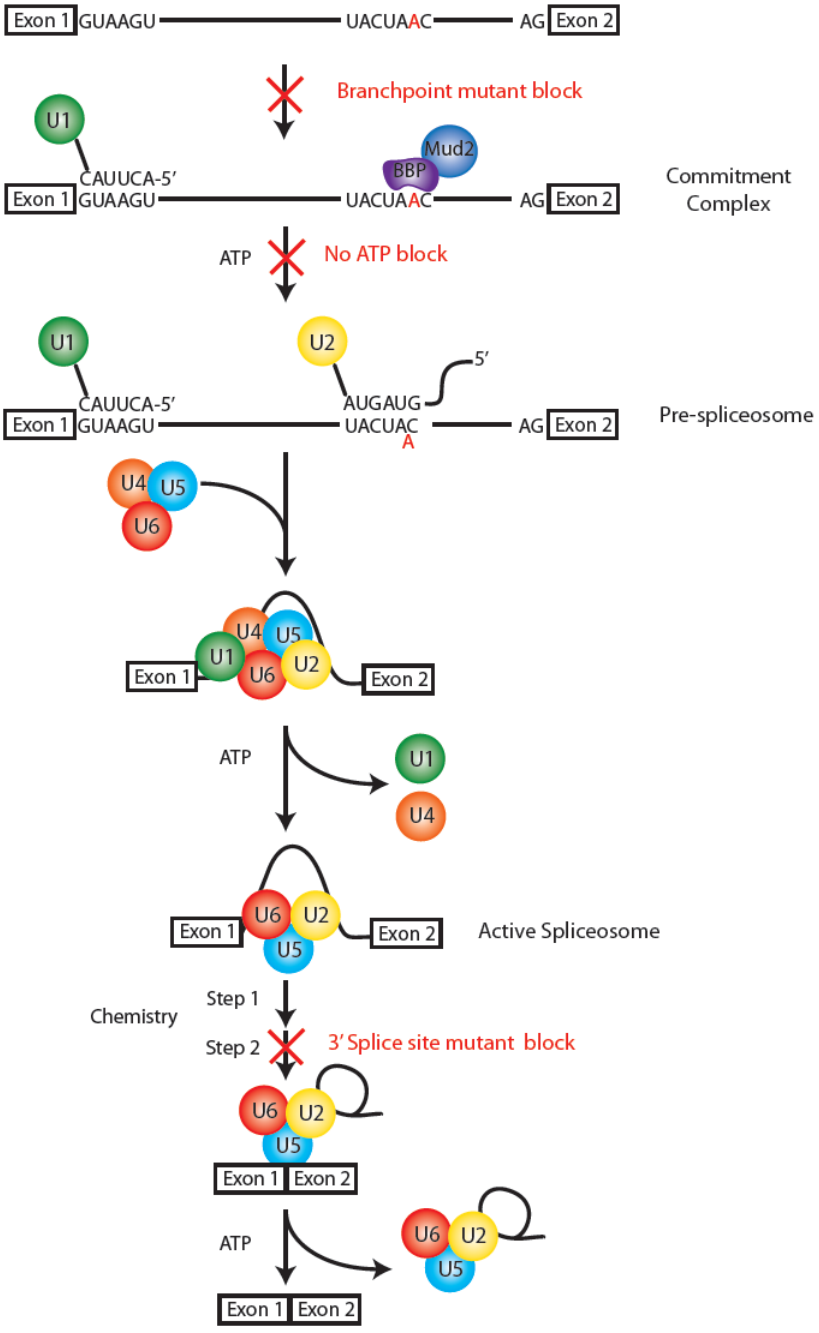


Figure 2. 1: Canonical spliceosome assembly pathway.

II.2 Materials and Methods

Microarray analysis of *prp2–1* versus wild–type.

RNA was extracted from *prp2-1* grown at the permissive temperature (2-1 RNA) and from *in vitro* splicing reactions in which the premRNA added to the reaction mixture was RNA isolated from *prp2-1* grown at the permissive temperature and then shifted to the nonpermissive temperature for 30 min. Splicing reactions containing 240 mg ml-1 of the RNA extracted from cells grown at the nonpermissive temperature were incubated for 30 min at room temperature with (+ATP RNA) and without ATP (- ATP RNA). cDNA synthesized from the 21 control RNA was labeled with Cy3 and and cDNA from +ATP RNA and ATP RNA with Cy5 as described in Pleiss et al. Splicing microarrays were hybridized with a mixture of 2-1 cDNA and +ATP cDNA or 2-1 cDNA and - ATP cDNA. After hybridization the microarrays were washed and analyzed for the ratio of +/- ATP to 2-1. For this analysis we only considered hybridization to the set of oligonucleotides specific for the mRNAs of genes containing introns. The order of the genes in Table 1 was determined by subtracting the rank orders of +ATP/2-1 minus -ATP/2-1 for each gene (for example Ubc4 was in the 64th percentile in the (+)ATP ratios and in the 0.8th percentile in the (-)ATP ratios. Ubc4 prem-RNA is second in this list and the canonical substrate for yeast *in vitro* pre-mRNA splicing, actin pre mRNA is 20th. Characterization of this microarray assay is described in Pleiss et.al[58]. Pleiss et.al. show that the splicing of most yeast introns is inhibited in the mutant *prp2–1* at the non–permissive temperature.*UBC4* was selected from among the top candidates from this microarray.

Synthesis and activity of Ubc4 pre–mRNA substrates.

The two oligonucleotides that comprise the Ubc4 pre–mRNA substrate; the 5′ segment of 76 nt and the 3′ segment of 59 nt, each containing a single 5–amino–allyl–uridine, were synthesized by Dharmacon. Mutant versions of these oligonucleotides were also synthesized. Half of each synthesis at a time was deprotected and then purified by electrophoresis on a denaturing, 7 M urea, 6% (w/v) polyacrylamide gel. The bands were identified by UV shadowing and eluted by overnight rotation in 0.3 M NaOAc pH 5.5, 1 mM EDTA, 0.1% SDS. Following elution the gel fragments were removed by centrifugation, the supernatant extracted with phenol chloroform, and the RNA ethanol precipitated. It is important to remove all Tris in this step because it can interfere with dye conjugation. For coupling, 2–5 nanomoles of each fragment were resuspended in 40 □l of 0.1 M sodium bicarbonate buffer, pH 9.0, and incubated for 60 min at 60 ^{o}C with one dye pack of the N–hydroxysuccinimidyl ester dissolved in DMSO. The 5′ and 3′ fragments were reacted with Cy3 and Cy5 N–hydroxysuccinimidyl ester, respectively. After dye coupling, the conjugated fragments were ethanol precipitated, washed until ethanol soluble dye was removed and then purified on a denaturing, 7 M urea, 20% (w/v) polyacrylamide gel. The conjugated oligonucleotide was eluted from the gel by overnight rotation and subsequently purified as described above. The overall yield of coupled oligonucleotide was generally ~2–5 nanomoles. The purified and conjugated oligonucleotides were ligated as described[59] using T4 RNA ligase I and a DNA splint (dSplint) from Integrated DNA Technologies. The ligated fragments were separated by gel electrophoresis and eluted as described above. The overall yield in this step is only 10–20%, but the resulting 200–500 picomoles suffice for a large number of smFRET experiments.

The 135 nt version of the *Ubc4* pre–mRNA with a 95 nt intron and 20 nt 5′ and 3′ exons has the 5-allyl-amine U residues in the 5’ exon and in the 3’ exon (positions, -7 and +4, respectively) In the 3′SS mutant, the bold underlined guanine was replaced with a cytosine. In the BP mutant the italicized underlined adenosine is replaced by a cytosine. The splicing activity of this fluorescently labeled Ubc4 pre–mRNA was determined using standard *in vitro* assays, wherein 5–10 fmoles of Ubc4 pre–mRNA at a final concentration of 0.2 nM was incubated in 40% (v/v) yeast whole cell extract containing 8 mM Hepes–KOH, pH 7.0, 2 mM $MgCl_2$, 0.08 mM EDTA, 60 mM $K_i(PO_4)$, 20 mM KCl, 8% (v/v) glycerol, 3% (w/v) PEG, 0.5 mM DTT, supplemented with 2 mM ATP, and the products were resolved via denaturing polyacrylamide gel electrophoresis and scanned on a Typhoon variable mode imager.

Preparation of yeast cell extract.

Splicing active whole cell extract was prepared as described previously from BJ2168 yeast cells[60] with the following exceptions. Cells were grown in 8 to 16 L of YPD medium to an OD600 of 2.5–3. The cells were then harvested and washed. Pellets of a thick slurry of the extract were obtained by dripping the slurry into liquid nitrogen. These pellets can be frozen at –70 ^{o}C for further use. The cell pellets were disrupted by grinding in a ball mill grinder (RMM301, Retsch) for 5 cycles of 3 min at 10 Hz. Between each cycle the canisters were cooled in liquid nitrogen. The frozen powder was thawed in an ice bucket and centrifuged at 34,950 g in a Ja–25 Beckman rotor; the supernatant was then centrifuged at 100,000g in a Ti–60 rotor for 1 h. The clear middle layer was removed using a syringe. This extract was dialyzed for 4 h against 20 mM Hepes–KOH, pH 7.0, 0.2 mM EDTA, 0.5 mM DTT, 50 mM KCl, 20% (v/v) glycerol with one buffer exchange. 40% (v/v) cell extract was used in all assays and in all single molecule studies.

ATP–depletion was achieved by adding 1 mM glucose to the cell extract and incubating at room temperature prior to each experiment.

Single molecule FRET.

Slides were prepared using a protocol modified from previous published methods[61,62]. Quartz slides were reacted with aminopropyltriethoxysilane (APTES) in acetone for 30 min to generate an amino functionalized surface, which was reacted overnight with a 10:1 mixture of succinimidyl ester functionalized O–methyl–PEG and biotin–PEG to PEGylate the surface. Sulfo–disuccinaimidlytartarate (sulfo–DST) was reacted for 30 min with the remaining unreacted amines, thus ensuring the surface does not carry a positive charge. Slide coverslips undergo a similar procedure. The slides were then rinsed, dried and a single flow channel per slide was assembled. A solution of 0.2 mg ml^{-1} streptavidin in buffer (50 mM Tris–HCl, pH 7.5, 50 mM NaCl) was added to the channel and incubated for 10 min at room temperature.

The doubly labeled pre–mRNA was heat annealed to a 2′–O–methyl–RNA capture (tether) strand complementary to the 17 3′ terminal nucleotides of the 3′ exon by incubating at 70 ˚C for 2 min and cooling to room temperature for 10 min. The capture strand carries a 5′ biotin, which binds to the streptavidin on the surface of the slide to immobilize the pre–mRNA. Following tether annealing, the hybrid was diluted to a concentration of ~50 pM in 100 μl splicing buffer and flowed into the slide and incubated for 10 min. Using a prism–based total internal reflection fluorescence (TIRF) microscope as described[38] data were collected from single molecules in splicing buffer (8 mM Hepes–KOH, pH 7.0, 2 mM $MgCl_2$, 0.08 mM EDTA, 60 mM $K_i(PO_4)$, 20 mM KCl, 8% (v/v) glycerol, 3% (w/v) PEG, 0.5 mM DTT), 40% (v/v) cell extract depleted of ATP or 40% (v/v) cell extract supplemented with 2 mM ATP, both in splicing buffer. An oxygen scavenger system composed of protocatechuate dioxygenase (PCD) and

protocatechuate (PCA) was added to splicing buffer and cell extracts to limit photobleaching[45], Trolox was also added to the solution to limit fluorophore blinking[63]. The Cy3 donor was exited using a 532 nm laser and emission by the Cy3 and Cy5 fluorophores was recorded at 100 ms time resolution using an intensified CCD camera (Princeton Instruments, I–Pentamax). A FRET value was then calculated by dividing the intensity of acceptor emission by the total emission from both donor and acceptor. It should be noted that our ability to observe FRET transitions has a lower boundary imposed by our TIRFM time resolution of 100 ms and an upper boundary derived from the time window before either of the fluorophores photobleaches (~100 s). This dynamic range of ~4 orders of magnitude resolves many, but probably not all kinetic differences among the pre–mRNAs, especially in light of the slow splicing kinetics *in vitro*. To further extend our observation time window, we therefore performed experiments involving extended pre–incubation with splicing active extract before commencing high–resolution (100 ms) smFRET detection.

Anti–correlation was detected in two ways. First, trajectories to be studied were pre–filtered by searching for the presence of any substantial anti–correlation by visual inspection. Second, transition events found within the pre–filtered FRET index trajectories were screened for localized anti–correlation in the donor and acceptor traces. The post—filtering system required the use of the Hidden Markov Modeling tools in the QuB software suite on the pre—filtered data (available at http://www.qub.buffalo.edu/soft.php). The raw donor, acceptor, and FRET indices were independently modeled to determine transition boundaries using a global fitting routine that simultaneously analyzes all data points in a given experimental condition (**Fig. 2.2**). The post—filter algorithm was coded and executed in a MATLAB environment. Alternatively, we carried out a preliminary analysis of individual trajectories in which we

explicitly optimized the rates of FRET transitions as in previous single—molecule FRET analysis[64,65]. Consistently, this analysis identified similar sequence and ATP–dependent changes. However, the large number of FRET states and trajectories in our data sets and the apparent kinetic heterogeneity ultimately resulted in a lower level of confidence in the resulting transition models.

Verification of splicing *in situ* by RNase H probing.

Each Ubc4 construct (WT, BP, and 3′SS, mRNA) was immobilized on a quartz slide, incubated with splicing buffer, ATP depleted extract, or ATP supplemented extract and monitored by smFRET as described above. Extract was washed out with 100 µl 1 M NaCl solution and incubation over 2 min, repeated three times. Following the final wash, the sample chamber was equilibrated with 100 µl RNase H buffer (20 mM HEPES–KOH, pH 7.8 at 25°C, 1 mM DTT, 50 mM KCl, 50% (v/v) glycerol, 0.2 mg ml^{-1} BSA) for 5 min, followed by incubation for 20 min with 100 µl of 250 µM Deoxyoligonucleotide dRH which is complementary to the Ubc4 intron in RNase H buffer. Next, a movie was taken for 50 frames with direct excitation of both Cy3 and Cy5 (′before RNase H′). The slide was then washed with 100 µl 250 µM dRH and 10 U RNase H (Promega) in RNase H buffer, was incubated for 30 min, and washed again with 100 µl 250 µM dRH in RNase H buffer. A second movie was taken as above (′after RNase H′). Movies collected before and after RNase H treatment were quantified using an in–house spot finder algorithm, which utilizes a second–order non–linear polynomial equation to correctly align the Cy3 and Cy5 channels, and quantified using a reference bead slide for calibration[39].

Secondary Structure Prediction of Ubc4 pre-mRNAs.

The lowest-energy secondary structures for each substrate was predicted using the by the Vienna software package(available at http://rna.tbi.univie.ac.at/). This software also calculates a partition

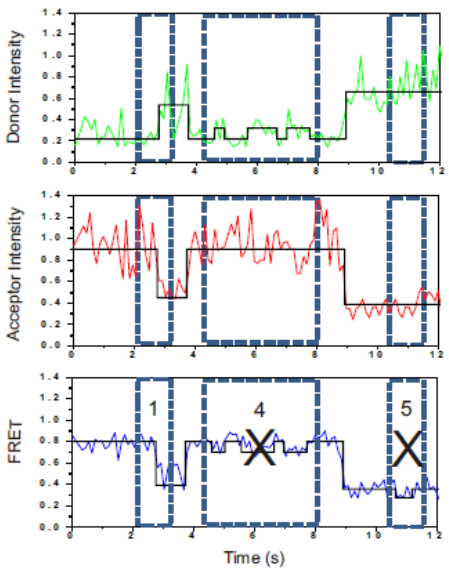
function of secondary structures. The software suboptimal structures that were within 5% of the Minimum Free Energy (MFE) for the WT and the two mutant pre-mRNAs. The suboptimal

a

Transition Scoring Classification:

Score	Event
1	Single transition in both channels in interval
2	Multiple transitions in one channel, single transition in other channel in interval
3	Multiple Transitions in both channels in interval
4	Transitions in one Channel, but not in the other in interval
5	No transitions in either channel

b



c

Substrate	Buffer (Percent of Scores1 -3/ Percent of Scores4 -5)	(-) ATP (Percent of Scores1 -3/ Percent of Scores4 -5)	(+) ATP < 15 min (Percent of Scores1 -3/ Percent of Scores4 -5)	(+) ATP > 60 min (Percent of Scores1 -3/ Percent of Scores4 -5)
WT	83.94 / 16.06	70.89 / 29.11	75.07 / 24.93	66.91 / 33.09
BP	82.89 / 17.11	68.99 / 31.01	63.50 / 36.5	64.65 / 35.35
3'SS	80.92 / 19.08	73.08 / 26.92	70.54 / 29.46	74.36 / 25.64

Figure 2. 2: Post-filtering analysis of Hidden Markov Modeling (HMM) data.

(a) A logical scoring criteria based on the number of transitions found in the Donor, Acceptor, and FRET channels was used to identify transitions of interest. In this study, transitions that had a score of 1-3 were used for further analysis.(b) Hidden Markov fits (black) are shown for independently analyzed Donor, Acceptor, and FRET channels. Transitions in the FRET channel are scored (inset) based on the criteria in Fig. 2.2a.(c) The percent of transitions that

were retained for analysis (Scores 1-3) or discarded (Scores 4-5) are displayed for each substrate and condition.

structures thus generated by the program represent a Boltzmann weighted population. The suboptimal structures thus produced were analyzed for the secondary structure distance between the fluorophores using the program developed by Rogic *et.al.,* with minor modifications to enable one to specify the fluorophore positions in the pre-mRNA sequence at the command prompt.

HMM analysis - Model Selection and Scoring Regime.

Multiple models with differing state numbers were used to determine the underlying FRET states; the entire data set for each condition was analyzed by the iterative application of the Viterbi and BaumWelch algorithms to generate idealized trajectories. The number of states assumed in the idealization was varied from 5 to 11 and the corresponding fits

were evaluated using the Bayesian information criterion (BIC). The number of states that resulted in the best BIC score was used in our analysis. BIC penalizes models with extra states that do not result in a significant improvement in the LogLikelihood; this allows us to select the most appropriate model of our data analysis by balancing goodness of fit and model parsimony. After idealization, the postfilter algorithm classified each FRET transition by counting the number and direction of transitions found in the donor and acceptor trajectories, within one quarter of the dwell time preceding the FRET index transition and one quarter of the dwell time after, or a minimum of 0.3 seconds in either direction. Logical classification was performed, scoring each transition based on the metric shown in **Figure 2.2**. In this work, transitions with scores of three or lower were used for further analysis with transition density plots. Additionally FRET transitions with a step size smaller than 0.1 were not included in our analysis.

II.3 Results

Identification and functional characterization of a suitable pre–mRNA

In *Sacchromyces cerevisiae*, most *in vitro* splicing assays have utilized the ~1,400 nucleotide (nt) actin pre–mRNA (with an intron of 308 nt)[66]. To develop our smFRET approach, however, we sought to identify a smaller pre–mRNA efficiently spliced *in vitro* that would be amenable to chemical synthesis. To this end, we analyzed pre–mRNAs from the temperature sensitive yeast mutant *prp2–1*[24].

In this mutant, splicing is blocked before the first catalytic step when cells are shifted to the non–permissive temperature, resulting in the efficient accumulation of most yeast pre–mRNAs[58,67]. Total RNA was extracted from *prp2–1* cells grown first at permissive temperature (30 ^{o} C), and then shifted for 30 min to the non–permissive temperature (37 ^{o} C). Previous experiments had shown that this RNA preparation contains pre–mRNAs from most genes containing introns and a small background of mature mRNAs, spliced during the permissive growth period[58]. This RNA mixture was subjected to *in vitro* splicing conditions in yeast cell extract, isolated, and analyzed using a splicing microarray that could detect mature, spliced mRNAs[58]. Several efficiently spliced candidates with small introns were identified . We ultimately chose the pre–mRNA for *UBC4* as the most suitable substrate for our smFRET analysis.

The entire *UBC4* gene is 542 nt long with a short intron of 95 nt. We synthesized a number of Ubc4 pre–mRNAs with different exon lengths and assayed splicing activity. In a trade–off between ease of synthesis and splicing activity of the substrate we chose to develop a

variant of the Ubc4 pre–mRNA for smFRET consisting of 20–nt long exons and the 95–nt intron for a total size of 135 nt. This truncated Ubc4 pre–mRNA was prepared through a two–piece ligation of synthesized RNAs (see Methods). We post–synthetically coupled the donor fluorophore Cy3 to the 5′ exon and the acceptor Cy5 to the 3′ exon, seeking to place the fluorophores as close as possible to the 5′ and 3′ splice sites, respectively, for maximum distance sensitivity without interfering with splicing activity. We prepared pre–mRNAs with dyes located at uridines in multiple positions, all of which are spliced with a similar efficiency of ~30–40% (**Fig. 2.3a**). The substrate used in this work carries Cy3 in exon 1, 7 nt from the 5′ splice site, and Cy5 in exon 2, 4 nt from the 3′ splice site (**Fig. 2.3a**). Introduction of 3′SS and BP mutations into this smFRET substrate was found to have the expected effects[68] of blocking splicing prior to the second and first steps of splicing, respectively (**Fig. 2.3b**).

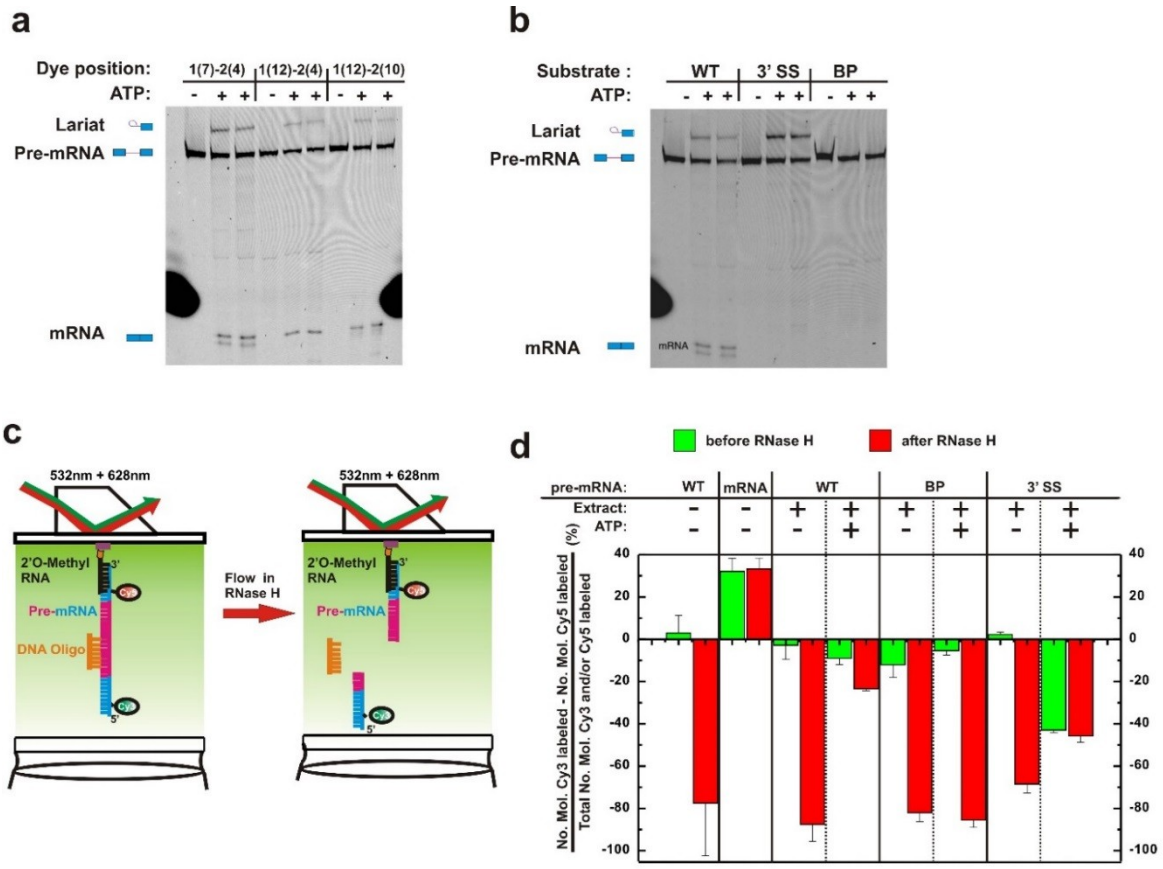


Figure 2. 3: Splicing activity of different Ubc4 pre–mRNA variants.

(a) Positioning the dyes in different sites in the 20–nt exon of Ubc4 pre–mRNA results in similar splicing efficiency. In this experiment splicing was assayed at room temperature as described in Methods. The three lanes for each pre–mRNA are (–) ATP, (+) ATP for 15 min and (+) ATP for 30 min. All pre–mRNAs examined have an apparent splicing efficiency of first plus second step [(mRNA+lariat intermediate)/(pre–mRNA+ mRNA+lariat intermediate)] of between 30–40%. The Cy5 fluorescence scan is shown. (b) The splicing assay of wildtype (WT), 3' splice site mutant (3′SS) and branchpoint mutant (BP) Ubc4 pre–mRNAs shows that the 3′SS mutant RNA only carries out step 1, leading to a lariat–intron intermediate. The BP mutant is inactive in splicing. The experiment was carried out as described above for panel A. The Cy5 fluorescence scan is shown. (c) Design of an in situ assay to probe for the presence of introns in immobilized (pre–)mRNAs. (d) Results of RNase H probing for the presence of the intron after incubation of WT, BP, and 3′SS pre–mRNA in yeast splicing extract. Control experiments on WT pre–mRNA and mature mRNA (leftmost 4 columns) show that an intron containing substrate strongly loses Cy3 signal after incubation with a complementary DNA oligonucleotide and RNase H, but not the intron–free mRNA. Loss of the intron to splicing is thus indicated by a small difference in signal before and after RNase H treatment. Error bars indicate 1 s.d. from the mean.

In our smFRET assays the 3′ end of the labeled pre–mRNA was immobilized through hybridization to a short 2′–O–methyl capture RNA with a 5′ biotin, which was bound to a streptavidin coated and PEG passivated quartz slide (**Fig. 2.4a**). Hybridization of the capture RNA to Ubc4 pre–mRNA does not affect the efficiency of splicing in ensemble assays (data not shown). The Cy3 and Cy5 fluorophores were detected by total internal reflection fluorescence microscopy (TIRFM) in real–time to determine the smFRET efficiency as described previously[36,37,69–71]. By utilizing TIRFM we minimized the background fluorescence of the splicing extract, while capturing all sufficiently long–lived conformational states of the pre–mRNA (even in a fully extended pre–mRNA of ~50 nm length the distal Cy3 dye is still well within the ~100 nm deep evanescent light field of TIRFM[8]). An enzymatic oxygen scavenger system (OSS) was used to limit photobleaching, and Trolox was added to suppress fluorophore blinking (see Methods). Ensemble splicing assays showed that the presence of OSS and Trolox in the reaction mix does not inhibit splicing (data not shown). Under these conditions, we observed limited quenching/blinking and photobleaching with rate constants of ~0.01 s^{-1}. These effects are the same for each of the three pre–mRNAs, and we found no appreciable change in the quenching/blinking and photobleaching upon addition of ATP.

To verify that the spliceosome assembles on the immobilized pre–mRNA to remove its intron we performed an *in situ* RNase H sensitivity assay under our TIRFM conditions. Briefly, a DNA oligonucleotide complementary to the intron was added together with RNase H to slides on which WT pre–mRNA had been immobilized and incubated in either ATP–depleted [(–)ATP] or ATP supplemented [(+)ATP] yeast splicing extract for an hour (Methods and **Fig 2.3c**). Only pre–mRNA is expected and experimentally found to be susceptible to RNase H degradation, leading to the specific loss of the 5′ exon

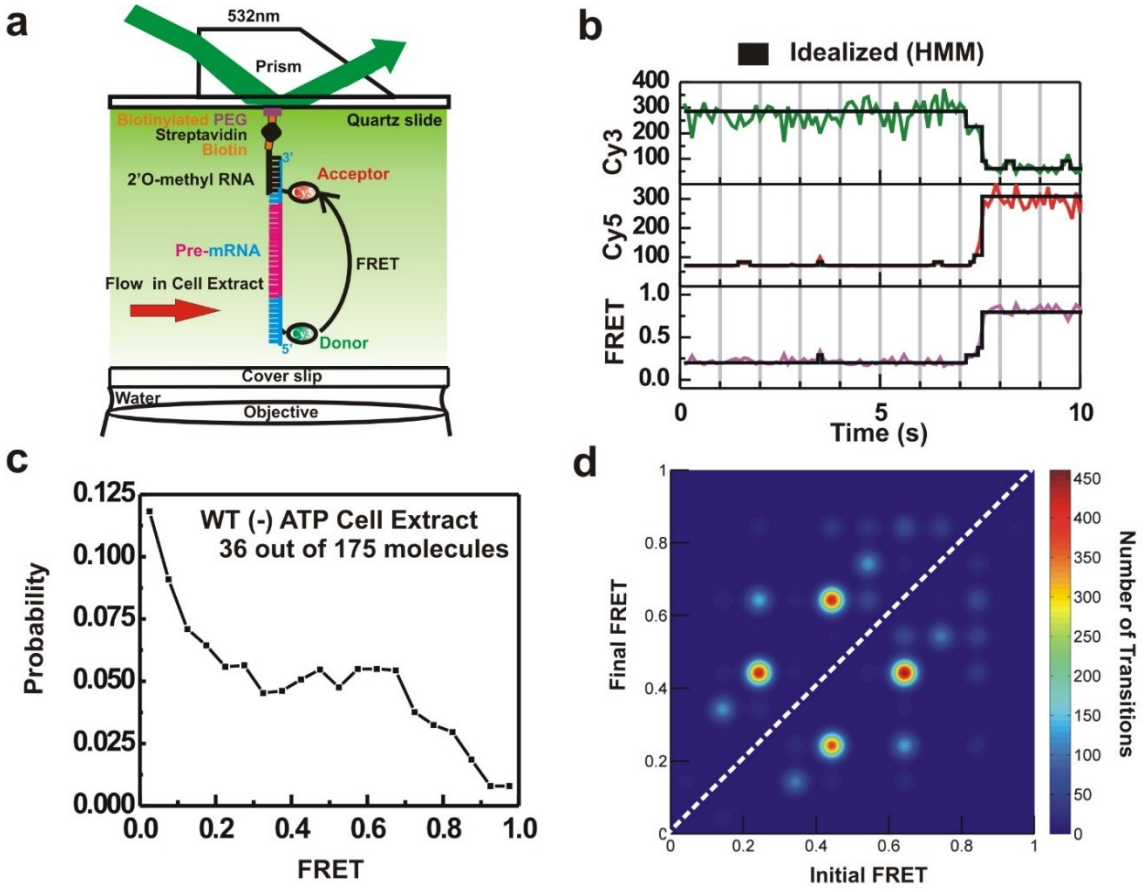


Figure 2. 4: Data analysis and examples of analysis.

(a) Synthetic Ubc4 pre mRNA is hybridized via a 17–nt 2′–O–methyl RNA tether to the 3′ exon, and attached via biotin to a streptavidin coated quartz slide. (b) Raw Cy3 (green), Cy5 (red), FRET (Magenta) time trajectories are analyzed using Hidden Markov Modeling (HMM) algorithms to yield idealized trajectories (black) as described in Methods. (c) The first 10 s of the raw FRET trajectories of a subset of (36 out of 175) WT molecules incubated in ATP–depleted cell extract were analyzed to determine the ensemble distribution of FRET states within the population of molecules analyzed. (d) Transition Density Plots (TDPs) utilize the idealized FRET trajectories to determine the entire set of transitions for a given set of molecules. The number of times a transition occurs is represented as a heat map whose index is defined by the color bar. The fact that this analysis, of a subset (36 of 175 molecules as in panel c) of the data shown in Figs. 5a and 6a, provides qualitatively the same result obtained for the full data set attests to the convergence of the analysis.

and thus Cy3 signal from the slide surface, while the mature mRNA is RNase H resistant (first four columns **of Fig 2.3d**). We find that after incubation in (–)ATP cell extract 75% of all molecules lose Cy3 but not Cy5 upon RNase H challenge, while in (+)ATP this fraction is negligible (**Fig. 2.3d**). Performing the same assay with the BP mutant results in extensive loss of Cy3 signal upon addition of RNase H both before and after incubation in (+)ATP extract (**Fig. 2.3d**), consistent with the expected block of the first step of splicing. The 3′SS mutant displays an intermediate picture as much of its Cy3 signal is lost during the high salt washes used to remove the cell extract (**Fig. 2.3d**), consistent with a block after the first step of splicing when the 5′ exon is no longer covalently attached to the intron. These results strongly support the notion that the immobilized WT pre–mRNA is splicing active to at least the extent observed in standard ensemble splicing assays in solution (~30–40%, see above). The apparent increase in splicing efficiency of the immobilized WT substrate may be the result of the more favorable stoichiometric ratio of limiting splicing factors to pre–mRNA under our single molecule conditions. Additionally, effects such as RNA aggregation and precipitation are known to inhibit reactions under *in vitro* conditions that favor intramolecular and thus inevitably intermolecular interactions[72]. Immobilization of isolated pre–mRNAs for single molecule studies is expected to suppress such inhibition and may thus provide ideal conditions for studying splicing.

Dissecting conformational states and dynamics in complex smFRET trajectories

Our smFRET approach monitors the relative dynamic positioning of the two exons within the spliceosomal complex. Results are presented in three ways. First, we show smFRET time trajectories with their corresponding donor and acceptor traces, which reveal multiple discrete FRET values or states that reflect the complexity of pre–mRNA conformational rearrangements during spliceosome assembly (**Fig. 2.4b**). Second, FRET occupancy histograms display the

probability of observing a particular FRET value (**Fig. 2.4c**). In these histograms we plot frame–by–frame FRET values for a set of molecules collected over the first 100 frames (10 s of real–time) of all FRET trajectories in a given data set, which is proportional to the probability that any single molecule will have that FRET value.

Third, we use traditional Transition Density Plots (TDPs), weighted by the number of times a particular FRET transition occurs, as well as novel POpulation–weighted and Kinetically–Indexed Transition density (POKIT) plots to comprehensively represent all observed discrete FRET states, their transitions and kinetics. To first reliably identify genuine FRET states and their transition kinetics in naturally noisy data sets we use Hidden Markov Modeling (HMM), which traces the most probable sequence of FRET states and derives their dwell times[23] (Methods and **Fig. 2.4b**). To eliminate from our analysis possible effects of the high concentrations of proteins directly on the fluorophores we additionally evaluate the donor and acceptor signals by HMM and introduce a scoring function selecting for anti–correlated donor and acceptor changes that reflect changes in FRET (see Methods and **Fig. 2.2**). As a further control, we monitored the donor and acceptor signals from a mature (synthetic) mRNA under the same conditions used for the pre–mRNA and found a largely steady FRET signal (**Fig.2.5**). These observations indicate that any smFRET dynamics we observe for pre–mRNA are not caused by spurious photophysical artifacts but by the expected changes in fluorophore distance. To determine the most likely number of states in our smFRET trajectories we analyzed whole data sets of 50–200 single molecule traces by our scored HMM approach. The best global fit to the entire trace set for each condition generally found 10 or 11 distinct FRET states (to be evaluated as distinct two states needed to exhibit a FRET difference of at least 0.1, see Methods), with each molecule adopting only a subset of states (**Fig. 2.6**). With knowledge of the number of

states in each data set we identified the transitions between these states, summarized as TDPs[65] and POKIT plots (see below). The transition from an initial FRET state of 0.8, for example, to a final FRET state of 0.3 is one point in these plots. There can be a maximum of N(N–1) different transitions between N states. We observed many fewer than the

possible 110 transitions. To test the robustness of our analysis we randomly selected a subset of 36 out of a total of 175 WT substrate molecules in (–)ATP extract and found this subset to yield a similar FRET occupancy histogram and TDP as the entire data set (compare **Fig. 2.4c** and **2.4d** with Figs. **2.10a** and **2.10a**, respectively), suggesting that our molecule sample is of sufficient size for the HMM model to converge.

The pre–mRNA substrate dynamically folds and unfolds in splicing buffer

When incubated in standard splicing buffer alone, all three pre–mRNAs exhibit substantial global dynamics with fluctuations between high and low FRET states (**Fig. 2.7a**). The several second long dwell times in high FRET states indicate that the pre–mRNAs must transiently form secondary structures that bring the distal 5′ and 3′ splice sites into close proximity. The distribution of FRET states is not identical between the pre–mRNAs, as the WT and BP mutant favor high FRET states (shorter distances) relative to the 3′SS mutant. By comparison, the 3′SS spends relatively more time in low FRET states as evident from the cumulative FRET occupancy histograms in **Fig. 2.7b**. These differences are also apparent in the corresponding TDPs, where the WT and BP pre–mRNAs exhibit more transitions in the high (>0.7) and fewer in the low (<0.4) FRET range than the 3′SS. Notably, there are some transitions with long dwell times found in all pre–mRNAs (**Figs. 2.7a**,b,**c**). Computational predictions suggest that in the ensembles of lowest free energy states distinct secondary structures separate the 5′ and 3′ splice

sites in the WT and BP mutant less than in the 3′SS mutant, qualitatively correlating with their experimentally observed relative bias towards higher FRET values (**Fig. 2.8**).

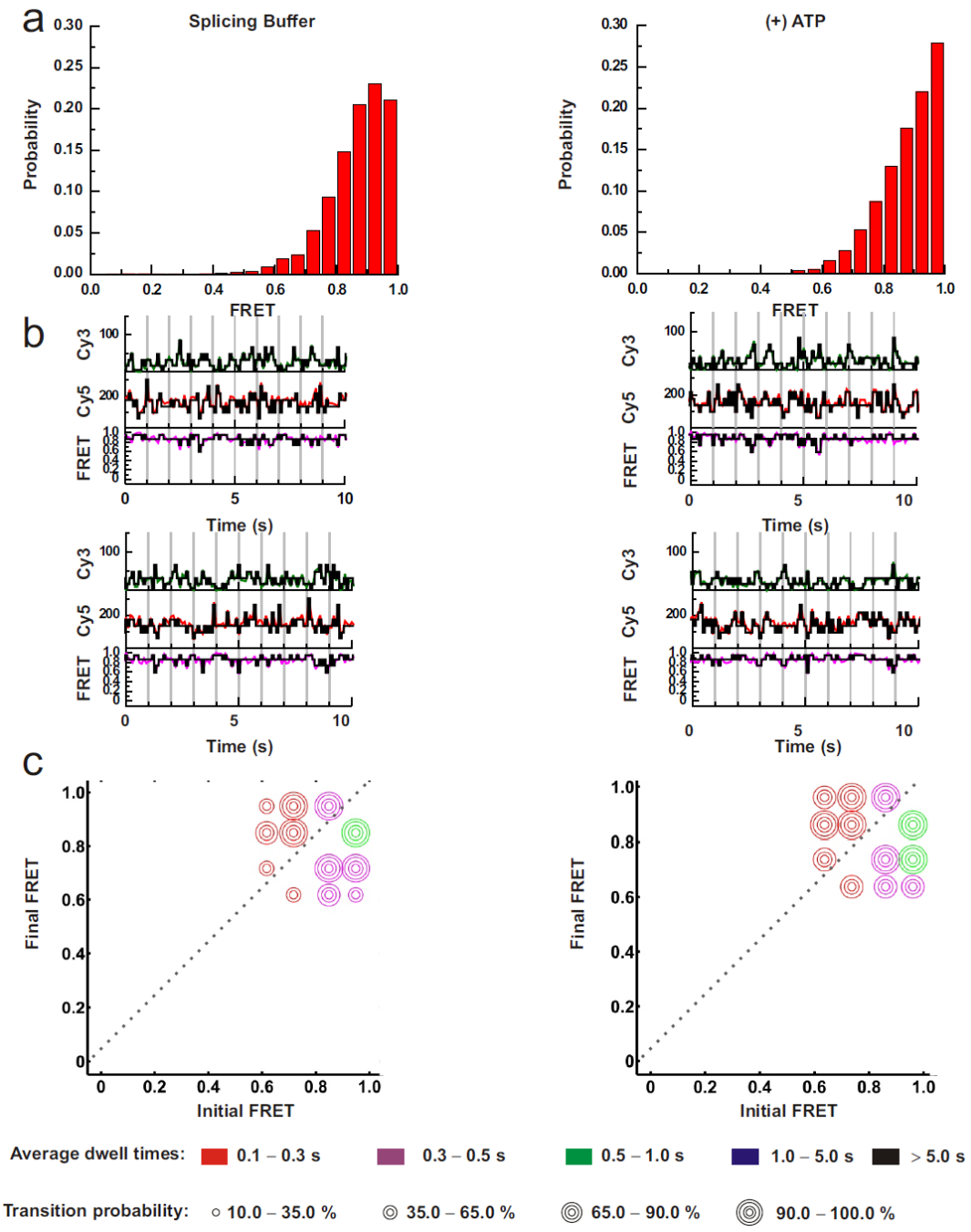


Figure 2. 5: Conformational dynamics of mRNA in splicing buffer and (+) ATP cell extract.

(a) Histograms of enzymatically ligated exons in buffer and extract. (b) Sample traces of the same mRNA in bothconditions. (c) POKIT plots of both conditions.

Since all RNA molecules assume transient secondary structures[25], it is not surprising that the Ubc4 pre–mRNA exhibits transient folding yielding high FRET states. In the case of Ubc4 pre–mRNA even single base changes affect these structures. That we can observe these single base differences by smFRET attests to the power of this method. Furthermore, this kinetic analysis of the substrate RNA in buffer is necessary to distinguish the pre–mRNA dynamics caused by the formation of transient secondary structures and those induced by spliceosome assembly. To achieve a maximal dynamic time range for observing spliceosome related kinetics, in the following we performed 100–ms time resolution smFRET experiments at various instances after addition of splicing extract.

Sequence and time dependent dynamics of single pre–mRNAs in splicing extract

The addition of yeast splicing extract depleted of ATP [(–)ATP] results in a profound difference from the FRET histograms observed in buffer alone (compare **Figs. 2.7b** and **2.9**). Lower FRET values are relatively favored for all three pre–mRNAs. The effect is most pronounced for the WT, which becomes biased toward particularly low FRET values. These changes to lower FRET distributions show that the relatively stable high FRET states observed in buffer alone are disfavored in extract.

The addition of ATP to extract [(+)ATP] leads to a time dependent change in the FRET distributions of the WT and 3′SS mutant pre–mRNAs, but not that of the BP mutant (**Fig. 2.9c**). In the case of the 3′SS mutant, the overall FRET distribution changes within 15 min and after 60 min of incubation in (+)ATP extract to increasingly more prevalent mid and high FRET states (**Fig. 2.9b**). In the case of the WT, the prevalent low FRET states are lost over 60 min of incubation (**Fig. 2.9a**), while the BP mutant shows no substantial change in the overall FRET distribution (**Fig. 2.9c**), in agreement with the requirement of an intact branch point for ATP–

dependent spliceosome assembly[68]. Thus, the time dependent changes in FRET distribution upon addition of ATP require a pre–mRNA substrate that can undergo the first step (3′SS) or both steps (WT) of splicing. The time–, ATP– and mutant dependent changes in the FRET histograms observed upon incubation in extract strongly suggest that we are observing splicing dependent changes at the single molecule level, but FRET histograms cannot tell us about the dynamics of those changes. To track even subtle changes in these conformational dynamics upon addition of splicing extract we turned to HMM and TDP plot analysis.

Spliceosome dependent dynamics of pre–mRNA substrates

The TDP of the WT pre–mRNA shows that upon addition of (–)ATP cell extract fewer transitions are observed when compared to the same substrate in buffer and the most prevalent transitions are on average lower on the FRET scale (compare **Figs. 2.7c** and **2.10a**). Within the first 15 min after the addition of (+)ATP cell extract we find that the distribution of transitions changes yet again, with more transitions originating from the 0.8 FRET state. After a 60 min incubation in (+)ATP cell extract, FRET transitions appear in an even higher FRET region (0.75–0.95), indicating that conformations are favored where the two exons are in close proximity (**Fig. 2.10a**). Most observed FRET transitions are reversible, leading to the relative mirror symmetry of the TDPs with respect to their main diagonal (**Fig. 2.10a**). Reversible FRET transitions imply conformational reversibility, a property that is characteristic of all substrates and extract conditions (**Figs. 2.10–2.12**).

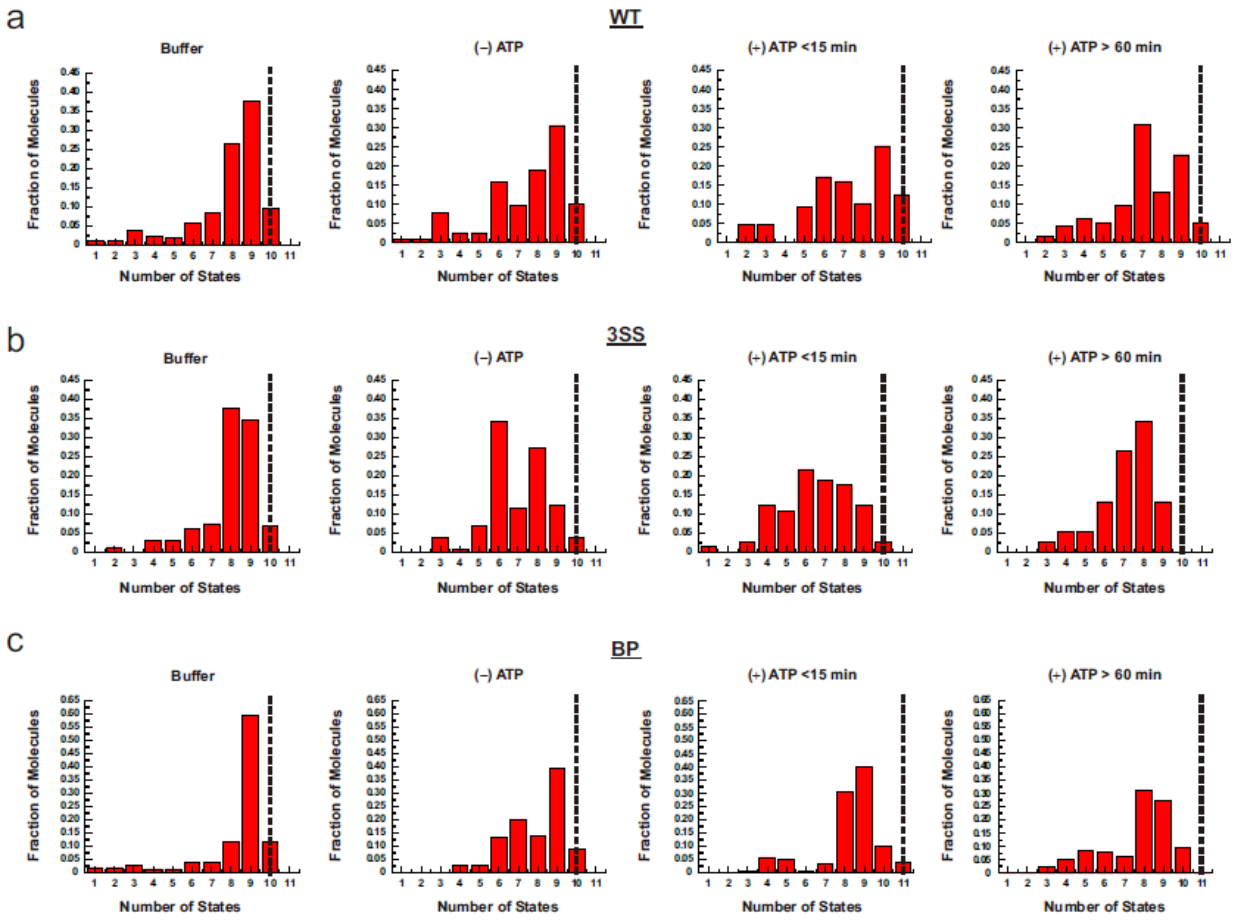


Figure 2.6: Distribution of idealized FRET states.

The number of idealized states for single molecules was binned, and plotted in histograms demonstrating the distribution of states per molecule for the WT **(a)**, 3'SS **(b)**, and BP **(c)** substrates in each experimental condition. Dashed lines indicate the number of states in the Markov model required to fit the entire data set.

Upon addition of (–)ATP extract the 3′SS pre–mRNA, like the WT, displays fewer, still symmetric (i.e., reversible) transitions than in buffer (compare **Figs. 2.7c** and **2.11a**). After a 15 min incubation in (+)ATP cell extract (compare **Figs. 2.10** and **2.11**) there is a shift towards higher FRET states. Most notably, transitions that become more prevalent after 60 min incubation in (+)ATP extract generally fall into a mid FRET regime (**Fig. 2.11a**), unlike for the WT where a shift occurs towards high FRET transitions (**Fig. 2.10a**). The BP mutant does not enter the spliceosome assembly pathway and while we observe transitions between a substantial number of FRET states, unlike the WT and 3′ SS substrates there are few changes when comparing the FRET states in (–)ATP extract and (+)ATP extract. The prominent set of symmetric transitions between 0.45 and 0.65 FRET are present before and following the addition of (+)ATP extract; **Figs. 2.12a–c**). After 60 min in (+)ATP extract the number and diversity of transitions sampled by each BP mutant molecule is only slightly changed.

To assess the kinetic differences between substrates we turn to POKIT plot analysis. **Figure 2.13** provides a side–by–side comparison of these plots with the more common TDPs to illustrate their close relationship; however, POKIT plots provide two additional, comprehensive pieces of information for each transition. First, they present as a number of concentric circles the fraction of molecules in the entire trace set that exhibits a specific FRET transition at least once (**Fig. 2.13**). This feature complements TDPs that are weighted by number of times a specific transition is observed overall[65], which emphasizes fast (frequent) transitions exhibited by possibly only a small fraction of molecules (**Fig. 2.13**). Second, our POKIT plots provide the average dwell time for each transition, or time spent in the initial state before transition to the final state, in the form of circle colors (**Fig. 2.13**), facilitating the rapid visual comparison of the kinetics of our various smFRET data sets as a function of ATP concentration, incubation time,

and intron signals. The POKIT plots of the WT pre–mRNA in buffer and (–) ATP extract highlight the fact that incubation with cell extract induces longer dwell times in low rather than high FRET states (**Fig. 2.14**). Within the first 15 min after the addition of (+) ATP cell extract the distribution of transitions then changes to include a higher fraction of molecules with FRET transitions in the 0.6–0.9 range (**Fig. 2.14**). After a 60 min incubation in (+)ATP cell extract, this trend continues toward even longer dwell times in an even higher FRET region (0.75–0.95, **Fig. 2.14**), indicating that relatively stable conformations are favored where the two exons are proximal. These long dwell times in the highest FRET states are observed only for the WT substrate and require an extended incubation with (+)ATP cell extract (**Fig. 2.14**). The fact that similar transitions are strongly dominant for the mature (synthetic) mRNA control (**Fig. 2.5**), but are found in only ~20% of all WT molecules (**Fig. 2.14**), suggests that only a fraction of the total WT pre–mRNA observed after 60 min in (+)ATP cell extract adopts a conformation similar to that of the synthetic mRNA. These findings indicate that most pre–mRNA molecules are spliced reversibly, possibly because our 2′–O–methyl RNA tether hybridized to the short 3′ exon blocks the binding of Prp22 and disrupts spliceosome disassembly[73]; and/or that the mRNA derived from activity of bound splicing factors exhibits a FRET signature distinct from that of the synthetic mRNA, which is not expected to bind splicing factors.

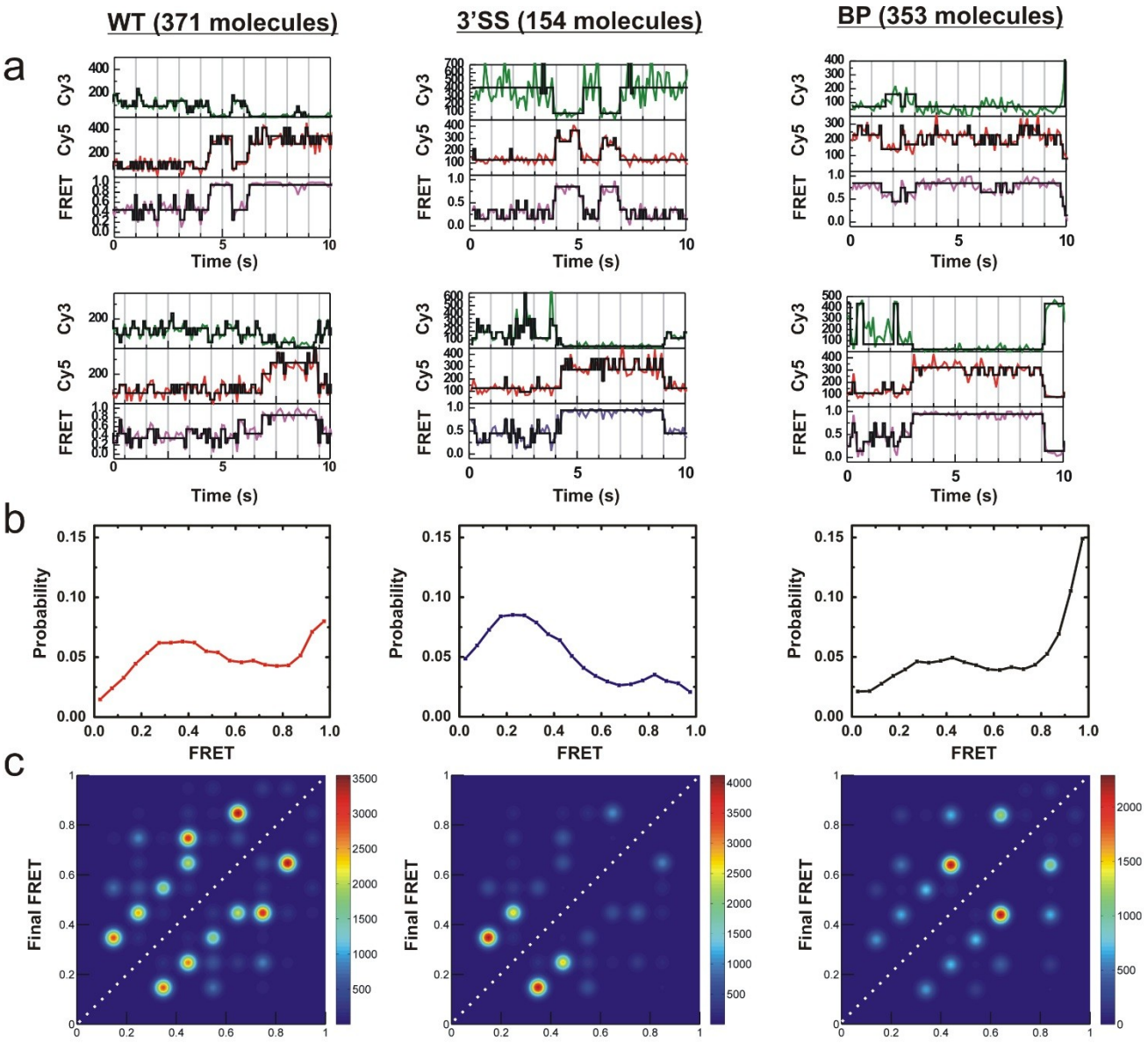


Figure 2.7: Conformational dynamics of wildtype (WT), 3′splice site mutant (3'SS), and branchpoint mutant (BP) pre–mRNA substrates in splicing buffer.

(a) Sample traces of all three substrates in splicing buffer, showing raw donor (Cy3, green), acceptor (Cy5, red), and FRET (blue) trajectories and their idealized HMM models (black). (b) FRET histograms of the three pre–mRNAs in splicing buffer. (c) TDPs for all three pre–mRNAs in splicing buffer.

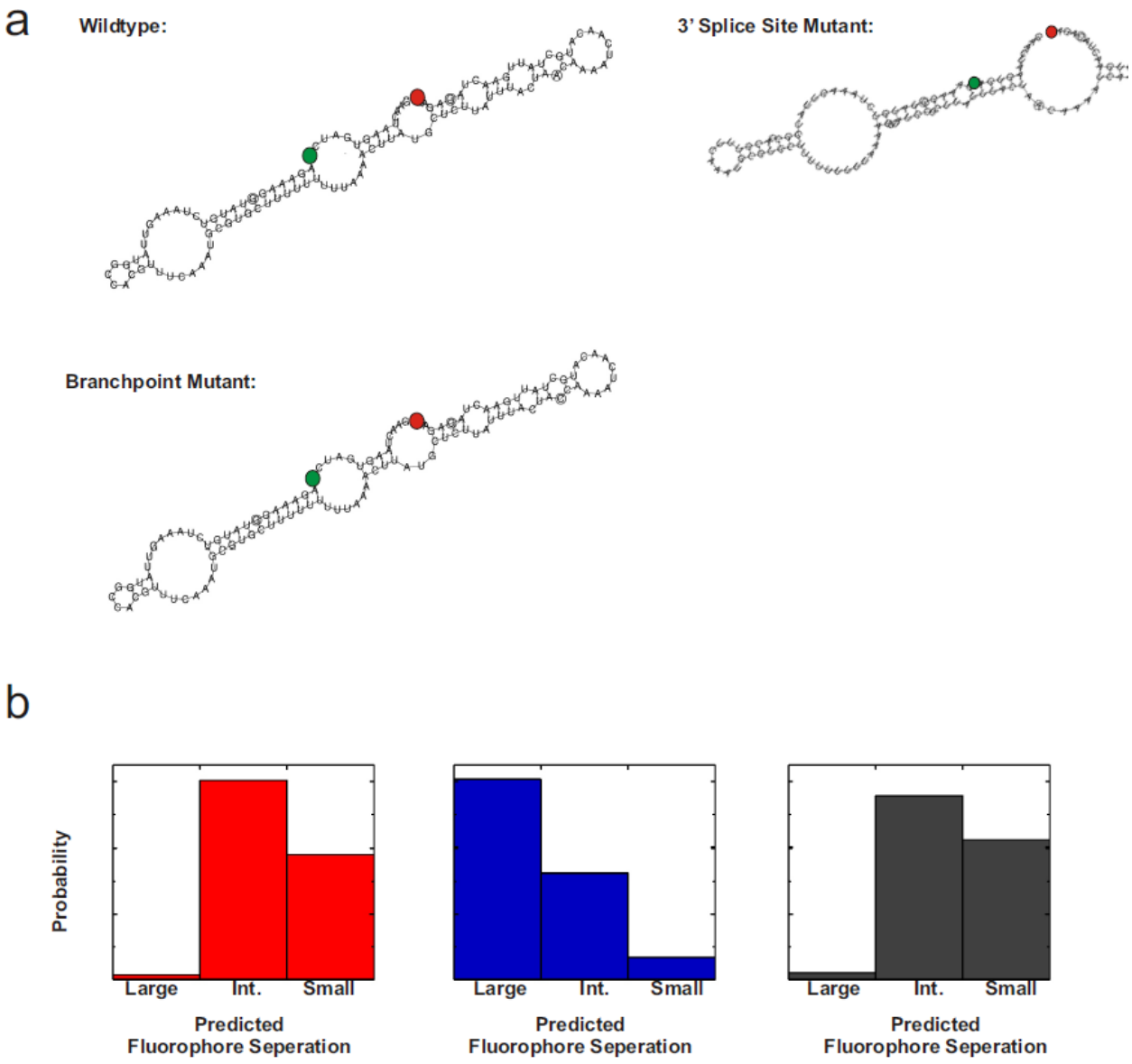


Figure 2.8: Transient secondary structures place fluorophores within FRET range in buffer.

(a) Predicted lowest free energy structure (RNAfold) for WT, 3'SS mutant, and BP mutant substrates. The position of fluorophores is indicated by a green circle and red circle for Cy3 and Cy5, respectively. The sequence of exon-2 that is used for tethering to the slide surface was excluded from this analysis. **(b)**The distribution of predicted inter-fluorophore distances based on secondary structure analysis is show for each pre-mRNA sequence. Small, step size of 0-10. Intermediate (Int), step size of 10-20. Large, step size >20.

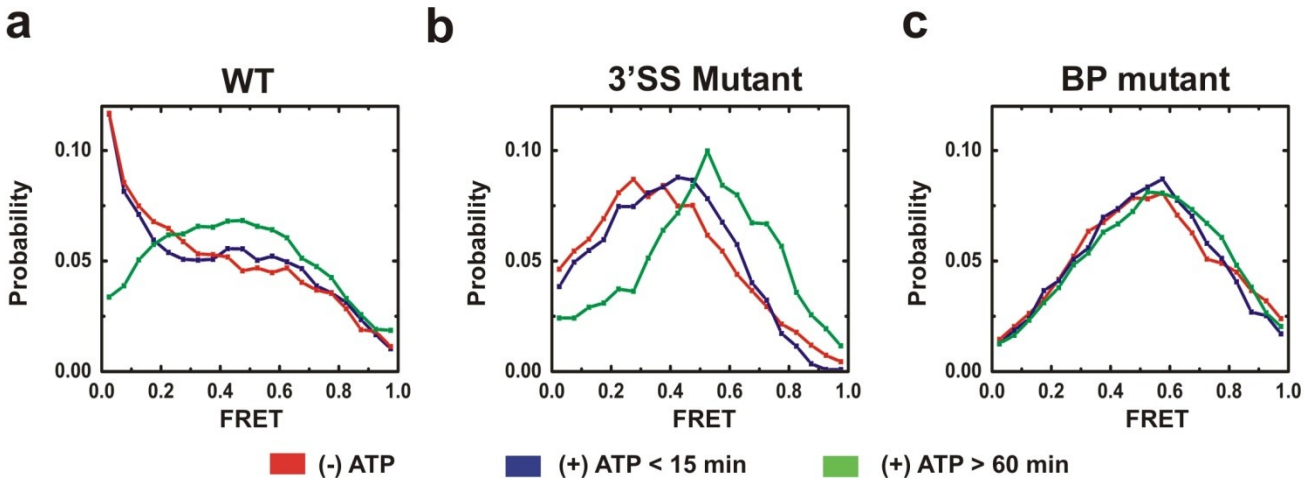


Figure 2.9: ATP–dependent conformational dynamics of the wildtype (WT), 3′splice site mutant (3'SS), and branchpoint mutant (BP) pre–mRNAs in yeast cell extract.

(a) Probability distributions of FRET states for the WT substrate for each experimental condition. (b) Probability distributions of FRET states for the 3′SS substrate for each experimental condition. (c) Probability distributions of FRET states for the BP substrate for each experimental condition.

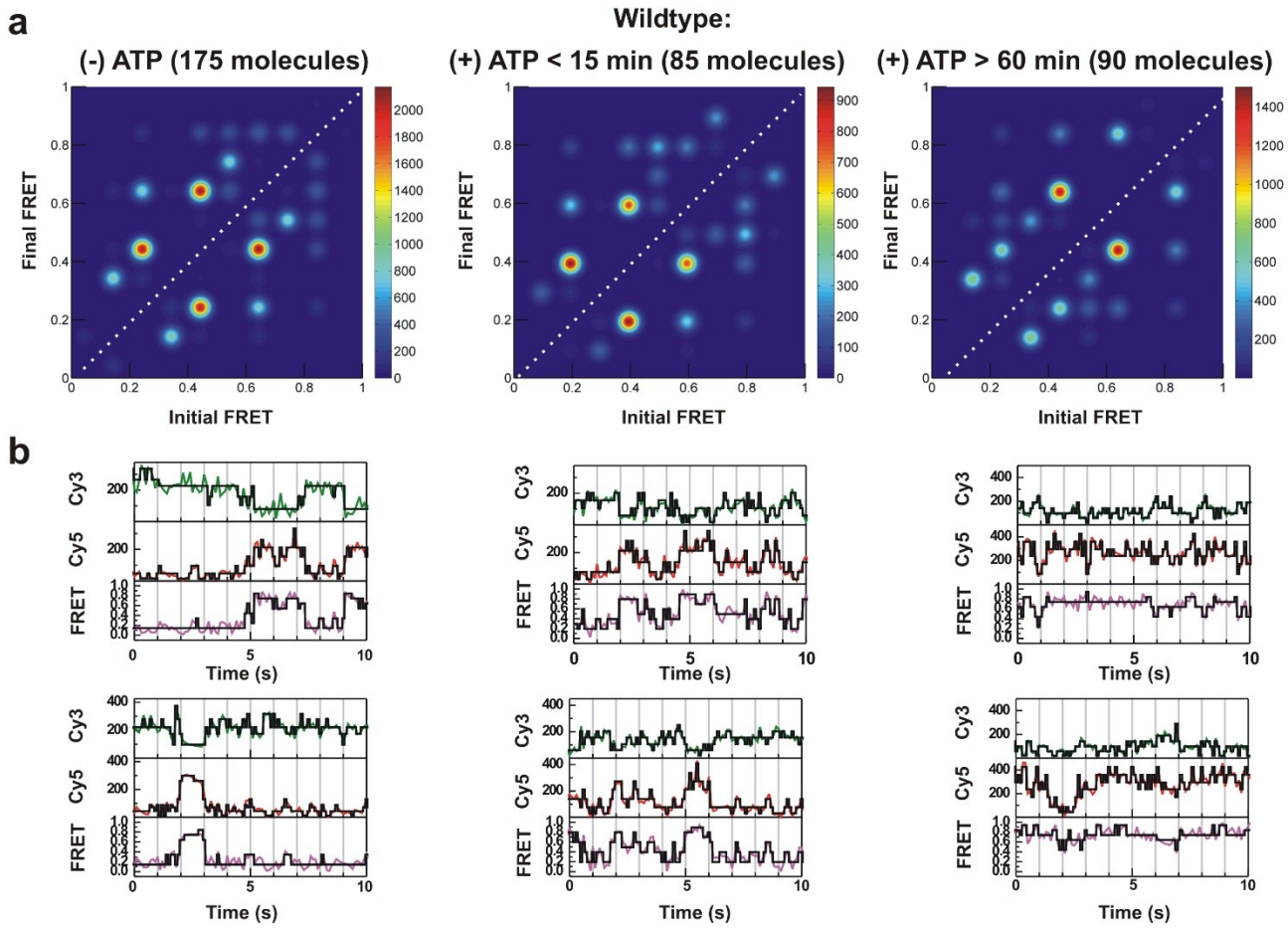


Figure 2.10: Mapping conformational changes of the wildtype (WT) pre–mRNA during spliceosome assembly and splicing in vitro.

(a) TDPs for WT substrate; in ATP–depleted extract, within 15 min after ATP addition to cell extract, and after 60 min incubation in (+) ATP cell extract. (b) Representative trajectories (donor, green; acceptor, red; FRET, magenta) and idealized HMM models (black) for WT substrate.

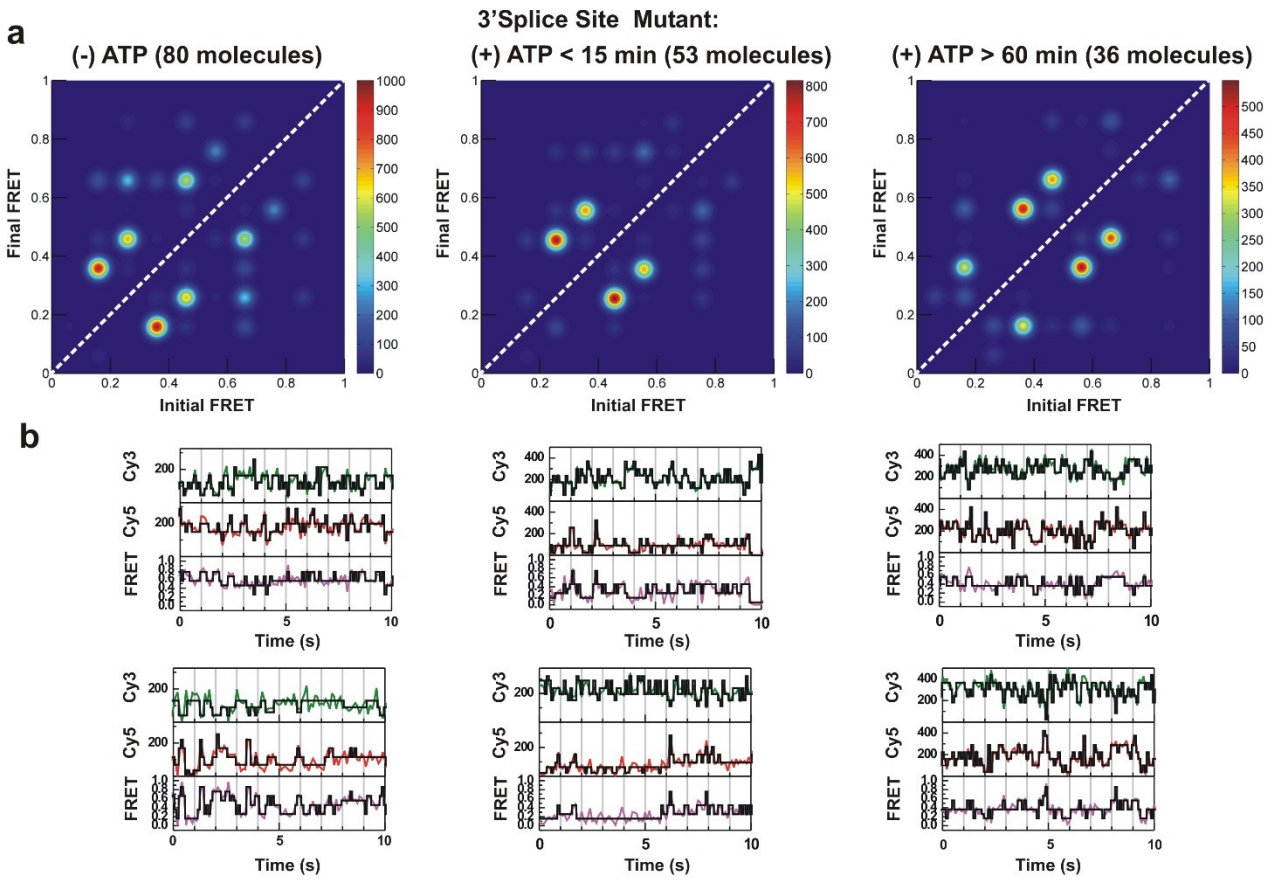


Figure 2.11: Mapping conformational changes of the 3' splice site mutant (3′SS) pre–mRNA spliceosome assembly and splicing in vitro.

(a) TDPs for 3′SS substrate; in ATP–depleted extract, within 15 min after ATP addition to cell extract, and after 60 min incubation in (+) ATP cell extract. (b) Representative trajectories (donor, green; acceptor, red; FRET, magenta) and idealized HMM models (black) for 3″SS substrate.

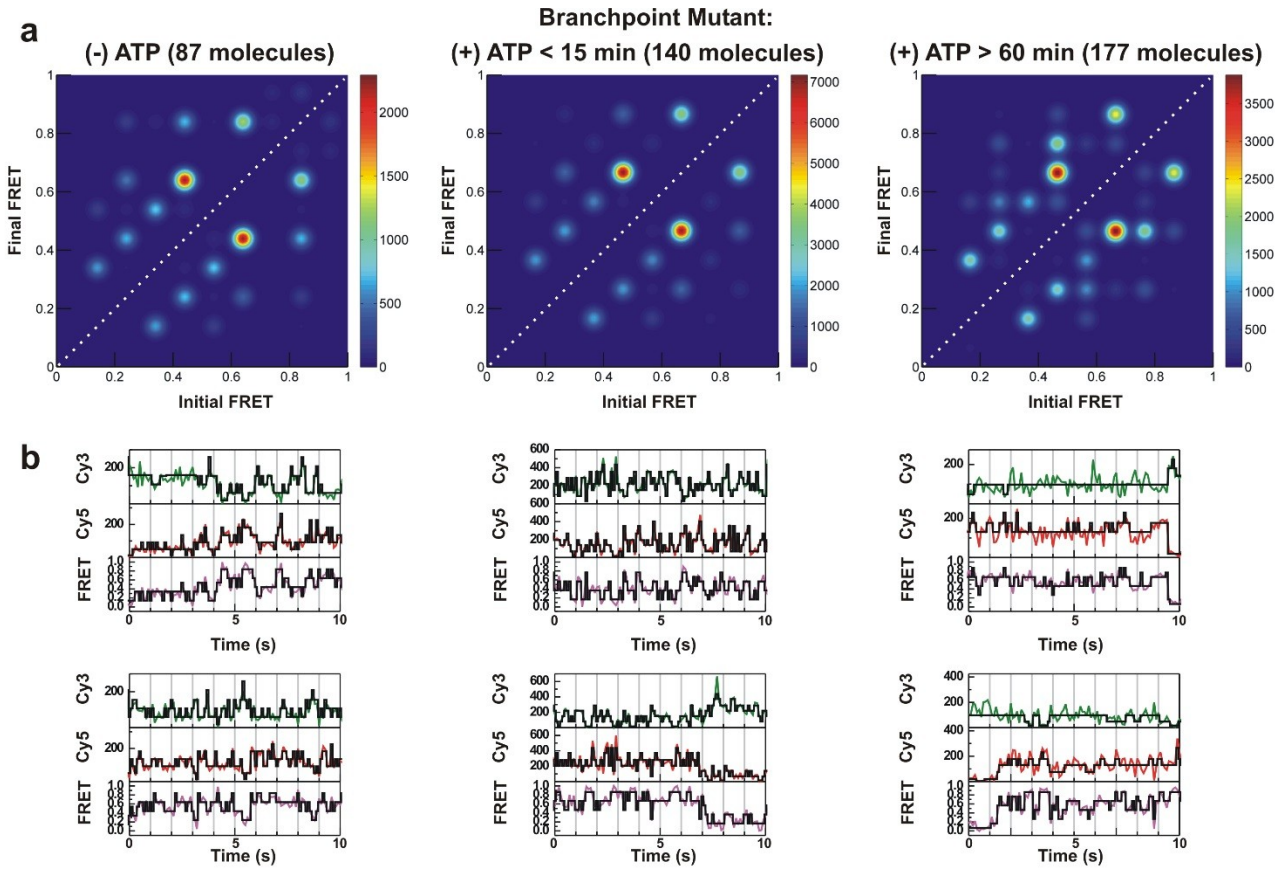


Figure 2.12: Mapping conformational changes of the branchpoint mutant (BP) pre–mRNA spliceosome assembly and splicing in vitro.

(a) TDPs for BP substrate; in ATP–depleted extract, within 15 min after ATP addition to cell extract, and after 60 min incubation in (+) ATP cell extract. (b) Representative trajectories (donor, green; acceptor, red; FRET, magenta) and idealized HMM models (black) for BP substrate.

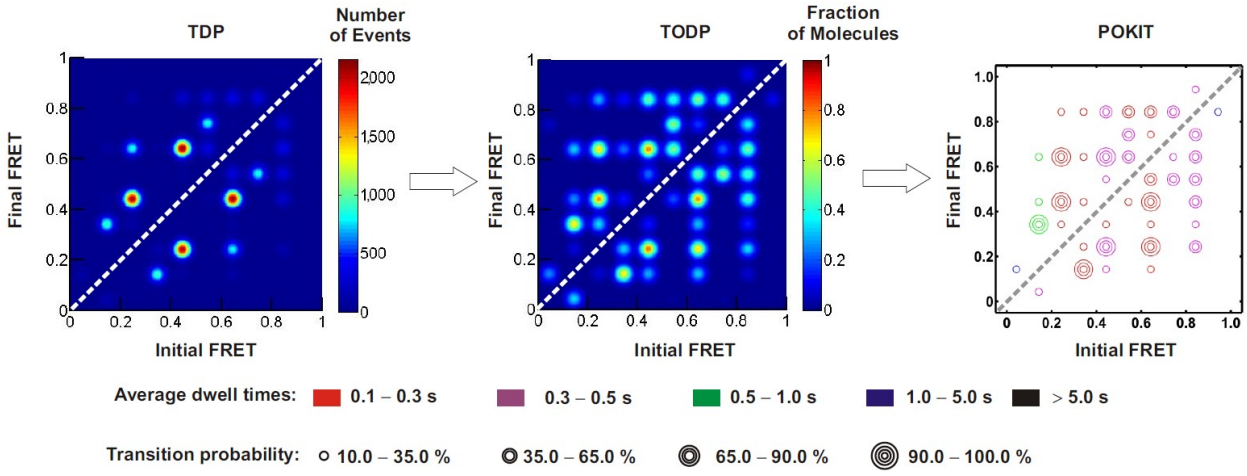


Figure 2.13: Transition Density Plot (TDP), Transition Occupancy Density Plot (TODP), Population and Kinetically indexed Transtion Density Plot (POKIT).

TDPs are weighted by the number of times a transition is seen, and therefore must be corrected for differences in kinetics due to an inherent bias against slower transitions. TODPs scale the data by the population of molecules exhibiting a particular transition. This allows for transitions that are equally probable to appear within a set of molecules to be weighted to the same degree. POKIT plots scale transitions based on their probability, but also display kinetic information of each transition.

In summary, our data suggest that *in vitro* the WT pre–mRNA samples both low and high FRET conformational states even in the absence of splicing extract. The addition of extract leads to a change in the distribution and kinetics of these FRET states, first by assembling with spliceosomal components in the absence of ATP and, second, by extended incubation with ATP–enriched cell extract over a time period where it undergoes both steps of splicing (**Fig. 2.3**).

Additionally, WT pre–mRNA after 60 min in splicing extract adopts relatively stable high FRET conformations found in an mRNA control but not in mutant substrates when placed under the same conditions (**Figs. 2.10–2.12** and **Fig. 2.5**). The BP mutant largely lacks these time and ATP–dependent changes. The 3′SS mutant is generally more similar to the WT than the BP mutant, but does not achieve the stable high FRET states after 60 min incubation observed for the WT substrate. Most conformational changes of the pre–mRNAs are readily reversible and their relative abundance and kinetics both ATP and intron signal dependant, implying that they are relevant for splicing activity. Additionally, WT pre–mRNA after 60 min in splicing extract adopts relatively stable high FRET conformations found in an mRNA control but not in mutant substrates when placed under the same conditions (**Figs. 2.10–2.12** and **Fig. 2.5**). The BP mutant largely lacks these time and ATP–dependent changes. The 3′SS mutant is generally more similar to the WT than the BP mutant, but does not achieve the stable high FRET states after 60 min incubation observed for the WT substrate. Most conformational changes of the pre–mRNAs are

readily reversible and their relative abundance and kinetics both ATP and intron signal dependant, implying that they are relevant for splicing activity.

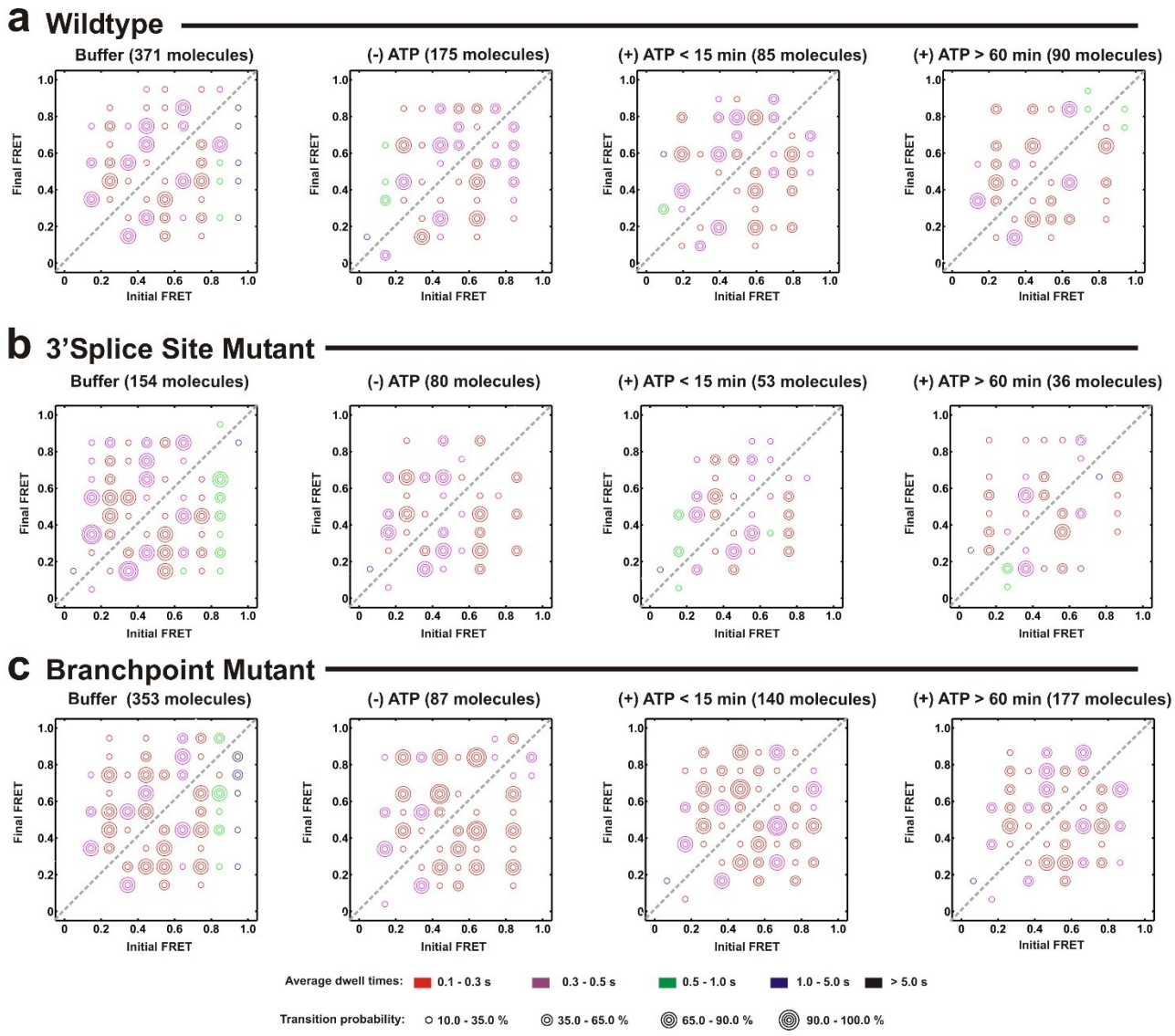


Figure 2.14: Detailed comparison of kinetic and conformational profiles of pre–mRNAs during spliceosome assembly and splicing in vitro.

(a) POKIT plots for WT substrate; in ATP–depleted extract, within 15 min after ATP addition to cell extract, and after 60 min incubation in (+) ATP cell extract. (b) POKIT plots for 3′SS substrate; in ATP–depleted extract, within 15 min after ATP addition to cell extract, and after 60 min incubation in (+) ATP cell extract. (c) POKIT plots for BP substrate; in ATP–depleted extract, within 15 min after ATP addition to cell extract, and after 60 min incubation in (+) ATP cell extract.

II.4 Discussion

Establishment of smFRET to monitor spliceosome assembly

In this study we have employed smFRET to observe the conformational dynamics of a pre–mRNA during spliceosome assembly. We identified Ubc4 as a natural intron–containing transcript that is both small – allowing ease of chemical synthesis – and very efficiently spliced *in vitro* as well as *in situ* during our smFRET experiments (**Fig. 2.3**). We placed the donor fluorophore Cy3 in exon 1 near the 5′ splice site and the acceptor Cy5 in exon 2 near the 3′ splice site. With this substrate under standard *in vitro* splicing conditions we observed a splicing efficiency of 30–40%, which *in situ* may be even higher (**Fig. 2.3**). This result is not unique to the Ubc4 substrate; recently Crawford et al. demonstrated that a different fluorescently labeled pre–mRNA bound to a slide for TIRFM studies undergoes splicing (measured as the release of a fluorescent intron) upon addition of splicing extract[45]. The authors did not, however, monitor the structural dynamics of the pre–mRNA during spliceosome assembly and action as we have done here.

Several criteria establish that the smFRET changes we observe in splicing extract reflect dynamic changes in the substrate during spliceosome assembly. These criteria are that the dynamic changes be time– and ATP–dependent, and that they be affected as expected by mutations in the substrate. The FRET occupancy histograms in **Fig. 2.9** illustrate that time and ATP–dependent changes indeed are observed for the WT substrate and the 3′SS mutant but, as expected, less so for the BP mutant (see also below). These results demonstrate that we monitor conformational changes associated with genuine spliceosome assembly processes.

The pattern of WT conformational dynamics is altered by intron mutations

When ATP–depleted extract is added, substantial differences in FRET histograms and FRET transitions of all three pre–mRNAs are observed. More specifically, the smFRET transitions observed in splicing buffer alone are replaced by a somewhat smaller set of transitions with on average shorter dwell times (**Figs. 2.7, 2.10–2.12**). This initial shift in FRET distribution of all pre–mRNAs upon extract addition is consistent with the presence of a large number of highly concentrated RNA binding proteins in yeast extract, e.g., the RNP protein NPL3[74]. These proteins are expected to alter transient secondary structures, but they are not expected to distinguish between RNA molecules differing by a single base change. Notably, the WT adopts a set of low FRET conformations with slow transitions that are not found in the 3′SS and BP mutants. As evident from **Fig. 2.1** the behavior of WT may reflect the ATP–independent interaction of the U1 snRNP with the 5′SS, BBP (SF1 in humans) and Mud2 (U2AF in mammals) with the branchpoint and polypyrimidine tract to form CC, the commitment complex. In any event, since an intact branchpoint is required for the resistance of a labeled pre–mRNA template to challenge by excess cold substrate[75], we predicted that the BP mutant would be blocked at the earliest stage of spliceosome assembly. This is fully consistent with our observation of substantial differences in the TDP and POKIT plot patterns of WT and 3′SS mutant on the one hand and BP mutant on the other (**Figs. 2.10–2.12**).

The 3′SS mutant was predicted to largely mimic the WT pattern, given that the 3′ splice site sequence is thought to be required only for the second chemical step. This prediction is consistent with the greater similarity of the TDP and POKIT plots of the WT and 3′SS mutant compared to that of the BP mutant. Curiously the 3′SS mutant, unlike the WT, does not adopt stable low FRET transitions in the absence of ATP (**Fig. 2.13**). These stable transitions appear in

the 3′SS mutant only after addition of (+)ATP supplemented extract, and are never observed in the BP mutant. It is tempting to speculate that our detailed kinetic smFRET analysis reveals differences between WT and 3′SS mutant that are not detectable by conventional assays. Other discrepancies to the canonical model of spliceosome assembly (**Fig. 2.1**) have been noted previously and will likely expand with the emergence of new experimental tools[76,77].

The spliceosome, like the ribosome, functions near thermal equilibrium

As described in the introduction, a dominant and well–known feature of splicing is the highly dynamic nature of the spliceosome, which is driven forward by (in yeast) eight RNA–dependent ATPases of the DEAD–box family[2]. It is thought that each of the eight DEAD–box dependent steps in splicing provide fidelity check–points[6,7]. Evidence for this model is found in our observation that within the first 15 min of incubation of WT pre–mRNA in ATP–enriched extract the molecule population shifts towards higher FRET transitions compared to the transitions in ATP–depleted extract (**Figs. 2.10 and 2.13**). This indicates that the assembly and rearrangements of spliceosomal components initiated by ATP bring the two exons closer to each other, as envisioned in the model to achieve splicing–specific conformations. The effect is more pronounced after a long incubation (60 min) with ATP–enriched extract (**Figs. 2.10 and 2.13**), when a substantial fraction of mRNA product has been generated (**Fig. 2.3**).

Historically, it has been assumed that the spliceosomal ATPases would drive the pathway unidirectionally. Our observation of the dynamics of single pre–mRNAs indicates that the substrate does not follow a simple unidirectional pathway. Instead, pre–mRNAs sample a variety of reversible conformational states guided by the assembly of ATP dependent spliceosomal components. Recently, it has been demonstrated that the chemical steps are themselves reversible, at least under certain *in vitro* conditions[35,78]. A reversal of the chemical steps of splicing requires

that the conformation of the substrate and spliceosome assembly be reversible as well. Such a model is supported by our observation of reversible smFRET transitions. To our knowledge, our data provide the first direct glimpse of such reversible conformational changes of the pre–mRNA throughout the splicing process. This realization prompts a provocative re–evaluation of the role of ATP in splicing, in which the function of hydrolysis is largely to improve the accuracy of the pathway, driving the "correct" substrate to completion while "incorrect" alternatives are rejected. Interestingly, this picture fits well with recent evidence from the ribosome, where spontaneous intersubunit rotation has been observed in the absence of GTP[43]. Thus it is tempting to speculate that both macromolecular machines carry out complex sets of conformational changes close to thermal equilibrium.

Outlook

We have shown that during spliceosome assembly a pre–mRNA′s conformations are highly dynamic and largely reversibly interchanging. These changes are reflected in a large number of complex transitions. By locating the dyes at different pre–mRNA positions, by placing the dyes in snRNAs or proteins, and by examining the effects of mutant extracts it now seems possible to assign particular sets of FRET transitions to specific steps in spliceosome assembly. Ultimately, approaches such as these will yield a comprehensive dynamic model of pre–mRNA splicing.

II.5 Acknowledgements

The authors wish to thank Reinhard Lührmann and Patrizia Fabrizio (Max Planck Institute for biophysical Chemistry – Göttingen, Germany), Jonathan Staley (University of Chicago – Chicago, IL) and Beate Schwer (Weill Cornell Medical College – New York City, NY) for providing splicing active yeast cell extracts at a moment, common in this field, in which we were having difficulty in making active extracts. The oligonucleotides synthesized at Dharmacon were

sometimes a lot longer than they usually make and their excellent quality was essential for this project. The work at the University of Michigan was supported by NIH grant GM062357 to N.G.W., NIH Cellular & Molecular Biology and Molecular Biophysics Training Grant fellowships to M.B. and M.A.D., respectively, as well as a Rackham Merit Fellowship to M.B. The work at UCSF was supported by an American Cancer Society Research Professor of Molecular Genetics award to C.G., by NIH grant GM021119 to C.G., by NIH postdoctoral fellowship GM077844 to C.M., and by a grant from the Agouron Institute to J.A. We acknowlege helpful comments on the manuscript from Greg Whitworth, Jonathan Staley and Haralambos Hadjivassiliou.

Author Contributions

J.A. worked at the bench and led the development of the Ubc4 system at UCSF. J.A.P. and J.A. performed the microarray analysis, while D.R., T.V. and C.M. participated in various phases of the biochemistry at UCSF. M.B., M.A.D, F.F., and M.W. performed the smFRET experimentation and data analysis, whereas P.A. performed the secondary structure analysis at the University of Michigan. J.A., M.A.D., M.B., C.M., C.G. and N.G.W. wrote the manuscript.

Chapter III: A hierarchical clustering approach to single molecule FRET analysis - Dissecting pre-mRNA dynamics throughout the splicing cycle[2]

III.1 Introduction

The dynamic assembly of the spliceosome is a highly regulated process that ensures catalysis occurs only after the proper checkpoints have been satisfied. Central to this regulation is the recognition and interaction of the pre-mRNA with the spliceosome. Characterizing pre-mRNA dynamics at various stages of assembly provides us with information regarding the role of known assembly factors and ATP consuming helicases in the promotion of catalysis. Data from the few single molecule experiments available on the spliceosome revealed that the dynamics are central to its function.[41,48] Hoskins *et al.* have shown that the assembly of the spliceosome procedes in an ordered stepwise manner. Previously, this process has been depicted as an irreversible event; the data confirmed a dynamic process where every major assembly event was reversible. We have previously shown that pre-mRNA reversibly samples a variety of conformations during the ATP independent and ATP dependent assembly steps. The addition of ATP leads to a change in dynamics that results in formation of spliced mRNA. The multitude of different conformations sampled stochastically in the absence of molecular synchronization limited our ability to assign dynamics to specific parts of the assembly cycle. To refine our understanding of the dynamics at specific steps of the splicing cycle we here have utilized a combination of substrate mutants and

[2] Adapted from: Blanco, Mario*, Ramya Krishnan*, Matthew Kahlscheuer, Joshua Martin, Alain Laederach, Nils Walter. "A hierarchical clustering approach to single molecule FRET analysis: dissecting pre-mRNA dynamics during spliceosome assembly". (in preparation) The single molecule experiments were carried out by Ramya Krishnan, Mario Blanco, and Matthew Kahlscheuer. The hierarchical clustering approach was developed by Mario Blanco with help from Joshua Martin. *Contributed equally to this work

extract modifications to stall the spliceosome assembly process. In total, we utilized seven unique blocks that enrich for the early, middle, and late stages of the splicing cycle (Fig 3.1). However, previously established single molecule analysis tools that regard individual state transitions as independent from one another were insufficient for the detailed analysis required for such a diverse set of behaviors. The enrichment of complexes is not complete and therefore a quantitative assessment of the variety of behaviors present within the population of molecules for each condition is required. Clustering tools provide the necessary analytical power to help tease out individual subpopulations from a larger set of behaviors. Here we have employed hierarchical clustering tools to help extract the consensus behaviors of pre-mRNAs throughout the splicing cycle. This technique is generalizable for other data sets with complex dynamics whose analysis by traditional smFRET tools is insufficient to extract the full scale of information available.

III.2 Materials and Methods

Single molecule measurements

Single molecule measurements were carried out as described in (**Abelson, Blanco – see Chapter II). The pre-mRNA substrates used in this study were immobilized via a direct biotinylation at the 5'-end in place of being hybridized to a biotinylated 2'-O-methyl oligonucleotide hybridized at the 3'-end. Single molecules were selected for further analysis after verification that they had each a single donor (Cy3) and acceptor (Cy5) as evident from photobleaching analysis or direct excitation of the Cy5 dye with a 635 nm laser. Histograms were constructed by sampling the first 10 seconds of each trajectory. HMM analysis was carried out by applying a global model to the entire data set of molecules utilizing the vbFRET[79] program. The number of states allowed ranged from 1-10 and vbFRET was allowed to choose

the final number of states, found to be 7 states. The global fit ensures that the same idealized FRET values are sought in all data sets, allowing for a direct comparison of state abundance and state-to-state transitions.

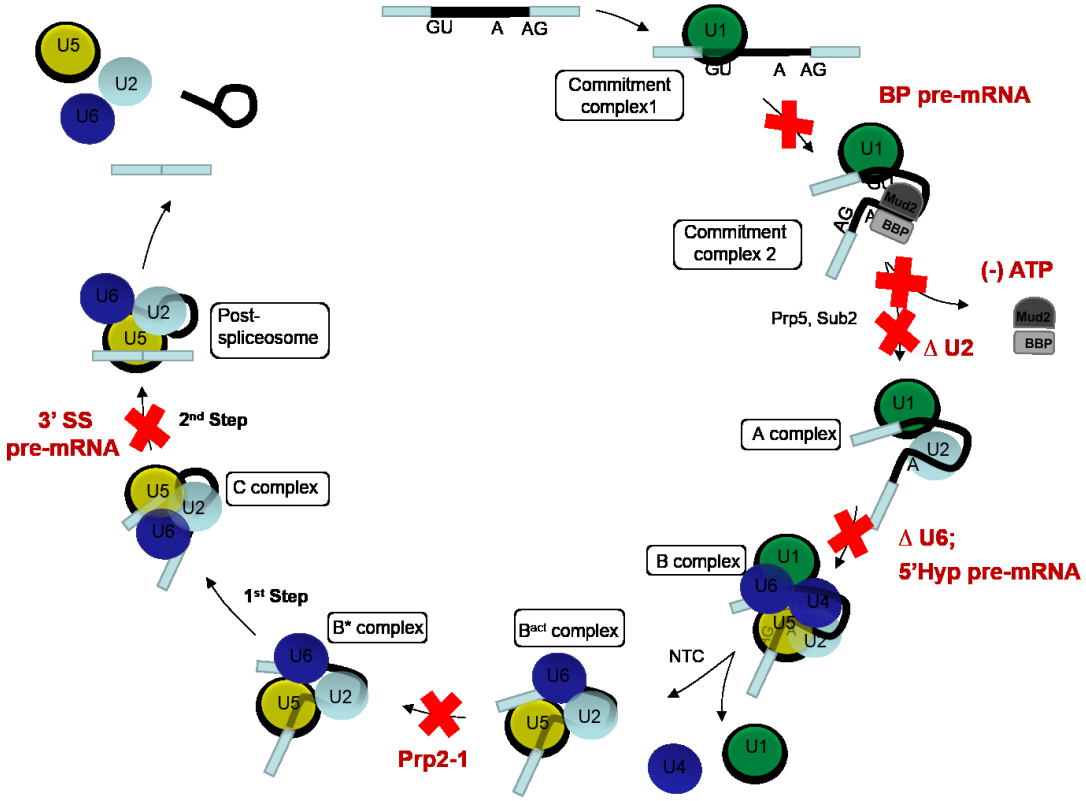


Figure 3. 1: Spliceosome assembly in budding yeast can be blocked at various stages of assembly.

Pre-mRNA splicing in budding yeast can be blocked through a variety of substrate mutants and modifications to the in vitro splicing extract. Splicing blocks enrich for intermediate complexes that altogether comprise the entire splicing cycle yielding spliced mRNA.

Extract preparations

Wildtype yeast splicing extracts were prepared as previously described.[41] U2 snRNA and U6 snRNA ablations were performed as previously described.[80,81] Prp2-1 extract was heat inactivated for 40 minutes at 37ºC to inactivate the Prp2 protein.[30] The ablation of activity and subsequent reconstitution of modified extracts was verified with ensemble splicing assays after each single molecule experiment.

Ubc4 pre-mRNA and mutant variants

The Ubc4 pre-mRNA sequence is the same as that previously described in Abelson *et al.*[41] The RNA has a 5'-biotin used for immobilization on the microscopy surface. Sequences for the mutants are given here (mutated sites are bolded and highlighted):

5'HYP:

5'GAACUAAGUGAUCAGAAAGGUA**A**GU**A**U**U**A**U**GUUAUGGCCACGUUUCAAAUGCG
UGCUUUUUUUUUAAAACUUAUGCUCUUAUUUACUAACAAAAUUCAACAUGCUAU
UGAACUAGAGAUCCACCUACUUCAUGUUT3'

BP:

5'GAACUAAGUGAUCUAGAAAGGUAUGUCUAAAGUUAUGGCCACGUUUCAAAUGC
GUGCUUUUUUUUUAAAACUUAUGCUCUUAUUUACUA**C**CAAAAUCAACAUGCUAU
UGAACUAGAGAUCCACCUACUUCAUGUUT3'

3'SS:

5'GAACUAAGUGAUCUAGAAAGGUAUGUCUAAAGUUAUGGCCACGUUUCAAAUGC
GUGCUUUUUUUUUAAAACUUAUGCUCUUAUUUACUAACAAAAUCAACAUGCUAU
UGAACUA**C**A**C**AUCCACCUACUUCAUGUUT3'

Creating an HMM similarity matrix and Clustering Analysis

The FRET data for all the single molecule's collected was fit with a global Hidden Markov Model model to determine and assign idealized states. The HMM analysis was performed with vbFRET allowing the number of states to range from 1-10. 7 states were determined by the vbFRET algorithm. Transition probability (TP) matrices were created for each individual molecule by using the path of idealized states determined by vbFRET and the hmmgenerate function in MATLAB. When a molecule did not sample a particular state that states position on the TP matrix was assigned a zero. This was necessary to make all the TP matrices the same size (7 states x 7 states). The FRET Similarity Matrix (FSM) combined the TP matrix with the fraction occupancy of a molecule at each state (1 x7, histogram), therefore the

final size of the FSM is 7 x 8. The FSM was then clustered using the hierarchical clustering toolbox in MATLAB. Before clustering molecules sampling only one state (static) were separated and grouped independently because their low complexity makes them easy to group together. A total of 5 different clusters of static clusters was identified. Briefly, the pairwise distance of each molecule's FSM was calculated using the Euclidean metric. These distances were then joined using the linkage function that created a hierarchical tree joined at nodes based on the median Euclidean distance between the trajectories. The hierarchical tree can be clustered by determining a height cutoff between pairs of nodes. We determined the appropriate cutoff by measuring the intra and inter cluster distances at several different thresholds. A threshold was chosen that resulted in 14 different clusters being made. These clusters were then joined with the 5 static clusters previously identified. Once the clusters were identified we could quantify the number of molecules present in each cluster per condition. To determine consensus behaviors for each cluster we calculated the mean TP matrix for each cluster and then determined the trajectory with the shortest Euclidean distance to that mean.

III.3 Results

Ensemble characterization of branchsite labeled Ubc4 pre-mRNA variants

In budding yeasts, the splice sites are nearly invariant throughout the genome. Point mutations at these splice sites lead to strong blocks that have been utilized to characterize their requirement at the various steps in spliceosome assembly.[68] We have previously characterized a doubly exon labeled Ubc4 pre-mRNA and substrate mutants via ensemble and single molecule experiments (Chapter II).[41] To interrogate the pre-mRNA dynamics near the site of first step of catalysis we have modified the position of the donor fluorophore (Cy3) so that it is at position +6 from the branchsite adenosine while maintaining the acceptor fluorophore (Cy5) at position -7 from the 5' splice site. The presence of Cy3 in the intron might affect the splicing efficiency of the Ubc4 pre-mRNA as well as how substrate mutants react in splicing extract; therefore it is important to verify the activity of this substrate through ensemble *in vitro* splicing assays. From the ensemble splicing assays we determined that the WT Ubc4 branchsite labeled substrate can undergo both steps of splicing efficiently when incubated in yeast cell extract supplemented with ATP (Fig. 3.2 top left) . After establishing that the labeled WT substrate can splice efficiently, we proceeded to introduce mutations at the conserved splice sites. The BP Ubc4 pre-mRNA has the branchsite adenosine replaced by a cytosine (A->C). This mutation has the effect of blocking nearly all spliceosome assembly, leading to a defect in both steps of splicing. The ensemble splicing assay verifies that neither first step nor second step products are produced in the presence of ATP (Fig. 3.2 top right). Another strong pre-mRNA variant is the 3'SS pre-mRNA. The 3'SS has a mutation at the 3'-end of the intron that blocks the second step of splicing. The mutation consists of a G-> C change at the exon ligation acceptor site: AG. The Ubc4 pre-

mRNA contains two consecutive AGs near the 3'-exon and mutation of just one led to the use of the adjacent AG; therefore this 3'SS mutant is a double G->C mutant. As expected, this mutant undergoes only the first step of splicing with accumulation of the intron-lariat intermediate but no mRNA formation (Fig 3.2 bottom right). Another mutation we have utilized in this study is the 5'SS hyperstabilized pre-mRNA (5'HYP). Unlike the BP and 3'SS substrates where mutations decrease the affinity of spliceosomal components to the splice site this mutation serves as an impediment to spliceosome assembly by increasing the complementarity between the U1 snRNA and 5'SS. This impedes spliceosome assembly by affecting the RNA switch at the 5'SS where the U1 snRNA is exchanged for the U6 snRNA through the action of the RNA dependent ATPase Prp28.[82] The 5'HYP substrate (Fig. 3.2 bottom left) is not a definitive block like the BP and 3'SS mutants and therefore allows both steps of splicing to proceed, albeit with decreased efficiency when compared to the WT substrate.

Modifications to yeast splicing extract provide additional points of enrichment

Spliceosome components, in addition to splice site mutations, can also serve as targets to block spliceosome assembly. The ordered assembly of the spliceosome in budding yeast has been well characterized biochemically.[23] The components are present at specific stable complexes that have been determined by over 25 years of study (Fig. 3.1). By targeting specific components of these complexes we can block assembly as specific steps leading to an enrichment of complexes prior to that block.

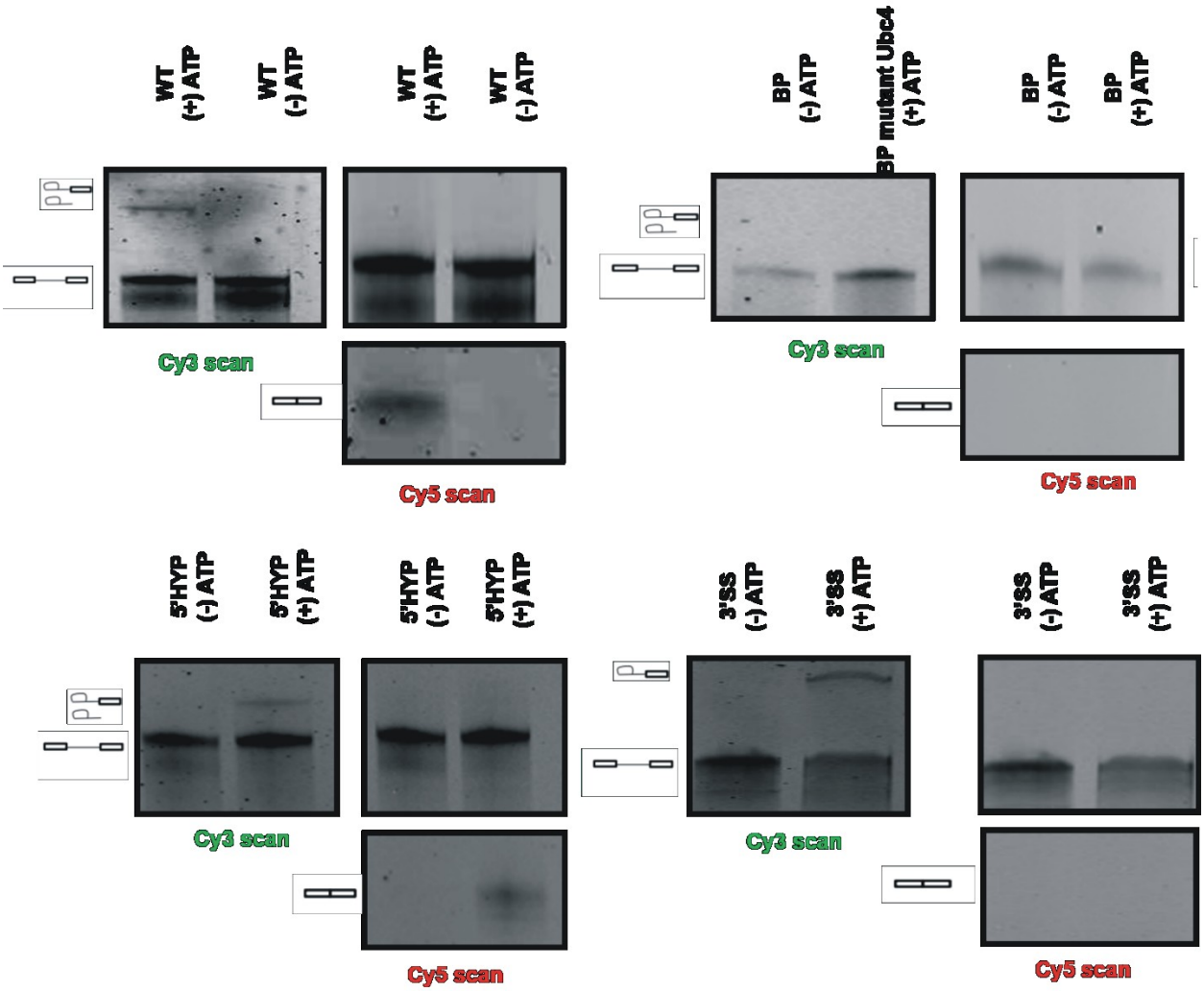


Figure 3.2: Ubc4 pre-mRNA mutants have the expected defects in splicing.

The substrate mutations used have been previously characterized and therefore have specific defects in splicing. As expected the WT substrate undergoes both steps of splicing only in extract supplemented with ATP. The BP substrate blocks early assembly steps and produces no products of chemistry. The 5'HYP substrate is capable of splicing albeit with decreased efficiency when compared to the WT substrate. The 3'SS substrate allows the first chemical reaction to proceed indicated by the presence of intron-lariat intermediate but blocks the second step resulting in no mRNA formation.

The snRNPs present in the spliceosome are recruited at distinct stages of assembly. The snRNA within each snRNP serves as the structural element that brings together the protein components and interacts with the conserved splice sites on the pre-mRNA. The U2 snRNP is assembled and interacts specifically at the branchsite. It represents the first ATP dependent step in the assembly process and, once assembled, is engaged throughout the entire splicing cycle.[28] The presence of RNase H in the *in vitro* splicing extract makes the snRNAs susceptible to cleavage if they are targeted by specifically designed DNA oligonucleotides. The activity of the U2 snRNP can be efficiently targeted via a DNA oligonucleotide (SRU2) complementary to the U2 snRNA.[80] This oligonucleotide can ablate splicing activity from an otherwise active yeast extract by incubating the splicing extract with the oligo at 30º for 30 min. We titrated SRU2 from 300 nM up to 700 nM and determined that 400 nM concentration was sufficient to disrupt splicing activity (Fig. 3.3a). To show that the U2 snRNP component is the only factor affected by SRU2 addition we reconstituted the extract with an *in vitro* transcribed (IVT), truncated form of U2 snRNA (U2Δ107) that has been shown to be sufficient for U2 reconstitution.[80] Pre-incubation of the U2 depleted extract with 50 nM U2Δ107 was sufficient to reconstitute both steps of catalysis. (Fig. 3.3b) Interestingly, 100 nM U2Δ107 had decreased activity, perhaps due to RNA aggregation effects at higher concentration. A similar procedure was performed to deplete the U6 snRNP with an oligonucleotide D1. This was performed as previously described.[81] In our extract, 300 nM D1 was sufficient to deplete the extract of U6 and splicing activity (Fig. 3.4a). This extract could be reconstituted with the addition of 10 nM IVT U6 snRNA (Fig. 3.4b).

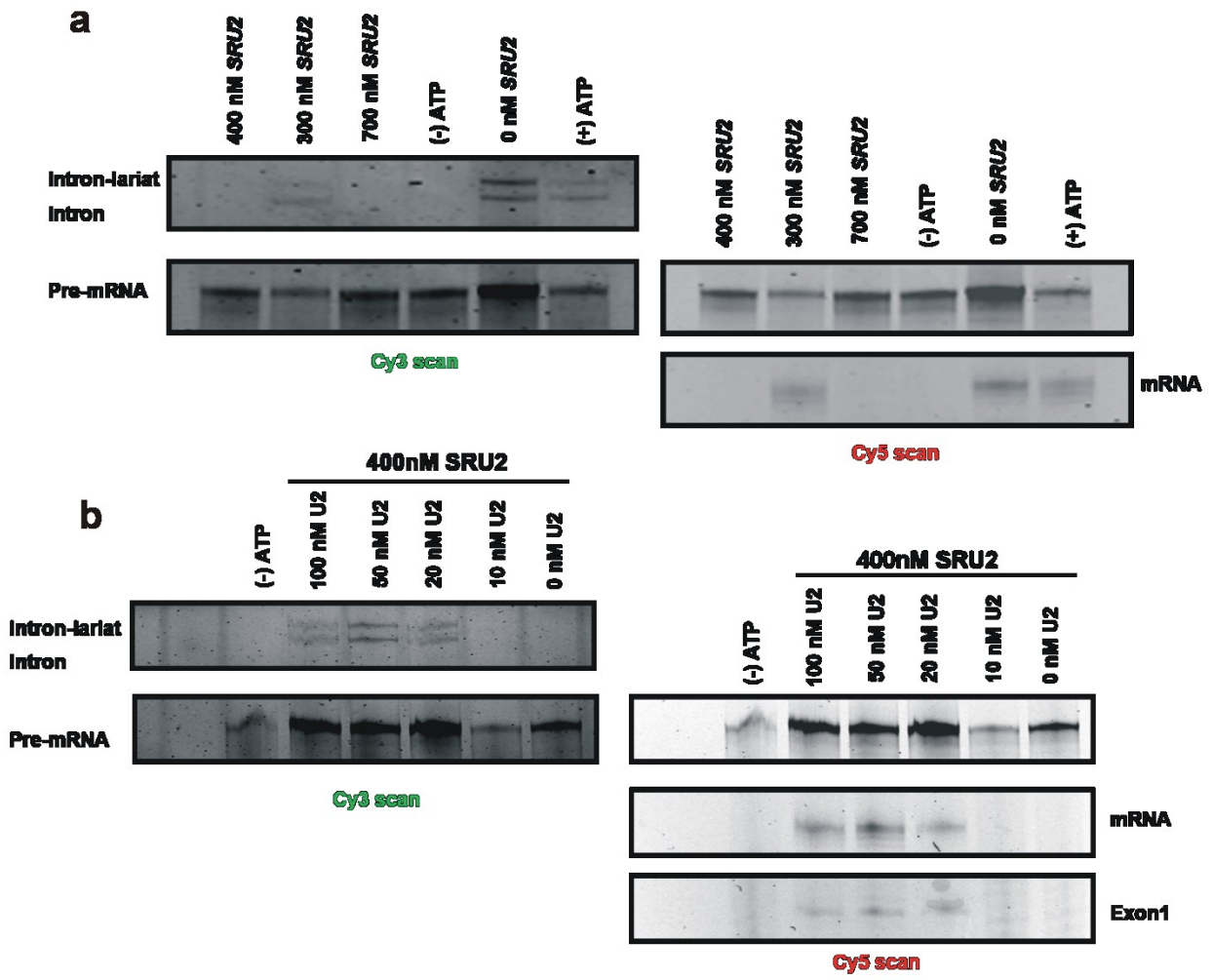


Figure 3.3: U2 snRNP can be efficiently depleted and reconstituted in vitro splicing extracts.

The DNA oligonucleotide SRU2 is complementary to the U2 snRNA and when incubated at a sufficiently high concentration can deplete WT extract of U2. (a) SRU2 was titrated into extract to determining the optimal concentration for U2 depletion as gauged by in vitro splicing activity of WT UBC4 pre-mRNA. 400 nM SRU2 was sufficient to abolish all splicing activity. (b) In vitro transcribed (IVT) U2Δ107 was supplemented to extract depleted of U2 via 400 nM SRU2 mediated degradation. 20 nM IVT U2Δ107 snRNA was sufficient to reconstitute activity. 50 nM IVT U2Δ107 gave the highest reconstitution activity. This verifies that depletion of U2 by SRU2 is specific and does not affect other components of extract.

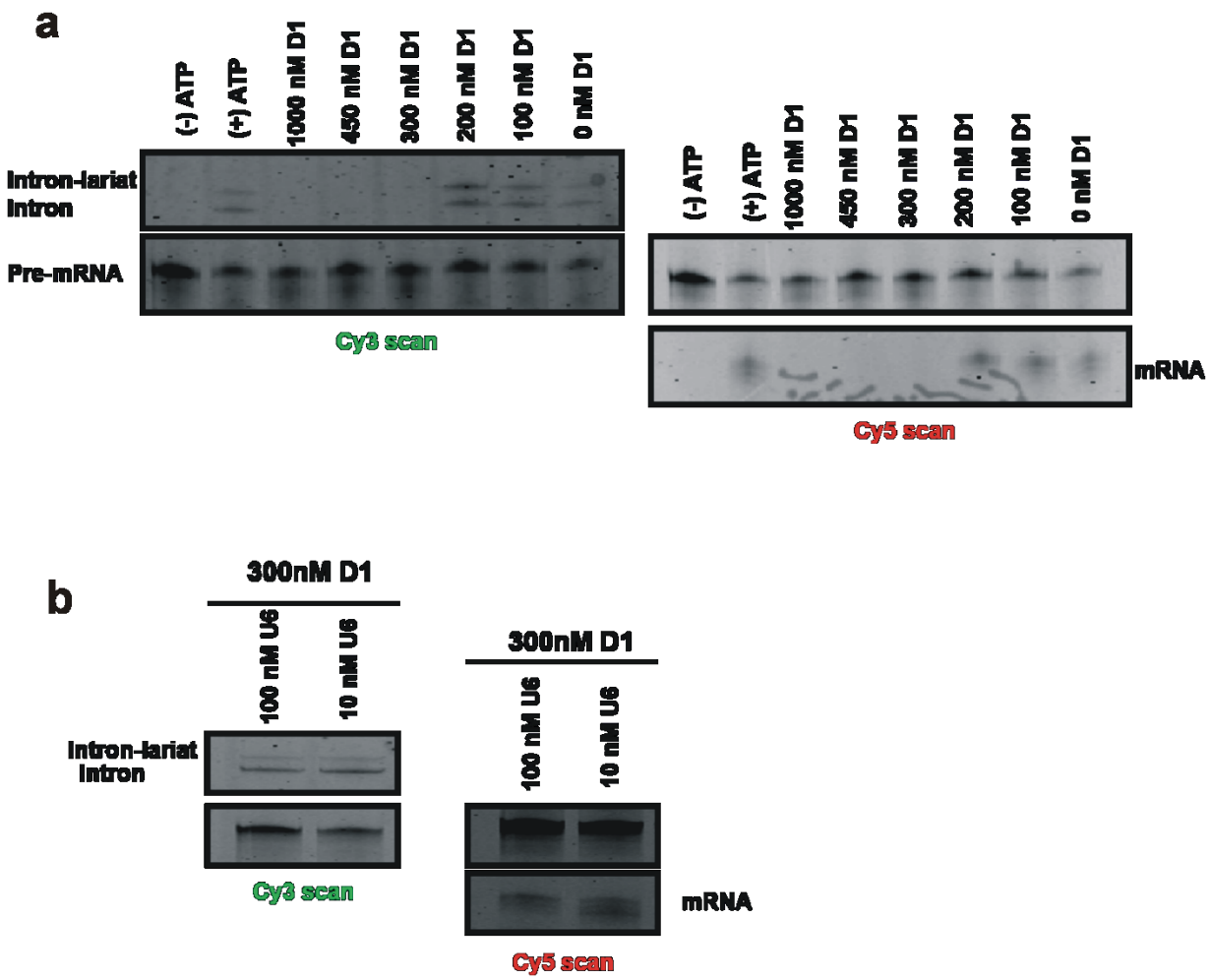


Figure 3. 4: U6 snRNP can be efficiently depleted and reconstituted in vitro splicing extracts.

The DNA oligonucleotide D1 is complementary to the U6 snRNA and when incubated at a sufficiently high concentration can deplete WT extract of U2. (a) D1 was titrated into extract to determining the optimal concentration for U6 depletion as gauged by in vitro splicing activity of WT UBC4 pre-mRNA. 300 nM D1 was sufficient to abolish all splicing activity. (b) In vitro transcribed (IVT) U6 was supplemented to extract depleted of U6 via 300 nM D1 mediated degradation. 10 nM IVT U6 snRNA was sufficient to reconstitute activity and no increase in activity was seen with 100nM IVT U6 gave the highest reconstitution activity. This verifies that depletion of U6 by D1 is specific and does not affect other components of extract.

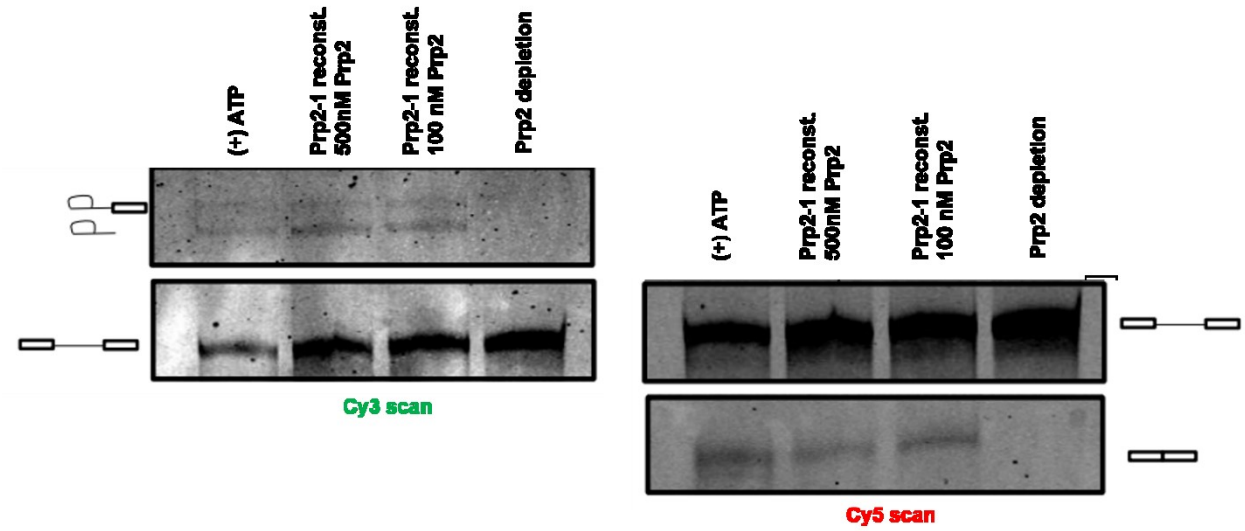


Figure 3. 5: Prp2-1 mutant extract can be heat inactivated and rescued by recombinantly expressed Prp2.

Prp2-1 extract carries a temperature sensitive mutation in the RNA dependent ATPase Prp2 that can be heat inactivated at 37 ºC for 40 min. This blocks any catalysis but allows assembly to proceed until just before the first step of catalysis. This block can be relieved by supplementing heat inactivated extract with just 100 nM recombinantly expressed Prp2 protein.

Budding yeast as a model organism for splicing affords us the ability to utilize the power of yeast genetics to manipulate individual splicing factors. One such example is the introduction of temperature sensitive alleles into genes that allow an extract to be inactivated by incubation at higher temperatures. The DexD/H box putative helicase Prp2 is an essential component in spliceosome catalytic activation. A temperature sensitive allele termed Prp2-1 makes this protein susceptible to denaturation both in live yeast and its cell extract through heat inactivation at 37ºC.[24] This allele has been used to enrich for the B_{act} complex in the spliceosome assembly pathway.[30] We utilized a yeast strain carrying the Prp2-1 allele to make splicing extracts. The Prp2-1 extract was then heat inactivated by incubation at 37ºC for 40 min. The heat inactivation blocks both steps of splicing (Fig 3.5) but allows for assembly. Reconstitution of the heat inactivated Prp2-1 extract through the addition of recombinantly expressed Prp2 protein ensures that only this component was affected, thus allowing the spliceosome to assemble and stall just prior to Prp2 activity (Fig 3.5).

Together the substrate mutations and extract modifications characterized here allowed us to collect single molecule FRET data on our branchsite labeled substrate with multiple points of blockage along the splicing cycle. We have previously established that the conformational behaviors of single pre-mRNAs undergoing splicing are complex with multiple interconverting states and a variety of kinetic behaviors.[41] The data collected with the branchsite labeled substrate also exhibited complex behaviors and therefore required a new approach to single molecule data analysis to extract the full information content in the data.

FRET Similarity Matrix (FSM): A Hidden Markov Modeling (HMM) derived similarity matrix suitable for clustering analysis

Substrate mutations or extract modifications limit the pre-mRNA dynamics to assembly steps prior to that block thus enriching certain behaviors and limiting others. We sought an approach capable of incorporating the complex, stochastic kinetic behavior of any given molecule and objectively determine how the dynamics of molecules collected during the various block conditions are similar or different. The large amount of data collected (>10,000 single molecules) and the complex behavior within the single molecule trajectories present a challenge when using traditional smFRET analysis routines[47]. Clustering analysis has been used in a variety of fields to help group large data sets (e.g., gene expression profiles) based on a similarity metric that helps distinguish similar from disparate behaviors.[83] Applying such an approach to our single molecule splicing data provides us with the analytical power to first classify, then assign dynamics along the splicing pathway. The ability to objectively cluster molecules with no *a priori* assumptions regarding their behaviors requires a similarity metric that fully captures the behaviors of all single molecules in the dataset. State A has a 75% probability of remaining in state A with only 25% chance of transiting to state B. The opposite holds true for State B. This distribution of probabilities leads to a molecule that spends the majority of the time occupying state A (Fig. 3.6b). In addition to the TP matrix, it is possible to extract a frequency distribution of states in the form of a histogram. The histogram describes the overall occupancy of states. In an effort to create a single molecule FRET similarity matrix (FSM) for clustering analysis, we combined the information from the TP and histogram information into a single matrix. A global HMM fit of all

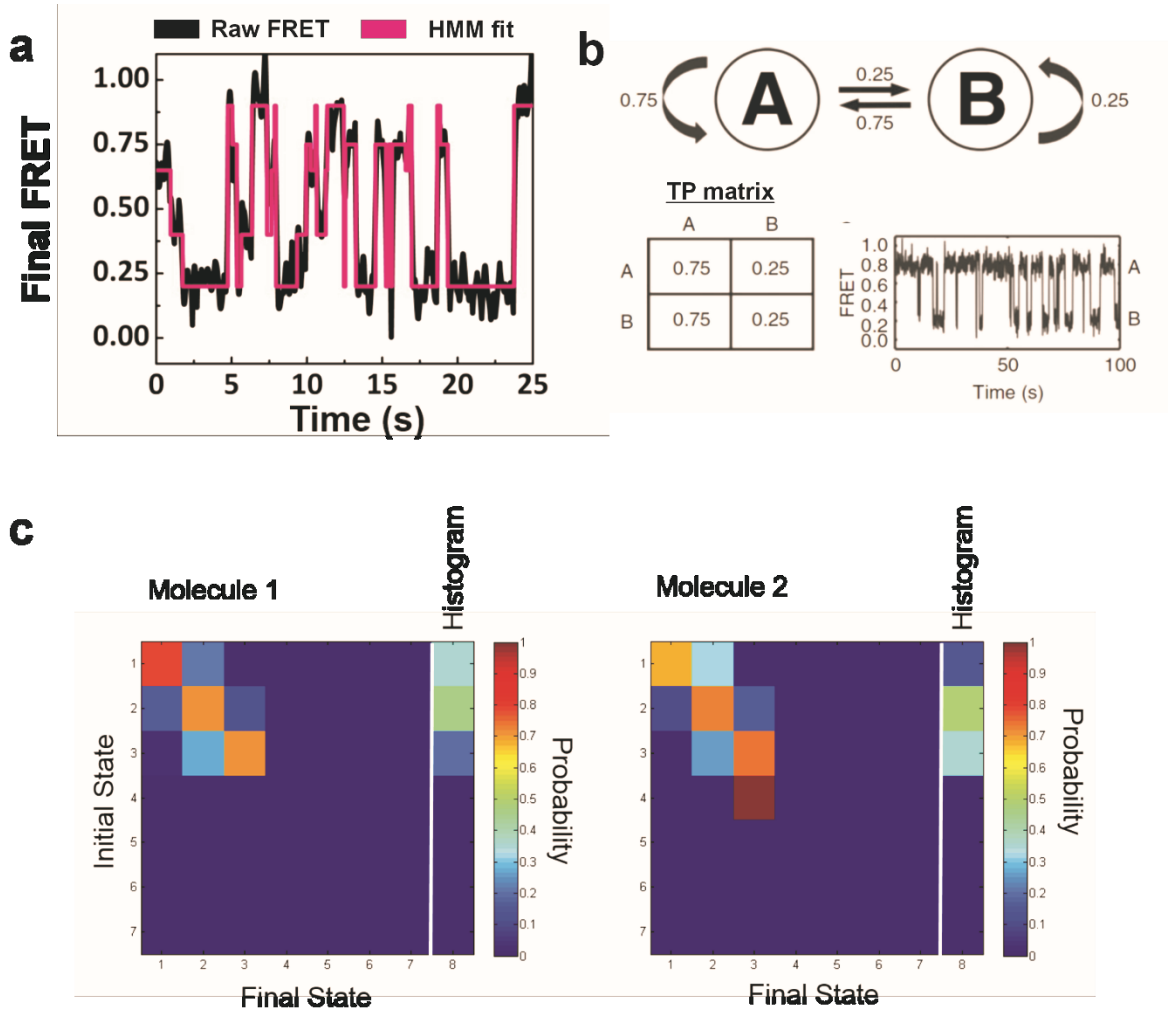


Figure 3. 6: FSM: A Hidden Markov Modeling derived similarity matrix.

(a) Hidden Markov Models (HMM) can objectively and accurately analyze single molecule FRET signals with complex behaviors. A complex single molecule trajectory with several underlying states is fit with an HMM algorithm that can identify change points within the trajectory. **(b)** A simplified 2 state HMM with defined transition probabilities. Transition probabilities describe the likelihood of any given state either transiting to another state or remaining in the same state. The Transition Probability (TP) matrix summarizes all the pairwise probabilities and can be used to describe the kinetic behavior of a single molecule. **(c)** A FRET similarity matrix (FSM) based on a TP matrix can be constructed for individual molecules and used as a clustering tool. The similarity matrix in this publication is a 7x7 matrix (7 idealized states) based on the TP matrix calculated for each molecule and a final column (column 8) which is the distribution of data points at each state, i.e. a histogram. The FSM of two similar molecules, depicted as heat maps, are shown. A vectorized form of the FSM is used for clustering analysis.

our data yielded a total of 7 discrete FRET states (that is, a 7x7 TP matrix). A Matlab routine was written to create an FSM for each single molecule trajectory collected. The FSM can now be used as input into clustering algorithms. Molecules with similar kinetic behaviors have similar values at each matrix position. (Fig. 3.6c). In addition to identifying molecules with varying occupancies in distinct FRET states, the FSM should be able to distinguish molecules occupying the same states but with different rates of inter-conversion. To test this expectation, Data were simulated from two HMM models differing only in their kinetic rates. 25 molecules with 200 data points sampled at 100 ms were simulated from each model (Fig. 3.7a). An FSM was created for each molecule and clustered (Fig. 3.7b). Four groups of molecules were identified from the clustering analysis. The clusters contained molecules that rapidly interconvert between the low and high FRET states (Fig. 3.7c top left) as well as slowly inter-converting molecules (Fig. 3.7c bottom left) and static molecules that did not interconvert during the 200 data points simulated (Fig. 3.7c top/bottom right). This example shows that the FSM allows for the efficient segregation of molecules with similar states but different kinetic behaviors.

Hierarchical Clustering can effectively partition subpopulations from a heterogeneous set of molecules

After establishing the FSM as an effective tool for clustering analysis, we created an FSM for each single molecule trajectory collected. The data sets collected for all the blocks were mixed and then a pairwise distance between all possible trajectories was calculated using the Euclidean distance metric. Static molecules (no transitions with occupancy in only one state) were then separated since their features are simple and can be treated separately. The Euclidean distances between FSMs for dynamic molecules were calculated and used to build a hierarchical cluster tree by utilizing the linkage function in Matlab.

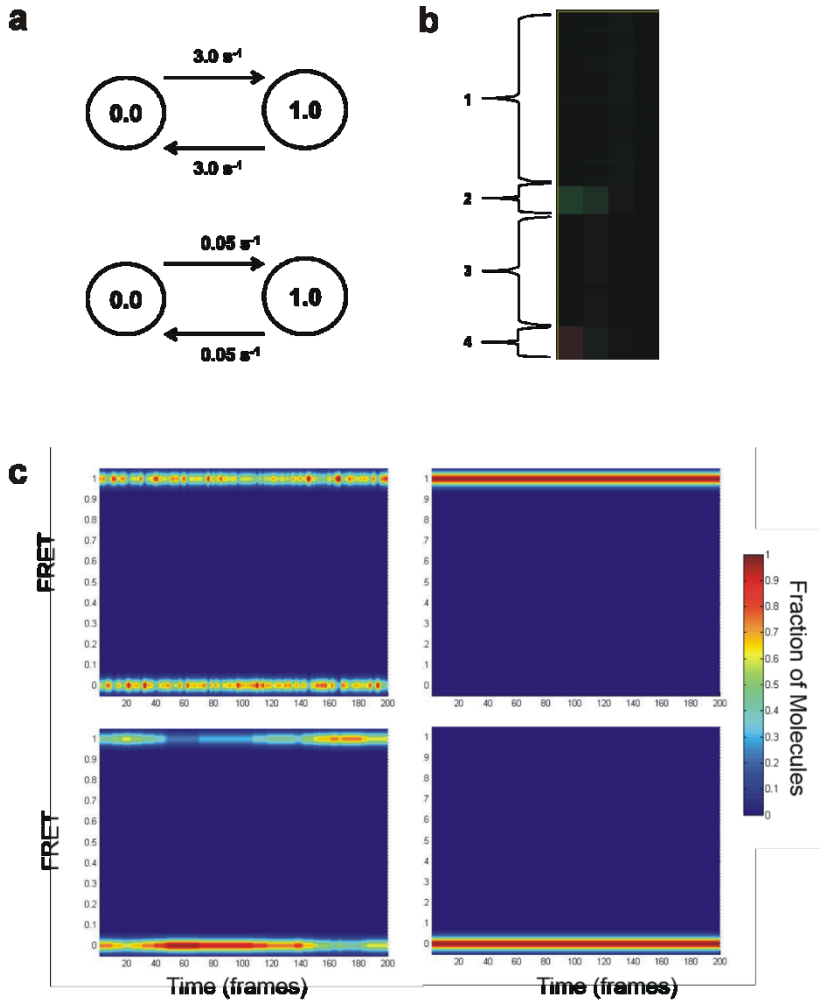


Figure 3.7: A HMM TP similarity matrix can distinguish molecules with the same states but differing kinetics.

(a) 50 simulated molecules were simulated from two HMMs (25 molecules of 200 data points each) with equal states (0.0 and 1.0) but different rates of interconversion. (b) Clustering was performed on simulated molecules using the HMM similarity matrix and four distinct clusters were identified. (c) The clusters identified molecules that were only in one state (top/bottom right) and molecules visiting both states but with either rapid (top left) or slow (bottom left) kinetics. This exemplifies the effectiveness of the HMM similarity matrix as a tool for describing single molecule kinetics.

Molecules with short distances (high similarity) were grouped together and connected to other groups of molecules based on their average inter-cluster distance (Fig. 3.8). A hierarchical tree can be cut (pruned) at various heights, leaving more or less clusters. A pruning threshold was selected based on the inter-cluster distance that minimized the intra-cluster distance (similarity within a cluster) and maximized inter-cluster distance (dissimilarity from other clusters). A threshold was selected that yielded 14 individual clusters (Fig. 3.9). An additional 5 clusters from the static molecules were also incorporated into the analysis, yielding a total of 19 clusters. It is important to note that data are not segregated by experimental condition prior to clustering analysis, allowing molecules with similar dynamics across conditions to be grouped objectively.

The 19 clusters created from the hierarchical clustering analysis can be visually assessed to verify that similar molecules are grouped together to create clusters of high homogeneity. One method of assessing this is by plotting the linearized form of the FSM (7x8 converted to 1x56) of each molecule in its assigned clusters as a heat map of the values at each position in the matrix (Fig. 3.10). The heat maps were organized as follows: each row represents a single molecule while the columns are the different values for each point in the FSM.

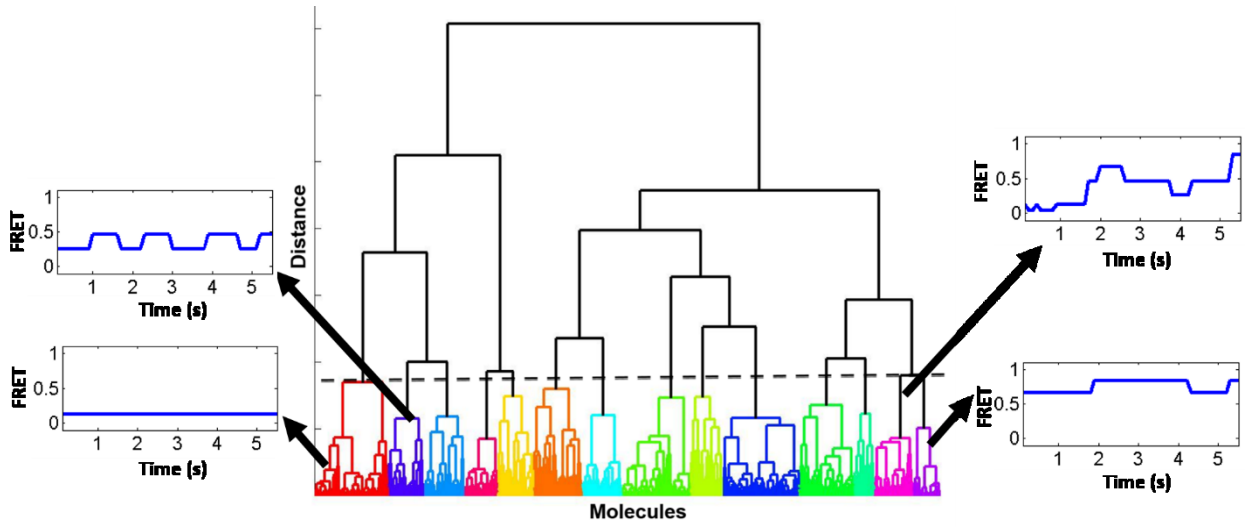


Figure 3. 8: Hierarchical clustering of single molecules.

Single molecule FRET data was collected for each substrate and extract mutant and globally fit to a 7-state HMM. An HMM similarity matrix was then constructed for each molecule and clustered. Molecules with only 1 state (static) were clustered separately and later incorporated as separate clusters. Hierarchical clustering was performed utilizing Matlab routines by calculating the pairwise Euclidean distance of every single molecule and building an agglomerative hierarchical cluster tree utilizing the unweighted average distance to compute distances between clusters. The tree was pruned by using a distance criterion which resulted in the formation of 14 well separated clusters.

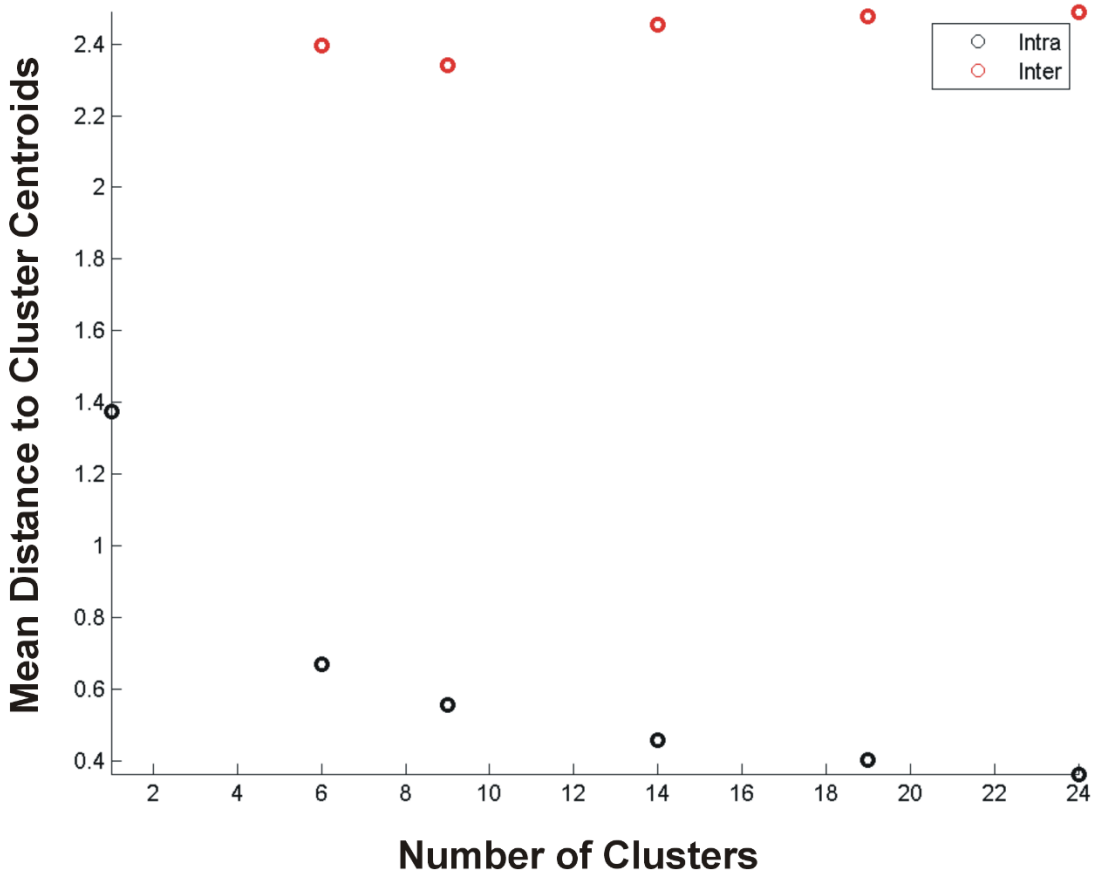


Figure 3. 9: Hierarchical tree pruning.

The hierarchical tree was cut several times varying the node height criterion to yield a range of clusters. Intra and inter-cluster distances were computed by measuring the mean Euclidean distance of the trajectories to the cluster centroids formed at the various cuts. 14 clusters were selected because there was minimal improvement in intra cluster distance beyond this point.

Each cluster projected has a distinct set of values for the FSM but high homogeneity within the clusters. The dynamics of the molecules within the clusters can be extracted from the clusters to visualize the behavior that was grouped. The HMM fit for 20 randomly selected molecules for each cluster was concatenated into a single plot with vertical bars separating individual molecules (**Figure 3.11**). The randomly selected molecules within clusters have similar dynamics and illustrate the behaviors that influenced the clustering results. The consensus behavior within a cluster was determined by calculating the mean behavior of all the molecules within a cluster and then determining the single molecule with the shortest distance to the mean. These consensus molecules represent the average behavior of all 19 clusters (**Figure 3.12**).

Clustering analysis groups biologically relevant data with no user bias

One consequence of clustering analysis is the ability to objectively compare data sets by comparing the abundance of each cluster for each condition. We performed a similar linkage analysis on the 18 different data sets collected. The similarity matrix used for analysis was the abundance profile (fraction of molecules in a given cluster) for each condition. The result of this analysis are represented in a hierarchical cluster tree whose leaf order is based on the similarity among conditions (Fig. 3.13). The first striking feature of this analysis is how the BP mutant data in (-)/(+) ATP are separated from the rest of the experimental conditions. This is in agreement with this perturbation being the most severe, blocking spliceosome assembly at Commitment Complex 1. Data collected with the WT, BP, and 5'HYP substrates in splicing buffer alone are most similar to each other and lie closer to the substrates that are able to assemble early spliceosomal components than the BP mutant data. The single molecules in splicing buffer are

less restricted than the BP mutant data in cell extract and often exhibit FRET dynamics that mimic dynamics

during spliceosome assembly. Similar behavior between pre-mRNAs in buffer and during spliceosome assembly was also observed for the doubly exon labeled Ubc4 initially characterized by smFRET.[41] We also detected differences in smFRET behaviors for substrate mutants, which could explain why the 3'SS mutant is so far removed from the other buffer conformations. This highlights the sensitivity of single molecule measurements and the clustering analysis to differences inherent because of differences in RNA structure. Finally, the WT(+)ATP data and conditions where extract depletions were reconstituted, allowing them to undergo splicing again, are grouped most closely to each other. This result is encouraging because, without any user input, the data that can undergo both steps of splicing have the highest degree of similarity.

In addition to being able to identify experimental conditions that are similar to each other, we can also identify clusters that correlate well with one another. This allows us to identify FRET behaviors that change together. A simple analogy to this would be identifying groups of genes with similar expression profiles, e.g., genes under the control of the same promoter. Indeed, similar analysis is performed on microarrays of genome wide expression patterns.[84] A clustergram is a heatmap of normalized intensity values that can identify groups of clusters that shift in a correlated manner. A clustergram was created based on the abundance profiles of the 19 clusters across all conditions. 12 subclusters (separated by vertical white bars) were identified from the original 19 clusters (Fig. 3.14). This form of analysis can identify dynamics that are dependent on each other. The 12 subclusters identified are summarized in Figure 3.15. There are 3 subclusters that contain correlated behaviors. For our data, the subclusters only minimally

reduced the number of dimensions; however, in other data sets the identification of subclusters may simplify the analysis and provide insight into behaviors that are dependent on the same experimental variable.

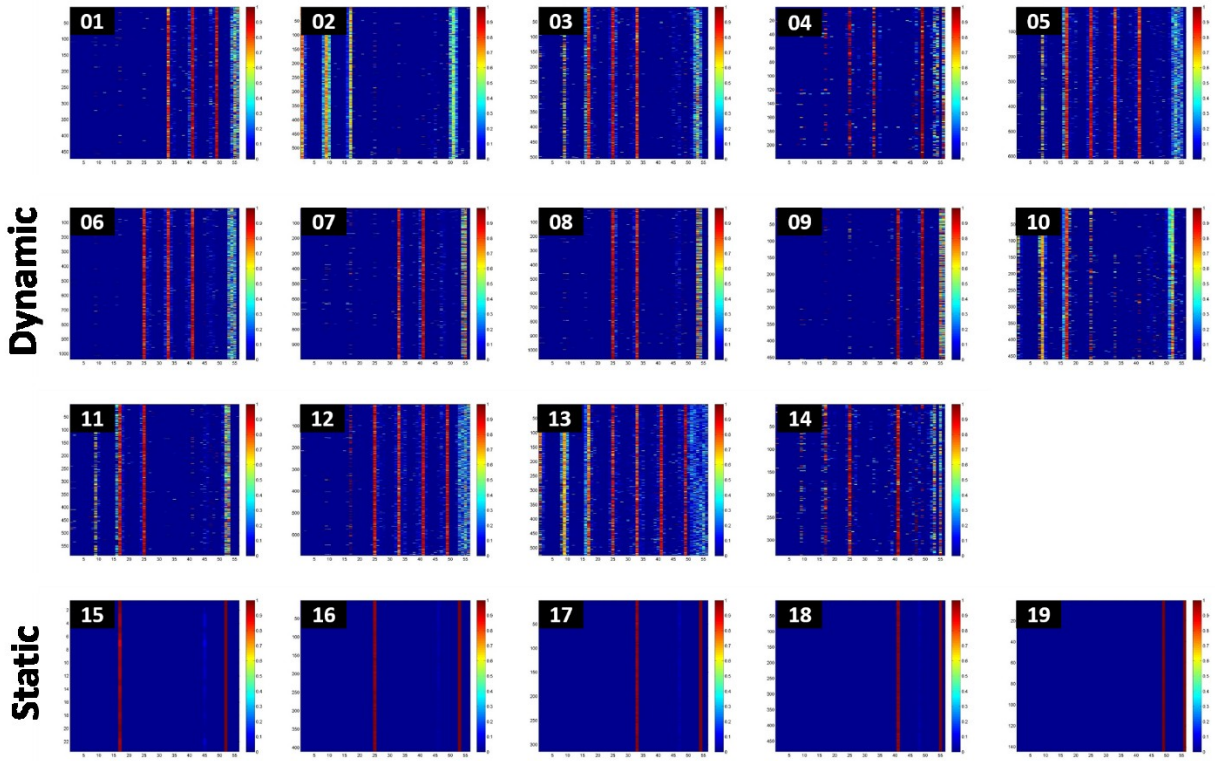


Figure 3. 10: Vectorized FSMs for each cluster.

A total of 19 clusters were derived from the clustering of single molecule splicing data. 14-dynamic clusters, and 5-static clusters. Heat maps of the vectorized HMM similarity matrices for the molecules assigned to each cluster are represented here. These vectorized matrices (columns) represent the features that were used for clustering the individual molecules (rows) by the hierarchical clustering routine.

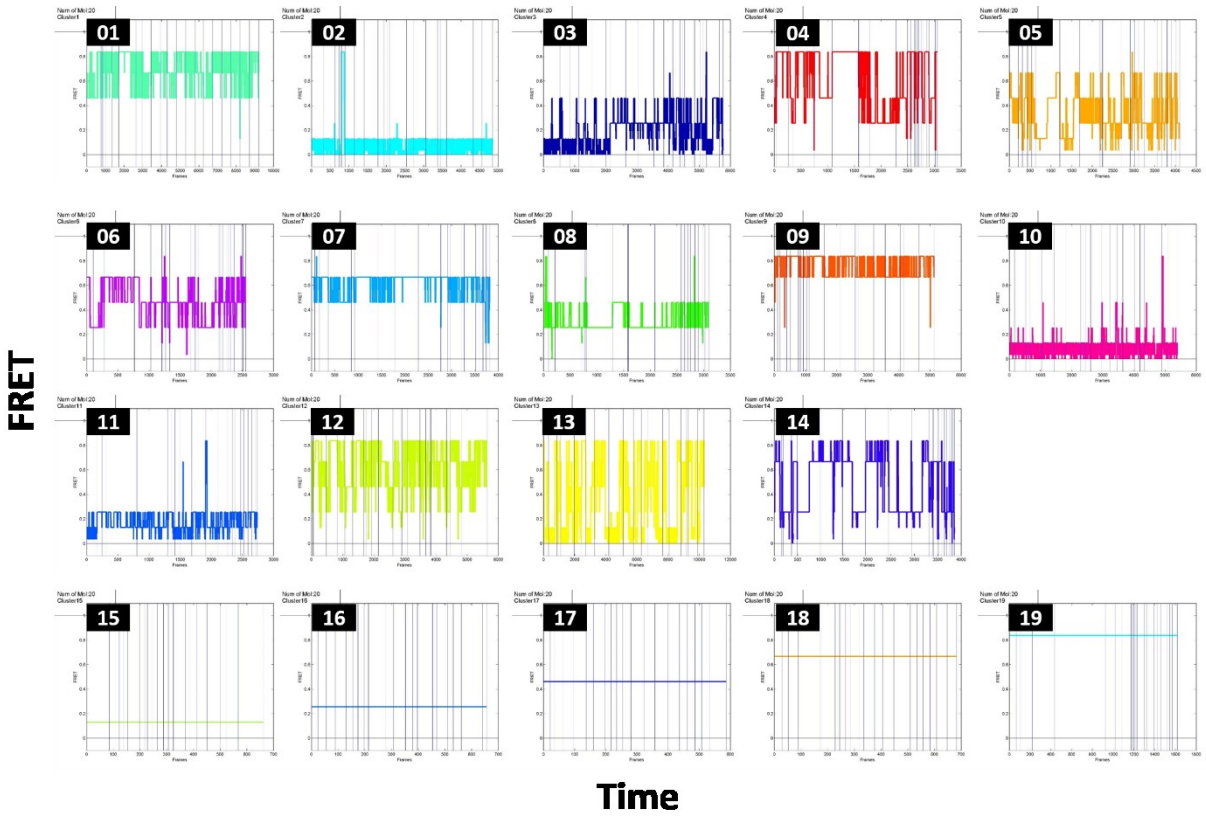


Figure 3.11: Representative molecules for identified clusters.

20 molecules were randomly selected from each of the 19 clusters and concatenated into a single plot. Vertical bars serve to separate the individual molecules. The randomly selected molecules exemplify the typical kinetic behavior of molecules with the assigned cluster. As expected, the behaviors within the clusters are homogeneous and distinct from those in other clusters. Cluster assignement given in the inset.

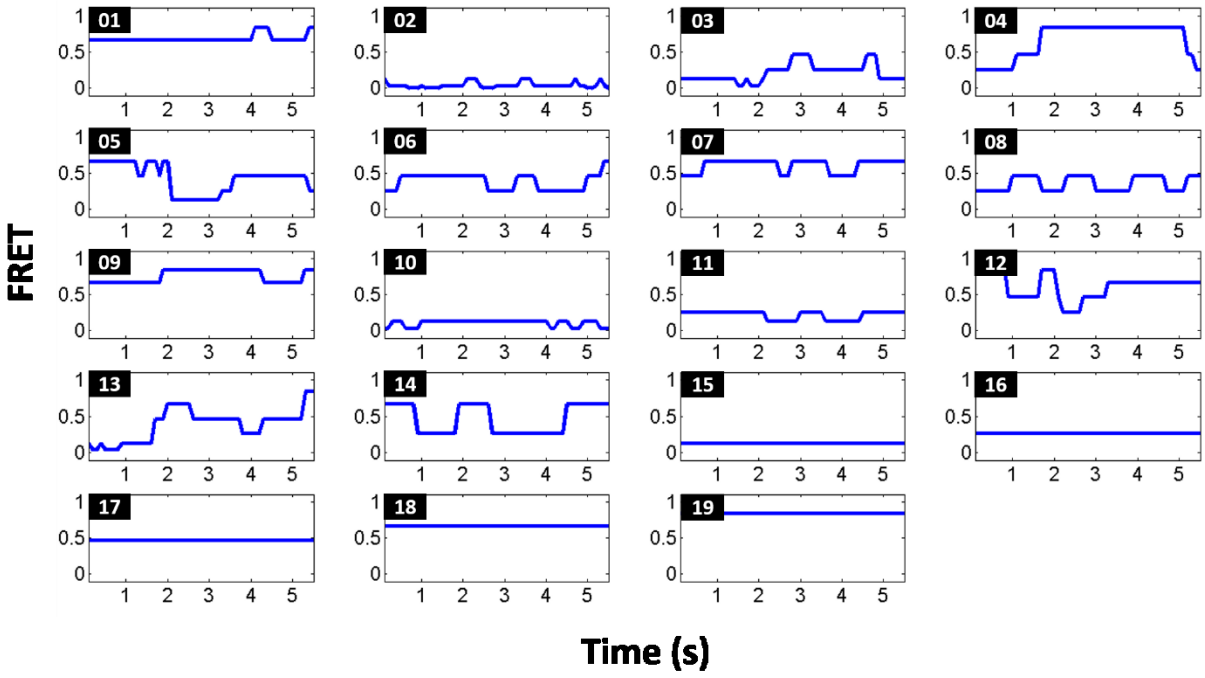


Figure 3.12: Consensus molecule for determined clusters.

The pair wise distance from each molecule to the centroid of its assigned cluster was determined and the molecule with the smallest distance (highest similarity) to the centroid was plotted. These molecules represent the average or 'consensus' behavior of the molecules within the given cluster (inset).

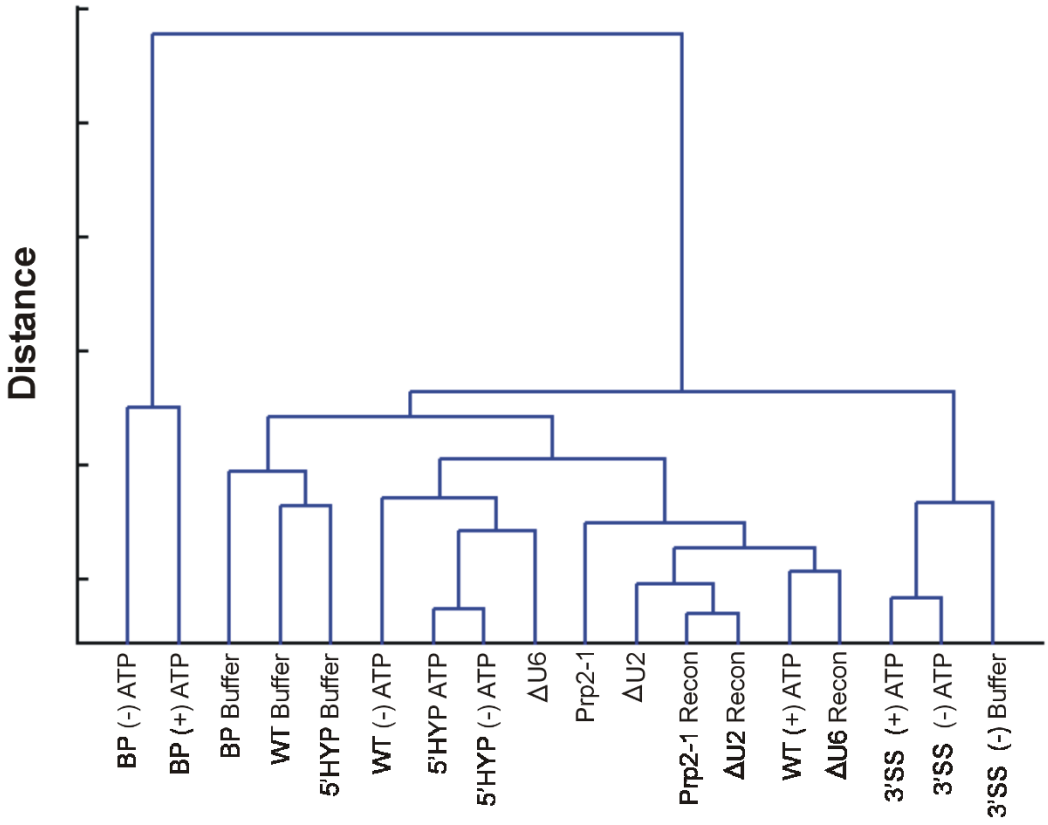


Figure 3. 13: Experimental conditions can be grouped based on cluster abundance profiles.

The abundance of each cluster for each condition was used as a metric for hierarchical clustering. Conditions are ordered based on similarity. The results of this clustering indicates that the BP(-)/ (+)ATP conditions are furthest away from any other condition. The data in splicing buffer collected for each substrate segregate to their own cluster. The WT(+)ATP and reconstitution datasets are also grouped together indicating high similarity between the conditions that can undergo both steps of splicing. Interestingly, the 3'SS datasets are also in their own cluster.

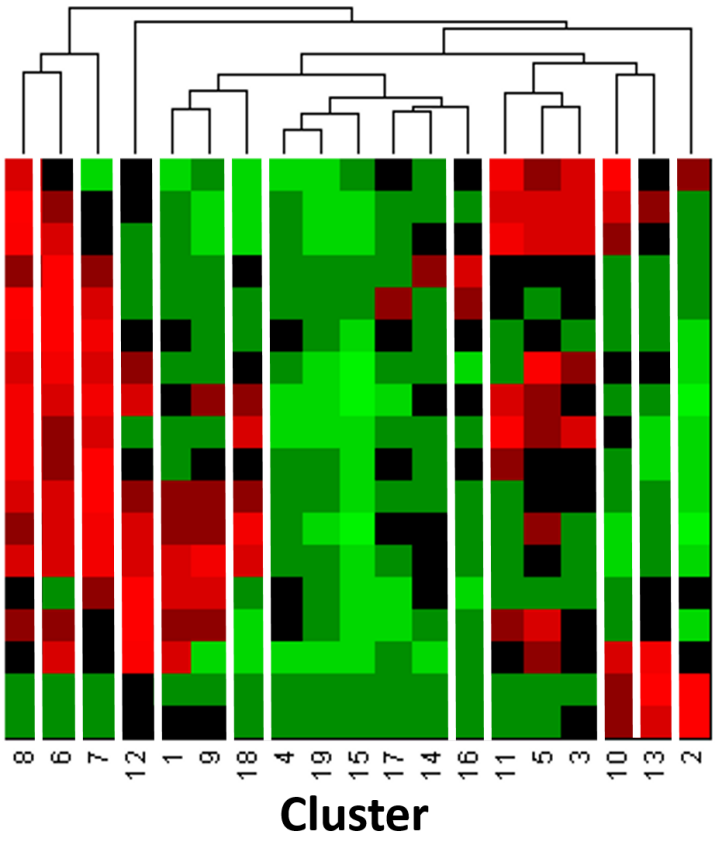


Figure 3. 14: Clusters can be group based on abundance profiles.

The abundance profiles for each cluster were constructed and clustered to determine if groups of clusters were shifting in groups, similar to groups of genes with similar expression profiles. Twelve sub-clusters were identified from the original 19 clusters (columns). A heat map organized with the standardized abundance profiles of all 19 clusters show that 12 subclusters, separated by white vertical bars, were identified. This procedure was performed using Matlab's clustergram function.

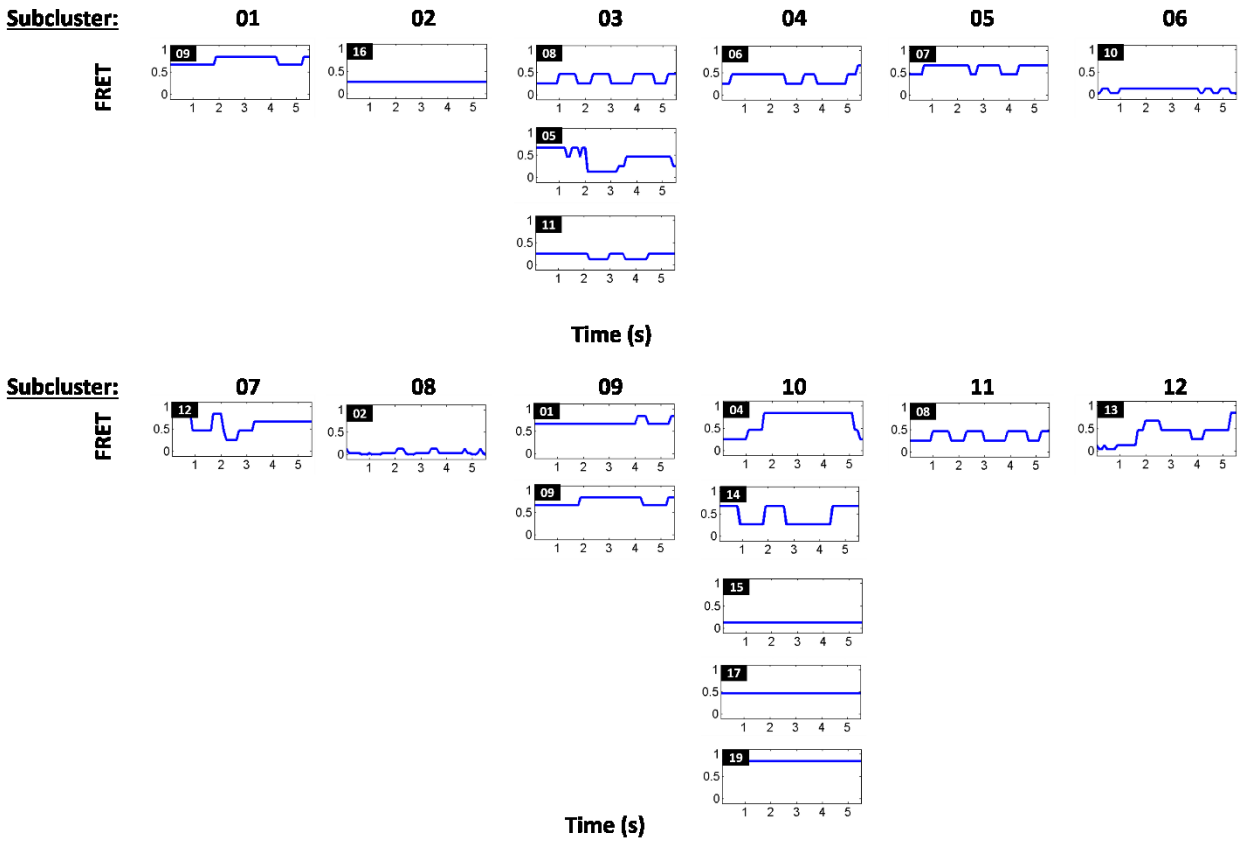


Figure 3.15: Consensus sequences of clusters within subclusters.

3 of the 12 subclusters identified by the clustergram analysis contain multiple clusters. Subcluster 3,9,10 all have more than a single cluster assigned to them. The consensus traces for each cluster is shown with cluster assignment as inset.

Kinetic and Conformational behaviors of substrate mutants during spliceosome assembly

The earliest block in the splicing cycle that we utilized is the BP mutant. This mutant blocks assembly by preventing binding of the Mud2/BBP complex at the branchsite.[85] Comparing histogram data collected in buffer of the BP mutant shows that it can adopt high FRET conformations. Upon addition of (-)ATP cell extract, the substrate becomes constrained and mainly adopts a low FRET distribution. The addition of (+)ATP does not have a dramatic effect on the FRET distribution with the most noticeable change manifesting as a slight increase in the abundance of high FRET states. We can also plot the abundance of clusters (organized by ascending mean FRET value from left to right) to determine what dynamics are leading to the differences between the conditions. The most abundant dynamics in the BP buffer are rapid inter-conversions in the high FRET states (Fig. 3.16). This shifts to dramatically low FRET states populated by 55% of the molecules in (-)ATP extract. The slight increase in the high FRET population with the addition of ATP is due to a redistribution of those molecules in cluster 2 to cluster 13, which exhibits the same low FRET state with a reversible excursion to the high FRET state (Fig. 3.17). Although the canonical spliceosome assembly pathway does not involve any ATP dependent steps during BP recognition, there have been genetic interactions that implicate the ATP dependent DeXD/H helicase Sub2p early in spliceosome assembly.[26] The slight difference in pre-mRNA dynamics near the branchsite of the BP mutant substrate upon ATP addition might be reporting on this interaction.

The 5'HYP substrate has limited splicing activity due to the increased interaction between the U1 snRNA and the 5'splice site. The stabilization of this interaction prevents the proper assembly of the spliceosome by affecting Prp28's ability to exchange the U1 snRNA for U6 snRNA.[82] If the spliceosome cannot be formed, the substrate is subject to proofreading and

discard.[86] The 5'HYP substrate in buffer has a FRET distribution that is lower in FRET than the BP mutant, suggesting that the 5'HYP mutation leads to a change in the conformational space sampled by the naked RNA (Fig. 3.18). The addition of (-)ATP extract diminishes the majority of high FRET dynamics leading to a lower FRET distribution. The addition of ATP has a minimal effect on the FRET distribution, which could be a result of the inefficient assembly of this substrate. The most abundant cluster in splicing buffer is cluster 12, which has rapid dynamics in the high FRET range. In both (-)/(+)ATP conditions, cluster 8 becomes the most abundant, exhibiting relatively slow 0.25-0.50 FRET transitions. This inability of this substrate to assemble efficiently in (+)ATP leads to an accumulation of early assembly complexes.

The 3'SS substrate blocks exon ligation through a mutation in the AG acceptor site near the 3'exon. It allows spliceosome assembly to proceed to just before exon ligation. When examined under single molecule conditions the 3'SS mutant undergoes a similar shift towards lower FRET behaviors from higher FRET after the addition of ATP depleted extract. This seems to be a general phenomenon as all three substrate mutants adopt lower FRET behaviors during early assembly steps. In the case of the 3'SS mutant, the increase in low FRET population originates from molecules now occupying cluster 8 as opposed to cluster 6, the predominant cluster in buffer (Fig. 3.19). Cluster 8 has dynamics similar to the cluster 6 found in buffer except that the transition to a higher FRET state is disallowed, suggesting the early assembly events are preventing the substrate from reaching the high FRET state, therefore leading to the lower FRET distribution. The addition of ATP leads to a slight shift towards high FRET that can be attributed to an increase in cluster 1, which exhibits rapid transitions in the high FRET regime and a decrease in the abundance of molecules in clusters with low FRET state consensus sequences (Fig. 3.20).

WT substrate dynamics stalled at various stages of assembly

The extract modifications allow us to take the WT substrate through early and late steps of splicing assembly. Like the other substrates the WT pre-mRNA undergoes a shift towards lower FRET states from the high FRET states observed in buffer upon the addition of (-)ATP extract (Fig. 3.21). As we progress from the early assembly steps towards the late assembly stages, the most abundant cluster behavior changes with a non-monotonic increase towards high FRET for the WT(+)ATP data (Fig. 3.23). A comparison of the most abundant cluster for the WT(+)ATP data and the reconstitution conditions reveals that the most abundant cluster for all these conditions is cluster 7 (Fig. 3.22 and Fig. 3.23).

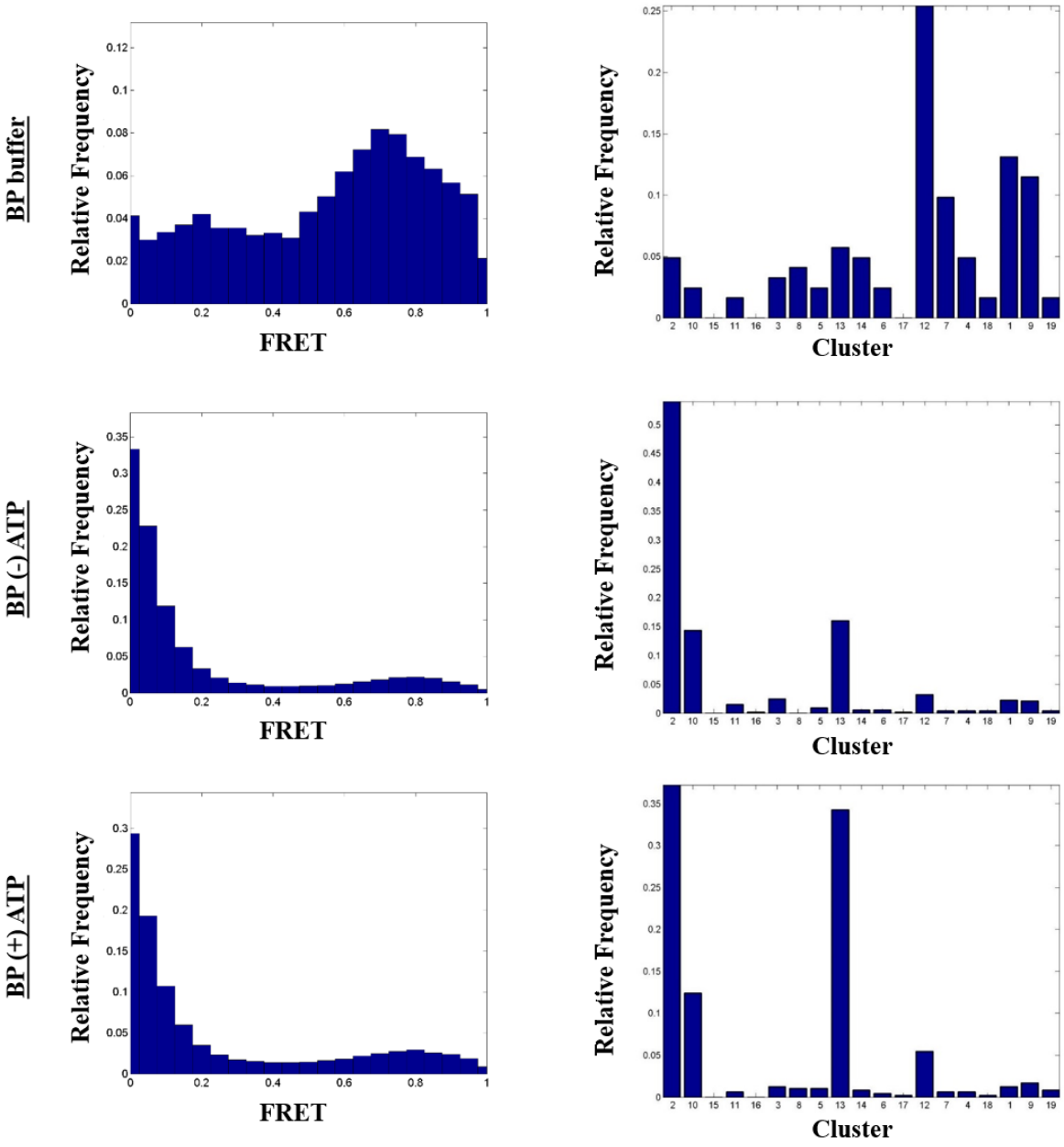


Figure 3.16: BP single molecule dynamics.

FRET histograms of the BP pre-mRNA in splicing buffer/(-)ATP/(+)ATP are compared. Abundance profiles of the original 19 clusters identified from the hierarchical clustering. Clusters are reorganized based on mean FRET value within the cluster so the plot goes from lowest mean FRET (left) to highest mean FRET (right).

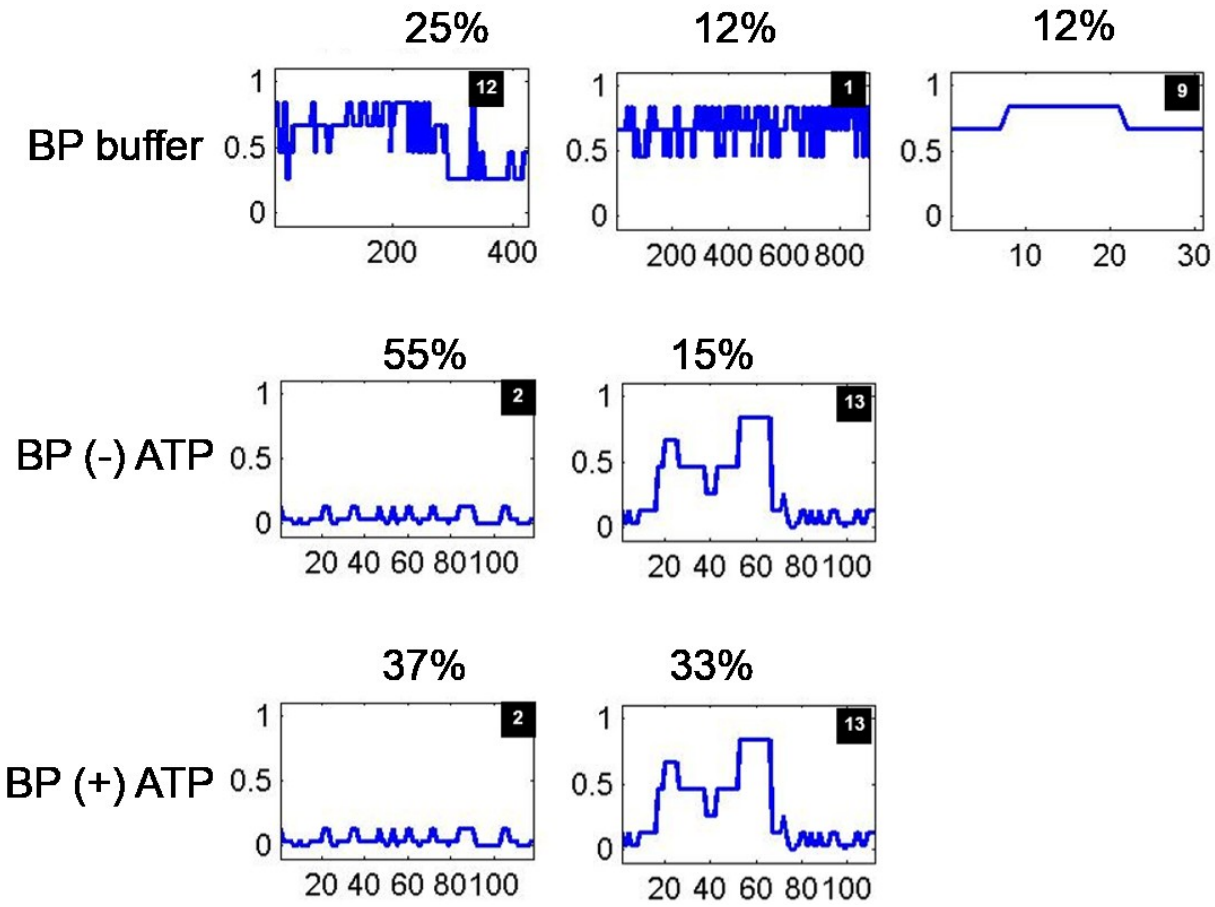


Figure 3. 17: BP consensus behaviors.

The most abundant dynamics present for the BP mutant conditions.

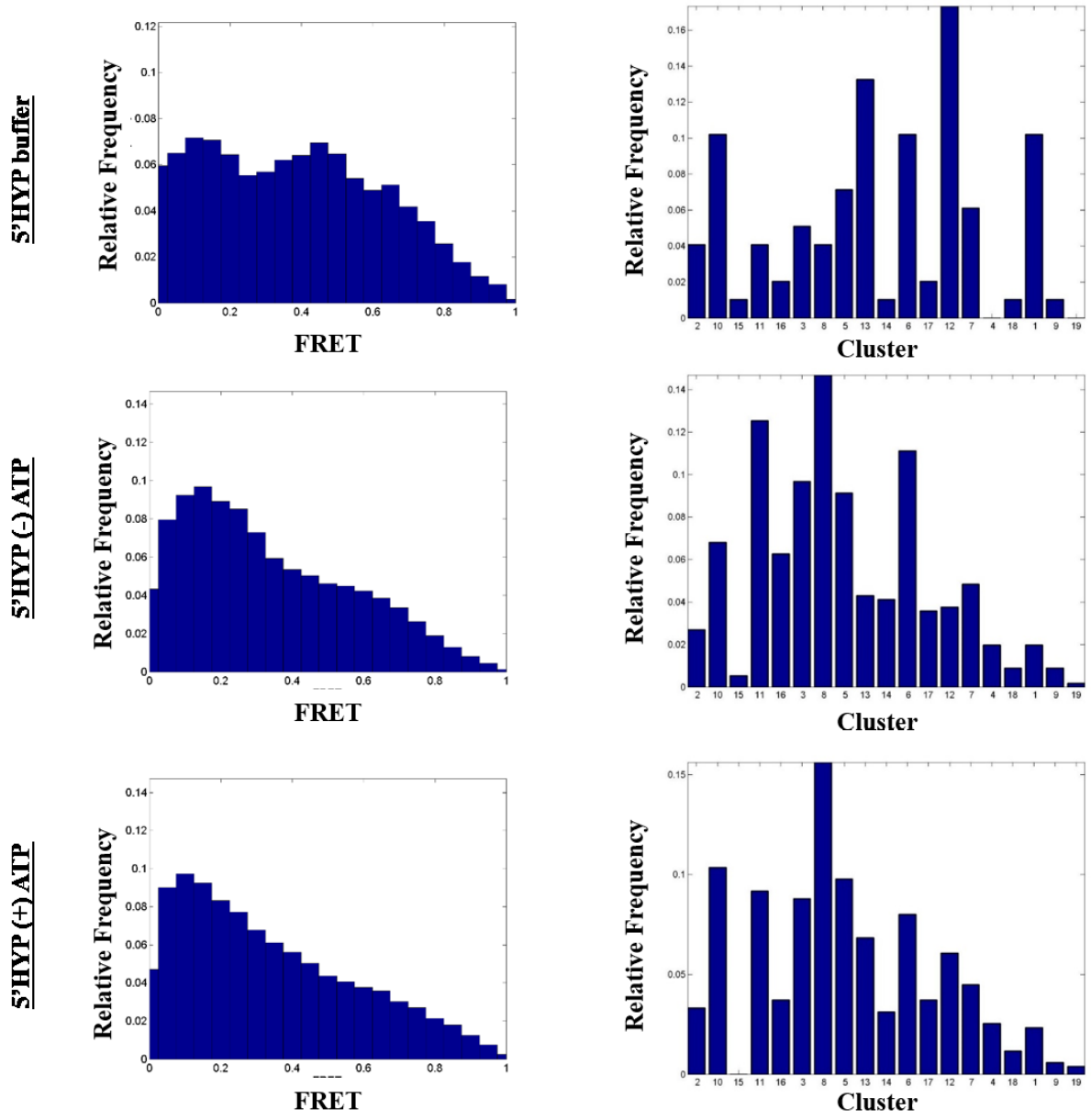


Figure 3. 18: 5'HYP single molecule dynamics.

FRET histograms of the 5'HYP pre-mRNA in splicing buffer/(-)ATP/(+)ATP are compared. Abundance profiles of the original 19 clusters identified from the hierarchical clustering. Clusters are reorganized based on mean FRET value within the cluster so the plot goes from lowest mean FRET (left) to highest mean FRET (right).

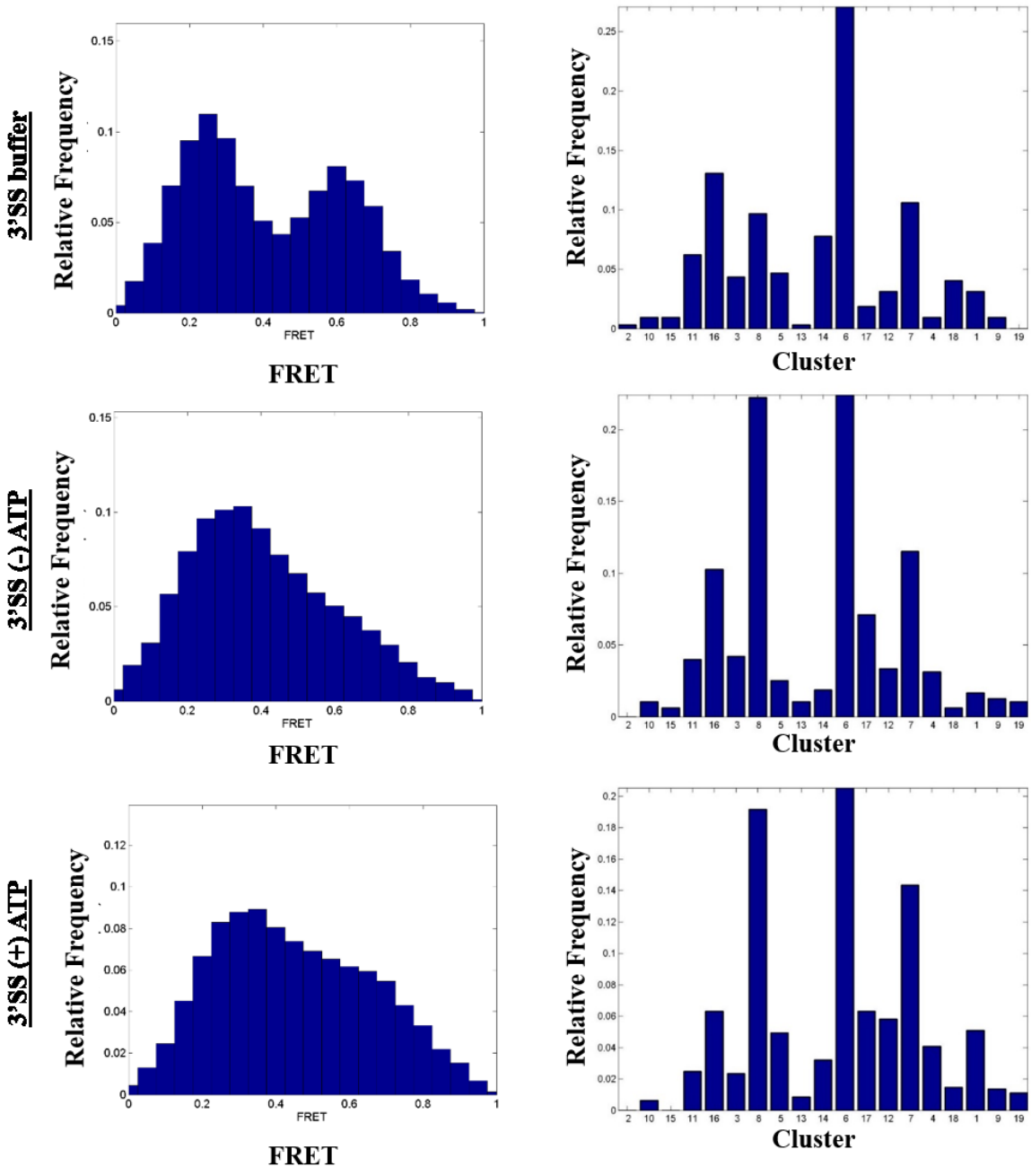


Figure 3.19: 3'SS single molecule dynamics.

FRET histograms of the 3'SS pre-mRNA in splicing buffer/(-)ATP/(+)ATP are compared. Abundance profiles of the original 19 clusters identified from the hierarchical clustering. Clusters are reorganized based on mean FRET value within the cluster so the plot goes from lowest mean FRET (left) to highest mean FRET (right).

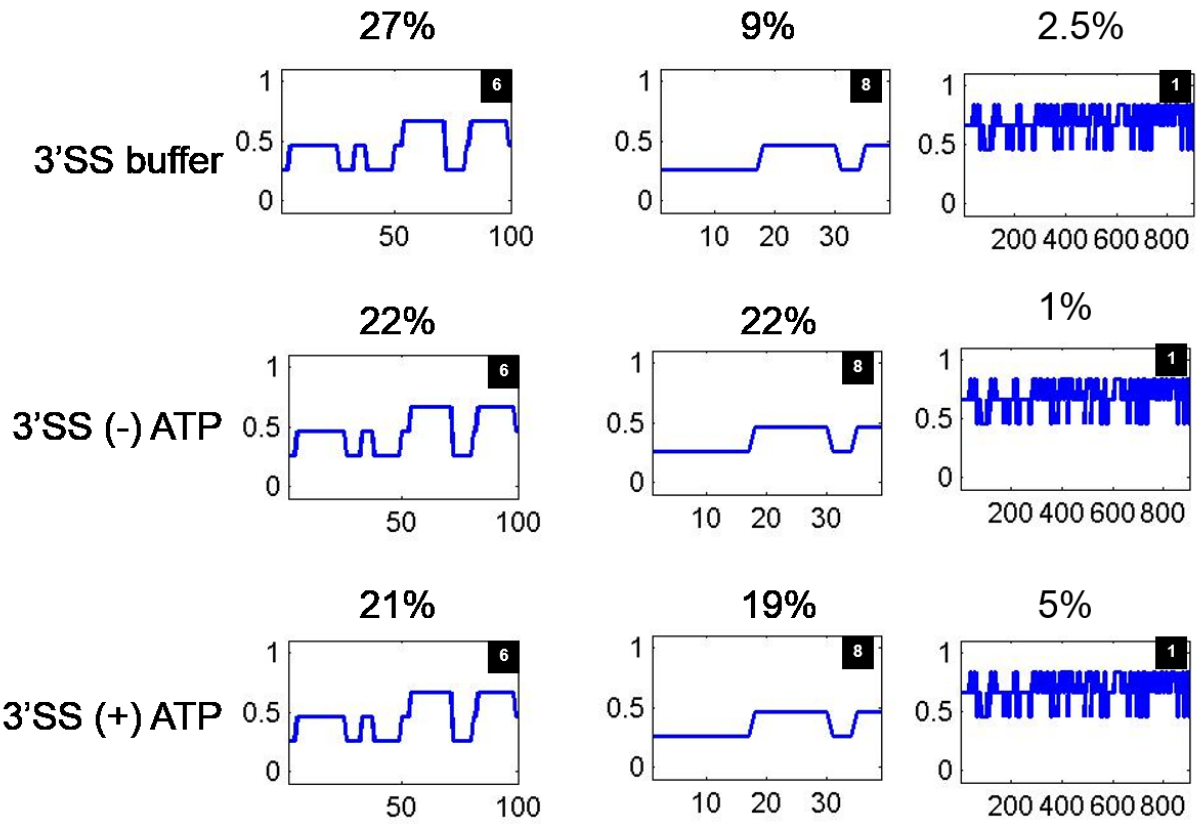


Figure 3.20: 3'SS consensus behaviors.

The clusters with the greatest fold change are presented for the 3'SS mutant.

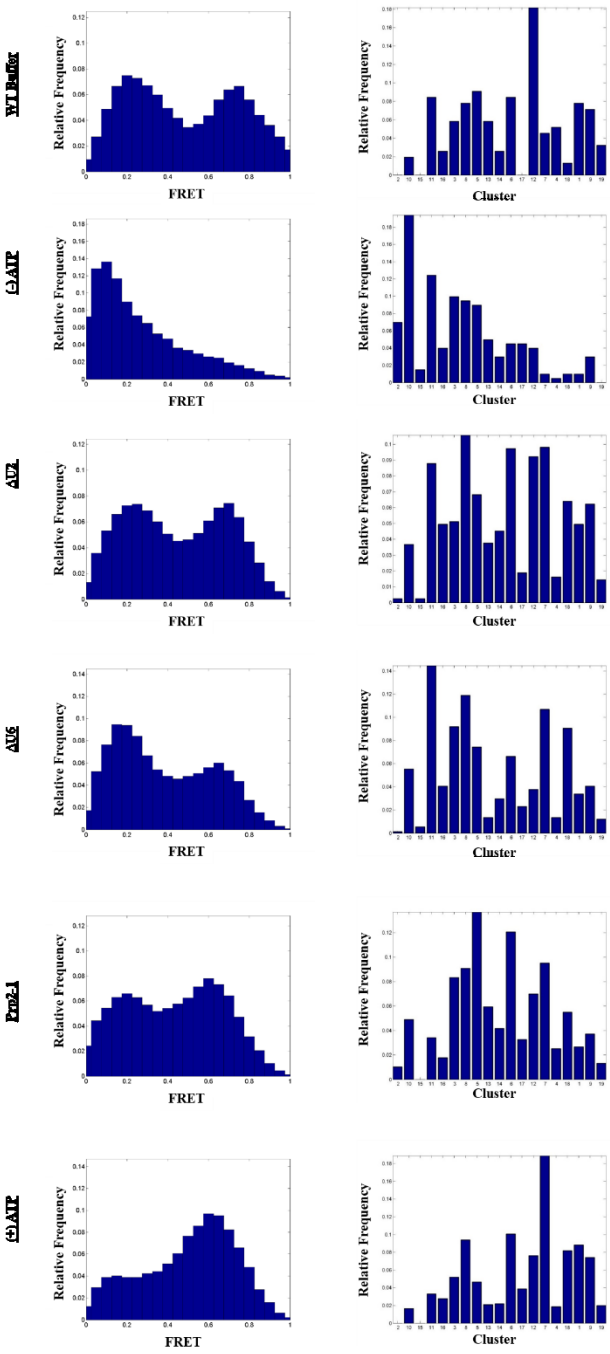


Figure 3.21: Comparison of WT substrate stalled at various points of assembly.

FRET histograms of the WT pre-mRNA. Abundance profiles of the 19 clusters identified from the hierarchical clustering. Clusters are reorganized based on mean FRET value within the cluster so the plot goes from lowest mean FRET (left) to highest mean FRET (right).

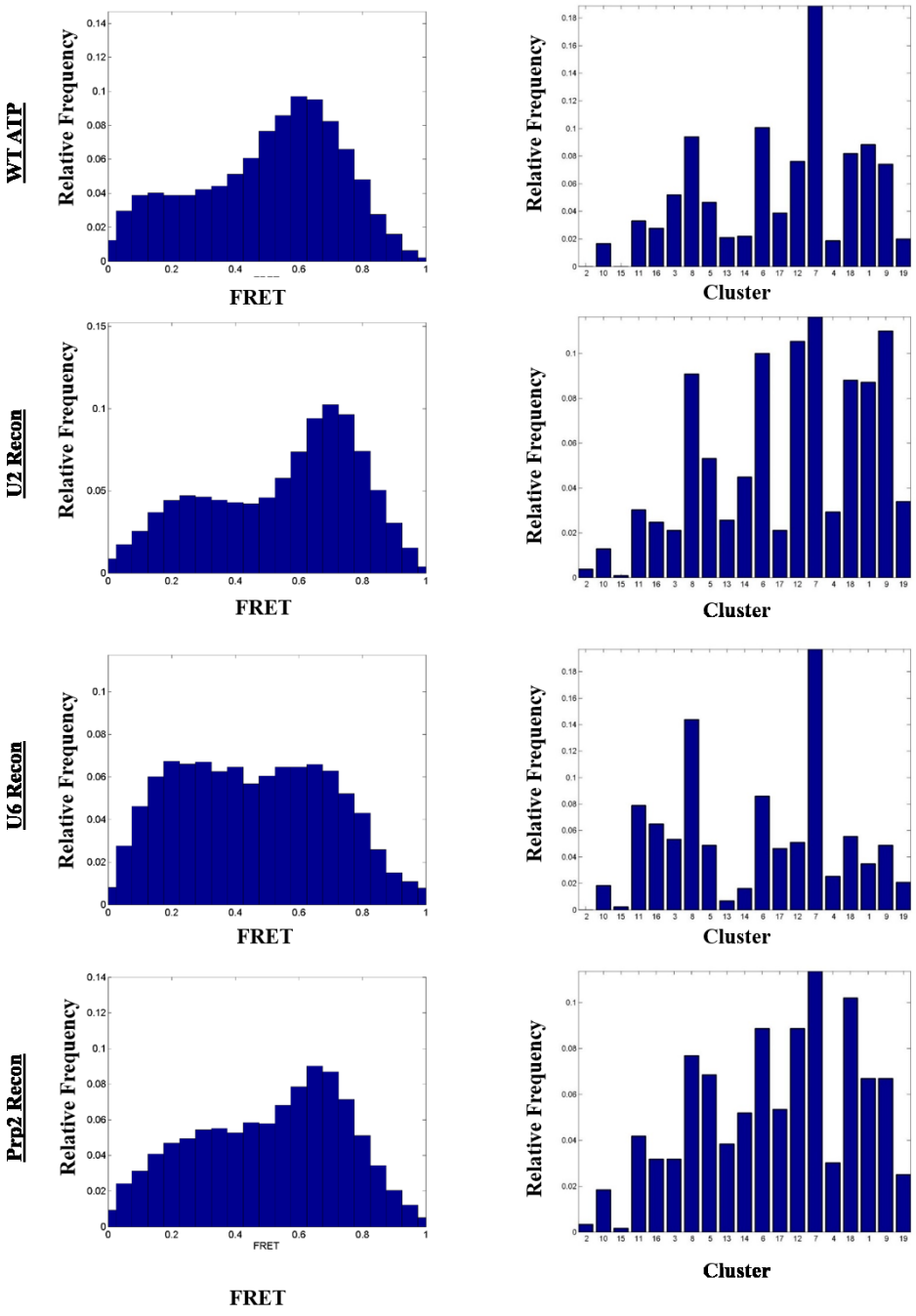


Figure 3.22: Comparison of splicing competent conditions.

FRET histograms of the WT pre-mRNA in (+) ATP, ΔU2 reconstituted, ΔU6 reconstituted and Prp2-1 reconstituted conditions are compared. Abundance profiles of the original 19 clusters identified from the hierarchical clustering. Clusters are reorganized based on mean FRET value within the cluster so the plot goes from lowest mean FRET (left) to highest mean FRET (right).

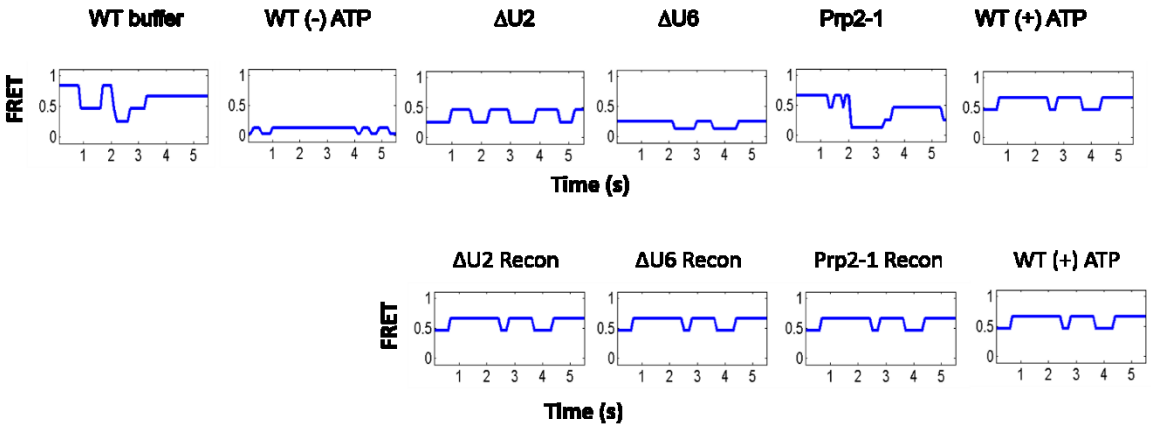


Figure 3.23: Consensus behaviors of splicing blocks.

The most abundant clusters for each WT substrate data set and extract modifications organized by their order on the splicing cycle. Cluster 7 is the most abundant cluster for the WT(+)ATP and reconstitution data sets.

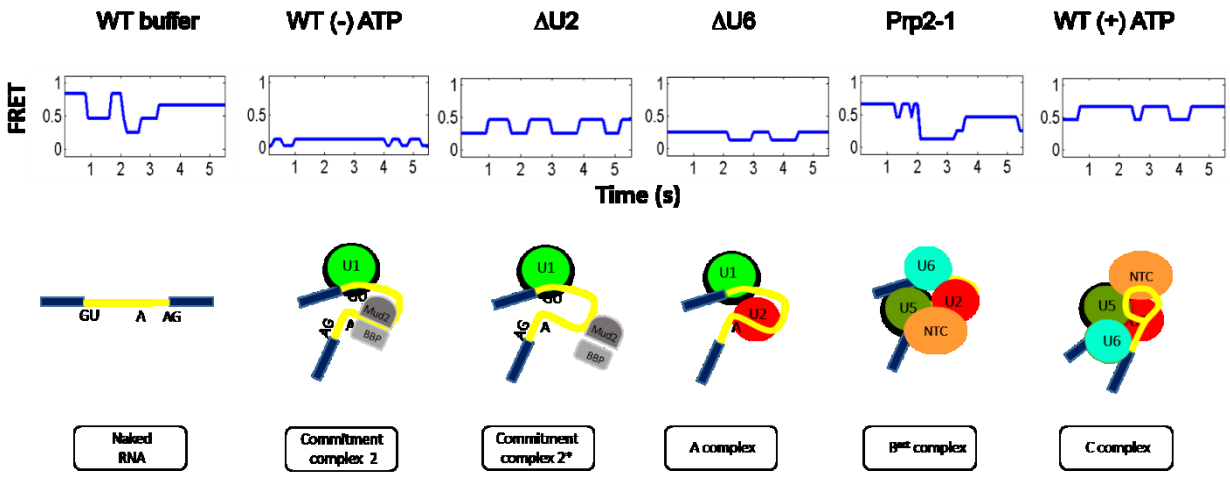


Figure 3. 24: Consensus behaviors aligned with splicing cycle.

The consensus behaviors of the most abundant cluster for each block is paired with the relevant complex enriched in the splicing cycle.

III.4 Discussion

In this study, we have collected single molecule data from pre-mRNAs labeled at the 5'SS and branchsite that can reach various steps of splicing assembly. To enrich for specific points along the splicing cycle we introduced substrate mutations and depleted specific splicing factors. Blocking specific points along the splicing pathway allows us to focus in on dynamics that are relevant during the early and late stages of assembly. The complex dynamic behavior of the molecules observed throughout the splicing cycle required the development of a novel approach to smFRET analysis. Hidden Markov Modeling has vastly improved the amount of information we are able to extract from smFRET data. It is a powerful tool for the unbiased extraction of FRET states and kinetics from single molecule trajectories. By utilizing the transition probability matrix that results from Hidden Markov Modeling, we created a FRET Similarity Matrix (FSM) that can be utilized in hierarchical clustering algorithms. Hierarchical clustering techniques allow us to group molecules with similar behaviors across conditions by applying the FSM as a measure of how similar two molecules are. The FSM can help distinguish molecules with similar FRET state levels and rates of inter-conversion.

By applying hierarchical clustering techniques, we identified consensus behaviors of molecules at various stages of splicing. First, we identified that the WT substrate as well as the mutant BP, 3'SS, 5'HYP substrates all undergo a shift to low FRET during ATP independent assembly events. The BP mutant, which is stalled during the earliest stages of assembly, can only bind the U1 snRNP in a stable fashion and adopts the lowest FRET conformation. The WT substrate can undergo all the necessary assembly steps to achieve both steps of catalysis. We controlled the assembly of the WT substrate by depleting the extract of ATP or ablating the activity of the U2 snRNP, U6snRNP, and Prp2 protein. Although the U2 snRNP ablation (ΔU2)

is expected to enrich the same complexes as the (-) ATP condition, we detected differences in the most abundant conformations present during the two blocks. The low FRET dynamics present in WT(-)ATP mirror those of the BP mutant, while the most abundant dynamics in the ΔU2 can now sample mid-FRET states from a low-FRET state, suggesting that the presence of ATP but absence of U2 has affects the assembly of components on the pre-mRNA. Genetic and biochemical evidence previously established a role for the ATP dependent helicases Sub2p and Prp5 near the site of U2 snRNP assembly.[27,28] It is possible that Sub2, which is located near the branchsite, is activated upon the addition of ATP to the extract but assembly fails to progress due to the absence of an assembled U2 snRNP. The ΔU6 block allows for the assembly of the U2 snRNP near the branchsite. This enriches for what is known as Commitment Complex 2 (CC2).[28] The U2 snRNP addition can promote cross-intron interaction between the branchsite and 5'SS through U1-U2 snRNP interactions.[87] Surprisingly, we detect the most common set of dynamics during the ΔU6 block as lower FRET than the previous ΔU2 block. Although the U2 snRNP may be bridging the intron, our fluorophores are generally farther away from each other. The ΔU2 block may allow more transient sampling of mid-FRET states due to the intrinsic pre-mRNA dynamics that are less constrained due to the lack of assembled U2 snRNP near the branchsite. The Prp2-1 condition allows the substrate to assemble up to the point just prior to catalytic activation.[25] At this stage, the most common pre-mRNA dynamics are molecules that sample low, mid, and high FRET states. Relieving this block and allowing splicing to proceed in the WT(+)ATP condition now makes the most common dynamics those sampling only mid-to-high FRET transitions with kinetics that favor the high FRET state. This suggests that the spliceosome components during the latter stages of splicing assembly are stabilizing conformations where the 5'SS and BS are in close proximity to promote catalysis.

A timeline of dynamics across the spliceosome assembly landscape has provided us with the ability to determine the effects the spliceosomal components on the juxtaposition of the 5'SS and branchsite. During early assembly steps, the two sites are restricted from interacting as evidence by the adoption of low FRET conformations. This restriction is partially relieved during late stages of assembly where high FRET states are sampled transiently. Allowing splicing to proceed to completion shows that fully active spliceosomes promote conformations with high-FRET states that are restricted from sampling low FRET states. Being able to dissect the complex set of dynamics using individual points of blockage was made possible by the use of a novel hierarchical clustering tool for smFRET data. This approach can be applied to complex data sets that are sure to become more common as we expand the use of single molecule tools to more complex macromolecular complexes.

CHAPTER IV:

Prp2 Mediated Conformational Change at the Spliceosome's Core Sets the Stage for Cwc25 Enhancement of First Step Catalysis[3]

IV.1 Introduction

The catalytic activation of the spliceosome requires a coordinated reorganization of the RNA-RNA and RNA-protein interactions present at its core.[88] These rearrangements are driven by a set of at least eight DExD/H box helicases. Prp2 has been identified as the DExD box helicase responsible for the structural transition that allows the spliceosome to catalyze the first step of splicing.[24] A temperature sensitive mutation in Prp2 (Prp2-1) can be utilized to heat inactivate this protein, a procedure that blocks catalysis but permits the assembly of an activated spliceosome (B^{act}).[30] The B^{act} complex contains the U2/U6/U5 snRNPs whose snRNA components come together to create a structure that resembles the self-splicing group II intron ribozymes.[89] Although the B^{act} complex contains all components required for chemistry, the absence of functional Prp2 leaves it in an inactive state. The addition of Prp2 with its protein cofactor Spp2 and ATP leads to a detectable change in the conformation of the B^{act} complex. This change is significant enough to

[3] Adapted from: Ramya Krishnan, Mario Blanco, Matthew Kahlscheuer, John Abelson, Christine Guthrie, Nils G. Walter. "Pre-mRNA conformational remodeling by Prp2 sets stage for Cwc25 mediated first step splicing." (in preparation) Mario Blanco helped optimize conditions for ensemble characterization of immunoprecipitated complex as well as performed clustering analysis of states, kinetic analysis of transitions, as well as helped write and create figures for manuscript. Ramya Krishnan preformed single molecule experiments and data analysis, created figures, and wrote paper. Matthew Kahlscheuer made splicing extract, expressed and labeled proteins.

cause a shift in the sedimentation coefficient of this complex from ~45S to ~40S; this new species is termed the B* complex.[30] Mass spectrometric analysis reveals that the protein composition of the B^{act} and B* complexes is nearly identical, signifying that the shift in sedimentation is caused by a conformational rearrangement and not a major loss of proteins. The B* complex allows for first step catalysis to occur, albeit at levels much lower than a fully functional spliceosome. It was found that an additional protein Cwc25 is required for complete reconstitution of first step catalysis.[30]

It is unclear what effect the Prp2 mediated conformational change has on the interactions near the sites of the first step of chemistry. The shift in sedimentation coefficient and changes in structure detected by electron microscopy[30] only provide information regarding the large-scale changes to the spliceosome during this activation step. The small extent of first step catalysis in the B* complex suggests that the catalytic core is rearranged in a manner that now allows still inefficient catalysis. Enhancement of first step catalysis by Cwc25 occurs through some undetermined mechanism. Cwc25 does not consume ATP and has no known role in spliceosome reorganization. Information regarding the state of the catalytic core during the B^{act} to B* transition would help elucidate the mechanism by which first step catalysis is controlled by Prp2/Spp2 and Cwc25. The dynamic properties of the pre-mRNA substrate we discovered while bound to the spliceosome[41] suggest that changes in spliceosome organization may be reflected by changes in the conformational dynamics of the pre-mRNA. The 5'splice site (5'SS) and branchsite (BS) on the pre-mRNA are the sites reacting in the first step of catalysis. At this point in spliceosome assembly (B^{act}), the 5'SS is interacting with the U6 snRNA and the BS with the U2 snRNA. In the current study, we have combined immunoprecipitation with single molecule fluorescence resonance energy transfer (smFRET) to track the proximity of these two sites

during the transition from B^{act} to B* and through the first step of catalysis. We have determined that the conformational change enacted by Prp2/Spp2 dramatically alters the conformation of the pre-mRNA, allowing the 5'SS and BS to transiently interact. We also present evidence for the mechanism of first step enhancement by Cwc25. Cwc25 alters the kinetics of the conformations present in the B* complex to favoring a state where the 5'SS and BS are in close proximity.

IV.2 Materials and Methods

Affinity purification of the B^{act} complex

Extracts were prepared from Prp2-1 Cef-1 TAP tagged yeast strain. The extract was heated at 37°C for 40 minutes to inactivate Prp2 protein and to stall the spliceosome at the B^{act} complex. 40% v/v of the heat inactivated extract was incubated with fluorescently labeled pre-mRNA (~50 pmoles) in the presence of 2mM ATP in splicing buffer (8 mM HEPES-KOH, pH 7.0, 2 mM $MgCl_2$, 0.08 mM EDTA, 60 mM $K_i(PO_4)$, 20 mM KCl, 8% (v/v) glycerol, 3% (w/v) PEG, 0.5 mM DTT) at 23°C for 35 minutes. For ensemble experiments, streptavidin coated magnetic beads (Dynabeads® MyOne™ Streptavidin C1) from Invitrogen were treated as per manufacturer's recommendation. For each splicing reaction of 135 µl, 200 µl of beads were used. The beads were equilibrated in 200 µl of T50 buffer (50 mM Tris-HCl, pH 7.5, 50 mM NaCl). Equal volume of 0.5 mg/ml IgG-biotin in T50 (ZyMAX™ Rabbit Anti-Mouse IgG (H+L) - BT (ZyMAX™ Grade) was added to the beads and incubated in a tube rotator at room temperature for 30 minutes. The beads were then pulled down using a magnet and the supernatant was discarded. In order to block any streptavidin not bound by biotin-IgG, the reaction was incubated with excess free biotin at 1.5mg/ml concentration (in T50 buffer) in a tube rotator at room temperature for 20 min. The beads were the pulled down with a magnet and washed in splicing

buffer. The assembled splicing reaction was then added to the beads and incubated in a tube rotator for 30 minutes at room temperature to allow the protein A group of the TAP tag in the spliceosome complex to bind the biotin-IgG molecules. The reaction was pulled down using a magnet and the unbound fraction is collected. The bound fraction was further washed 3X with buffer A (20 mM HEPES-KOH pH 7.9, 75 mM KCl, 0.01% NP40, 1.5 mM MgCl2, 5% glycerol) and 1X with splicing buffer to obtain the B^{act} complex. The reactions are scaled up for various combinations of reconstitution and split at this step. Prp2, Spp2 and Cwc25 were added (at 90 nM – 120 nM final concentration) in splicing buffer in the presence or absence of 2 mM ATP/AMPPNP/UTP and incubated in the tube rotator for 30-40 minutes for the reconstitution experiments. RNA was isolated and products of splicing were analyzed on a 15% 7 M urea-PAGE denaturing gel and scanned on a Typhoon variable mode imager.

Cloning, expression and purification of proteins

The full-length PRP2 gene was PCR-amplified from genomic DNA of *S. cerevisiae* and ligated into plasmid pRSETA (Invitrogen), thus fusing the protein to a C-terminal hexa-histidine tag. The N-terminally truncated form of SPP2 (coding for amino acids 37-185) containing a C-terminal hexa-histidine tag, and the full-length CWC25 gene were obtained from Reinhard Luhrmann (Max-Planck-Gesellschaft). CWC25 was PCR-amplified and ligated into pRSETa vector (Invitrogen). A short DNA sequence encoding a TEV cleavage site (GGGC), and a hexa-histidine tag was then inserted downstream of the gene, creating a C-terminally hexa-histidine tagged fusion with a single cysteine residue. The constructs were transformed into Rossetta II *E. coli* strain (Novagen). Cultures were grown in 2-4 liters of TB medium and induced with 125 uM IPTG. Cultures were then incubated at 20 ºC for 18h. Cells were harvested by centrifugation at 5,000 RPM for 30 min, washed, and the pellets stored at -80ºC.

Purification of Cwc25-His and Spp2-His was performed as described[30] with the following modifications: supernatants were applied to Ni^{+2} resin equilibrated in lysis buffer (50 mM HEPES-NaOH pH7.5, 600 mM NaCl, 2 mM B-mercaptoethanol (B-ME), 20 mM imidazole, and 10% (v/v) glycerol) and purified by gravity affinity chromatography. Nonspecifically bound proteins were removed by washing with 10 column volumes (CV) of lysis buffer followed by two CV of washing buffer (20 mM HEPES-NaOH pH 7.5, 2 M LiCl, 5% glycerol) followed by two CV of lysis buffer. Bound protein was then eluted with 10 CV of elution buffer (50 mM HEPES-NaOH pH7.5, 600 mM NaCl, 2 mM B-ME, 250 mM imidazole, and 10% (v/v) glycerol) and immediately dialyzed into storage buffer (20 mM HEPES-NaOH pH7.5, 200 mM NaCl, 2 mM DTT, 5% (v/v) glycerol). Protein purity was confirmed by 16% SDS-PAGE and proteins were then either used for fluorescent labeling or subsequently aliquoted, flash frozen in liquid nitrogen, and stored at -80C. Protein concentrations were determined by Bradford assay analysis and measurement at A280.

Prp2-His was purified as described[30] with the following modifications. Prp2-His cell pellets were resuspended by vortexing in 5 mL of buffer BC (20 mM Tris-HCl pH 7.9, 250 mM KCl, 10 mM imidazole 10% glycerol) per 1g of cells. Cell lysis was performed by sonication and insoluble material was removed by centrifugation for 40 minutes at 10,000rpm. Supernatants were loaded onto Ni^{+2} resin equilibrated in BC buffer. Columns were washed with BC containing 500 mM KCl (BC_{500}) followed by washing with buffer BC containing 100 mM KCl (BC_{100}) and 20 mM imidazole. Prp2 was eluted using buffer BC_{100} containing 150 mM imidazole. Eluted Prp2 was then loaded onto a polyuridylic acid-agarose column (Sigma) equilibrated with buffer BP_{50} (20 mM HEPES pH 7.9, 0.5 mM EDTA, 0.01% NP-40, 10% glycerol, 50 mM KCl). Columns were washed with BP_{50} and Prp2 was eluted using buffer BP containing 250 mM KCl. Protein purity

was confirmed by 10% SDS-PAGE and subsequently aliquoted, flash frozen in liquid nitrogen, and stored at -80C. Protein concentrations were determined by Bradford assay analysis and measurement at A280.

Fluorescent labeling of Cwc25 and distance estimation from FRET

Single-cysteine mutant Cwc25 was labeled with Cy5-maleimide (GE Healthcare). Labeling was performed using 0.150 umoles of Cwc25 in storage buffer and 0.5mg of dye containing 10 uM reducing agent tris (2-carboxyethyl) phophine (TCEP) (Sigma). Reactions were incubated at room temperature for 1h and then overnight at 4C. Free dye was removed by re-purification of protein on a Ni+2 column and dialysis back into storage buffer. The degree of labeling was determined using GE Healthcare's protocol and was found to be 70% of labeled Cwc25. Protein functionality was confirmed using an ensemble pull down assay as described above.

The fluorophore distance, R, was calculated from $Eapp = c[1 + \left(\frac{R}{R0}\right)^6]^{-1}$ where c=0.69 and R0=54A° as described[1]. The apparent FRET efficiency, Eapp was estimated from the equation $Eapp = \frac{ICy5}{ICy5+ICy3\times\frac{(\emptyset Cy5 \times \eta Cy5)}{(\emptyset Cy3 \times \eta Cy3)}}$[2], where φ and η signify the fluorophores quantum yields and detector channel efficiencies respectively. The donor and acceptor intensities ICy3 and ICy5, respectively, were corrected for leakage of donor photons into the acceptor channel.

smFRET on purified spliceosome

For the single molecule FRET experiments on affinity purified complex, we prepared slides using previously published procedures.[41] The surface of the slide was amino functionalized, PEGylated and reacted with 0.2 mg ml^{-1} streptavidin in buffer (50 mM Tris-HCl, pH 7.5, 50

mM NaCl) for 15 minutes at room temperature. 100µl of 0.5 mg/ml IgG-biotin in T50 (ZyMAX™ Rabbit Anti-Mouse IgG (H+L) - BT (ZyMAX™ Grade) was flowed onto the slide and incubated for 20 minutes followed by free biotin at 1.5 mg/ml concentration (in T50 buffer) for 15 minutes. Assembled and stalled B^{act} spliceosome complex (along with oxygen scavenger system) were then flowed onto the slide surface and incubated for 15-20 minutes in order to allow the TAP tag on the spliceosome NTC (Cef-1) to bind the biotin-IgG. The slide was further washed 3X with buffer A (20 mM HEPES-KOH pH 7.9, 75 mM KCl, 0.01% NP40, 1.5 mM MgCl2, 5% glycerol) and 1X with splicing buffer to remove non-specifically bound particles. For the various reconstitutions, Prp2/Spp2/Cwc25 (80-100 nM), ATP/AMPPNP were added in various combinations in splicing buffer and oxygen scavenger system and incubated 10-40 minutes before obtaining data. A prism-based TIRF microscope was used to collect data as described.[41] In order to obtain FRET data, we directly excited the Cy3 donor located near the branchsite of Ubc4 pre-mRNA with a 532-nm laser, and we recorded emission from Cy3 and Cy5 fluorophores at 100-ms time resolution using an intensified CCD camera (Princeton Instruments, I–Pentamax). The raw trajectories obtained from the experiment were visually pre-filtered for the presence of anticorrelation. We then screened transition events found within the pre-filtered FRET index trajectories for localized anticorrelation in the donor and acceptor traces. The post-filtering system required the use of the HMM tools vbFRET on the pre-filtered data.

Single Molecule Data Analysis

Cross-correlation analysis was carried out utilizing MATLAB scripts that rely on the built-in xcorr function. Time lags for determining the cross-correlation value ranged from 0-50. Molecules with a positive cross-correlation value greater than 0.15 at time lag 0 were filtered out from any further analysis (**Figure IV.9**). Histograms for data sets were constructed by sampling

the first 100 frames of data from each molecule. Hidden Markov Modeling (HMM) was done on trajectories retained after cross-correlation filtering utilizing the vbFRET software suite. Each trajectory was individually fit to models ranging from 1-5 states. The number of states fit to each trajectory was determined by the vbFRET algorithm. The inherent experimental fluctuations on the FRET signal from single molecule experiments leads to a slightly different state assignment for similar states across different molecules. A K-means clustering approach performed in MATLAB was utilized to group similar states into larger macro states (L1, L2, M and H). A matrix cataloging the HMM assigned FRET state, raw FRET level, and difference in Donor/Acceptor intensities for each HMM assigned event was utilized as input for the K-means algorithm (**Figure IV.10**). A total of 4 macro states were identified whose boundaries were used to re-assign the original HMM idealized FRET states. Transition Occupancy Density Plots (TODPs) were used to plot the fraction of molecules that contain any given HMM transition at least once. Molecules that do not contain any transitions were centered along the diagonal at the position of the state in which they reside. Transition Density Plots (TDPs) that they are scaled by the number of times a transition occurs irrespective of how many molecules exhibit that transition were used to extract dwell time and kinetic information as described[3]. A cumulative histogram scatter plot was then fit with a double exponential association equation in MicroCal Origin. (**Figure IV.11**) A weighted average of the two rate constants from the double-exponential fits was calculated based on the amplitude values A1 and A2 of the exponential equation. Post-synchronized histograms (PSHs) were constructed by synchronizing individual FRET events to the time where one of the macro-states (M, H) was achieved. The scale bar represents the fraction of FRET events which exhibit a certain FRET state at a given time.

IV.3 Results

Immunopurification of fluorescently labeled Ubc4 pre-mRNA in functional B^{act} complex

The inability to functionally reconstitute spliceosome activity *in vitro* from recombinantly purified components alone has necessitated the use of crude yeast extracts to investigate pre-mRNA splicing. Recently, the reconstitution of both steps of splicing has been achieved through the purification of an assembled pre-catalytic spliceosome.[30] This was accomplished through the use of a yeast strain containing a heat-sensitive mutation in the first step factor Prp2. The Prp2-1 allele allows for the assembly of an activated spliceosome, B^{act}, when the extract is heated to 37 ºC. To specifically purify this complex, a series of glycerol gradient purification steps were utilized. The B^{act} complex could then be chased through the first step of catalysis through the addition of three recombinant factors (Prp2/Spp2/Cwc25) and ATP. We sought an approach that allowed us to visualize this complex and its transition through the first step of catalysis at the single molecule level. We created a yeast strain carrying the Prp2-1 allele as well as a Tandem Affinity Purification (TAP) tag on Cef1, an NTC component that is present only in late stages of spliceosome assembly and that mass spectrometry[30] has verified is present in the B^{act} complex. The Prp2-1 allele allowed us to heat inactivate the extract, blocking assembly at the stage of Prp2 action while the Cef1-TAP serves an affinity tag for the enrichment of B^{act} without the need for glycerol gradient purification. Spliceosomes were allowed to assemble on a fluorescently labeled Ubc4 pre-mRNA in heat-inactivated Prp2-1 Cef1-TAP extract supplemented with 2mM ATP. We then isolated and immobilized this complex by incubating the entire reaction with streptavidin coated beads that had been pre-incubated with a biotinylated-rabbit anti-mouse (H+L) antibody (biotin-IgG). To verify that this protocol purified the B^{act} complex we proceeded

with the ensemble characterization of its activity. Our Prp2-1 Cef1-TAP immuno-purified B^{act} complex only

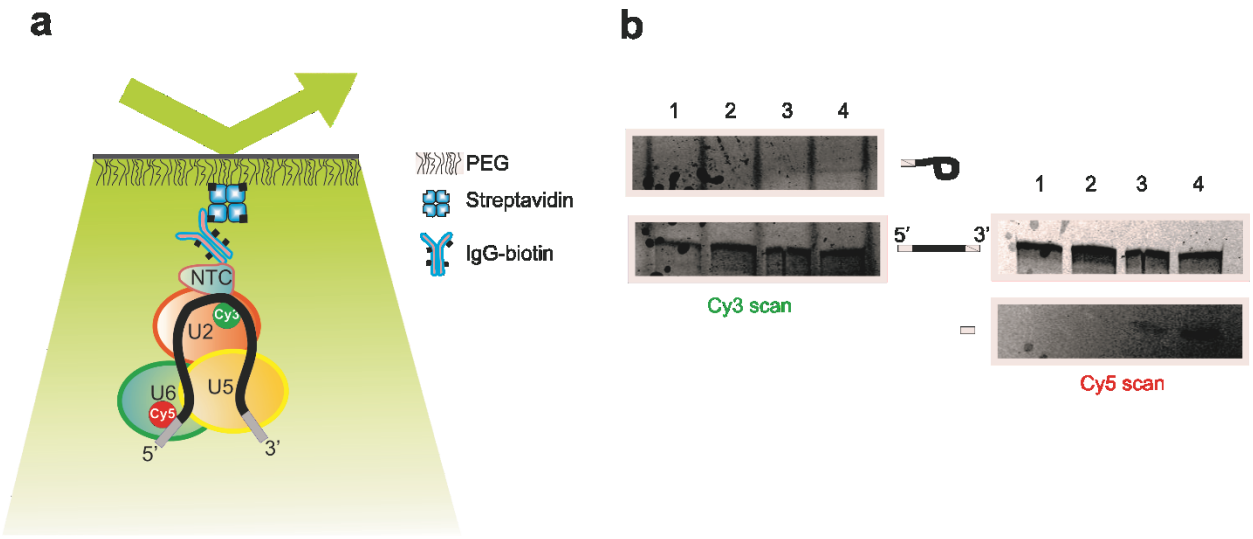


Figure 4. 1: Schematic of SiMPull set-up and verification of activity.

(a) Schematic showing the affinity purified B^{act} complex immobilized to a streptavidin coated quartz slide via biotinylated-IgG. The green and red circles on the pre-mRNA represent Cy3 and Cy5 flurophores respectively. **(b)** The affinity purified B^{act} complex is active and can be chased to 1st step of chemistry with the addition of Prp2, Spp2 and Cwc25. Urea-polyacrylamide gel scanned using a variable mode Typhoon imager showing pre-mRNA and 1st step products (left-Cy3 scan, right-Cy5 scan) in the following conditions- affinity purified B^{act} complex (lane 1) with Prp2 (lane 2), Prp2 and Spp2 (lane 3) and Prp2, Spp2, Cwc25 (lane 4) added to B^{act} complex with 2mM ATP. Splicing was assayed at room temperature as described (**Methods**). B^{act} was incubated with each of the proteins for 30-40 minutes for the reconstitution.

contained pre-mRNA indicating we successfully inactivated Prp2 (**Fig IV.1b, lane 1**). Furthermore, just as Warkocki *et al* had previously described the addition of recombinant proteins Prp2 and Spp2 and ATP resulted in a low level of first step splicing products (**Figure IV.1b, lane 3**). Efficient reconstitution only occurred when Cwc25 was added in addition to Prp2/Spp2 (**Figure IV.1b, lane 4**). We observed a 2-10 fold enhancement of first step products with the addition of Cwc25. In addition to validating the activity of the purified complex, this result allowed us to establish that the introduction of fluorescent tags on pre-mRNA did not interfere with the intricate assembly pathway.

Once we confirmed that the ensemble characteristics of our immuno-purified complex corresponded with those of the previously described B^{act} complex, we devised a strategy for single molecule observation. We assembled the B^{act} complex in the same manner as in the ensemble experiments, but rather than utilizing streptavidin coated magnetic beads for immobilization we used a streptavidin coated biotin-IgG conjugated slide suitable for microscopy (**Figure IV.1a**). The publication of a similar technique, single molecule pull-down (SiMPull), appeared while these experiments were ongoing[90]; for simplicity and to indicate our expansion toward the use of FRET, we call our technique SiMPull-FRET. The Ubc4 pre-mRNA assembled was chemically synthesized and labeled (see **Methods**) with a Cy3 donor fluorophore +6 nucleotides downstream from the BS adenosine and Cy5 acceptor fluorophore, -7 nucleotides upstream from the 5'SS (**Table IV.1**). These sites were chosen because these two sequences are involved during the first step of catalysis, thus allowing us to track any structural dynamics that occur during the transition from the B^{act} complex to the first step of splicing. The assembly of the spliceosome on this substrate has been previously characterized through the immobilization of the substrate via a biotin present at its 5' end (**Chapter III**). To prevent binding of the pre-

mRNA to the slide via the 5'-biotin-streptavidin interaction, free biotin in excess was added to the slide to block any streptavidin molecules not bound by biotin-IgG. We see very little binding of the pre-mRNA by itself to the surface saturated with excess free biotin when compared to the IgG-bio-TAP-Cef1 affinity purified B^{act} complex with the pre-mRNA pre-assembled (**Figure IV.12**). Additionally, we detected minimal non-specific binding of the purified complex in the absence of IgG-biotin treatment of the surface. A TAP-tag on a different protein, Prp4, which has been recently shown to be absent from the B^{act} complex[91] did not pull down functional B^{act} complex on the slide surface, as expected.

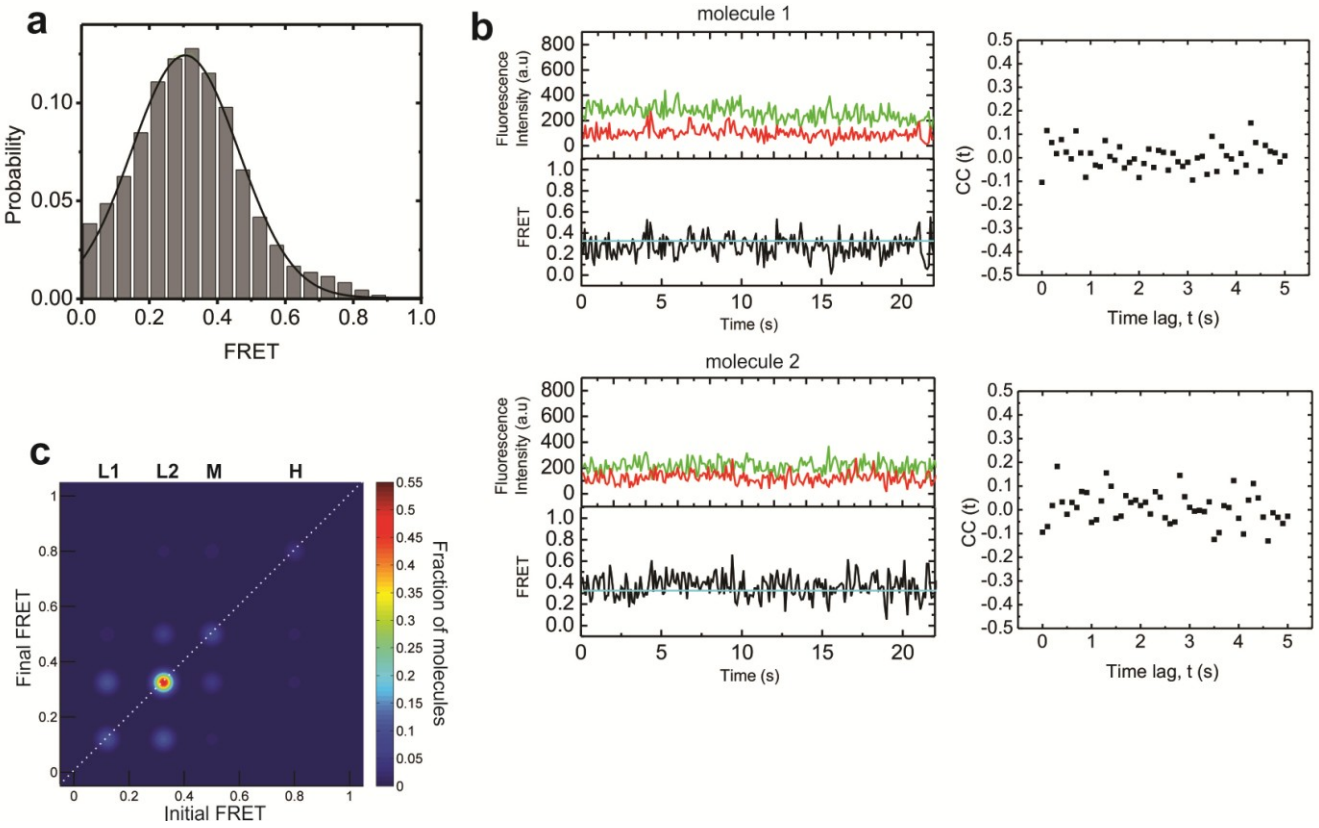


Figure 4. 2: Pre-mRNA in the stalled B^{act} complex is predominantly restricted to low L2 FRET states.

(a) FRET efficiency histograms generated by binning the first 10s of the raw smFRET profiles of the B^{act} complex. **(b)** Sample traces from the B^{act} complex with raw donor (Cy3, green), acceptor (Cy5, red), and FRET (black) trajectories and their idealized HMM models (cyan) in the stable L2 FRET state. The right panel of each molecule is the respective cross-correlation analysis of donor and acceptor intensities showing that the stable FRET state under B^{act} condition is a true feature of a system and not due to limited resolution**.** **(c)** Transition Occupancy Density Plots (TODP) showing the same trend for more than 50% of molecules. The dotted white line across the diagonal populates stable FRET states.

B^{act} complex holds the 5'SS and BS in an expanded conformational state

We assembled the B^{act} under single molecule conditions and collected data with it in the presence of ATP but in the absence of Prp2/Spp2. FRET efficiency histograms plotted by using the FRET value collected over the first 100 frames of 297 molecules indicated an average FRET efficiency centered at 0.3±0.15 for the population (**Figure IV.2**). Hidden Markov Modeling (HMM) was utilized to determine the underlying FRET states in the shot-noise dominated fluorescence signal. Due to experimental variability, the same FRET state across different molecules and experiments may not agree completely. To determine the true underlying FRET states and facilitate comparison across molecules, we utilized a K-means approach to cluster the HMM assigned states (see **Methods**) into four macro states: L1, L2, M, H encompassing 0.00 - 0.23, 0.23-0.42, 0.42-0.60 and 0.60-1.00, respectively (**Figure IV.10)**. The most common FRET behavior in the B^{act} complex is a stable low FRET state centered at the L2 (0.30) state (**Figure IV.2b**). Molecules in this state had very few transitions as detected by the HMM algorithm (**Figure IV.2b**, HMM fit, blue line). Small changes in FRET with dynamics that are approaching the integration time of the camera (100 ms for our experiments) are often undetected by HMM algorithms. To test if the stable L2 state in the B^{act} for any underlying dynamics we performed cross-correlation analysis using the donor and acceptor trajectories for each molecule. Any significant anti-correlation between the two channels would result in a cross-correlation plot with an asymptotic regression towards zero. Molecules with no significant dynamics and thus a stable L2 state will have cross-correlation plots centered at zero. Transition Occupancy Density Plots (TODPs) that are scaled to report on the transition found to be most common in a population, indicate that this stable L2 state represents the behavior of more than 50% of the B^{act} molecules (**Figure IV.2c**). Less than 11% of the molecules show excursions to a higher FRET state (H)

centered at 0.80. This could be reflective of the small population of molecules going through splicing related conformational changes from residual Prp2/Spp2 not completely washed off that result in activity not detectable at the ensemble level. Our data indicate that the pre-mRNA in the B^{act} complex has adopted an extended conformation where the 5'SS and BS are relatively far apart, making unlikely that catalysis is accessible from this state.

Prp2 mediates large amplitude NTP-dependent conformational remodeling of pre-mRNA

The DEAD box helicase Prp2 acts immediately upstream of the first step of pre-mRNA splicing and plays an essential role in catalyzing first step of splicing.[24] Spliceosomal binding of Prp2 is dependent on its interaction with the G patch domain of its cofactor protein Spp2.[92,93] The addition of Prp2 and Spp2 along with ATP transforms the pre-catalytic B^{act} 45S complex to the catalytically active slower migrating 40S B* complex resulting in low levels of first step splicing.[30] To investigate the role Prp2 mediated remodeling has on the conformation of the pre-mRNA we incubated the assembled B^{act} complex (treated as mentioned in **Methods**) with Prp2, Spp2 and 2 mM ATP (henceforth referred to as the B* complex/condition). This resulted in a shift of 45% of the molecules from a mean FRET value of 0.3±0.15 of the B^{act} complex to 0.7±0.11 as observed in the FRET efficiency histogram (**Figure IV.3a**). The predominantly static L2 population in the

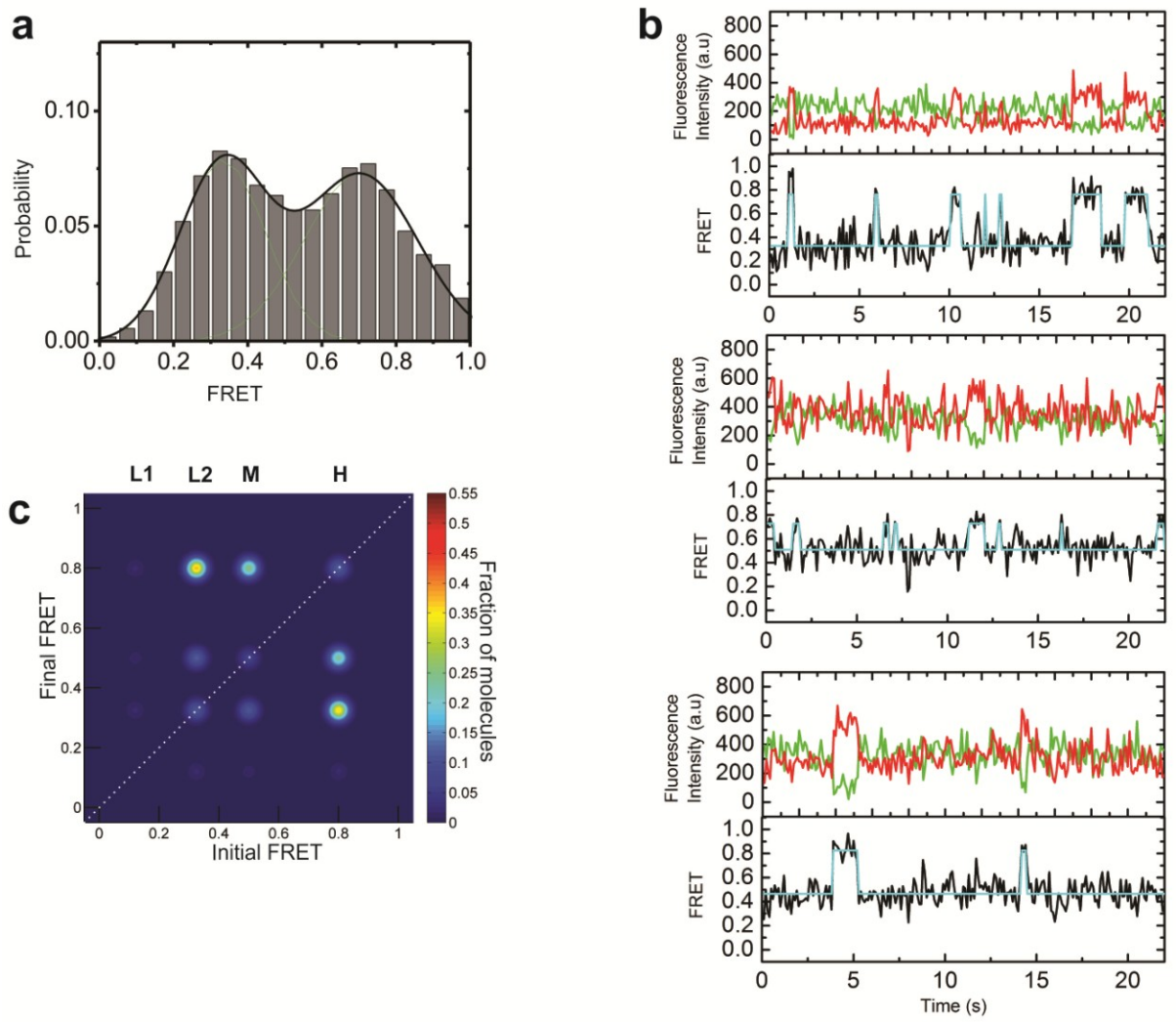


Figure 4. 3: Pre-mRNA is able to explore splice site proximity upon the addition of Prp2, Spp2 and ATP

(a) FRET efficiency histograms show higher FRET state sampling upon the addition of Prp2/Spp2 and ATP to the B^{act} complex. **(b)** Representative trajectories from this condition show molecules which were in the stable L2 state prior to the addition of Prp2, now transitioning transiently to the H state from either the L2 state (top panel) or the M state (2^{nd} and 3^{rd} panel). **(c)** TODPs show that 11% of the molecules retain B^{act} like behavior (stable L2 state), while a significant population explores higher FRET transitions. Approximately 12% of the molecules populate in the stable H state along the diagonal.

B^{act} complex now transits to conformations placing the 5'SS and BS in close proximity. The B* condition comprised three predominant FRET states - L2, M and H. We found that the H state could was predominantly sampled from either the L2 or M states (**Figure IV.3b**). TODP plots show that only 11% of the molecules retain the stable L2 state characteristic of the B^{act} complex while >35% of the molecules exhibit a L2->H transition (**Figure IV.3c**). Although the B* complex pre-mRNA can now sample high FRET transitions (H state) we find that these transitions are short lived in a majority of the molecules (**Figure IV.3b**). However, we also observed that 12% of the molecules now show a stable H state (**Figure IV.3c**), a fraction consistent with the fraction spliced under these conditions, supporting the notion that the stable H state is associated with first step splicing chemistry.

Both ATP dependent and independent roles have been proposed for many of the spliceosome DExD/H box helicases. Prp2 has been shown to bind the pre-mRNA directly even in the absence of ATP.[94] An ATP independent role of Prp2 has been shown to cause a significant conformational remodeling of the spliceosome as seen by the shift in sedimentation, while an ATP dependent role has been assigned to SF3b displacement.[91] In order to categorize Prp2-mediated pre-mRNA conformational changes as an ATP-dependent/independent activity, a non-hydrolyzable ATP analog, AMPPNP, was used instead of ATP in our B* complex conditions. The resulting FRET efficiency histogram replicated the overall behavior of the molecules seen in the B^{act} complex. This observation suggests that ATP hydrolysis is required to bring about the B* conformation where the BS and 5'SS are transiently interacting (**Figure IV.4**, compare panel 1 and 4).

Mass spectrometry studies have shown that the stalled B^{act} complex contains stoichiometric amounts of one other helicase, Brr2, a strict ATPase. By contrast, Prp2 utilizes all NTPs[95]. To

distinguish whether Brr2 or Prp2 are involved in substrate remodeling, we incubated the B^{act} complex in B* conditions with UTP instead of ATP and found this modification to result in a shift from the B^{act} stable L2 state into a similar FRET distribution to that seen in the presence of ATP (**Figure IV.4**, compare panel 5 with panel 6 and panel 1). This observation argues against a role for Brr2 in this conformational change and instead points to Prp2 with its broad NTPase specificity. There was an observable overall shift towards the lower FRET state with UTP addition when compared to ATP addition. This can be attributed to the slightly reduced efficiency of Prp2 (30-80%) in the presence of NTPs other than ATP[4][4].[95] The B^{act} complex was not changed when incubated with just Prp2 or Spp2 and ATP (**Figure IV.4**, compare panels 2, 3 and 4 with 1). Collectively, these results indicate that the NTPase activity of Prp2, along with its cofactor Spp2 causes a major structural reorganization at the core of the spliceosome bringing the pre-mRNA in close proximity in order to facilitate the first step of splicing. The reversible excursions to the H state may result in the small amounts of first step products detected in the ensemble assays.

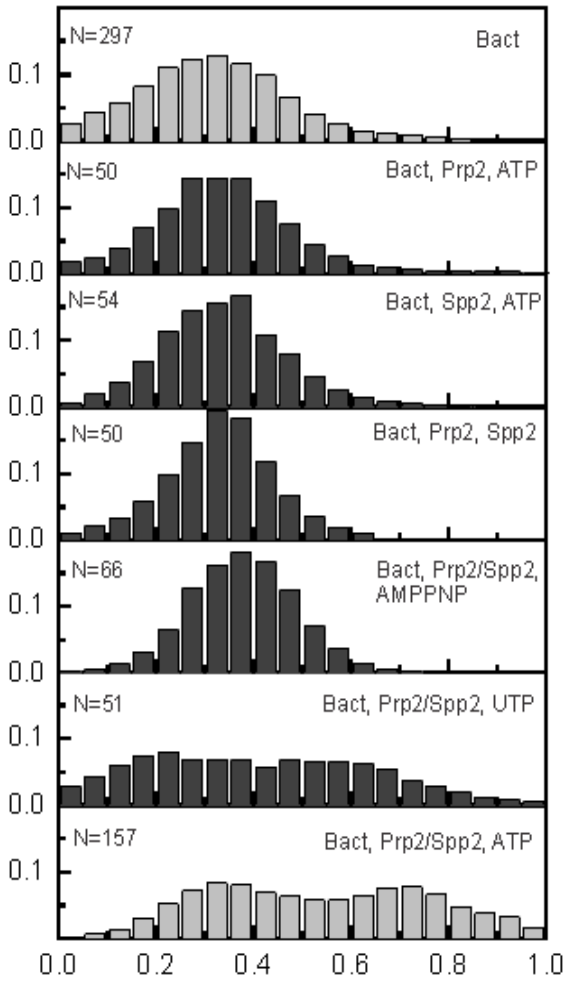


Figure 4. 4: NTPase activity of Prp2 brings about structural remodeling the pre-mRNA.

FRET probability histograms of different conditions stacked for comparison.

5'SS and BS dynamics in B^{act} complex incubated with Prp2, ATP (panel 2), with Spp2, ATP (panel 3), with Prp2, Spp2 (panel 4) and with Prp2, Spp2 and a non-hydrolysable AMPPNP (panel 5) overlap well with that of B^{act} under splicing conditions (panel 1). While, the addition of UTP (panel 6), shifts the histogram towards a bimodal distribution as seen when B^{act} complex chased to B* with the addition of ATP (panel 7).

Cwc25 enhances first step of splicing by increasing the residence time in the H state

Although the pre-mRNA is remodeled to a significant extent by the NTPase action of Prp2, it does not result in efficient first step catalysis. Addition of a protein, Cwc25, to the B* complex is necessary to enhance first step catalysis and reach the C complex state (**Figure IV.1b**). Little is known about the substrate or the mechanism of first step catalytic activation by Cwc25. We performed smFRET experiments on the purified B* complex (B^{act} with Prp2, Spp2 and ATP) supplemented with Cwc25. This resulted in a shift of 73% of the molecules to the high FRET distribution as observed in the FRET efficiency histograms (**Figure IV.5a**). Although the FRET states in the C complex were identical to those in the B* complex, the occupancy in the H state was significantly enhanced under C complex conditions (**Figure IV.5b**). TODP analysis estimated ~3-fold increase in the fraction of stable H state (**Figure IV.5c**, diagonal line) while the percent of molecules in the B* and C complex conditions transitioning from the L2->H and M->H states

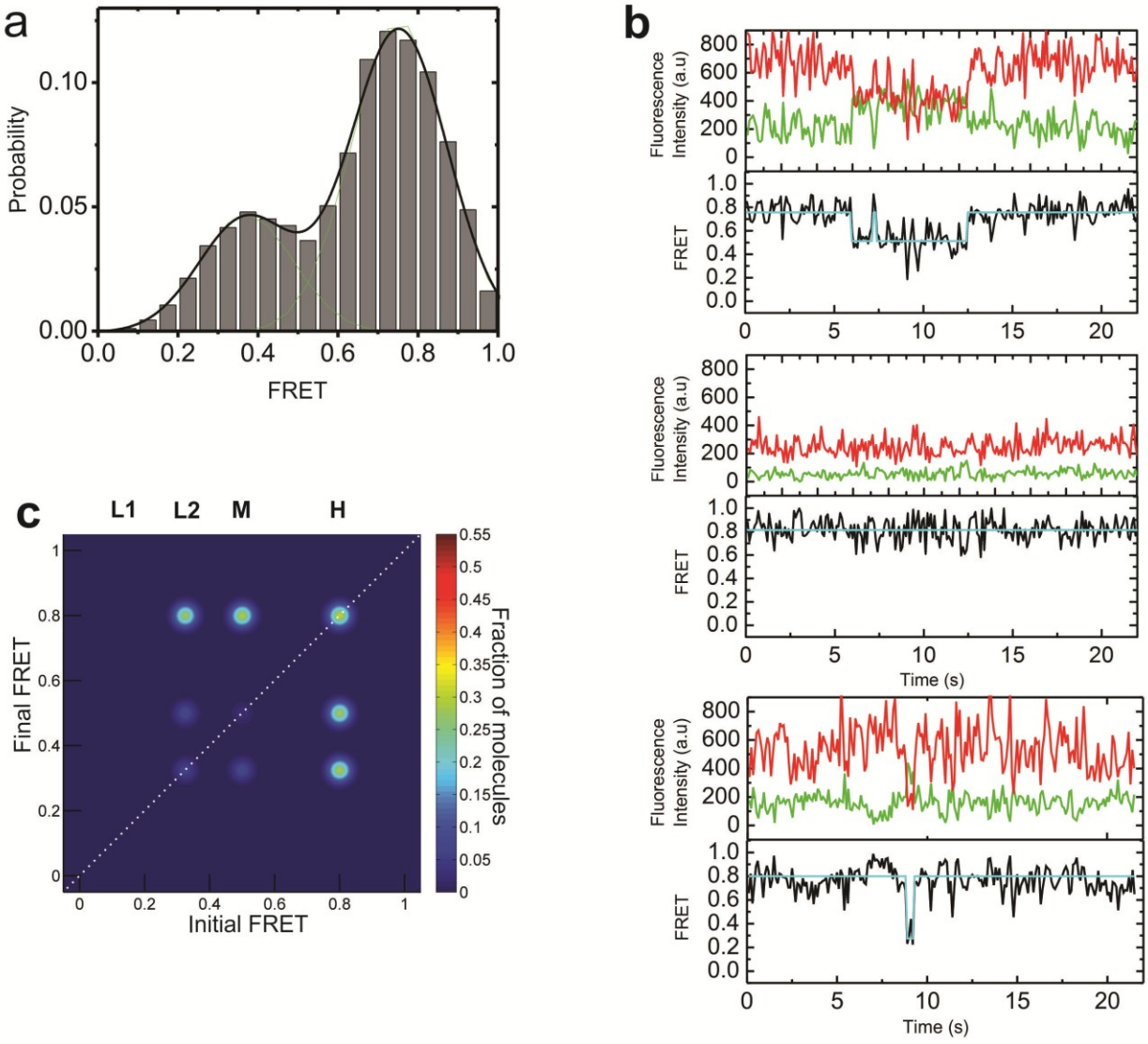


Figure 4. 5: Pre-mRNA explores long lived, stable high FRET states under C complex condition.

(a) FRET probability histograms plotted for the molecules under C complex conditions with the addition of Cwc25 resulted in a shift of 73% of the molecules to the higher FRET occupancy. **(b)** Representative trajectories from this condition with either M to H transition (top panel), a stable H state (mid panel) or an L2 to H transition (bottom panel), show an overall increase in the H state residence time. **(c)** TODPs under this condition show that about 35% of the molecules have a stable H state (diagonal), which is 3-fold higher than the B* condition although percent of molecules transitioning from the L2->H and M->H states are similar. White diagonal line populated stable FRET states that don't transition under our observation.

remained similar. Post synchronized histograms (PSHs) allow for the quick visual comparison of the relative stability of FRET states across different conditions. PSHs can be synchronized at the occurrence of an event (e.g., appearance of fluorescent signal in binding experiments) or when a given FRET state is visited. The relative decay in signal from the point of synchronization is related to the stability of the synchronization even, i.e., k_{off} for binding events or stability of conformation for FRET state synchronization. We plotted PSHs for the B* and C complex synchronized by the presence of the M FRET state. In C complex conditions, the molecules starting at the M state transition more frequently to the H state where they exhibit a higher residence time (**Figure IV.6a**, upper panels, B* and C condition). A similar comparison of transitions starting at the H state further emphasizes the stabilization of this state by Cwc25 (**Figure IV.6a**, lower panels, B* and C condition). An apparent increase in H-state could result from shorter fluorescence lifetimes of molecules in the C complex. We ruled out any such bias by comparing the lifetimes of molecules in the B* and C conditions that are both on the order of ~15 seconds.

A quantitative assessment of the effects of Cwc25 in the conversion of B* to C complex was performed through kinetic analysis on the observed L2->H and M->H transitions. Dwell times for the forward and backward transitions were extracted and plotted as a cumulative scatter plot, which was then best fit to a double exponential function (**Figure IV.11**). A comparison of the weighted apparent rate constant (k_w) for the L2->H transition showed similar forward and backward rates between conditions yielding K_{eq} values of ~0.80 (**Figure IV.11**). In contrast, the equilibrium of the M->H transition is biased in the presence of Cwc25. The presence of Cwc25 accelerates the forward rate and reduces the backward rate of this transition leading to a 2.8-fold preference for the H-state (**Figure IV.6b**). This suggests that Cwc25 acts kinetically to stabilize

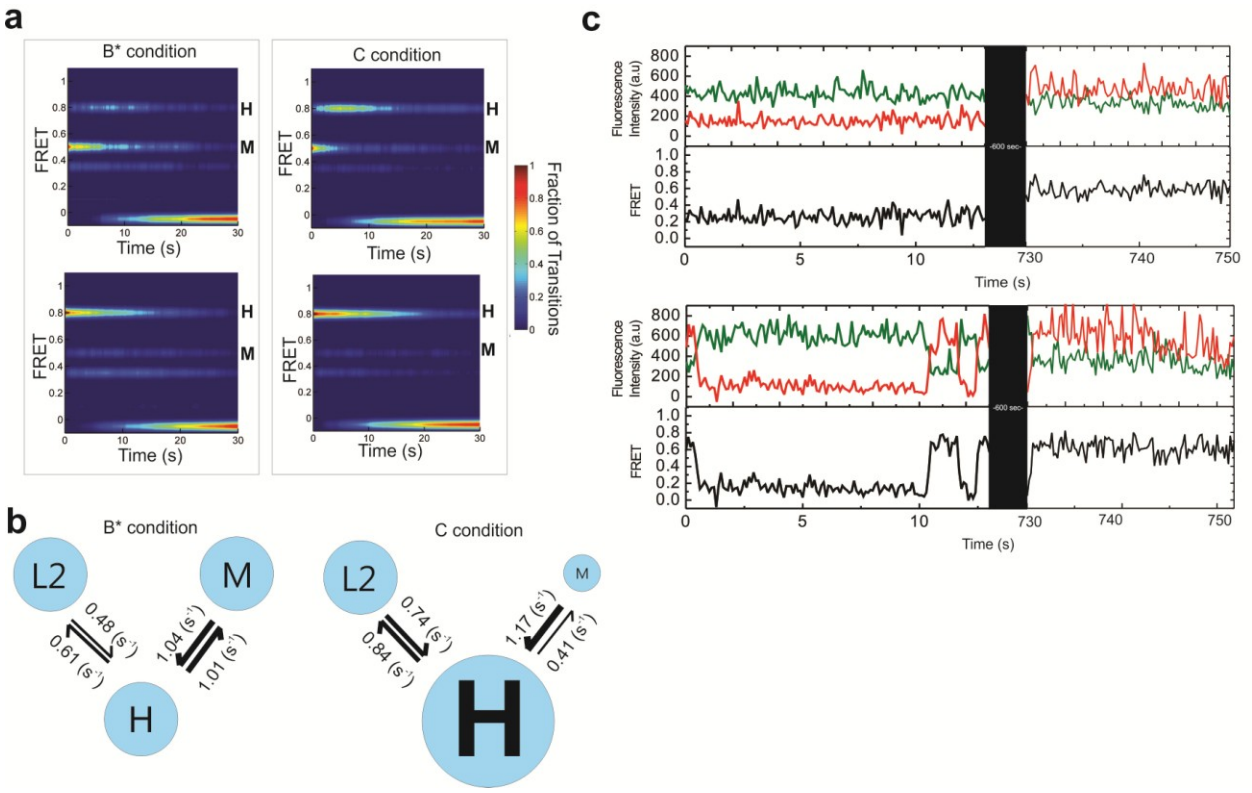


Figure 4. 6: Cwc25 enhances 1st step of splicing by stabilizing the H state.

(a) Comparison of condition before (B*) and after Cwc25 addition (C) via Post-synchronized histograms (PSH) reveal a higher occupancy in the H state in C complex condition when synchronized to a start state of M or H (top plots and bottom plots, B* and C complex respectively) . **(b)** Comparison of K_{for} and K_{rev} for the L2→H and M→H transition in the B* and C condition reveals ~3 fold increase in H state stabilization for the C complex. The thickness of arrows corresponds to the rate. **(c)** Representative trajectories with transition dynamics from the same molecules imaged before (B*) and after Cwc25 (C) addition. The dark box represents 10 minutes of incubation post-Cwc25 addition. These replicate the behaviors seen in the ensemble of molecules collected showing a switch to stable high FRET state upon Cwc25 addition.

the catalytically favorable conformation, thereby bringing about an enhancement in the first chemical step of splicing.

To test the hypothesis that Cwc25 can trigger a kinetic switch during first step activation, we tracked the same pre-mRNA molecules transitioning from B* to C complex. We were able to chase molecules through catalysis with the inclusion of a dark period during the addition of Cwc25 (**Methods**). We found that about 50% of the molecules observed shifted from a dynamic behavior between the high and low FRET state in the B* to a more stable and long-lived high FRET state upon the addition of Cwc25 (**Figure IV.6c**). 20% of these molecules displayed longer lived high FRET behavior even under B* condition. These molecules could represent the low levels of first step chemistry seen with the addition of just Prp2, Spp2 and ATP. 33% of the molecules showed faster switching in B* and slow or no switching to high FRET state in C conditions. These molecules could be the Prp2/Spp2 unbound population. ~8% of the molecules show switching behavior without any significant dwell in both B* and C conditions, representing a population unbound by Cwc25.

Our results suggest that Cwc25's role in first step catalysis is to stabilize the productive high FRET population by modulating the kinetics of inter-conversion between the available conformational states. It is not uncommon that in some cases, different conformational species may translate into the same FRET values. Nevertheless, they can be differentiated based on single molecule transition and rate information. Although our current placement of dyes on the pre-mRNA may limit our ability to distinguish the pre- from the post-catalytic H states, kinetic analysis presented here has helped us classify these as a dynamic or stable population, respectively. Hence, the H state comprises molecules either immediately prior to catalysis or in a post-catalytic state (H*). Evidence presented here points toward this possibility. We observe a 2-

10-fold enhancements of first step products in our ensemble assays (**Figure IV.1b**) with the addition of Cwc25. We observe a 3-fold increase in the M2 to H rate and similarly a 3-fold increase in the molecules in the stable H state under C complex conditions. The enhancement seen in the smFRET histogram is about 3-fold with the addition of Cwc25. Additionally, in the experiments where the switch from B* to C was observed in the same molecule., the number of molecules that showed the high FRET behavior in the C complex was 5-times that of the number of molecules in B* complex conditions with the same behavior. These numbers are consistent, within error, with the notion that the H state represents conformations proximal to or after first step splicing.

Cwc25 dynamically interacts with the pre-mRNA close to the branch site upon Prp2-mediated pre-mRNA remodeling

Prp2 mediated remodeling allows for Cwc25 to increase first step efficiency by several fold. Previous results have shown that the ATPase action of Prp2 results in the weakening or dissociation of several proteins including the U2 SF3a/b complex that are known to interact around the BS of the pre-mRNA.[91] This could result in the BS adenosine now becoming exposed to favor the nucleophilic attack involved in the first step of splicing. Additionally, a high affinity binding site for Cwc25 is created after Prp2 action[96]. Photo-crosslinking experiments have confirmed that Cwc25 can interact near the branchsite[97], but to date there is no dynamic information regarding the proximity of Cwc25 to its putative binding site. In order to understand the timing of Cwc25 binding and its proximity to the BS of the pre-mRNA, we used a C-terminally Cy5 labeled Cwc25 during SiMPull-FRET with pre-mRNA. Unlike the ribosome, there have been few single molecule protein-RNA FRET studies done in the spliceosome primarily because of the difficulties associated with obtaining a fluorescently labeled protein.

Cwc25 does not have any cysteines and thus presents an ideal choice for fluorescent labeling. The Cy5 fluorophore on the 5'SS of the

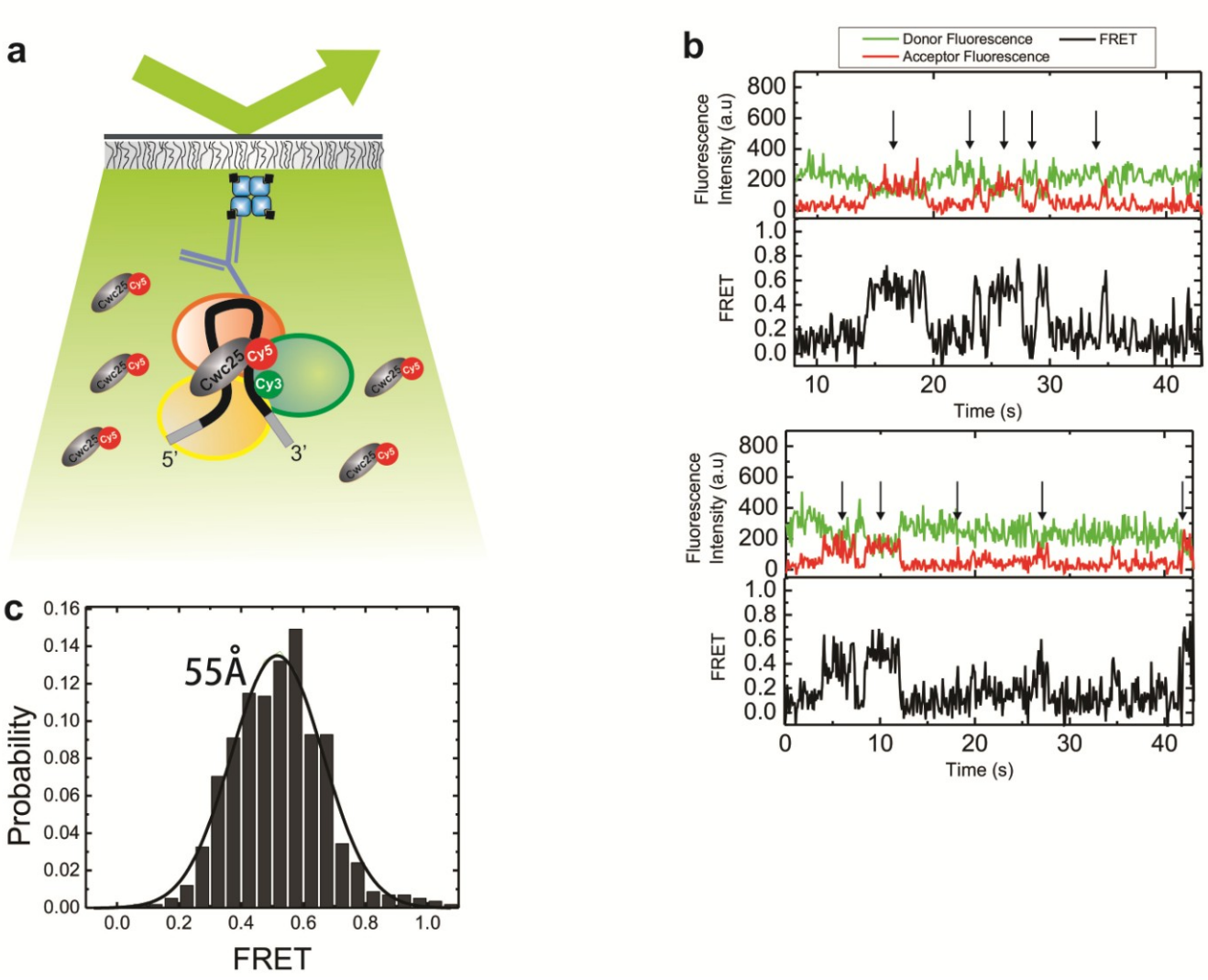


Figure 4. 7: Prp2 mediated pre-mRNA remodeling creates a binding site for Cwc25 near the BS.

(a) Schematic sm-FRET set-up showing the binding of Cwc25-Cy5 on the spliceosome assembled on pre-mRNA with a single active Cy3 fluorophore. **(b)** Representative smFRET trajectories showing the binding and associated FRET between Cwc25-Cy5 and BS Cy3 of the pre-mRNA. Arrows indicate binding events. **(c)** Histogram showing the average FRET value obtained from the experiment. The lower FRET peak is the background associated with the system while the broader peak centered at 0.4 is the average FRET between the fluorophores. The estimated distance between Cwc25-Cy5 and pre-mRNA Cy3-BS was estimated to be 55A° (**Methods**).

pre-mRNA was bleached and hence the pre-mRNA had a single fluorescent Cy3 label near the BS. The Cy5 label on Cwc25 did not interfere with its first step enhancement activity (data not shown). Single molecule FRET experiments were carried out with Cwc25-Cy5 added to the B^{act} or B* complex conditions. We observed binding and association of Cwc25-Cy5 under B* complex conditions (**Figure IV.7b**). From the observed FRET values, we estimate that the C-terminus of Cwc25 binds approximately 55 Å away from the Cy3 dye at position +6 from the BS adenosine (**Figure IV.7c** and **Methods**). We observed very low or no FRET between Cwc25 and the pre-mRNA BS in the absence of Prp2/Spp2 and ATP (data not shown). These results confirm that Cwc25 binds near the BS in a dynamic fashion only after Prp2 activity.

IV.4 Discussion

In this study, we combined single molecule FRET with affinity purification of a stalled spliceosome complex. By using a temperature sensitive mutant of Prp2 and a TAP tag on Cef1 protein, we were able to stall and purify a functional B^{act} complex. This has allowed us to successfully assign splice site conformational changes to the previously unknown roles of Prp2 and Cwc25.

Prp2 activates the spliceosome by inducing a conformational change at its core

DExD/H-box RNA helicases utilize the energy derived from the hydrolysis of ATP to promote structural changes in the spliceosome. Among them, the protein Prp2 has been shown to have an essential role in the first step of splicing along with its cofactor Spp2. In this study, we demonstrate how Prp2 functions in the first catalytic step by promoting a reorganization of the core of the spliceosome. A temperature sensitive mutation in Prp2 (Prp2-1) allowed us to assemble catalytically inactive spliceosomes (B^{act}) on a fluorescently labeled pre-mRNA. The

pre-mRNA was labeled near its 5'SS and BS allowing us to monitor the sites of the first step of splicing. Once assembled we isolated and immobilized this complex through the interaction between the TAP tag on Cef1, a protein present in B^{act}, and the biotinylated IgG conjugated to our single molecule surface. Interestingly, even though the B^{act} complex is at the latter stages of spliceosome assembly the splice sites are in a relatively extended conformation. Using our purified system, we found that the ATPase action of Prp2 results in a large scale remodeling of the pre-mRNA with the BS and 5'SS brought into close proximity, thus aiding catalytic activation for the first step of splicing. We further show that the juxtapositioning of the pre-mRNA 5'SS and BS for chemistry does not occur until the action of Prp2. One can imagine that the extended conformation in the B^{act} complex could serve as an intermediate for proofreading events just prior to catalytic activation. The exact mechanism by which Prp2 reconfigures the core of the spliceosome to allow for 5'SS and BS interaction is still not known. A commonly proposed mechanism for DExD/H box helicases is that they bind to single-stranded RNA regions and displace obstacles in a 3'to 5' direction. Prp2 binds to the 3' region of the branchsite and has a role in destabilizing the interaction of the SF3a/b complex with the branchsite region[91,97]. Whether or not this is the exact mechanism that Prp2 utilizes will require further investigation. Using the SiMPull-FRET approach described here, we can directly address this hypothesis.

Cwc25 promotes first step catalysis through a kinetic effect

Although the spliceosome is, in principle, capable of catalyzing first step chemistry with the addition of Prp2, an additional factor Cwc25 is required to promote efficient first step catalysis. Our data provide mechanistic insights into how Cwc25 enhances the first step of splicing. We find that, in the absence of Cwc25, the spliceosome is capable of reversibly exploring conformations favorable for chemistry (H state). The addition of Cwc25 results in the

stabilization of the conformation by preventing the H state from reverting back to a more extended conformation. Additionally, using protein-RNA single molecule FRET experiments, we are able to report on the proximity of Cwc25 in the catalytic center of the spliceosome. We find that Cwc25 binds approximately 55 Å from the BS of the pre-mRNA. Taken together, we propose a model in which, in the absence of Cwc25, the splice sites fluctuate between an active (H state) and an inactive conformation (M state). Cwc25 disfavors the conversion back to the M state by binding to the BS region in the H state conformer. This results in an equilibrium biased to favor the first step of chemistry (**Figure IV.8**). The ability to extract kinetic information

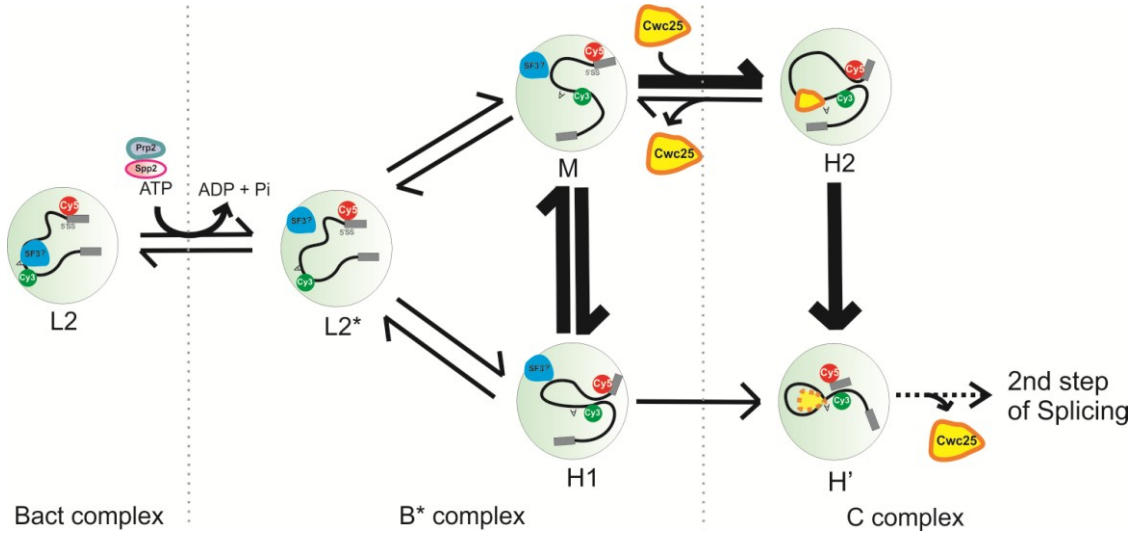


Figure 4. 8: Model for the mechanism of 1st step enhancement.

5'SS and BS of pre-mRNA in the B^{act} complex is predominantly in the low L2 conformation. Prp2 mediates a large scale conformational change allowing the splice sites to sample close conformational space (M and H states). We propose that SF3b complex is bound near the BS during this stage in the L2 and M conformations. While the H state(s) is the catalytically favorable conformation, with SF3b displaced from the BS, it can reversibly bind back and form the M state. The presence of Cwc25 reduces this conversion back by 3-fold thus favoring the conformation relevant to the 1st step of splicing.

from single molecule experiments proved key to our interpretation of these data. Our results invoke structural dynamics as essential in spliceosome catalysis.

Traditionally, single molecule experiments on the spliceosome have been performed using a pre-mRNA immobilized to a surface subsequently incubated with whole cell yeast extracts[41,45,48]. Utilizing SiMPull-FRET, we were able to isolate a specific complex along the spliceosome assembly cycle. There are several advantages to using such a purified system for studying single molecule pre-mRNA dynamics in the spliceosome. First, since reconstitution of the first step is achieved by the addition of single proteins, we were able to assign any kinetic shift in the pre-mRNA conformation to single components. In addition, we only observed those pre-mRNA molecules that have spliceosome components assembled on them and are all synchronized to the same stage in assembly. We also found that the affinity purified B^{act} complex shows much more defined dynamics than what is observed for a pre-mRNA immobilized with whole cell yeast extract that can adopt many stages of assembly (**Chapter III**). Furthermore, our purification method utilizes TAP-tag mediated immobilization that excludes the need for specific, customized antibodies for purification that may be difficult to produce. Thus, the technique presented here opens up a wide array of possibilities for targeting both spliceosomal and non-spliceosomal components from the TAP-tagged yeast collection for single molecule observation.

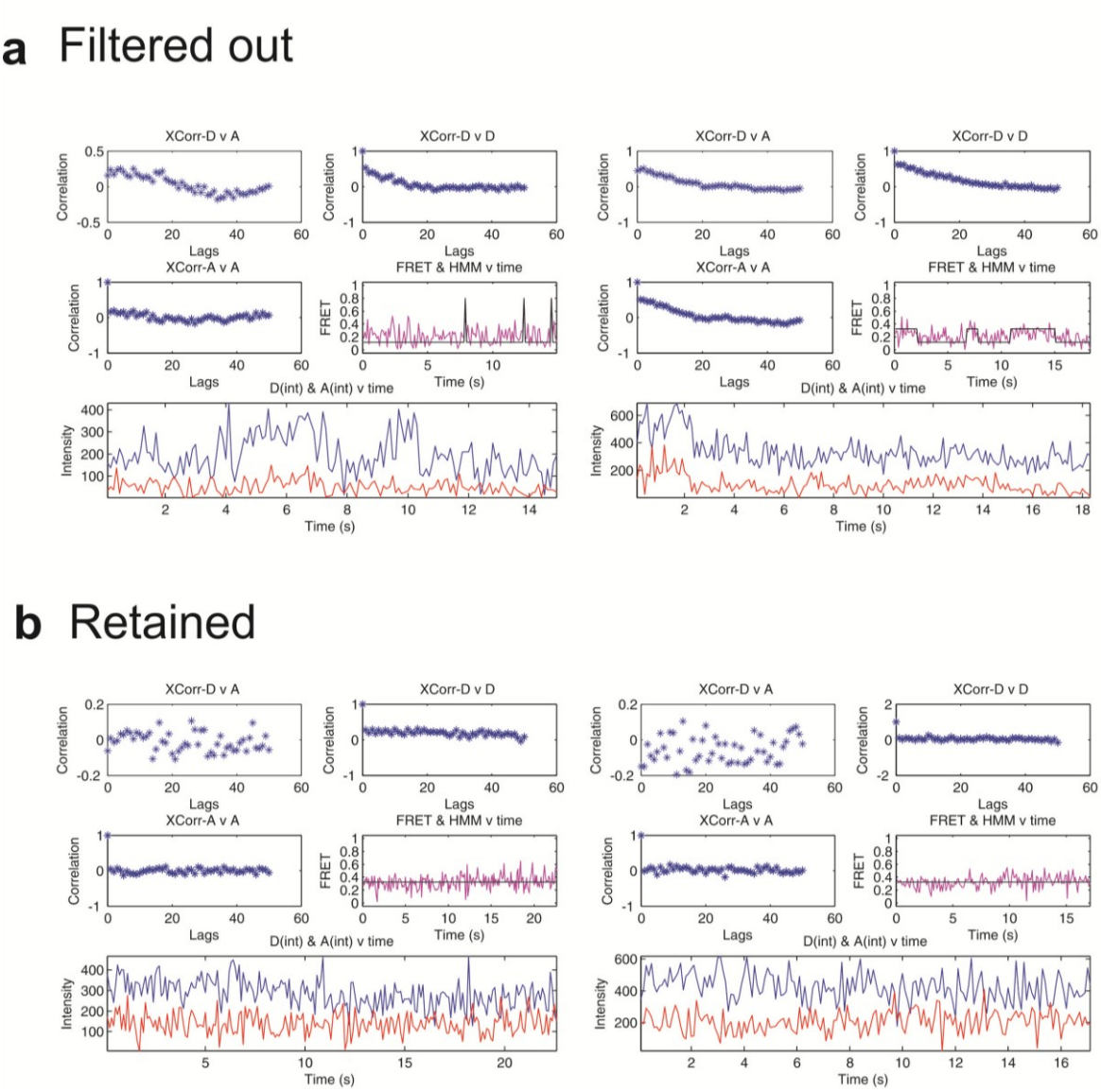
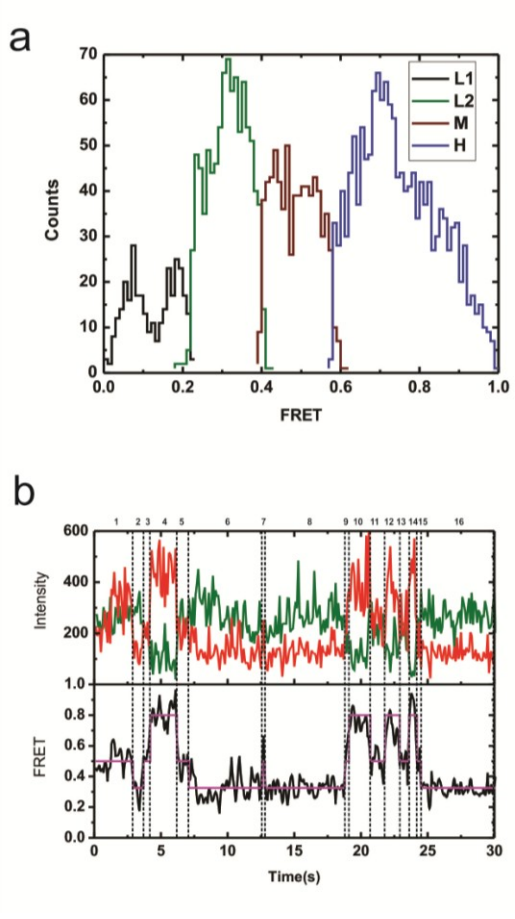


Figure 4. 9: Cross-correlation analysis of sm-FRET measurements

(a) Analysis of two representative molecules that were filtered out from further analysis because of significant positive cross-correlation between donor and acceptor signals at time lags of 0-50. **(b)** Analysis of two representative molecules in stable low FRET states that were retained because of negative cross-correlation at a time lag 0 and centered on 0.0 up to time lag 50. B^{act} molecules with no transitions detected by HMM do not have a decay in the cross-correlation signal indicating a lack of transitions beyond the sensitivity of the HMM algorithm

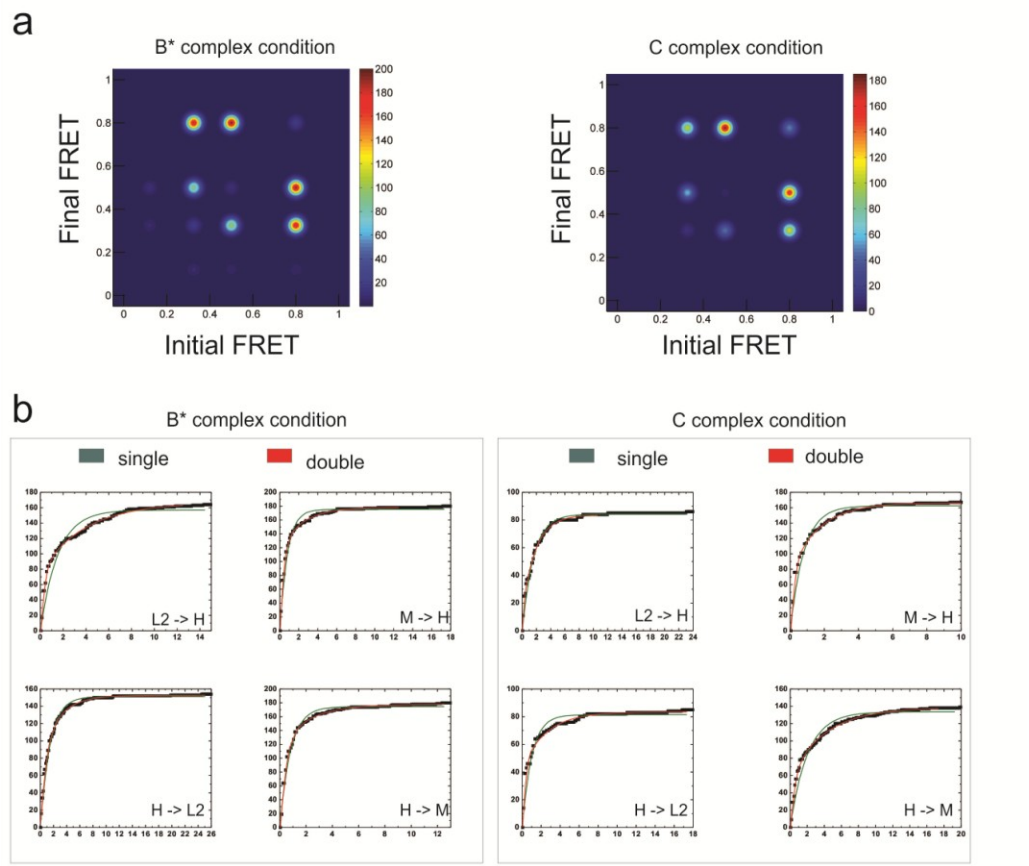


c

Segment	μ_{FRET}	HMM state	μ_{ACP}-μ_{DNR}/(μ_{ACP}+μ_{DNR})	Cluster
1	0.51	0.5	0.039	M
2	0.31	0.325	-0.37	L2
3	0.51	0.5	0.030	M
4	0.82	0.8	0.628	H
5	0.47	0.5	-0.075	M
6	0.32	0.325	-0.352	L2
7	0.56	0.5	0.123	M
8	0.32	0.325	-0.353	L2
9	0.55	0.5	0.111	M
10	0.76	0.8	0.525	H
11	0.50	0.5	0.012	M
12	0.72	0.8	0.454	H
13	0.52	0.5	0.061	M
14	0.87	0.8	0.747	H
15	0.45	0.5	-0.084	M
16	0.31	0.325	-0.366	L2

Figure 4. 10: K-means clustering of HMM assigned FRET states

(a) Histogram of HMM-idealized states for each of the k-means derived clusters L1, L2, M and H. **(b)** A representative molecule with HMM assigned state given a bin assignment based on clustering algorithm. **(c)** The values for each of the clustering parameters: the mean raw FRET state (2nd column), HMM state (3rd column), and normalized difference of mean acceptor and mean donor intensities for each segment (4th column).



	L2 -> H		H -> L2		M -> H		H -> M	
Parameter	B*	C	B*	C	B*	C	B*	C
A1	90.50 +/- 1.32	70.53 +/- 1.10	65.60 +/- 4.30	40.58 +/- 1.23	61.00 +/- 3.03	91.65 +/- 2.08	54.90 +/- 4.14	80.33 +/- 1.18
t_1 (s)	3.92 +/- 0.14	1.84 +/- 0.05	2.82 +/- 0.11	2.54 +/- 0.10	2.33 +/- 0.13	1.65 +/- 0.05	2.40 +/- 0.18	4.09 +/- 0.09
A2	94.40 +/- 2.75	27.76 +/- 5.61	89.70 +/- 4.01	55.51 +/- 4.01	135.84 +/- 3.24	105.75 +/- 3.40	140.41 +/- 4.10	68.46 +/- 1.59
t_2 (s)	0.29 +/- 0.02	0.08 +/- 0.03	0.78 +/- 0.04	0.21 +/- 0.02	0.35 +/- 0.02	0.16 +/- 0.01	0.43 +/- 0.02	0.45 +/- 0.02
t_w (s)	2.07	1.34	1.64	1.19	0.96	0.85	0.98	2.42
k_w (s^{-1})	0.48	0.74	0.61	0.84	1.04	1.17	1.01	0.41

Figure 4. 11: Kinetic rate determination for B* and C complex transitions.

(a) Transition Density Plots (TDPs) for the B* and C complex conditions scaled the number of transitions determined by HMM. The most prominent transitions in each condition are L2->H, H->L2, M->H, H->M. **(b)** Dwells times were extracted for each corresponding transition and plotted as a cumulative count plot and fit to either a single or double exponential rate equation. The double exponential function best approximates the data and was chosen for further analysis. **(c)** Parameters for the double exponential equations fitted to the dwell time data. To reduce the dimensionality of the data, a weighted average rate kw was calculated by utilizing the amplitudes associated with each time constant in the as weighting factors. kw was used for Keq calculations and rate comparisons between B* and C complex conditions.

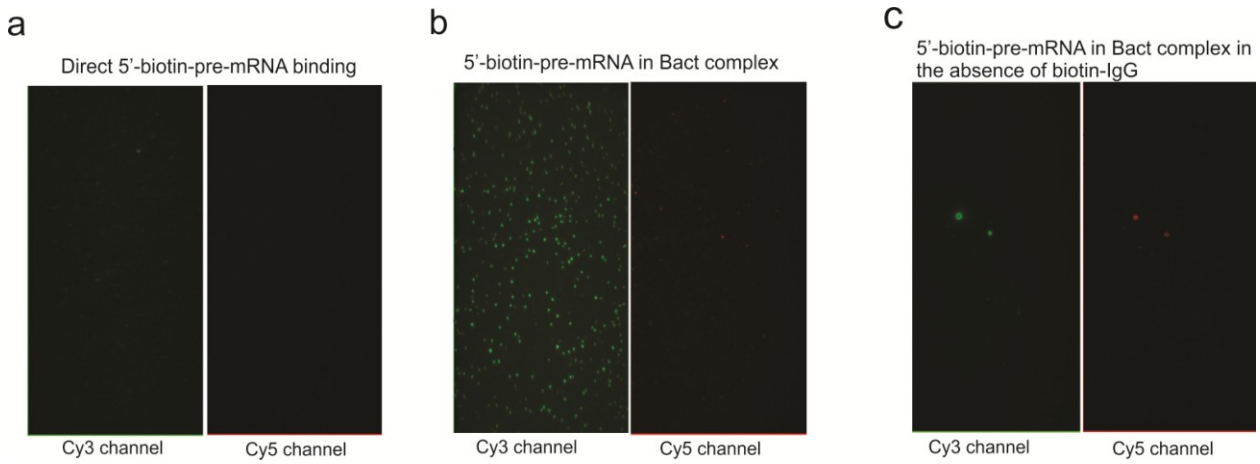


Figure 4. 12: Verification of specificity of the affinity purification of single spliceosome complexes.

(a) Field of view showing the binding of the 5' biotinylated pre-mRNA to the Streptavidin on the slide surface saturated with free biotin and biotin-IgG. **(b)** Field of view showing the binding of the immuno-affinity purified B^{act} spliceosome (with cef-1- TAP) to the Streptavidin on the slide surface saturated with free biotin and biotin-IgG. **(c)** Field of view showing the binding of the immuno-affinity purified B^{act} spliceosome (with cef-1- TAP) to the Streptavidin on the slide surface saturated with free biotin in the absence of IgG-biotin. Left and right panels are the Cy3 and Cy5 channels respectively. 532nm laser used for imaging in all conditions.

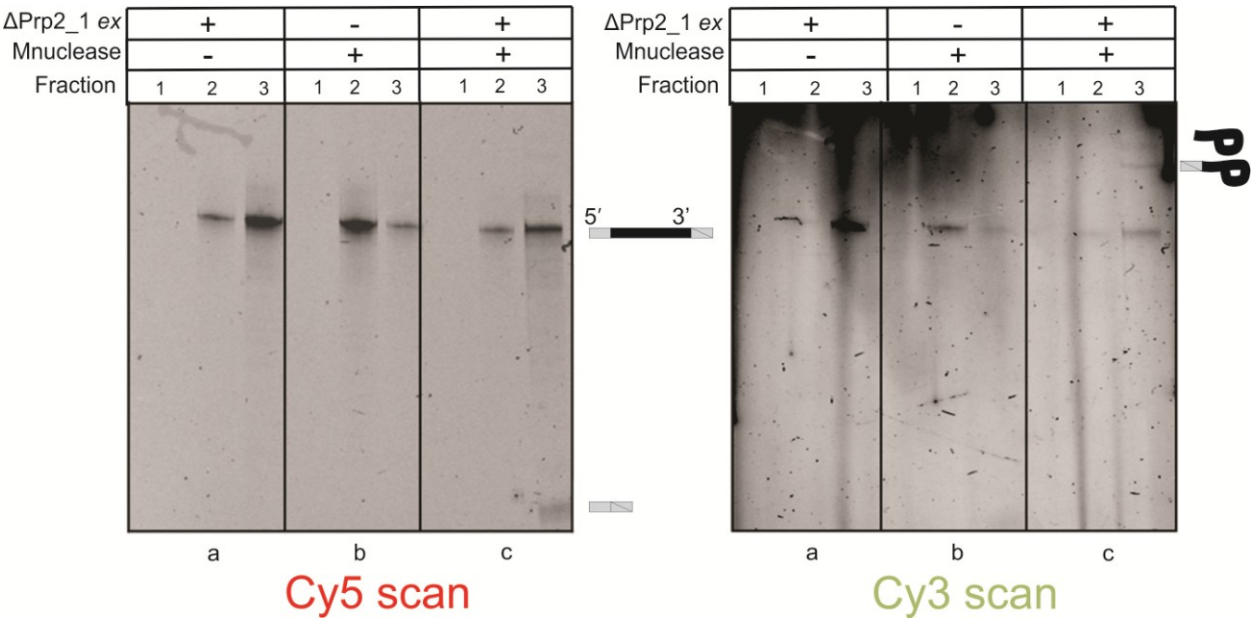


Figure 4. 13: Affinity purified B^{act} complex is active and can go through both steps of splicing.

Lanes 1, 2 and 3 represent fractions wash, unbound and supernatant respectively. Condition a, is wildtype pre-mRNA assembled in ΔPrp2-1,Cef-I TAP extract under splicing conditions with ATP at 23°C for 60 min. The B^{act} complex thus formed is affinity purified on Streptavidin coated magnetic beads bound by biotin-IgG. Unbound pre-mRNA is in lane 2 and the bead is further washed to remove non-specifically bound components (lane 1). The remaining fraction released from the bead and incubated in ATP and splicing buffer is in lane 3. Condition b, is wildtype pre-mRNA assembled in the absence of ΔPrp2-1,Cef-I TAP extract with splicing buffer and ATP at 23°C for 60 min. Most of the pre-mRNA is in the unbound fraction (lane 2). Condition c, is pre-mRNA assembled in ΔPrp2-1,Cef-I TAP extract to form the B^{act} complex as described in condition (a). The fraction bound to the magnetic bead is then reconstituted with Micrococcal Nuclease (MNuclease) treated extract and run in lane 3. The mature mRNA product is visualized in the Cy5 scan and the intron and intron-lariat is visualized Cy3 scan.

Ubc4 (20/20) 3' splice site mutant (3'ss)	5'-GAACUAAGUGAUC(5-N-U)AGAAAGGUAUGUCUAAAGUUAUGGCCACGUUUCAAAU GCGUGCUUUUUUUUUAAAACUUAUGCUCUUAUUUACUA**A**CAAAA(5-N-U)CAACAUGCUAUUGAACUACA**C**AUCCA CCUACUUCAUGUUT-3'
Ubc4 (20/20) Wildtype (WT)	5'-GAACUAAGUGAUC(5-N-U)AGAAAGGUAUGUCUAAAGUUAUGGCCACGUUUCAAAU GCGUGCUUUUUUUUUAAAACUUAUGCUCUUAUUUACUA**A**CAAAA(5-N-U)CAACAUGCUAUUGAACUAGAGAUCCA CCUACUUCAUGUUT-3'
DNA splint-dSplint	5' - GTTGATTTTGTTAGTA AATAAG(SP9)GTTTTAAAAAAAAAGCACGC -3'

Table 4. 1: Sequence information of oligonucleotides used in this study.

UBC4 intron is italicized, and the allyl-amine modified uridines are denoted as (5-N-U). The red and green color represents positioning of the Cy3 and Cy5 fluorophores respectively. In the 3'SS mutant, the bold and underlined cytosine replaced the guanine in the wildtype. The bold and underlined A is the branchpoint Adenosine. dSplint is the DNA splint used for template ligation to synthesize the pre-mRNA as described[16]. Sp9 denotes a 9 carbon link.

Initial FRET	Final FRET	Fraction of molecules		
		Bact	B*	C
0.12	0.12	0.11	0.00	0.00
0.12	0.33	0.11	0.02	0.00
0.12	0.50	0.02	0.01	0.00
0.12	0.80	0.01	0.02	0.00
0.33	0.12	0.12	0.02	0.01
0.33	0.33	0.52	0.11	0.06
0.33	0.50	0.06	0.12	0.08
0.33	0.80	0.01	0.39	0.29
0.50	0.12	0.01	0.01	0.00
0.50	0.33	0.05	0.11	0.08
0.50	0.50	0.09	0.06	0.03
0.50	0.80	0.02	0.26	0.31
0.80	0.12	0.01	0.02	0.00
0.80	0.33	0.01	0.37	0.30
0.80	0.50	0.01	0.24	0.29
0.80	0.80	0.04	0.12	0.32

Table 4. 2: TODP quantification for all data sets.

Molecules with at least one occurrence of the FRET transition given by the Initial and Final FRET states in columns one and two are counted and divided by the total number of molecules in that transition. Molecules that only occupy one state are accounted for in rows where the Inital and Final FRET states are equal. The majority of B^{act} molecules are in a stable 0.33 FRET state which diminishes greatly in the B* and C conditions. The most significant difference between the B* and C complex is the significant increase in the fraction of molecules occupying only the high FRET state (0.8->0.8).

IV.5 Acknowledgements

We would like to thank Reinhard Luhrmann's lab for kindly providing plasmids for expression of Prp2, Spp2, and Cwc25.

Chapter V: Conclusions and Outlook

V.1 Conclusions

The structure-function relationship in biology became evident once techniques like X-ray crystallography provided high resolution three-dimensional structures of biomolecules. More recently, a new paradigm has emerged as a result of the information gleaned from single molecule FRET (smFRET) experiments: the structure-dynamics-function relationship. The ability to monitor a molecule's dynamics and relate them to its function has provided insight into how dynamics relate to function from molecules as small as the hairpin ribozyme to as large as the ribosome.[38,43]

Until recently, the pre-mRNA splicing field had not benefitted from high-resolution crystal structures or smFRET data. Although a high-resolution structure of the fully assembled spliceosome is still not available, the recent emergence of structures of individual components has helped pave the way for a better understanding of what their roles may be in a fully assembled complex.[98] Perhaps the most intriguing structures are those of the human U1 snRNP and the recently published Prp8 structure.[99,100] The U1 snRNP structure provided insights into how the snRNA component is utilized to bring together a variety of proteins to create a stable snRNP. Additionally, the serendipitous interaction between two U1 snRNPs in the crystal created an RNA-RNA interaction that mimicked the U1-5'Splice site during early stages of spliceosome assembly.

Sequence-specific interactions between the spliceosomal snRNAs and pre-mRNA are present throughout the splicing cycle and this structure provides a framework for understanding those interactions. By comparison, Prp8 is part of the U5 snRNP and has been termed the "heart of the spliceosome".[101] Crosslinks between Prp8 and the 5' splice site, branchsite and 3' splice site have all been detected. It is one of the most conserved nuclear proteins with nearly 61% amino acid sequence identity between yeast and humans. Due to its large size (~279 kDa) the only high resolution structures of Prp8 have been limited to small truncations. The structure by Galej *et al* is the most complete Prp8 crystal structure to date. It revealed a structure with domains resembling the bacterial group II reverse transcriptase and a type II restriction endonuclease.[100] These functional domains reinforced the connections between the spliceosome and self-splicing group II intron.[102] The structure also helped explain the effect of previously characterized Prp8 mutations that affect first and second step catalysis by providing a structural context by which to interpret them. The increasing number of high-resolution structures will provide us with a new perspective to better understand spliceosome catalysis. Yet, we know that static structures alone are not sufficient to complete the picture. In this thesis, we utilized smFRET to address how pre-mRNA splicing dynamics play a central role in spliceosome activity.

Establishing an smFRET assay to assess pre-mRNA dynamics during splicing

The ability to track conformational changes in real-time through FRET provided a great opportunity for monitoring the structural dynamics present during splicing. We developed an smFRET based assay to monitor the relative position of conserved splice sites on the pre-mRNA during spliceosome assembly. Development of this assay was aided by a microarray based screen of yeast introns that was able to identify UBC4 as a candidate gene with a short intron that is

spliced well *in vitro*. The 95-nucleotide intron in this gene made it accessible for chemical synthesis, which allows for site-specific labeling with FRET dyes. Ensemble characterization of this substrate showed that even after fluorescent labeling it spliced with the same efficiency as other established introns and responded to splice site mutations in the same manner. We utilized the chemically synthesized UBC4 with fluorescent labels at the 5' splice site and 3' splice site to investigate the dynamics of exon-exon position. The smFRET experiments indicated that the substrate folds dynamically *in vitro* prior to spliceosome assembly and that single point mutations affect the ensemble of structures that it can visit. This hints at the possibility that point mutations within the intron could affect splicing by affecting interactions with the spliceosome (likely for highly conserved regions) or could affect the intrinsic dynamics of the substrate, which would in turn affect spliceosome assembly. In fact, the involvement of pre-mRNA structure during spliceosome assembly and intron recognition is becoming evident.[103,104] Another result from these sets of experiments was the identification of splicing dependent signals that required ATP and were not sampled when the substrate was mutated at either the branchsite or 3' splice site. The most striking feature of the dynamics was their reversibility throughout spliceosome assembly. Every accessible conformational state was sampled reversibly. Similarly, splicing catalysis has been shown to be reversible, and the prevailing model for proofreading is based on the assumption that the spliceosome is constantly sampling active and inactive conformations.[33,35] Our data provide the first direct evidence of conformational flexibility and suggest that the spliceosome, like the ribosome, operates at near thermal equilibrium, relegating the role of ATP in the splicing reaction to changing that equilibrium.

A quantitative analysis of the conformational pathway during the first step of splicing

Having established smFRET as a tractable tool for investigating splicing dynamics, we sought to map the conformational pathway leading to the first step of catalysis. We site-specifically labeled the UBC4 intron near the sites of the first step of chemistry, the 5' splice site and branchsite. We introduced a series of mutations into the substrate that affect various stages of the splicing cycle and allow us to limit the extent to which the spliceosome can assemble on the substrate. In addition, we were able to use previously characterized ablations of specific spliceosome components to block assembly while maintaining wildtype splice sites. To be able to extract the full extent of information from a complex set of dynamics from ~10,000 molecules, we developed a novel approach to analyze smFRET data. The approach relies on the hierarchical clustering of dynamics based on the use of a Hidden Markov Modeling derived FRET Similarity Matrix (FSM). The FSM allows for an efficient clustering of molecules in an objective and efficient manner. Importantly, the FSM can distinguish molecules that inhabit different FRET states as well as those that are in the same FRET states but have different rates of inter-conversion. We were able to identify the consensus behavior of molecules assembled in the various experimental conditions and placed them in an ordered fashion based on what is known about the blocks to spliceosome assembly used. We found that during early stages of spliceosome assembly, the intrinsic structure of the pre-mRNA is disrupted, leading to an extended conformation between the 5' splice site and branchsite. The 5' splice site and branchsite are then brought closer together through the addition of ATP once the spliceosome has been activated by the loss of U1 and U4 snRNPs. The latter stages of assembly limit the number of extended conformations that the substrate can sample, thus promoting the first step of catalysis. The clustering approach was key to our ability to extract specific sets of dynamics

responsible for differences observed among the various experimental conditions tested. The analysis routine is generalizable to any set of smFRET data and we believe that it will become a commonly used tool as the behavior of increasingly complex systems is investigated by smFRET.

A Prp2 dependent conformational change at the core of the spliceosome sets the stage for Cwc25 assisted first step catalysis

The achieve a catalytically active conformation, the spliceosome must undergo a conformational change that is dependent on the splicing factor Prp2.[30] This new conformation is capable of catalyzing the first step of splicing inefficiently and requires the presence of the protein Cwc25 to reconstitute full first-step activity. The conformational change set forth by the ATPase activity Prp2 was previously detected by sedimentation analysis and electron microscopy studies.[30] These techniques only provide information on large-scale changes to the spliceosome during this activation step. During this stage of spliceosome assembly, the pre-mRNA substrate has been integrated into the splicing core. The 5' splice site is base paired with U6 snRNA and the branchsite is bound by U2 snRNA. By utilizing the UBC4 pre-mRNA construct fluorescently labeled at the 5' splice site and branchsite, we assessed the effect the Prp2 mediated conformational change has on the catalytic core deep inside the spliceosome. Combining immuno-purification techniques with our smFRET approach, we purified a spliceosome complex that is stalled just prior to the Prp2 conformational change. The subsequent conformational change can be induced by the addition of Prp2, its protein cofactor Spp2 and ATP. Our smFRET data established that Prp2 action induces a large reorganization of the pre-mRNA substrate that is dependent on Prp2's ATPase activity. The hydrolysis of ATP thus allows the pre-mRNA substrate to sample conformations where the 5' splice site and branchsite are in close proximity. These interactions are transient and reversible. We then chased this complex

through first step catalysis with the addition of Cwc25. The presence of Cwc25 in the complex did not introduce a new set of conformations. Instead, we detected differences in the kinetics of inter-conversion between the conformations being sampled after the Prp2 induced conformational change. The kinetics now favor a scenario where the 5' splice site and branchsite are proximal, thus favoring catalysis. These sets of experiments helped elucidate the role of Prp2 in the activation of the spliceosome. Its ATPase activity is utilized to affect a large-scale conformational rearrangement at the core of the spliceosome. Additionally, we present evidence for a kinetic mechanism underlying the enhancement of first step catalysis by Cwc25. These results highlight the strength of single molecule measurements on the spliceosome by providing a direct link between structure, dynamics, and function.

V.2 Outlook

The establishment of an smFRET based approach to splicing has provided the field with a new tool for investigating the various stages of splicing in more mechanistic detail. Being able to extend this approach to create a comprehensive picture of the structural dynamics involved will require the continued development of current strategies and development of new experimental approaches and analysis methods.

In this thesis, our main focus was on FRET pairs placed on the pre-mRNA substrate, but future studies should not be limited to this approach. The labeling of spliceosomal protein components is possible and has been used to monitor spliceosome assembly in real-time.[48] We have evidence that by fluorescently labeling a protein, Cwc25, we can effectively monitor FRET between a spliceosome component and the pre-mRNA. This opens up the possibility for a large set of experiments where we utilize known protein:pre-mRNA crosslink information to specifically monitor the dynamic interactions between the spliceosome and its pre-mRNA

substrate. An immediate target of interest would be the SF3a/b complex that is bound near the branchsite and thought to be affected by the Prp2 dependent conformational change during first step activation.[91] Attempts have been made in our lab to label the SF3a/b component Cus1 with protein-based fluorescent labeling tags in a strain background where the Prp2-1 allele is present to enrich for the catalytically inactive B^{act} complex. This strain has proven difficult to create, possibly due to the synthetic lethality of these combined modifications. As an alternative, I propose an approach where we use the already established extract where Cus1 has been labeled with a SNAP tag in combination with dominant negative proteins. There are well characterized dominant negative (DN) proteins for both Prp2 and Prp16 that, when expressed recombinantly, can be added to WT extract to stall splicing just prior to the first step and right after the first step, respectively. Adding these proteins to the Cus1-SNAP extract and incubating this extract with a UBC4 construct fluorescently labeled at the branchsite will allow us to answer two questions. First, we will be able to assess whether or not Cus1 is present in a post-first step complex by comparing the amount of colocalization it has with the pre-mRNA in the Prp2DN reaction with that in the PRP16DN. Secondly, if Cus1 is present in both complexes, we will be able to determine if there is any change in the FRET signal (value or kinetics) between Cus1 and the branchsite before and after catalysis. It has been suggested that Prp2's role is destabilizing this interaction and allowing for Cwc25 to bind near the branchsite. We have already established that the immuno-purification of the Prp2 stalled complex is possible through a TAP-tag on one of the NTC complex proteins (Chapter IV). A biotin on the 5' end of the pre-mRNA could easily replace this tag if the assembled and stalled complexes are purified through a glycerol gradient; a similar approach utilizing an MS2 hairpin loop on the 5' end the actin pre-mRNA was used in the initial purification and characterization of this complex.[30]

Another example of a potential protein-pre-mRNA FRET pair is found near the 5' splice site during early intron recognition. The U1 snRNA is base paired with the 5' splice site in one of the first steps of yeast splicing. The U1-70 k protein stabilizes this interaction. Genetic evidence in yeast for a role of this protein in 5' splice site recognition was strengthened when the crystal structure of the human U1 snRNP was determined. In this structure the U1-70k protein is positioned near the site of U1snRNA:5' splice site interaction.[99] Labeling this protein and monitoring its FRET efficiency with a 5' splice site labeled substrate will provide us with information on the interactions of the U1 snRNP during intron recognition. Fluorescently labelling U1-70k can be done by labelling its single cysteine residue via a thiol reactive dye. We have already utilized this exact approach to measure FRET between Cwc25 and the branchsite (Chapter IV). The exchange of U1 snRNA for the U6 snRNA at the 5' splice site is critical for spliceosome activation and is dependent on the ATPase activity of Prp28. By combining the labeled U1 snRNP protein with characterize Prp28 mutants and substrate mutants that affect the stability of the U1:5' splice site interaction, we can achieve a better understanding of the structural rearrangements involved in this step of splicing.

In addition to labeling spliceosomal proteins, we can fluorescently label spliceosomal snRNAs. The snRNAs come together to form an intricate web of RNA:RNA interactions with each other and with the substrate. The ability to deplete specific snRNAs and reconstitute the activity of their corresponding snRNPs with *in vitro* transcribed RNAs has been demonstrated (Chapter III). [80] The size of most of the snRNAs makes them inaccessible to chemical synthesis. The site-specific incorporation of fluorescent dyes is critical to FRET studies, therefore we must rely on approaches that allow the site-specific modification of *in vitro* transcribed RNAs at the 5' end or 3' end. Nearly all the interactions of the pre-mRNA with complementary snRNA

sequences are near the 5' end of the snRNA, therefore the 5' end would be the preferential labeling site. The interaction between the U1 snRNA and the 5' splice site is at the very 5' end of the snRNA and the interaction of the branchsite sequence with U2 is approximately 30 nucleotides from U2 snRNA's 5' end. Labeling the 5' end of these RNAs can be achieved by the incorporation of an excess of guanosine-5'-O-monophosphorothioate during an *in vitro* transcription reaction. This modified nucleotide can only be incorporated at the 5' end and subsequently can be labeled by a thiol reactive dye. The U6 snRNA is only 112 nucleotides in length, smaller than our currently chemically synthesize substrate, which means that site-specific incorporation of dyes does not have to be limited to the termini. The possibility to monitor FRET between the snRNAs and pre-mRNA will allow us to directly observe rearrangements caused by the RNA dependent ATPases present in the spliceosome that are postulated to affect these interactions. Furthermore, by labeling the U1 snRNA and the U6 snRNA we can monitor and extract the kinetics of the exchange of these snRNAs at the 5' splice site, a key step in the formation of the spliceosome's catalytic core.

The number of potential targets for single molecule experiments are great. The wealth of genetic and biochemical data available in the budding yeast system serves as a superb resource from which to narrow our focus. As more high-resolution structures of the spliceosome become available, they will also provide further specific targets for interactions to monitor. The work presented in this thesis pioneers a framework for further advances in what is sure to expand into the exciting and fruitful field of single molecule splicing.

Appendices

Appendix A : Tracking pre-mRNA conformational changes throughout the entire splicing cycle

A. 1 Experimental approach for visualizing dynamics on the *in vitro* splicing timescale

The lifetime of fluorophores used in single molecule experiments is much shorter than the minutes timescale required for in vitro splicing experiments. This limits our ability to observe the entire set of conformational changes that lead to mRNA formation. To circumvent this limitation, we have utilized shuttered time-lapse illumination to increase our observation window. This is performed by introducing dark periods during the imaging process, which cuts down our time resolution but allows us to observe a single pre-mRNA for an extended amount of time. The number of illuminated frames and length of the dark period can be adjusted in an attempt to find a compromise between time resolution and the length of the observation window. By introducing 50 s dark periods with 10 s of illumination, it is possible to image a molecule over 1 hour (**Figure A.1)**.

Another challenge of *in vitro* splicing experiments is the relative inefficiency of the process. We are mainly interested in the dynamics that lead to successful mRNA formation. To access these dynamics, we combined the time-lapse imaging with a method to accurately distinguish spliced from unspliced RNAs at the level of individual molecules. For mRNA detection, we designed a DNA probe that binds to the pre-mRNA with 9 bp of complementarity and the mRNA with 11 bp of complementarity. The dissociation rate (k_{off}) of hybridized probes is exponentially dependent on the number of base pairs with which it binds.[105] We labeled our probe with a fluorophore so that we can visualize its association and dissociation through the appearance and disappearance, respectively, near the surface bound RNA. This approach

provides us with distinct signatures for the pre-mRNA and mRNA (**Figure A.2**). As a result, we first can measure the dynamics of single pre-mRNA molecules by smFRET and then determine whether or not these dynamics lead to mRNA formation.

We applied the shuttered illumination and mRNA detection scheme to group the dynamics we observed based on whether or not they led to mRNA formation. To capture data in extract and negate the effect of photobleaching, we collected only 0.1 s (one video frame) of smFRET data at 45 second intervals. We utilized a 5'SS-branchsite labeled substrate incubated in ATP supplemented extract for a total of 15 min. The position of the Cy3 dye in the intron of this substrate provides an additional method of mRNA detection. Molecules that do not splice within the 15 minute time period still have Cy3 signal at the end of the time period (**Figure A.3a**). Spliced and disassembled substrates lose the intron and therefore lose the Cy3 signal (indicated by the black arrow), but retain the Cy5 signal on exon 1 because it is still tethered to the microscopy surface via the 5' biotin on the RNA (**Figure A.3b**).

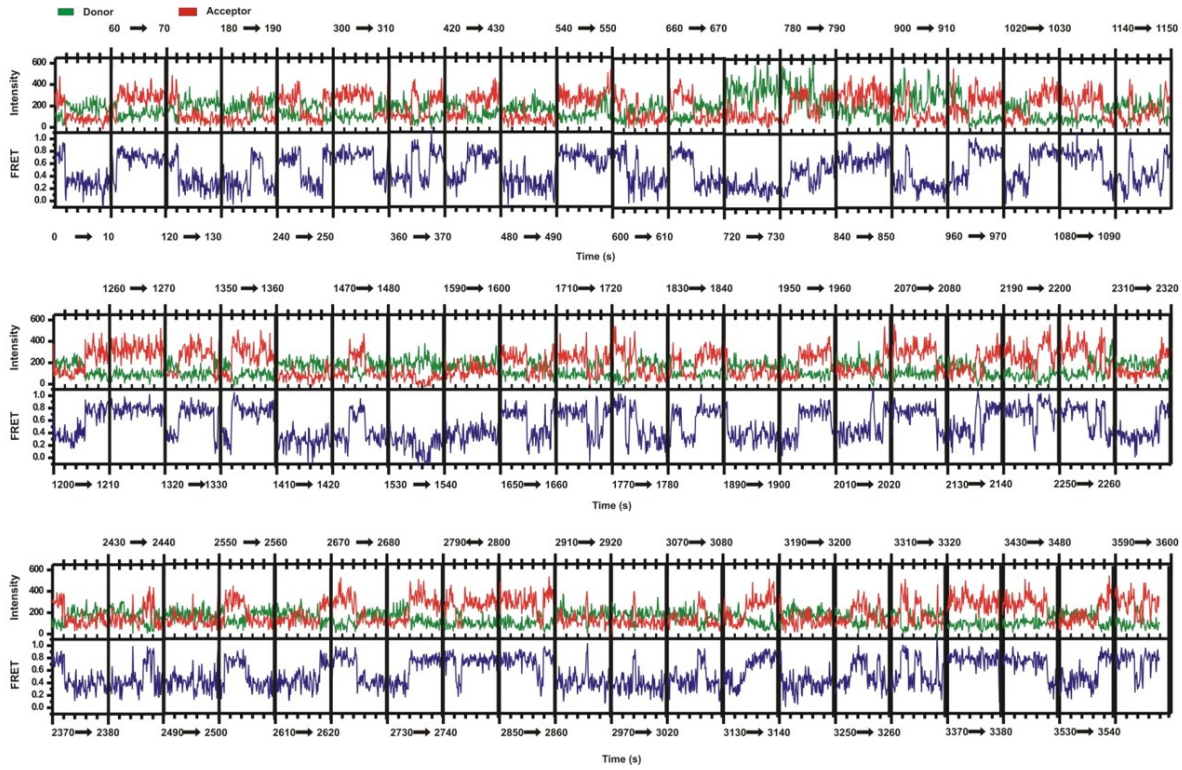


Figure A.1: Single pre-mRNA imaged with shuttered illumination.

Shuttered illumination was used to observe a single pre-mRNA in splicing buffer for 60 min. The pre-mRNA was illuminated at 10 second intervals interrupted by 50 second dark periods represented by black vertical bars.

During the splicing time course, the spliced molecule began at a low FRET state (with the splice sites far apart) to conformations where the 5'SS and BS are brought in close proximity in a reversible fashion before mRNA formation. Altogether, a combination of these approaches will provide an experimental test of whether or not the 5'SS and BS are juxtaposed during the entire splicing cycle. To complete these investigations we would need to collect more data on the false positive and false negative rates of the mRNA binding probe. This can be accomplished by using control substrates with either pre-mRNA or mRNA sequences. Once these rates are shown to be reasonably low, more confidence can be assigned to the dynamics correlated with mRNA formation.

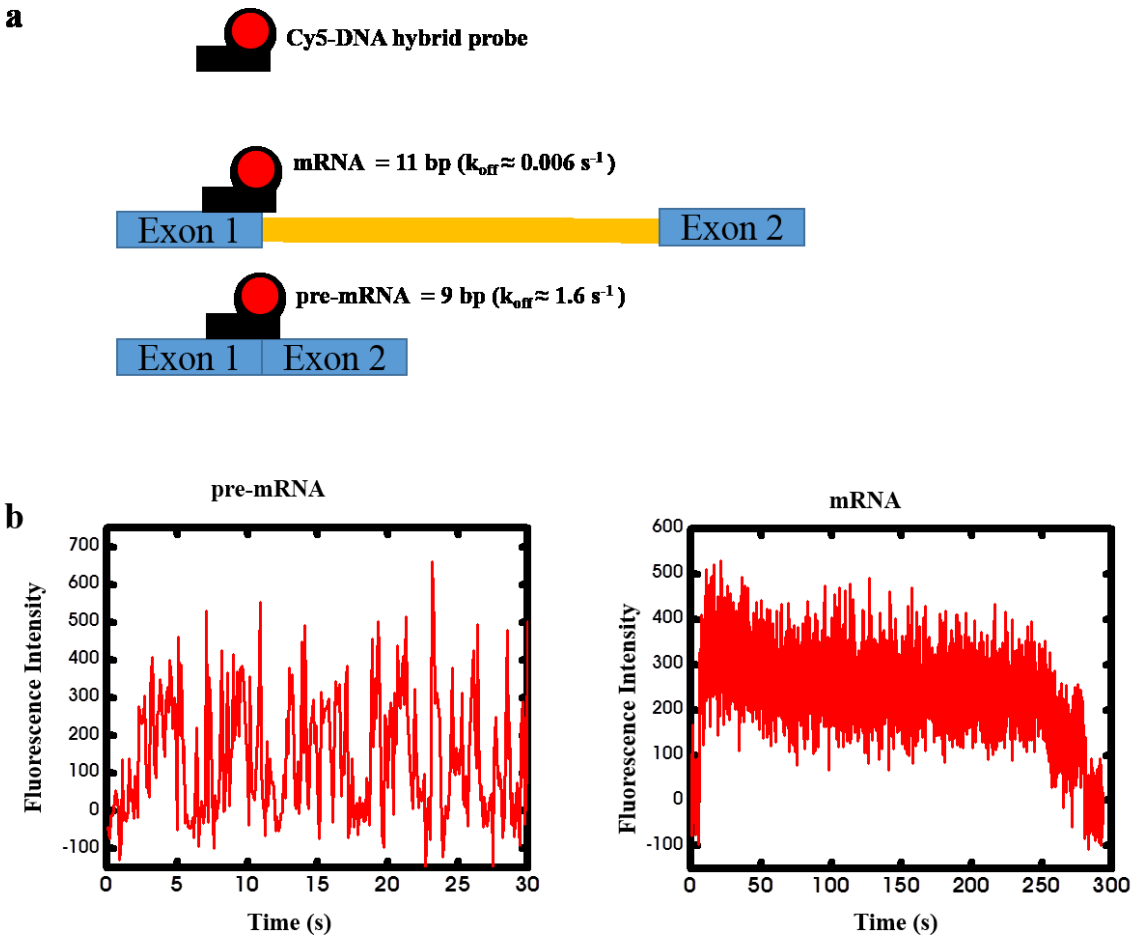


Figure A.2: Detection of spliced mRNAs.

(a) A Cy5 labeled DNA probe was designed so that it has 9bp of complementarity to the pre-mRNA and 11bp with the mRNA. This affects the dissociation rate and allows for the detection of either species of RNA. (b) Example binding and unbinding events of the Cy5-DNA oligo from a control sample containing either pre-mRNA or chemically synthesized mRNA. The pre-mRNA unbinding events are short while those on the mRNA are slow and easily distinguished from those on the pre-mRNA.

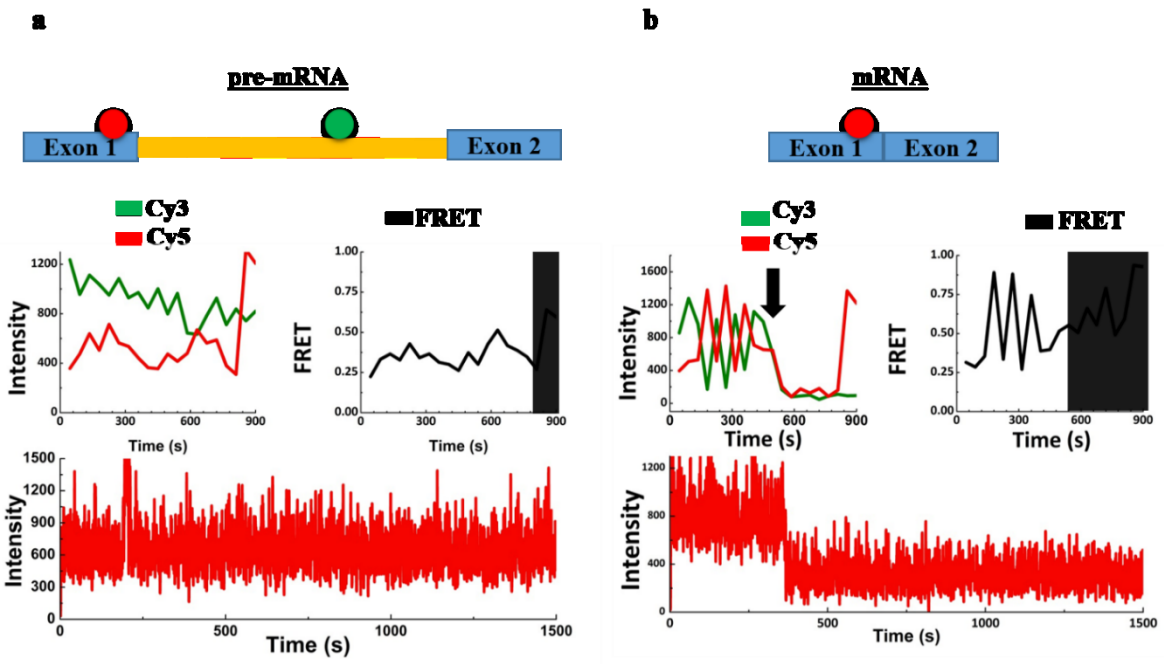


Figure A.3: Shuttered illumination combined with mRNA detection.

Pre-mRNAs tracked throughout a single molecule splicing reaction. One frame of data (0.1s) was collected at 45 s intervals for a total of 15 minutes. **(a)** A molecule that was determined to be a pre-mRNA from the fast unbinding rates (bottom panel). The fluorescence intensity trace shows that the Cy3 fluorescence which would be lost upon mRNA formation is still present at the end of the 15 minutes and that at every time point we imaged it was in a low FRET conformation. **(b)** Slow dissociation kinetics that are indicative of mRNA formation (bottom panel).The fluorescence intensity traces shows that we lose our intron at about the 7 minute mark, indicated by the black arrow and that during the splicing time course this molecule began from a low FRET state (with the splice sites far apart) to conformations where the 5'SS and BS are brought in close proximity in a reversible fashion before mRNA formation.

A. 2 Materials and Methods

Single molecule imaging

The UBC4 substrate and single molecule setup are the same as those described in Chapter III and Chapter IV. Images were acquired using Micro-Manager[106] because it collects data at the I-Pentamax's camera 12-bit range allowing the use of a higher concentration of mRNA detection oligonucleotide during the hybridization part of the experiment. A single frame of data at 100 ms was taken every 45 s for a total of 15 min. The Cy5 on the pre-mRNA was excited directly at the end of the imaging period to distinguish molecules in a low FRET state from those whose Cy5 has bleached.

mRNA detection

After the pre-mRNA has been imaged for 15 minutes the extract in the microfluidic chamber was replaced by 100 uL of stop buffer A with formamide (50 mM Tris pH 7.2, 20mM NaCl, 20 mM EDTA pH 8.0, 4.5% formamide). Stop buffer A with formamide was then replaced by Stop buffer A with formamide and 0.5 % dextran sulfate. This solution was left incubating on the slide for 10 min to block the surface with the negatively charged dextran sulfate. Preliminary experiments with salmon-sperm DNA, another blocking agent, had higher non-specific background binding than the dextran sulfate treatment. During the 10 min incubation the Cy5 fluorophore was directly excited to photobleach any remaining fluorescence from this dye. This is done because our mRNA detection oligonucleotide is also labeled with Cy5. After the 10 min incubation the channel was flushed with stop buffer B with formamide (50 mM Tris pH 7.2, 500 mM NaCl, 20 mM EDTA, 4.5 % v/v formamide) and then once again with stop buffer B but with no formamide. A solution of 100 nM the mRNA detection oligo (RC1 – 5'-Cy5-CTCTTTCTAGA-3') in hybrid buffer (500mM NaCl, 100mM Tris pH 7.2, 20 mM EDTA) was

then flowed on imaged while simultaneously exciting both Cy3 and Cy5 at 250 ms exposure time for a total of 30 minutes. Binding events were visualized MATLAB scripts written to analyze the TIFF images from MicroManager and utilizing pre-mRNAs as fiduciary marks for mapping the molecules from the smFRET movie to the mRNA detection movie.

Appendix B : Combining computational and experimental approaches to model pre-mRNA structures

B. 1 Structural constraints induced by hybridization

There is mounting evidence to suggest that pre-mRNA structure and dynamics have an effect on splicing.[103,104] Single molecule FRET (smFRET) based splicing assays allow us to directly monitor a pre-mRNA's structural dynamics as well as its folding path through spliceosome assembly and catalysis. The binding of spliceosomal components to the pre-mRNA affect the number of conformations that can be sampled by occupying sites that could be utilized for intramolecular base pairing. We utilized oligonucleotides targeted to specific pre-mRNA sequences to mimic the effect of spliceosomal assembly at those sites. The secondary structures of the naked and hybridized pre-mRNA structures can be modeled with the use of Vienna RNA package.[107] We are currently collaborating with the group of Peter Stadler to help create secondary and tertiary structure predictions of constrained and unconstrained RNAs.

Our preliminary experiments have utilized three different oligonucleotides that target different parts of the pre-mRNA (**Table B.1**). Oligonucleotide (Oligo) 1 and 2 both bind near the site occupied by the U1 snRNA complex during the first stages of assembly. Oligo 1 was chemically synthesized with 2'-O-methyl modifications, allowing it to be made shorter than a DNA oligonucleotide because of the increased stability of 2'-O-methyl:RNA hybrids. Oligonucleotide 3 was designed to bind near the branchsite where Mud2/BBP are bound during intron recognition.

We collected smFRET data for wildtype RNAs heat annealed in the presence or absence of each oligonucleotide. Histograms showing the overall occupancy of FRET states for each condition were generated by sampling the first 10 seconds from each molecule observed (**Figure B.1**). Oligos 1 and 2 shift the FRET distribution to a dramatically lower FRET population when compared to the RNA heat annealed in the absence of any oligonucleotide. The overall distributions of Oligo 1 and 2 are very similar, which is re-assuring considering they have similar binding sites. This shift in distribution from high FRET states for the naked RNA and low FRET states with the U1 binding site occluded is similar to the shift we see for the same substrate in extracted depleted of ATP (**Chapter III**). U1 snRNP binding is the earliest intron recognition event and it occurs in an ATP independent manner. The smFRET distribution of the RNA in the presence of Oligo 3 does not change much. This may be because of incomplete saturation of the naked RNA but could also reflect a new set of structures that adopt similar FRET states.

By targeting specific sites with oligonucleotides we can change the distribution of structures sampled by the pre-mRNA much like the spliceosome. We will rely on our collaborators to provide us with potential structures for these RNAs and find ways to reconcile experimental evidence with computational predictions through additional experiments that target new sites and by modifying the modeling parameters. Before proceeding, however, it is important to verify the binding of all the oligonucleotides to the RNA. This can be done at the single molecule level by incubating the hybridized RNAs with RNase H, an enzyme that targets RNA:DNA hybrids. The oligonucleotides are designed so that RNase H cleavage of the RNA:DNA hybrid will lead to the loss of Cy3 fluorescence because the cleaved RNA is only tethered at the 5'-end where exon 1 is labeled with Cy5. A simple count of the molecules with both fluorophores before and after

RNase H incubation will give a quantitative result regarding the percent of hybridized pre-mRNAs.

Oligo 1 (2'-O-methyl modified)	5'-mCmUmUmUmAmGmAmCmAmUmAmC-3'
WT premRNA/ Oligo1	5'- GAACUAAGUGAUCUAGAAA**AG*GUAUGU***CUAAAGUUAUGGCCACGUUUC *AAAUGCGUGCUUUUUUUUUAAAACUUAUGCUCUUAUUUACUAA**C**AAAAU* *CAACAUGCUAUUGAACUA***G**AGA**U**CCACCUACUUCAUGUU-3'
Oligo 2	5'-CTTTAGACATACCTTTC-3'
WT premRNA /Oligo2	5'- GAACUAAGUGAUCUA**GAAAG*GUAUGUCUAAAG***UUAUGGCCACGUUUC *AAAUGCGUGCUUUUUUUUUAAAACUUAUGCUCUUAUUUACUAA**C**AAAAU* *CAACAUGCUAUUGAACUA***G**AGA**U**CCACCUACUUCAUGUU-3'
Oligo 3	5'-GGTAGTAAATAAGAGC-3'
WT premRNA /Oligo3	5'- GAACUAAGUGAUCUAGAAAG*GUAUGUCUAAAGUUAUGGCCACGUUUC AAAUGCGUGCUUUUUUUUUAAAACUUAU**GCUCUUAUUUACUAAC***AAAA *UCAACAUGCUAUUGAACUA***G**AGA**U**CCACCUACUUCAUGUU-3'

Table B. 1: Sequences of pre-mRNA and oligonucleotides used for hybridization. Sequences of the oligonucleotides used in this study. m=2'-O-methyl modification. Nucleotides highlighted in yellow are the position of the fluorophores. Regions highlighted in red are regions of complementarity between the WT pre-mRNA and the oligonucleotide indicated.

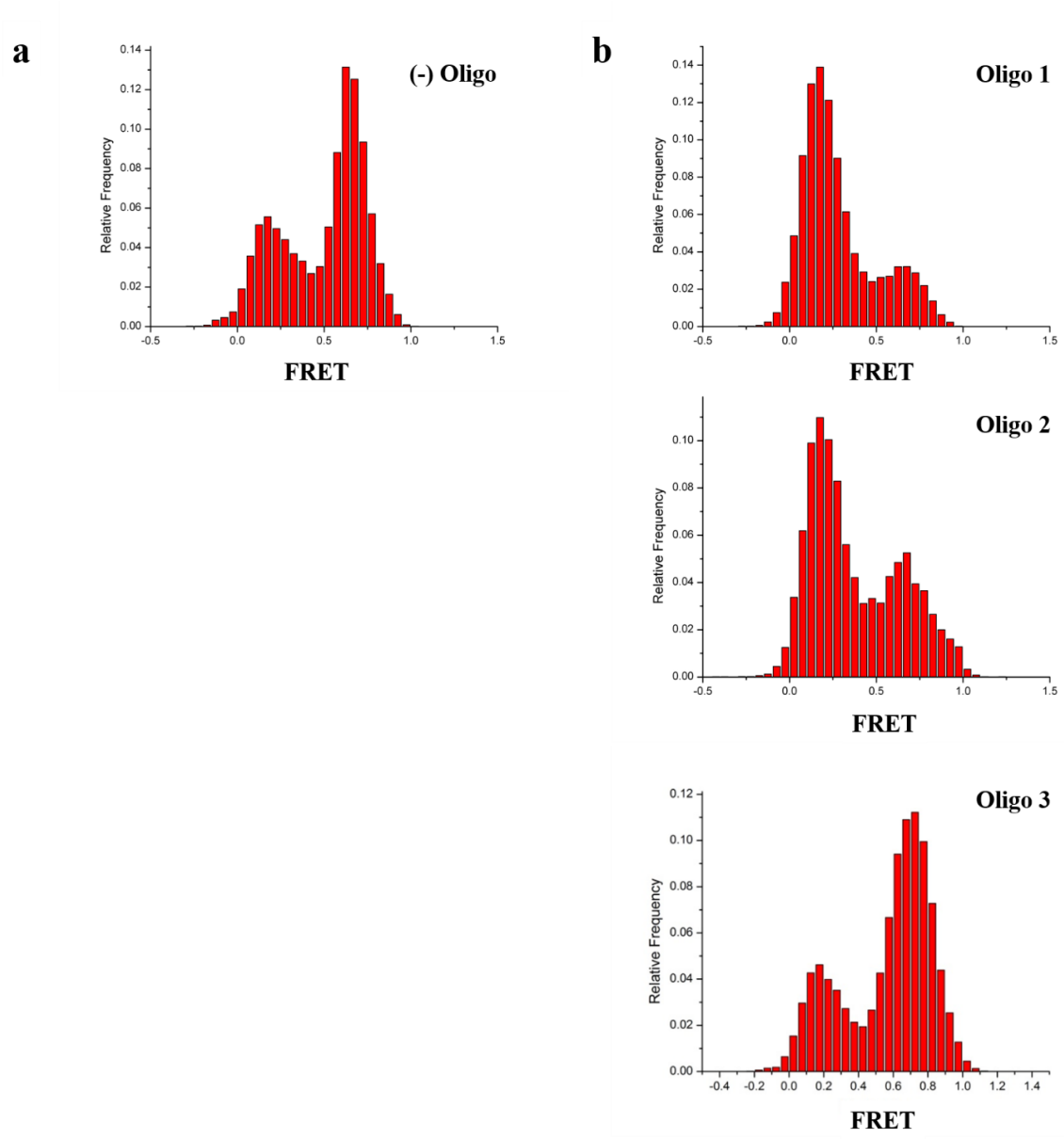


Figure B.1: FRET distribution of free and hybridized pre-mRNAs.

(a) FRET distribution of the wildtype with no hybridizided oligonucleotides. (b) FRET distributions of wildtype pre-mRNA hybridized with the oligonucleotide indicated. Oligonucleotide 1 and 2 have overlapping sites and adopt similar distributions as the unbound pre-mRNA. Oligonucleotide 3 adopts a conformation that style adopts a FRET conformation similar to the unbound pre-mRNA.

B. 2 Materials and Methods

Hybridization of pre-mRNA to oligonucleotides

The pre-mRNA and oligonucleotides were mixed at a 1:100 ratio (50nM pre-mRNA:5000nM oligonucleotide) in a 10 uL volume in T250 buffer (250mM NaCl, 50mM Tris pH 7.2, 5 mM EDTA). This was heated in an 70ºC copper bath for 90 s then cooled at room temperature for 10 min and then placed on ice for another 10 min. Before single molecule imaging this solution was diluted 200 fold in T250 buffer with OSS and flowed onto a streptavidin coated PEGylated slide.

B. 3 Acknowledgments

I would like to thank Chanrith Siv for help with collecting and analyzing the smFRET data.

References

1. Saxena, A. & Carninci, P. Long non-coding RNA modifies chromatin: epigenetic silencing by long non-coding RNAs. *Bioessays* **33,** 830–839 (2011).
2. Dinger, M. E., Mercer, T. R. & Mattick, J. S. RNAs as extracellular signaling molecules. *J Mol Endocrinol* **40,** 151–159 (2008).
3. Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136,** 701–718 (2009).
4. Kim, Y. & Kim, V. N. MicroRNA Factory: RISC Assembly from Precursor MicroRNAs. *Molecular Cell* **46,** 384–386 (2012).
5. Zappulla, D. C. & Cech, T. R. RNA as a Flexible Scaffold for Proteins: Yeast Telomerase and Beyond. *Cold Spring Harb Symp Quant Biol* **71,** 217–224 (2006).
6. Tsai, M.-C. *et al.* Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* **329,** 689–693 (2010).
7. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190,** 576–581 (1961).
8. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561–563 (1970).
9. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12,** 1–8 (1977).
10. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5’ terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* **74,** 3171–3175 (1977).
11. Gilbert, W. Why genes in pieces? *Nature* **271,** 501 (1978).
12. Fedorova, L. & Fedorov, A. Puzzles of the human genome: Why do we need our introns? *Current Genomics* **6,** 589–596 (2005).
13. Rogozin, I. B., Carmel, L., Csuros, M. & Koonin, E. V. Origin and evolution of spliceosomal introns. *Biol Direct* **7,** 11 (2012).
14. HOROWITZ, N. H. The one gene-one enzyme hypothesis. *Genetics* **33,** 612 (1948).
15. Lynch, K. W. Consequences of regulated pre-mRNA splicing in the immune system. *Nat Rev Immunol* **4,** 931–940 (2004).
16. Mattick, J. S. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* **2,** 986–991 (2001).
17. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448,** 83–86 (2007).
18. Ladewig, E., Okamura, K., Flynt, A. S., Westholm, J. O. & Lai, E. C. Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res.* **22,** 1634–1645 (2012).
19. Wang, G.-S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics* **8,** 749–761 (2007).
20. Havens, M. A., Duelli, D. M. & Hastings, M. L. Targeting RNA splicing for disease therapy. *Wiley Interdisciplinary Reviews: RNA* n/a–n/a (2013). doi:10.1002/wrna.1158
21. Brody, E. & Abelson, J. The ‘spliceosome’: yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228,** 963–967 (1985).
22. Valadkhan, S. & Jaladat, Y. The spliceosomal proteome: At the heart of the largest cellular ribonucleoprotein machine. *PROTEOMICS* **10,** 4128–4141 (2010).
23. Cheng, S. C. & Abelson, J. Spliceosome assembly in yeast. *Genes Dev.* **1,** 1014–1027 (1987).

24. Vijayraghavan, U., Company, M. & Abelson, J. Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. *Genes Dev.* **3,** 1206–1216 (1989).
25. Chen, H. & Cheng, S. Functional roles of protein splicing factors. *Bioscience Reports* **32,** 345–359 (2012).
26. Wang, Q., Zhang, L., Lynn, B. & Rymond, B. C. A BBP-Mud2p heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast. *Nucleic Acids Research* **36,** 2787–2798 (2008).
27. Kistler, A. L. & Guthrie, C. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase. *Genes Dev.* **15,** 42–49 (2001).
28. Newnham, C. M. & Query, C. C. The ATP requirement for U2 snRNP addition is linked to the pre-mRNA region 5' to the branch site. *RNA* **7,** 1298–1309 (2001).
29. Pyle, A. M. The tertiary structure of group II introns: implications for biological function and evolution. *Critical Reviews in Biochemistry and Molecular Biology* **45,** 215–232 (2010).
30. Warkocki, Z. *et al.* Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nat Struct Mol Biol* **16,** 1237–1243 (2009).
31. Koodathingal, P., Novak, T., Piccirilli, J. A. & Staley, J. P. The DEAH Box ATPases Prp16 and Prp43 Cooperate to Proofread 5′ Splice Site Cleavage during Pre-mRNA Splicing. *Molecular Cell* **39,** 385–395 (2010).
32. Tseng, C.-K., Liu, H.-L. & Cheng, S.-C. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17,** 145–154 (2011).
33. Horowitz, D. S. The splice is right: Guarantors of fidelity in pre-mRNA splicing. *RNA* **17,** 551–554 (2011).
34. Mayas, R. M., Maita, H. & Staley, J. P. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol* **13,** 482–490 (2006).
35. Tseng, C.-K. & Cheng, S.-C. Both Catalytic Steps of Nuclear Pre-mRNA Splicing Are Reversible. *Science* **320,** 1782–1784 (2008).
36. 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 Meth* **5,** 475–489 (2008).
37. Pereira, M. J. B. *et al.* Single VS Ribozyme Molecules Reveal Dynamic and Hierarchical Folding Toward Catalysis. *Journal of Molecular Biology* **382,** 496–509 (2008).
38. Zhuang, X. *et al.* Correlating Structural Dynamics and Function in Single Ribozyme Molecules. *Science* **296,** 1473–1476 (2002).
39. Chen, J., Tsai, A., O'Leary, S. E., Petrov, A. & Puglisi, J. D. Unraveling the dynamics of ribosome translocation. *Current Opinion in Structural Biology* **22,** 804–814 (2012).
40. Mihalusova, M., Wu, J. Y. & Zhuang, X. Functional importance of telomerase pseudoknot revealed by single-molecule analysis. *Proceedings of the National Academy of Sciences* **108,** 20339–20344 (2011).
41. Abelson, J. *et al.* Conformational dynamics of single pre-mRNA molecules during in vitro splicing. *Nat Struct Mol Biol* **17,** 504–512 (2010).
42. Noller, H. F., Hoffarth, V. & Zimniak, L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256,** 1416–1419 (1992).
43. Cornish, P. V., Ermolenko, D. N., Noller, H. F. & Ha, T. Spontaneous Intersubunit Rotation in Single Ribosomes. *Mol Cell* **30,** 578–588 (2008).
44. Lee, T.-H., Blanchard, S. C., Kim, H. D., Puglisi, J. D. & Chu, S. The role of fluctuations in tRNA selection by the ribosome. *Proc Natl Acad Sci U S A* **104,** 13661–13665 (2007).

45. Crawford, D. J., Hoskins, A. A., Friedman, L. J., Gelles, J. & Moore, M. J. Visualizing the splicing of single pre-mRNA molecules in whole cell extract. *RNA* **14,** 170–179 (2008).
46. Johnson-Buck, D. A. E., Blanco, D. M. R. & Walter, P. N. G. in *Encyclopedia of Biophysics* (Roberts, G. C. K.) 2329–2335 (Springer Berlin Heidelberg, 2013). at <http://link.springer.com/referenceworkentry/10.1007/978-3-642-16712-6_492>
47. Blanco, M. & Walter, N. Analysis of Complex Single Molecule FRET Time Trajectories. *Methods Enzymol* **472,** 153–178 (2010).
48. Hoskins, A. A. *et al.* Ordered and Dynamic Assembly of Single Spliceosomes. *Science* **331,** 1289–1295 (2011).
49. Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* **12,** 5–14 (2003).
50. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: review motors, clocks, springs, and things. *Cell* **92,** 315–326 (1998).
51. Burgess, S. M. & Guthrie, C. 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 (1993).
52. Couto, J. R., Tamm, J., Parker, R. & Guthrie, C. A trans-acting suppressor restores splicing of a yeast intron with a branch point mutation. *Genes & development* **1,** 445–455 (1987).
53. Xu, Y.-Z. & Query, C. C. Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Molecular cell* **28,** 838–849 (2007).
54. Furman, E. & Glitz, D. G. Purification of the spliceosome A-complex and its visualization by electron microscopy. *Journal of Biological Chemistry* **270,** 15515–15522 (1995).
55. Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. & Moore, M. J. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *Rna* **8,** 426 (2002).
56. Kent, O. A. & MacMillan, A. M. Early organization of pre-mRNA during spliceosome assembly. *Nature Structural & Molecular Biology* **9,** 576–581 (2002).
57. Newman, A. J., Teigelkamp, S. & Beggs, J. D. snRNA interactions at 5'and 3'splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *Rna* **1,** 968 (1995).
58. Pleiss, J. A., Whitworth, G. B., Bergkessel, M. & Guthrie, C. Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS biology* **5,** e90 (2007).
59. Stark, M. R., Pleiss, J. A., Deras, M., Scaringe, S. A. & Rader, S. D. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* **12,** 2014–2019 (2006).
60. Stevens, S. W. & Abelson, J. Yeast pre-mRNA splicing: Methods, mechanisms, and machinery. *Methods in enzymology* **351,** 200–220 (2002).
61. Ha, T. *et al.* Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature* **419,** 638–641 (2002).
62. Van Oijen, A. M. *et al.* Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science* **301,** 1235–1238 (2003).
63. Rasnik, I., McKinney, S. A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Methods* **3,** 891–893 (2006).
64. Munro, J. B., Altman, R. B., O'Connor, N. & Blanchard, S. C. Identification of two distinct hybrid state intermediates on the ribosome. *Molecular cell* **25,** 505–517 (2007).

65. McKinney, S. A., Joo, C. & Ha, T. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophysical Journal* **91,** 1941–1951 (2006).
66. Lin, R. J., Newman, A. J., Cheng, S.-C. & Abelson, J. Yeast mRNA splicing in vitro. *Journal of Biological Chemistry* **260,** 14780–14792 (1985).
67. Clark, T. A., Sugnet, C. W. & Ares, M. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296,** 907–910 (2002).
68. Vijayraghavan, U. *et al.* Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. *The EMBO Journal* **5,** 1683 (1986).
69. Rueda, D. *et al.* Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 10066–10071 (2004).
70. Ditzler, M. A., Rueda, D., Mo, J., H\aakansson, K. & Walter, N. G. A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic acids research* **36,** 7088–7099 (2008).
71. De Silva, C. & Walter, N. G. Leakage and slow allostery limit performance of single drug-sensing aptazyme molecules based on the hammerhead ribozyme. *RNA* **15,** 76–84 (2009).
72. Bartel, D. P. Isolation of New Ribozymes from a Large Pool of Random. *Tetrahedron Lett* **32,** 4904 (1991).
73. Schwer, B. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Molecular cell* **30,** 743–754 (2008).
74. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425,** 737–741 (2003).
75. Seraphin, B. & Rosbash, M. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59,** 349–358 (1989).
76. Stevens, S. W. *et al.* Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Molecular cell* **9,** 31–44 (2002).
77. Maroney, P. A., Romfo, C. M. & Nilsen, T. W. Functional recognition of the 5′ splice site by U4/U6. U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Molecular cell* **6,** 317–328 (2000).
78. Smith, D. J. & Konarska, M. M. Mechanistic insights from reversible splicing catalysis. *RNA* **14,** 1975–1978 (2008).
79. Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L. & Wiggins, C. H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophys J* **97,** 3196–3205 (2009).
80. McPheeters, D. S., Fabrizio, P. & Abelson, J. In vitro reconstitution of functional yeast U2 snRNPs. *Genes Dev.* **3,** 2124–2136 (1989).
81. Fabrizio, P., McPheeters, D. S. & Abelson, J. In vitro assembly of yeast U6 snRNP: a functional assay. *Genes Dev.* **3,** 2137–2150 (1989).
82. Staley, J. P. & Guthrie, C. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* **3,** 55–64 (1999).
83. D'haeseleer, P. How does gene expression clustering work? *Nat Biotech* **23,** 1499–1501 (2005).
84. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *PNAS* **95,** 14863–14868 (1998).

85. Berglund, J. A., Chua, K., Abovich, N., Reed, R. & Rosbash, M. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89,** 781–787 (1997).
86. Yang, F. *et al.* Splicing proofreading at 5′ splice sites by ATPase Prp28p. *Nucl. Acids Res.* (2013). doi:10.1093/nar/gkt149
87. Xu, Y.-Z. *et al.* Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO J* **23,** 376–385 (2004).
88. Chang, T.-H., Tung, L., Yeh, F.-L., Chen, J.-H. & Chang, S.-L. Functions of the DExD/H-box proteins in nuclear pre-mRNA splicing. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* doi:10.1016/j.bbagrm.2013.02.006
89. Fedorova, O. & Zingler, N. Group II introns: structure, folding and splicing mechanism. *Biol. Chem.* **388,** 665–678 (2007).
90. Jain, A. *et al.* Single Molecule Immuno Pull Down Assay (SiMPull) for Studying Protein-Protein Interactions. *Biophysical Journal* **96,** 25 (2009).
91. Lardelli, R. M., Thompson, J. X., Yates, J. R. & Stevens, S. W. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16,** 516–528 (2010).
92. Silverman, E. J. *et al.* Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing. *Molecular and cellular biology* **24,** 10101–10110 (2004).
93. Roy, J., Kim, K., Maddock, J. R., Anthony, J. G. & Woolford, J. L. The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. *Rna* **1,** 375–390 (1995).
94. Teigelkamp, S., McGarvey, M., Plumpton, M. & Beggs, J. D. The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. *The EMBO journal* **13,** 888 (1994).
95. Kim, S.-H., Smith, J., Claude, A. & Lin, R.-J. The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. *The EMBO journal* **11,** 2319 (1992).
96. Ohrt, T. *et al.* Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* **18,** 1244–1256 (2012).
97. Liu, H.-L. & Cheng, S.-C. The Interaction of Prp2 with a Defined Region of the Intron Is Required for the First Splicing Reaction. *Mol. Cell. Biol.* **32,** 5056–5066 (2012).
98. Will, C. L. & Lührmann, R. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* **3,** (2011).
99. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458,** 475–480 (2009).
100. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493,** 638–643 (2013).
101. Grainger, R. J. & Beggs, J. D. Prp8 protein: At the heart of the spliceosome. *RNA* **11,** 533–557 (2005).
102. Marcia, M. & Pyle, A. M. Visualizing Group II Intron Catalysis through the Stages of Splicing. *Cell* **151,** 497–507 (2012).
103. Rogic, S. *et al.* Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in Saccharomyces cerevisiae. *BMC genomics* **9,** 355 (2008).
104. Gahura, O., Hammann, C., Valentová, A., Půta, F. & Folk, P. Secondary structure is required for 3′ splice site recognition in yeast. *Nucl. Acids Res.* **39,** 9759–9767 (2011).

105. Jungmann, R. *et al.* Single-Molecule Kinetics and Super-Resolution Microscopy by Fluorescence Imaging of Transient Binding on DNA Origami. *Nano Lett.* **10,** 4756–4761 (2010).
106. Stuurman, N., Amdodaj, N. & Vale, R. Micro-Manager: Open Source software for light microscope imaging. *Microscopy Today* **15,** 42–43 (2007).
107. Hofacker, I. L. RNA secondary structure analysis using the Vienna RNA package. *Current Protocols in Bioinformatics* 12–2 (2004).