Biochemical and single molecule studies of backbone branched RNAs and lariat debranching enzyme

by

Sourav Kumar Dey

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Carnegie Mellon University

2016

To my grandmother and

in loving memory of my parents

Acknowledgements

There are many people whom I would like to thank for their support, knowledge, time, help, and care throughout my graduate studies at Carnegie Mellon.

Firstly, I would like to express my sincere gratitude to my advisors Dr. Subha R. Das and Dr. Linda A. Peteanu for their continuous support during my PhD studies. I could not have imagined having better advisors and mentors for my PhD studies. I would like to thank both of them for their patience, motivation, and immense knowledge to successfully carry on an interdisciplinary project such as mine. Because I was jointly advised by both of them, I was able to use both of their expertise in two separate fields which made my doctoral experience an exciting and enjoyable one. Their guidance helped me in all the time of research and writing of this thesis as well. Their confidence in my abilities and their high expectations of my research were critical to successfully carry on projects which involves two highly specialized disciplines. The many freedoms they gave me to pursue my curiosities have benefitted me tremendously. I believe that the experience I have gained by working in both of their labs will help me throughout my career to become a better scientist.

I would like to thank my PhD advisory committee members, Dr. Marcel Bruchez and Dr. Gordon Rule for all their suggestions and helpful discussions throughout my PhD studies. Their excellent comments and critical analysis of my projects made me a more through and better researcher. I would also like to thank the external member of my PhD committee, Dr. Alex Deiters from University of Pittsburgh for his comments and constructive criticism of my thesis.

I would like to thank all the Das lab members, past and present, for all their helpful discussions and suggestions making my experience at Carnegie Mellon a very enjoyable one. I would like to especially thank Dr. Eduardo Paredes and Molly Evans for the initial studies on lariat debranching enzyme which motivated a major part of my thesis. I would also like to thank Dr. Debasish Grahacharya for developing the photo-protected phosphoramidites and also teaching me how to make them. His knowledge and expertise in synthetic organic chemistry helped me in many different projects and I believe it will continue to help me throughout my career as a scientist. I would also like to thank Stephanie Mack who performed many different experiments with me and moved the projects where they are today. I was also very fortunate to mentor Andrea Topacio, an undergraduate student in Das lab who synthesized modified DNAs for single molecule studies. I would also like to thank Dr. John Pettersson who recently joined our labs as a postdoctoral researcher and will continue the projects that I have worked on. I also want to thank Munira Fouz and Sushil Lathwal in Das lab with whom I have worked on projects related to polymer-nucleic acid conjugates.

Next I would like to thank all the Peteanu lab members, past and present, for all their help with the microscopy experiments and scientific discussions and suggestions. I especially want to thank Eric Wu for setting up the confocal microscope, without which my single molecule experiments will be incomplete. I would also like to thank Christian Legaspi for maintaining all the computers in the Peteanu lab properly running which made all our life so much easier. I would also like to thank Woong Young So and Sikandar Abbas Wattoo for all their help with both the confocal and TIRF microscope. I was also fortunate to mentor Priyanka Nandakumar, an undergraduate in the Peteanu lab,

who helped me with development of the TIRF microscope. Finally I would like to thank all the Das and Peteanu lab members for their help with my projects and for their friendship which made my stay in Pittsburgh a very enjoyable one.

I would like to thank my collaborators who made this work possible. I would like to thank Elizabeth Ransey and Dr. Mark Macbeth for initially expressing and purifying the lariat debranching enzyme (Dbr1p). I would also like to thank our current collaborator Ryan Korbo and Dr. Andrew VanDemark for their help with Dbr1p related studies. I would also like to thank Dr. Saadyah Averick and Dr. Krzysztof Matyjaszewski for their collaboration with polymer-nucleic acid conjugate related projects. These projects helped me to expand my scientific knowledge and the breadth of my research experience. I would also like to thank Dr. Brigitte Schmidt who helped me a lot with synthesis of Alexa dyes and also other organic synthesis as well. I would also like to thank Dr. Fred Lanni for all his teaching and help with setting up and running the microscopes.

I would also like to acknowledge Dr. Rea Freeland for all her help and suggestions throughout the years which made my journey through the graduate school an easy one. I also want to express my sincere gratitude to Valerie Bridge who was always happy to help me since my admission to CMU till the very last day. Additionally I would like to thank all the other Chemistry department staff Sara, Brenda, Patsey and Tim who made my life so much easier in every aspect of my studies.

Finally I would like to thank my family back in India for all their support and enthusiasm throughout the years. Their unconditional love and support were very critical for my success throughout my studies. First I would like to thank my parents, Late

Bikash Kumar Dey and Late Kalyani Dey, who taught me the values of honesty, love, generosity and hard work in life. Then I would like to thank my grandmother Lakshmimoyee Dey and my aunt, Anjana Sarkar who dedicated a great part of their life towards my upbringing. Their love and support has always driven me towards success even in difficult times. I would especially like to thank my beautiful wife Moumita, who always motivated me to be a better researcher and always stood by my side through tough times. I am grateful for her unwavering love, encouragement and support especially in such challenging times. Finally I would like to thank my friends back in India and also in Pittsburgh, especially Husain, Matha, Dia, Sriram, Sushil, Santosh, Christian, Lisa, Saumya, Matteus, Yogi among many others who made my stay in Pittsburgh an enjoyable one. Their unconditional friendship and support helped me throughout my graduate studies and I will always cherish the happy memories of their friendship throughout my life.

Abstract

The intronic lariat RNA generated during pre-mRNA splicing includes a branch-point adenosine residue that is linked through the 2'-O position to the 5'-end of the RNA sequence. This 2'-5'-phosphodiester linkage in the lariat RNA backbone is cleaved (debranched) by the lariat debranching enzyme (Dbr1p). Following this debranching reaction, some introns can participate in highly important biological processes like snoRNA biogenesis, microRNA pathways etc. Although Dbr1p has structural resemblance to some non-specific nucleases, it still remains elusive how Dbr1p can selectively cleave the 2'-5' phosphodiester bond in lariat RNAs without affecting the 3'-5' phosphodiester bonds. The major roadblock towards understanding Dbr1p mechanism has been easy access to its substrate.

We developed a branching phosphoramidite with 3'-photoprotecing group which gives us facile access to 2'-5' phosphodiester linked backbone branched RNAs (bbRNAs) which are mimics of the lariat RNAs. Our method of bbRNA synthesis is versatile and allows incorporation of both internal and terminal modifications. To demonstrate this, several bbRNAs with modified 2'-branch were synthesized using this method which showed different debranching activity. To examine the kinetics of the debranching reaction, we synthesized a dual-fluorescently-labeled bbRNA with Cy3 and Cy5 dyes that acts as donor and acceptor in a FRET based assay. By following the increase in the donor emission with time for this dual-labeled substrate, the debranching reaction can be followed in real time. Using this assay we have been able to find the kinetics parameters of the Dbr1p enzyme for the first time. We have also synthesized a non-cleavable

analogue of the native bbRNA substrate where the 2'-5' phosphodiester bond is replaced by a triazole (click) linkage. Using the FRET based kinetics we show that this 'click' branched RNA (cbRNA) is a competitive inhibitor of Dbr1p. This suggests that the cbRNA binds the Dbr1p enzyme in a similar fashion as the native substrate.

The unique structure of 2'-5' phosphodiester linked bbRNAs could play an important role in the selective cleavage of bbRNA substrates by Dbr1p. While synthetic access to bbRNAs has enabled biochemical and kinetics studies of the debranching reaction, these studies do not inform regarding the structure and dynamics of the bbRNAs. Towards this goal, we performed single molecule FRET (smFRET) experiments using dual-labeled (Alexa488 Alexa594) bbRNA and its non-cleavable analogue, cbRNA. SmFRET experiments using confocal microscopy revealed a single broad distribution of cbRNA conformations, whereas the native bbRNAs showed two distinct populations. For both species, addition of divalent metal ions (Mn^{2+} and Mg^{2+}) induces a high FRET population indicating that the stem and the branch strands of the bbRNA are in close proximity. Additionally, conversion of the stem to a duplex gives rise to a single narrow FRET distribution meaning that both the native bbRNA and cbRNA lose conformational flexibility. Finally, using TIRF microscopy, we showed that cbRNAs undergo a conformational switch between a high and a medium FRET state. However the importance of these conformations and their fluctuations, for the debranching reaction remains to be investigated. Together this two-pronged approach of biochemical and single molecule studies will help the reaction mechanism of the elusive lariat debranching enzyme to be elucidated in the future.

Table of contents

List of Figures

List of Tables

List of Schemes

Chapter 1

Introduction to backbone branched RNA and lariat debranching enzyme

The central dogma of molecular biology proposed by Francis Crick depicted the role of RNA as carrier of genetic information which is then translated into proteins.[1,2] The proteins were thought to be the main workhorse of cell carrying out all essential functions. However over the years it has become more and more evident that apart from translating genetic information, RNA carries out a broad range of important cellular functions. The total number of functional genes in humans has been identified to be around 25000, which comprises only 1.5% of the entire human genome that is actually translated into proteins.[3] The rest of the genome consists of non-coding RNAs (nc-RNA), regulatory sequences, introns and non-coding DNAs. The non-coding RNA sequences produce transcripts that function as structural, regulatory and catalytic elements, rather than producing protein sequences.[4,5] Some RNAs, despite of being consisting of the similar four nucleotides, can form distinct three dimensional structures which give them the capability of acting in a similar manner to proteins or enzymes.[6] The discovery of pre-messenger RNA (pre-mRNA) splicing was one of the very first steps towards understanding complexity of life in terms of RNA. Sharp's and Robert's initial finding that large non-coding RNA sequences or introns interrupt the coding sequences or exons in genes led to the discovery of alternative splicing. This process can result in one gene producing more than one messenger-RNA (mRNA) for protein synthesis which can account for some level of complexity in higher organisms.[7,8] The main focus of this thesis

work is backbone branched RNA which is an intermediate and also a final product during the process of pre-mRNA splicing.

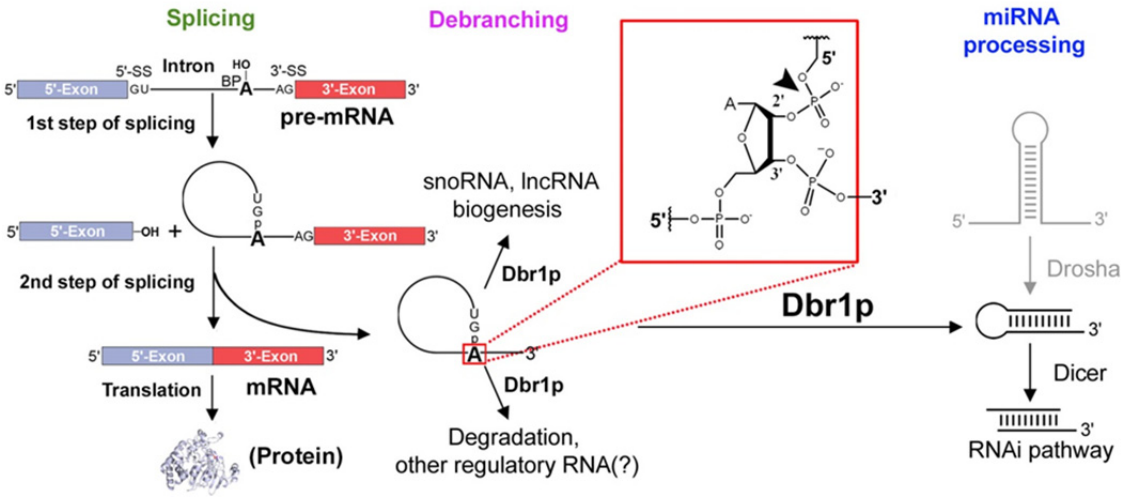


Figure 1.1. Splicing, debranching and microRNA (miRNA) pathways. The coding sequences (exons) in pre-mRNA are joined together through two phospho-transesterification steps in splicing. The noncoding sequences (introns) are removed as a 'lariat' RNA. This 2'-5'-branched lariat RNA is then debranched by the lariat debranching enzyme (Dbr1p) before degradation or incorporation into other regulatory pathways. Some debranched introns are cleaved by Dicer to generate short double stranded RNAs which can take part in RNA interference (RNAi). Figure adapted from ref. 21.

1.1. Backbone branched RNAs are produced during pre-mRNA splicing

During the process of pre-mRNA splicing an intron that is flanked by two exons, is removed from the pre-mRNA (Fig. 1.1) and the two exons are joined to form the mature mRNA. This removal and ligation is mediated by the macromolecular spliceosomal apparatus that consists of five small nuclear RNAs (snRNAs) and more than hundred proteins. Together with its RNA and protein components, the spliceosome is a highly efficient and dynamic ribonucleoprotein machine.[9–11] Previous biochemical studies and recent outburst of electron microscopy data of the spliceosome structure shows that the

active site of the spliceosome consist of only RNA suggesting that it is a highly efficient ribozyme.[12,13] Key locations within the pre-mRNA substrate are indicated in Figure 1.1. The 5'-splice site (5'SS) exists at the exon-intron junction at the 5' end of the intron, while the 3'-splice site (3'SS) exists at its 3'-end.

The branchpoint (BP) includes a crucial adenosine (A) residue, and a conserved sequence towards the 3'-end of the intron. The first step of the splicing reaction is a nucleophilic attack by the 2'-hydroxyl (OH) group of the BP adenosine on the phosphate group at the 5'SS, which leads to the formation of a lariat intermediate containing a 2'-5' phosphodiester (Fig. 1.1, inset) bond and a free 5'-exon. This is followed by the second step, nucleophilic attack on the phosphate group at the 3'SS by the 3'-OH group of the cleaved 5'-exon. As a result, the 5'SS and the 3'SS are ligated, generating a lariat intron byproduct.

1.2. Lariat RNA and the lariat debranching enzyme in regulatory process

The lariat RNA is a RNA that in addition to the normal 3'-5'-phosphodiester linked backbone, includes a branch-point (BP) adenosine residue that is linked at the 2'-position to the 5'-end of the RNA sequence (Fig. 1.1, inset).[14] This unique feature of the lariat RNA protects it from degradation by cellular 5'-exoribonucleases. The 2'-5'-phosphodiester linkage in the lariat RNA backbone is cleaved by a unique lariat debranching enzyme (Dbr1p) after its release from the spliceosome (Fig. 1.2).[15,16] Debranching of the lariat intron by Dbr1p converts the lariat into a linear RNA with 5'-phosphate (Fig. 1.2). This initiates the degradation of most of the intronic RNA by cellular exonucleases.[17] While deletion of the DBR1 gene causes severe growth defects in

S. pombe, deletion of the gene in higher organism is lethal.[18,19] This is presumably due to the higher number of intronic RNAs present in higher eukaryotes. Although the turnover of intronic lariat RNAs has not been studied directly, studies of the lariat intermediate elucidated the pathways of branched RNA turnover in cell.[17,20] Following debranching, the linearized lariat intermediates are degraded by the cytoplasmic 5'-3' exoribonuclease Xrn1 (Rat1 in budding yeast). This was quite a surprising finding as the lariat intermediates are formed in the nucleus of the cell and Xrn1 is a cytoplasmic nuclease. Other turnover pathways for degradation of the lariat intron have also been discovered, although key components are yet not fully identified. Some lariat introns have also been identified which evade the canonical degradation pathway and have more stability than the regular mRNA transcripts.[17]

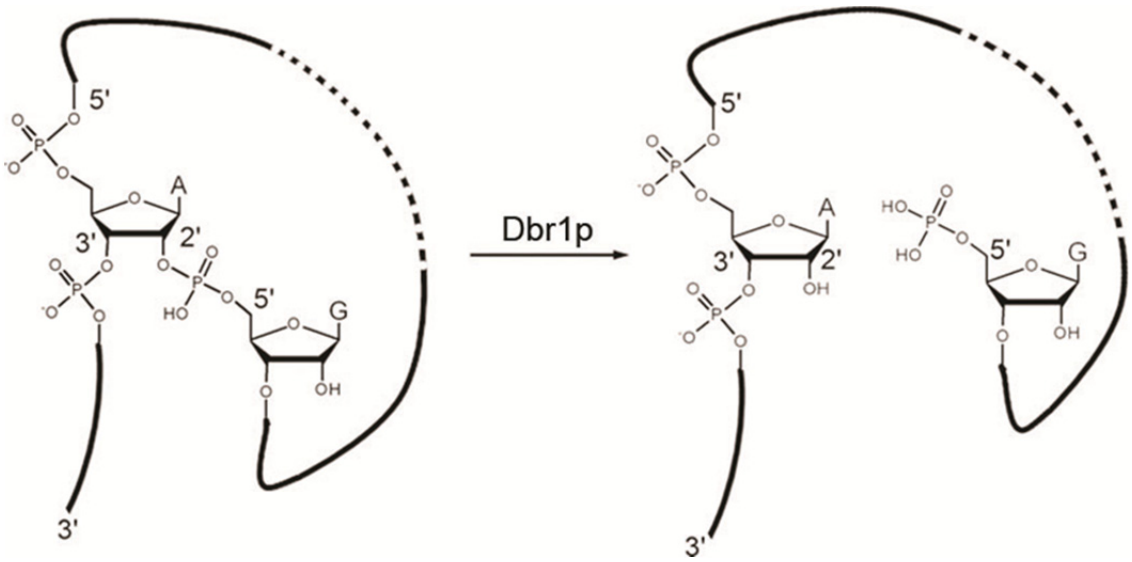


Figure 1.2. The debranching reaction. Dbr1p selectively cleaves the 2'-5' phosphodiester bond of the lariat intron without affecting the 3'-5' phosphodiester bond making the intron linear. The debranching reaction generates a 5'-phosphate and a 2'-OH at the branchpoint. The dotted line indicates that Dbr1p does not require the intact lariat for cleavage of the 2'-5' phosphodiester bond. A backbone branched RNA (bbRNA) with a 2'-5' phosphodiester bond where the loop is absent is still a substrate for Dbr1p. The dashed line shows that even in absence of closed loop, Dbr1p can still cleave the bbRNAs.

Although most of the intronic RNAs do not have any cellular function, some lariat introns encode regulatory RNA sequences within them (Fig. 1.1). Following debranching, some lariat introns are shown to be further processed into key regulatory RNAs such as small nucleolar RNAs (snoRNAs), long non-coding RNAs (lncRNAs) and most recently into microRNAs (miRNAs).[21] Dbr1p is essential for the processing of box C/D snoRNAs in yeast (*S. Cerevisiae*).[22] It has been shown that, in humans, the intron of the UGH gene encodes for eight snoRNAs while the exon has no known function.[23] Dbr1p is also implemented in retrotransposition.[24,25] The DBR1 gene was first discovered during a genetic screen to identify mutants defective in Ty1 retrotransposition.[16] Although the exact role of Dbr1p in retrotransposition remains elusive, it is an essential component for efficient retrotransposition. Consistent with this idea, Dbr1p has also been shown to be essential for replication of HIV-1 retrovirus in human cells.[26] Retroviruses and retrotransposons are though to share a common ancestral origin because of their common genetic makeup and replication mechanism. Therefore understanding how Dbr1p is involved into these processes will improve our understanding of the retrovirus life cycle. Loss of Dbr1p was also shown to reduce the toxic effect of the RNA binding protein TDP-43 in yeast model of amyotropic lateral sclerosis (ALS).[27] The authors suggested that the pool of the lariat intron accumulated due to lack of debranching, sequester the toxic TDP-43 protein preventing it from aggregating.

Lariat introns generated and released in the process of splicing have recently been associated with RNA interference (RNAi) pathways.[28,29] RNAi and microRNAs (miRNAs) are among the most prominent discoveries in gene regulation in recent times.[30,31] The regulatory microRNAs (miRNAs) are derived from longer precursor

RNAs after the cleavage by endonucleases like Drosha that generate the pre-miRNA stem loop structure (Fig. 1.1). Such stem- loop-structured RNAs are further cleaved by another endonuclease, Dicer, that generates short (21-22 mer) double stranded RNAs which then participate in the RNAi process. Although most miRNAs derive from non-coding genes, about one third of them are mirtrons (miRNAs from introns) coming from the introns of mRNA-coding genes.[28,29,32] These mirtron sequences are located as isolated sequences flanked by exons (Fig. 1.1) and thus require efficient splicing for their excision from the main transcript. Once spliced and debranched, these intronic RNAs form hairpin structures with 3'-overhang similar to those generated by Drosha (Fig. 1.1). Therefore, these pre-miRNA transcripts require Dbr1p activity but do not require Drosha activity for maturation. Following debranching, mirtron derived pre-miRNAs are exported and processed like traditional miRNAs. All these regulatory processes described here require the debranching of the lariat intron by cleavage of the 2'-5' phosphodiester bond for further processing of the intronic RNA. Therefore understanding how Dbr1p selectively cleaves the 2'-5' phosphodiester bond is essential to improve our knowledge of these highly important biological pathways and also to exploit these pathways for therapeutic intervention.

1.3. Biochemical and structural properties of Dbr1p

Dbr1p is a Mn^{2+} dependent phosphodiesterase that selectively cleaves the 2'-5' phosphodiester bond without affecting the 3'-5' phosphodiester bond.[15,33,34] Dbr1p also hydrolyzes the 2′-5′ phosphodiester linkage found in Y shaped branched RNA and multicopy single-stranded DNA (msDNA).[33,35] Hydrolysis of the branchpoint by Dbr1p yields a 5'-phosphate end and a 2'-OH at the branch attachment site (Fig. 1.2). Dbr1p is

well conserved among eukaryotes; the *S. cerevisiae* null DBR1 mutant phenotype can be alleviated by expression of *S. pombe*, *C. elegans*, human, and mouse Dbr1p.[18,36,37] Dbr1p has significant sequence homology to other metallo-phosphodiesterases such as λ phosphoprotein phosphatase, calcineurin and Mre11 that belong to the metallophosphoesterase superfamily.[34,38] Mre11 is a nuclear protein involved in homologous recombination, telomere length maintenance, and DNA double-strand break repair.[39] It has 3'-5' exonuclease and endonuclease activity.

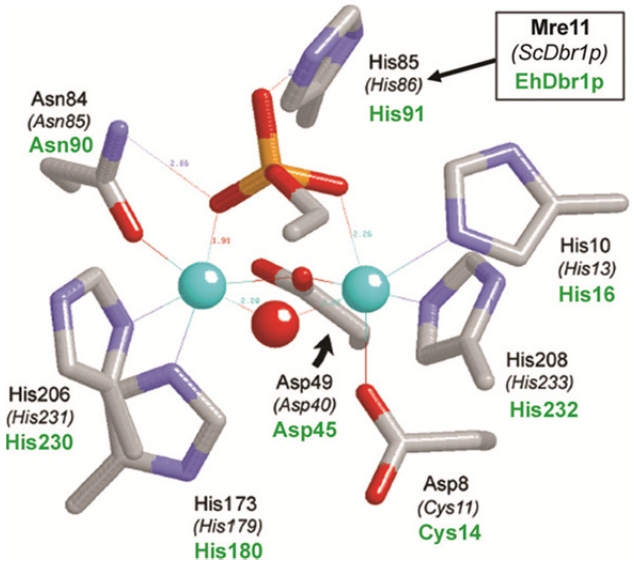


Figure 1.3. Proposed active site similarities between Mre11 and Dbr1p. The active site of the Pyrococcus furiosus Mre11 with bound Mn^{2+} and 5'-dAMP is shown here (PDB 1II7). The amino acid side chains which coordinate the two catalytically important Mn^{2+} ions are also conserved in Dbr1p except the Asp8 which is a Cys in Dbr1p. The corresponding amino acids of yeast Dbr1p (ScDbr1p) are indicated in parentheses and the same for *Entamoeba histolytica* Dbr1p (EhDbr1p) are shown in green. The manganese ions are cyan spheres and water is the red sphere. Figure modified from ref. 33.

Structural studies of the Mre11 superfamily show that conserved His, Asp, and Asn side chains of the signature motif bind the two metal ions while conserved His and Asn side chains of the signature motif contact the phosphate backbone of the substrate (Fig. 1.3).[40–42] Debranching enzymes from *S. cerevisiae*, *E. histolytica* and other species

show similar conservation of this motif suggesting that the structure of the active site and the coordination of the manganese ions may be similar in Dbr1p.[34] Figure 1.3 shows the active site similarities between Mre11 and two different Dbr1p variants. For metal-ion coordination, six of the seven corresponding amino acids in Dbr1p (His13, Asp40, Asn85, His179, His231, and His223 for ScDbr1p) are conserved (Fig. 1.3) and essential for debranching activity.[34,43] It has also been proposed that ScDbr1p residues Asn85 and His86 along with the manganese ions coordinate the 2'-5'-scissile phosphate similar to Asn84, His86 and manganese ion coordination of the 3'-5'-scissile phosphate in Mre11. This proposed coordination is supported by the fact that Asn85 and His86 are intolerant of conservative substitutions.[34,43] Additional support is the lack of catalytic activity of the R_P 2'-5'-phosphorothioate substrate; the larger sulfur atom would disrupt coordination by Asn85 and manganese ion on the left (Fig. 1.3), leading to non-optimal enzyme-substrate interactions and thus loss of catalytic activity.[44]

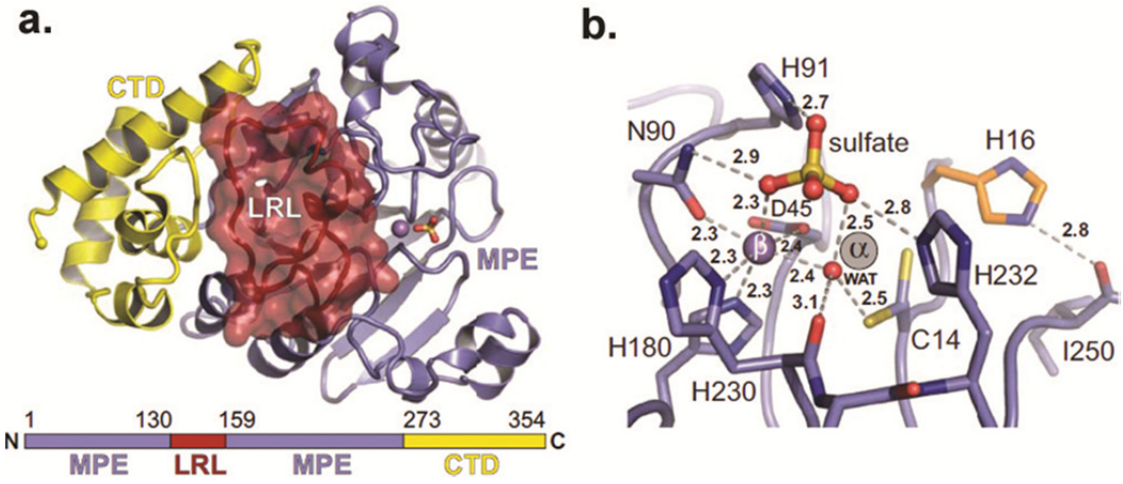


Figure 1.4. Crystal structure of EhDbr1p (PDB 4PEF). **a**. EhDbr1p contains three major domains; MPE domain containing the active site (blue), a LRL domain for lariat recognition and a CTD which has important contacts with the LRL domain. **b.** The active site of EhDbr1p with a sulfate anion shows the conserved amino acid residues described in Fig. 1.3. Although the β pocket contained a Mn^{2+} ion coordinated with the conserved amino acids, the α pocket was devoid of any metal ion. Figure adapted from ref. 42.

The crystal structure of Dbr1p from *Entamoeba histolytica* (EhDbr1p) was published recently by Montemayor *et al.*[43] The structure of Dbr1p can be divided into three major domains; a metalophosphoesterase (MPE) domain, a lariat recognition loop (LRL), and a C-terminal domain (CTD) (Fig. 1.4a). The N-terminus of EhDbr1p forms the MPE fold with a pocket for binding two metal ions (α and β). Surprisingly in this study the authors were unable to see any metal ion in the α pocket. The β pocket however had a Mn^{2+} ion which was coordinated to the conserved Asp45, Asn90, His180, His230 residues, a water molecule and a sulfate ion (Fig. 1.4b). The MPE domain is interrupted by a stretch of 30 amino acids which forms a loop positioned near the active site (Figs. 1.4a and 1.4b). This loop element referred as the lariat recognition loop (LRL), interacts strongly with the CTD through an extensive network of hydrogen bonding. In this study, three analogue of the branched RNA substrate was co-crystalized with the Dbr1p. One of the substrate analog containing a 2'-5' phosphorothioate linkage and two nucleotides in the branch closely resembles the branched RNA substrate. However the phosphorothioate linkage in this analogue did not bind to the active site; instead, a 3'-5' phosphate linkage showed contact with the β metal ion in the co-crystal structure. Although the active site and several important contacts between the substrate analogue and Dbr1p were identified in this study, the exact nature of interaction between the Dbr1p and branched RNA substrate required for debranching still remains elusive.

1.4. Scope of this thesis work

Although the Dbr1p activity was initially identified more than thirty years ago for efficient intron turnover, additional roles of Dbr1p in different regulatory processes were identified over the years.[17] Central to all these is the cleavage of 2'-5' phosphodiester

bond in the lariat intron. The central question related to Dbr1p is; despite having structural similarities with nonspecific nucleases (Mre11), how Dbr1p can selectively recognize the 2'-5' phosphodiester bond and cleave it without affecting the 3'-5' phosphodiester bonds. To answer such a question one requires detailed biochemical and kinetic analysis of Dbr1p. Although Dbr1p was biochemically characterized to some extent, the branched RNA substrates used for those studies were non-native.[33,34] In some studies the substrates were RNA-DNA hybrid, whereas in others the substrates had same sequence in the 2'- and the 3'-arm of the branched RNA. Therefore synthesis of natural sequences is essential for the biochemical studies of Dbr1p. With such lack of easy access to 2'-5' linked branched RNA substrates; part of this thesis work will focus on synthesis of native substrates for Dbr1p using photoprotected phosphoramidites (chapter 2). These 2'-5' phosphodiester linked and backbone branched RNA substrates will be referred as backbone branched RNA (bbRNA) for the rest of this thesis work. Additionally, using the developed protocol, several substrates with modified 2'-branch nucleotide were synthesized and biochemically characterized (chapter 3). Along with the biochemical characterization, the structure-function relationship for an enzyme can only be established with stuudies of the enzyme's binding and cleavage kinetics. So far there has been no report of the kinetics parameters (k_{cat} and K_m) of Dbr1p. Therefore we have developed a Foster Resonance Energy Transfer (FRET) based real time assay for the analysis of debranching reaction using a dual-fluorescently labeled bbRNA substrate. We were able to determine the kinetics parameters of three different Dbr1p variants using this method (chapter 4). Additionally, using this kinetics we have also identified a non-cleavable substrate analogue which is a competitive inhibitor of Dbr1p.

Although the crystal structure of Dbrp1p is recently reported, there is very little information available in the literature regarding the structure of branched RNA substrates and its role in recognition by Dbr1p enzyme. To gain insight into the structure of the 2'-5' linked branched RNAs, we turned to single molecule spectroscopy. Unlike bulk experiments, single molecule experiments are ideal to identify multiple conformations in a complex system. Single molecule experiments can also be used to observe the dynamics of individual molecules without the necessity to synchronize them. We have used dual-fluorescently labeled branched RNAs and single molecule FRET (smFRET) to identify conformations of both native and non-native branched RNA using confocal microscopy (chapter 7). Additionally, we have also looked into the conformational changes of these branched RNAs under different physiological condition. Although confocal microscopy with freely diffusing molecules can give us information about RNA conformation, it cannot give us any idea about the dynamics of the bbRNAs. Therefore I have set up a TIRF microscope where the bbRNAs can be immobilized on surface and can be observed for longer time (in the order of minutes). Finally I have looked into the dynamics of these branched RNA using smFRET under different physiological conditions using the TIRF microscope.

The development of the branched RNA synthesis methodologies and the kinetics studies will ultimately allow us to establish a structure-function relationship between the Dbr1p enzyme and the bbRNA substrates. Such studies will be of utmost importance to understanding how Dbr1p can specifically cleave the 2'-5' phosphodiester bond. The single molecule studies will help us to identify conformations of bbRNAs which are important for binding with Dbr1p and how these change under different physiological

conditions. The study of the dynamics of bbRNAs before and after binding with Dbrp1p will also shed light on to the binding pathways. This two pronged approach of biochemical and single-molecule studies will ultimately enable us to uncover the reaction mechanism of the lariat debranching enzyme.

References

1. Crick, F. On Protein Synthesis. in *The Symposia of the Society for Experimental Biology* **XII,** 138–166 (1958).

2. Crick, F. H. C. Central Dogma of Molecular Biology. *Nature* **227,** 561–563 (1970).

3. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2001).

4. Mattick, J. S. A new paradigm for developmental biology. *J. Exp. Biol.* **210,** 1526–1547 (2007).

5. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell* **157,** 77–94 (2014).

6. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* **8,** 776–790 (2007).

7. Sharp, P. A. Split Genes and RNA Splicing Nobel Lecture. *Cell* **77,** 805–815 (1994).

8. Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **84,** 1–33 (2015).

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

10. Papasaikas, P. & Valcárcel, J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends in Biochemical Sciences* **41,** 33–45 (2016).

11. Cate, B. J. H. D. A Big Bang in spliceosome structural biology. *Science* **351,** 1390–1392 (2015).

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

13. Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. *Science* **349,** 1191–8 (2015).

14. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **38,** 317–331 (1984).

15. Ruskins, B. & Green, M. R. An RNA Processing Activity that Debranches RNA Lariats. *Science* **229,** 135–140 (1985).

16. Chapman, K. B. & Boeke, J. D. Isolation and Characterization Encoding Yeast Debranching of the Gene Enzyme. *Cell* **65,** 483–492 (1991).

17. Hesselberth, J. R. Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA* **4,** 677–691 (2013).

18. Nam, K., Lee, G., Trambley, J. & Devine, S. E. Severe Growth Defect in a Schizosaccharomyces pombe Mutant Defective in Intron Lariat Degradation. *Mol. Cell. Biol.* **17,** 809–818 (1997).

19. Wang, H., Hill, K. & Perry, S. E. An Arabidopsis RNA Lariat Debranching Enzyme Is Essential for Embryogenesis. *J. Biol. Chem.* **279,** 1468–1473 (2004).

20. Hilleren, P. J. & Parker, R. Cytoplasmic Degradation of Splice-Defective Pre-mRNAs and Intermediates. *Mol. Cell* **12,** 1453–1465 (2003).

21. Dey, S. K., Paredes, E., Evans, M. & Das, S. R. in *From Nucleic Acids Sequences to Molecular Medicine* (eds. Erdmann, A. V. & Barciszewski, J.) 475–501 (Springer Berlin Heidelberg, 2012). doi:10.1007/978-3-642-27426-8_19

22. Ooi, S. L., Samarsky, D. A., Fournier, M. J. & Boeke, J. D. Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme : intron length effects and activity of a precursor snoRNA. *RNA* **4,** 1096–1110 (1998).

23. Tycowski, K. T., Shu, M.-D. & Steitz, J. A. A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379,** 464–466 (1996).

24. Karst, S. M., Ru, M. & Menees, T. M. The Yeast Retrotransposons Ty1 and Ty3 Require the RNA Lariat Debranching Enzyme , Dbr1p , for Efficient Accumulation of Reverse Transcripts. *Biochem. Biophys. Res. Commun.* **268,** 112–117 (2000).

25. Cheng, Z. & Menees, T. M. RNA Branching and Debranching in the Yeast Retrovirus-like Element Ty1. *Science* **303,** 240–243 (2004).

26. Ye, Y., De Leon, J., Yokoyama, N., Naidu, Y. & Camerini, D. DBR1 siRNA inhibition of HIV-1 replication. *Retrovirology* **2,** 63 (2005).

27. Armakola, M. *et al.* Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat. Genet.* **44,** 1302–1309 (2012).

28. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448,** 83–86 (2007).

29. Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. *Cell* **130,** 89–100 (2007).

30. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136,** 642–655 (2009).

31. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16,** 421–433 (2015).

32. Lin, S., Miller, J. D. & Ying, S. Intronic MicroRNA (miRNA). *J. Biomed. Biotechnol.* **2006,** 1–13 (2006).

33. Nam, K. *et al.* Yeast Lariat Debranching Enzyme: Substrate and sequence specificity. *J. Biol. Chem.* **269,** 20613–20621 (1994).

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

35. Sutton, R. E. & Boothroyd, J. C. Evidence for Trans Splicing in Trypanosomes. *Cell* **47,** 527–535 (1986).

36. Hyung-cheol, K., Kim, G., Yang, J. & Kim, J. W. Cloning, Expression, and Complementation Test if the RNA LAriat Debranching Enzyme cDNA from Mouse. *Mol. Cells* **11,** 198–203 (2000).

37. Kim, J. W. *et al.* Human RNA lariat debranching enzyme cDNA complements the phenotypes of Saccharomyces cerevisiae dbr1 and Schizosaccharomyces pombe dbr1 mutants. *Nucleic Acids Res.* **28,** 3666–3673 (2000).

38. Koonin, E. V. Conserved sequence pattern in a wide variety of phosphoesterases. *Protein Sci.* **3,** 356–358 (1994).

39. Stracker, T. H. & Petrini, J. H. J. The MRE11 complex: starting from the ends. *Nat. Rev. Mol. Cell Biol.* **12,** 90–103 (2011).

40. Hopfner, K. P. *et al.* Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105,** 473–485 (2001).

41. Kissinger, C. R. *et al.* Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* **378,** 641–644 (1995).

42. Goldberg, J. *et al.* Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* **376,** 745–53 (1995).

43. Montemayor, E. J. *et al.* Structural basis of lariat RNA recognition by the intron debranching enzyme Dbr1. *Nucleic Acids Res.* **42,** 10845–10855 (2014).

44. Mourani, R. & Damha, M. J. Synthesis, characterization, and biological properties of small branched RNA fragments containing chiral (Rp and Sp) 2′,5′-phosphorothioate linkages. *Nucleosides, Nucleotides and Nucleic Acids* **25,** 203–229 (2006).

Chapter 2

Solid phase synthesis of backbone branched RNAs

2.1. Introduction

The 2'-5' phosphodiester linked backbone branched RNAs (bbRNAs) occur in nature in lariat RNAs produced during the process of pre-mRNA splicing. The intronic RNA excised as a lariat shows greater nuclease resistance compared to linear 3'-5' linked RNAs.[1–3] This nuclease resistant component was shown to be the branchpoint (BP) adenosine containing a novel 2'-5' phosphodiester bond. The 2'-5' phosphodiester linkage is formed during the first step of splicing when the 2'OH group of the conserved branchpoint (BP) adenosine attacks the 5' splice site (5'SS). Soon after the discovery of the 2'-5' phosphodiester linkage in lariat RNAs, a debranching activity was also discovered in HeLa cell extract which specifically cleaves the 2'-5' phosphodiester linkage in lariat RNAs.[4] Since then there have been numerous efforts to synthesize 2'-5' phosphodiester linked bbRNAs to study splicing and debranching.[5] The initial studies of RNA debranching were performed using lariat RNAs and multicopy single-stranded DNAs (msDNAs, a type of DNA-RNA hybrid that consists of a single-stranded DNA molecule covalently linked via a 2'-5' phosphodiester bond to the guanosine of an RNA sequence) extracted from cells.[6] For additional studies, it was very important to develop a more robust method which allowed incorporation of chemical and nucleobase modifications. Therefore a lot of efforts were put to develop chemical synthesis of 2'-5' phosphodiester linked bbRNAs. The very first chemical synthesis of 2'-5' phosphodiester linked RNA was described by Damha and Ogilvie where they synthesized a branched

trinucleotide using solution phase synthesis.[7] Since then there have been numerous efforts towards synthesis of bbRNAs using both solution phase and solid phase chemistry. However the solid phase synthesis methods became more popular as they are automated and also require minimum purification. Before describing solid phase synthesis of bbRNAs, it is important to understand solid phase synthesis of linear RNAs. Figure 2.1 describes the solid phase synthesis of linear RNAs using phosphoramidite chemistry.

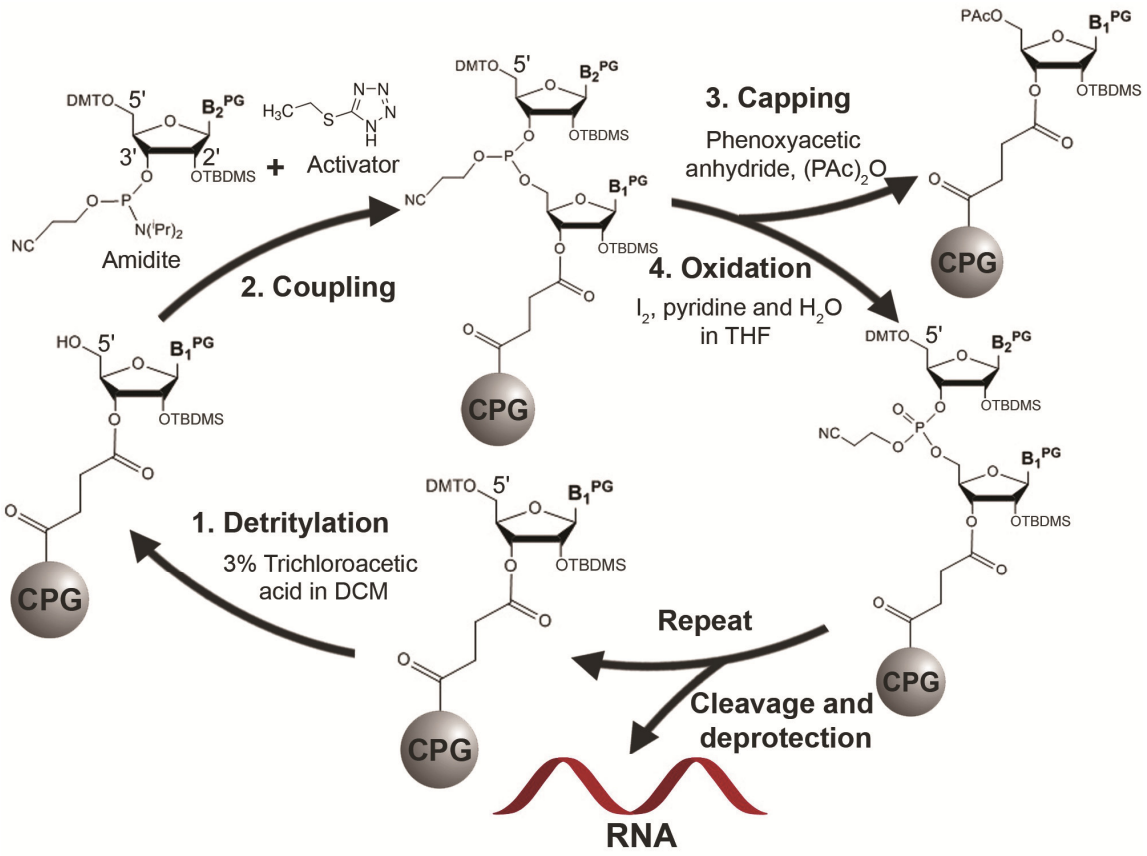


Figure 2.1. Solid phase synthesis of linear RNAs using phosophoramidite chemistry.

During the solid phase automated synthesis of RNA, the sequence is synthesized in 3' to 5' direction (Fig. 2.1). The synthesis of a linear RNA starts from the first 3'-nucleotide (B_1) attached to the controlled pore glass (CPG) solid support. The 5'-DMT protecting group of the first nucleotide is deblocked using trichloroacetic acid to generate

a 5'-OH group. Then the next monomer amidite (B_2) is introduced which contain a 3'-phosphoramidite along with 2'-O-tert-Butyldimethylsilyl (2'-*O*-TBDMS) and 5'-*O*-DMT protecting group. The 5'-OH group of the deblocked first base is coupled to the 3'-phosphoramidite group of the next base which is activated using the 5-ethylthio tetrazole (ETT) activator to generate a phosphite. As the efficiency of the phosphoramidite coupling is not 100%, a portion of the first nucleotide (B_1) remains unreacted. The 5'-OH of this unreacted first nucleotide is then capped using phenoxyacetic anhydride to stop any growth from this nucleotide during the next round of phosphoramidite coupling. Following this the phosphite linkage is oxidized from P(III) to P(V) using a mixture of iodine, water and pyridine in THF to generate a phosphotriester. The oxidation step completes one full round of synthesis for incorporation the second base (B_2). This cycle is repeated as many times as required using the required phosophoramidites of the different nucleotides as dictated by the sequence of the RNA. As the coupling yield in each step is less than 100%, the yield of the synthesized RNA decreases with increase in the length of the RNA. Once the desired sequence is synthesized the CPG beads are taken outside the synthesizer for cleavage and deprotection. First the CPG beads are treated with a basic solution (usually a mixture of ammonia and methylamine) which cleaves the RNA strands from the solid support and also removes the nucleobase protecting groups and cyanoethyl protecting groups from the phosphate backbone. Then RNAs are treated with a fluoride reagent (usually triethylamine trihydrofluoride, TEA.3HF) to selectively remove the 2'-*O*-TBDMS protecting groups generating functional RNAs with 2'-OH groups. This solid phase synthesis of the linear RNAs using 2'-*O*-TBDMS protected monomers is a very important intermediate step for solid phase synthesis of bbRNAs and

will be described in detail in later part of this chapter. The following section describes previous efforts towards solid phase synthesis of bbRNAs using both solid phase and solution phase methods.

2.2. Previous efforts towards synthesis of bbRNAs and lariats

2.2.1. Solution phase synthesis of bbRNAs

The initial methods developed for bbRNA synthesis were based on solution phase synthesis which involved non-standard protecting groups and reagents. The first report of branched trinucleotide synthesis by Damha and Ogilvie used methyl phosphoramidites and also resulted in sequence where both the 2'- and 3'-nucleotides were the same.[7] Following this report another method of branched tetranucleotide synthesis was described by Caruthers and coworkers which utilized 2'-*O*-TBDMS protected phosphoramidites.[8] Although this method resulted in bbRNA sequences which are naturally occurring, the solution phase synthesis required multiple steps even for short oligonucleotides. Another method of branched tri- and tetra-nucleotide synthesis was developed by Chottopadhaya *et al.* using phosphotriester chemistry in solution.[9,10] Since then there have been other efforts towards solution phase synthesis of bbRNAs using phosphotriester, phosphoramidite and H-phosphonate chemistry.[11–16] These solution phased based methods have also been used for the synthesis of lariat RNAs with tri- or penta-nucleotide loops.[17–19] Although the synthesis of the small branched oligonucleotides using solution phase chemistry was novel, longer sequences were required for more detailed biochemical studies of bbRNAs. Such sequences were not accessible using

solution phase synthesis because they required a large number of synthesis and purification steps.

2.2.2. Biochemical synthesis of bbRNAs and lariats

Lariat RNAs containing 2'-5' phosphodiester bond is produced during splicing and they has been isolated from cellular extract for biochemical studies.[4,6] However synthetic access to bbRNAs has been limited due to difficulties associated with their synthesis. Therefore researchers have sought to other routes for the synthesis of bbRNAs. Wang and Sileverman identified a 40 nucleotide long deoxyribozyme (DNA enzyme) using in vitro selection which can synthesize bbRNAs and lariats.[20,21] The deoxyribozyme can join the internal 2'-OH of an RNA strand with the 5'-terminal triphosphate of another strand in presence of Mn^{2+} to generate a 2'-5' phosphodiester linked bbRNA (Fig. 2.2a). The lariat RNA can be synthesized by the same deoxyribozyme when the internal 2'-OH and the 5'-triphosphate came from the same RNA strands (Fig. 2.2b). Following this work, Pratico and Silverman also reported identification of another deoxyribozyme which can tolerate all four nucleotide at the branchpoint.[22,23] Recently the same group has also reported identification of another set of deoxyribozymes with better metal ion and pH requirements.[24] Although the synthesis of bbRNAs and lariat RNAs using the deoxyribozyme is a useful method, it has not been used extensively to prepare bbRNAs for biochemical studies due to certain restrictions on the bbRNA sequence around the branchpoint. Therefore a better method of bbRNA synthesis is required which can tolerate different modifications.

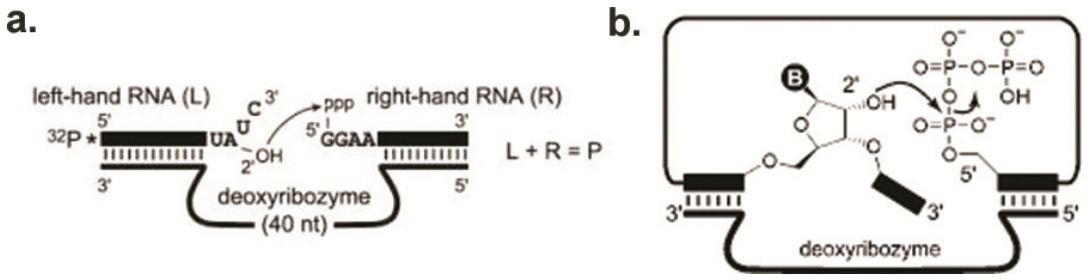


Figure 2.2. Synthesis of bbRNAs and lariats using deoxyribozyme. **a.** Sythesis of 2'-5' phosphodiester linked bbRNA using deoxyribozyme. **b.** Synthesis of lariat RNA using deoxyribozme. This figure is adapted from ref 20 and 23.

Another strategy for synthesis of lariat RNAs was demonstrated by Damha *et al.* using chemical ligation.[25,26] In this strategy a template RNA strand is used to bring the 2'-phosphate and 5'-OH of a reacting RNA strand in the desired 'lasso-like' configuration. Then the terminal OH group and the 2'-phosphate of the branchpoint nucleotide are joined together using cyanogen bromide mediated chemical ligation producing a lariat RNA. Although these biochemical methods are used for the synthesis of different bbRNAs and lariats, some of them are not amenable to chemical modifications such as modified nucleotides and other functional groups. Therefore solid phase based synthesis strategies which can incorporate a diverse set of modifiers will be a more desirable strategy for synthesis of these bbRNAs.

2.2.3. Solid phase synthesis of bbRNAs

The discovery that the lariat RNA intermediate contained a 2'-5' phosphodiester bond happened during the same time when solid phase synthesis of linear DNA and RNA were in the development.[27,28] Because solid phase synthesis of RNA is automated and requires no purification after introducing each monomer, it became the method of choice for bbRNA synthesis. The very first solid phase based synthesis of bbRNA was described by Damha and Zabarylo.[29–31] The key synthetic intermediate in this method was a 2',3'-

bisphosphoramidite derivative of the branchpoint adenosine. This bisphosphoramidite was used to join two growing chains of oligonucleotide and form a branch which was then grown in the 5' direction. Although this method generated bbRNAs of longer length, the 2'- and the 3'-arm of the branchpoint adenosine contained exact same sequence which is not observed in naturally occurring lariat RNAs. Following this a modified method of bbRNA synthesis was also developed where two different strands of RNA was grown from the same solid support one after another.[32] Then the bisphosphoramidite was used to join those two strands. Although this method produced bbRNAs with different 2'- and 3'-branch sequence, it also produced the symmetric sequences with same 2'- and 3'-branch sequence due to lack of any selectivity of the bisphosphoramidite between the two different strands.

Another approach that has been explored by researchers is the use of orthogonal 2'-OH protecting groups. In this approach, first a linear strand of the oligonucleotide is synthesized where the 2'-OH of the branchpoint adenosine is protected using a different protecting group compared to the rest of the oligonucleotide. Then the 2'-OH of the branchpoint adenosine is selectively removed by using a reagent which does not affect the rest of the protected 2'-OH groups. Then the 2'-5' branch is synthesized using reverse DNA amitides which contains 5'-phosphoramidite and a 3'-*O*-DMT protecting group. The first application of this approach was demonstrated by Damha *et al.* by regiospecific synthesis of 2'-5' linked branched DNAs using reverse DNA amidites.[31,33] A similar method for regiospecific synthesis of branched RNAs was also demonstrated by Sporat *et al.* using reverse RNA amidites.[34] However this method required synthesis of multiple phosphoramidite reagents with non-conventional protecting groups which limited the

ease and speed of the synthesis. The strategy of using orthogonal 2'-OH protecting group has also been utilized to make branched DNA-RNA hybrid molecules by Damha *et al.*[35] The synthesis began by assembly of the RNA stem using sequential additions of 2'-*O*-[1-(2-fluorophenyl)-4-methoxypiperidin-4-yl] (Fpmp) protected monomers with a 2'-*O*-TBDMS protected monomer at the branch point. Then the solid support was treated with fluoride reagent which selectively deprotected 2'-*O*-TBDMS group without affecting the Fpmp protecting groups generating a free 2'-OH group at the branchpoint. Following this the branch was grown in 5' to 3' direction using commercially available reverse DNA amidites. The synthesis of 2'-5' linked RNA branch was difficult at that time due to easy access to reverse RNA phosphoramidite monomers.

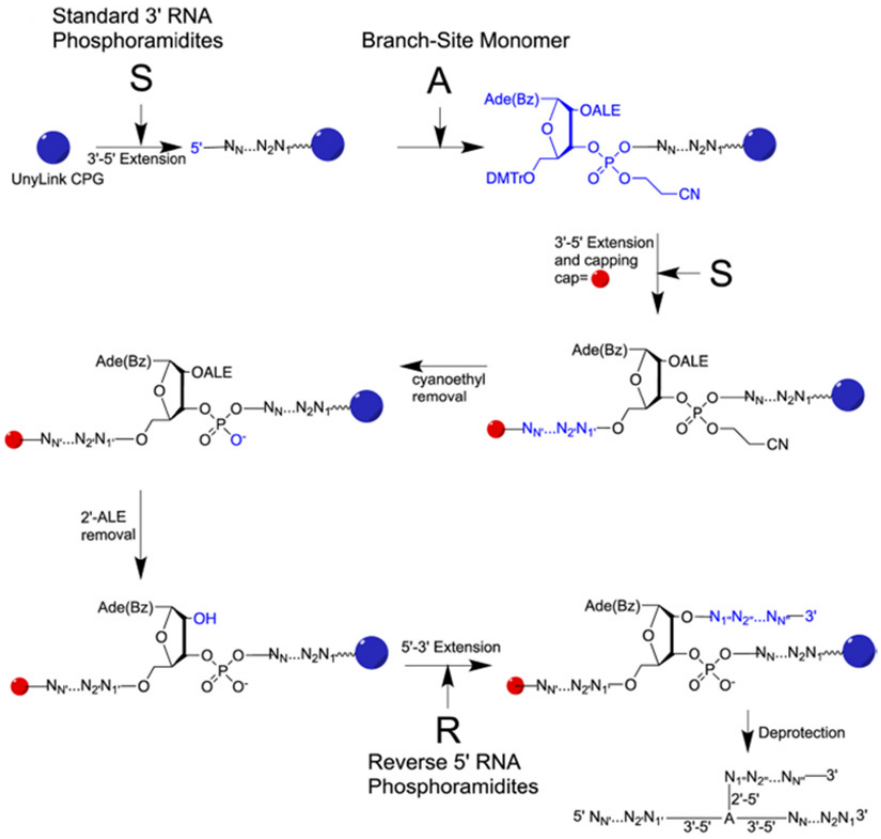


Figure 2.3. Regiospecific solid phase synthesis of 2'-5' phosphodiester linked bbRNAs by Damha *et al.* This figure is modified from ref 36.

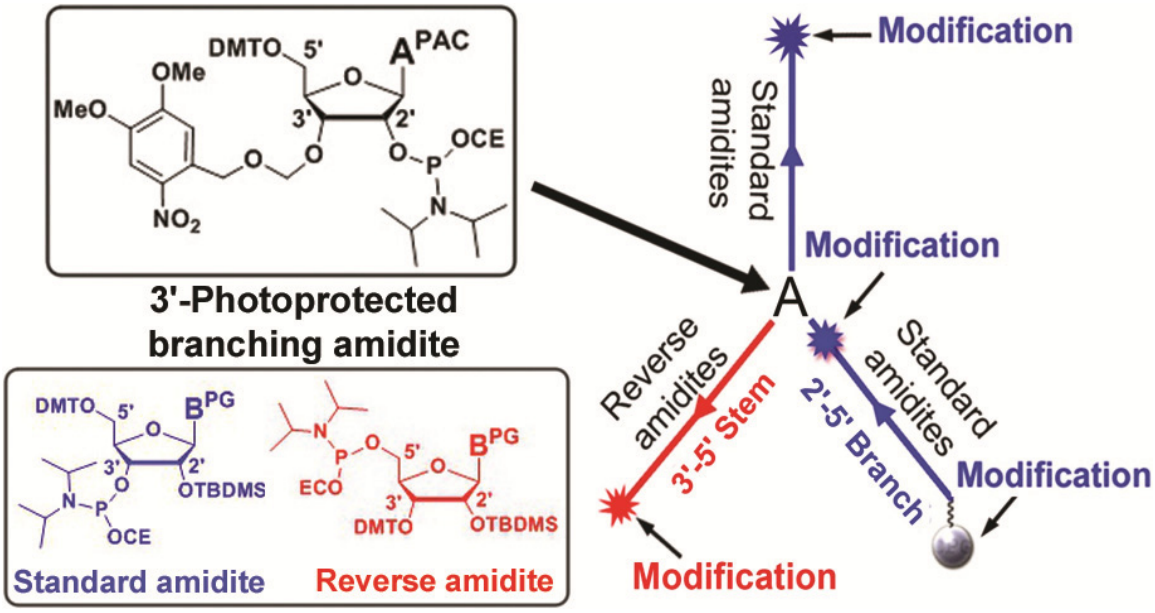
Very recently another regiospecific synthesis of bbRNA is reported by Damha *et al.* which also uses the orthogonal protecting group strategy.[5] In this method the branch point adenosine is introduced as a 2'-acetal levulinic ester (2'-*O*-ALE) monomer, where the other monomers were standard 2'-*O*-TBDMS protected (Fig. 2.3). Then following the selective deprotection of branchpoint 2'-*O*-ALE group using hydrazine, the 2'-5' branch was synthesized using commercially available reverse RNA amidites (Fig. 2.3). Reverse RNA amidites were initially developed for synthesis of linear RNAs in the 5' to 3' direction.[36] However the synthesis of 2'-5' linked bbRNAs has now become easy and fast due to commercial availability of the reverse RNA monomers.

This method of bbRNA synthesis described by Damha *et al.* enjoys all the benefits of solid phase synthesis and also could be used to synthesize any sequence. However this method suffers from one disadvantage that the 2'-branch is synthesized using reverse amidites. Therefore this method cannot be used to incorporate any modified nucleotide in the 2'-branch because the reverse amidite for any modified nucleotide is not commercially available. To avoid this shortcoming, I have developed a solid phase synthesis of bbRNA based on photoprotected phosphoramidtes which can be used to incorporate modified nucleotide in the 2'-branch.

2.3. Solid phase synthesis of native bbRNAs using photoprotected branching phosphoramidite

To understand the structure function relationship of any enzyme, one requires making modified substrates and observing how those modifications change the enzymatic

reactions. However the regiospecific method of bbRNA synthesis described by Damha *et al.* does not allow incorporation of modified nucleotide in the branch due to commercial unavailability of modified reverse amidites.[5] Therefore we have developed a solid phase synthesis method based on a photoprotected phosophoramidite which allows incorporation of modified nucleotide in the branch (Scheme 2.1). In this method the synthesis starts from the 3' to 5' direction using standard 2'-*O*-TBDMS protected amidites (blue). Then a 3'-photoprotected branching amidite is incorporated at the branchpoint which contains a 2'-phosphoramidite (Scheme 2.1). Following this the RNA synthesis is continued in the 3' to 5' direction again. Once the 5' end of the sequence is completed, then the photoprotecting group is removed from the branchpoint nucleotide using UV light generating a free 3'-OH group. The 3' end of the stem is then synthesized using reverse amidites (red) which contains a 5'-phosphoramidite.

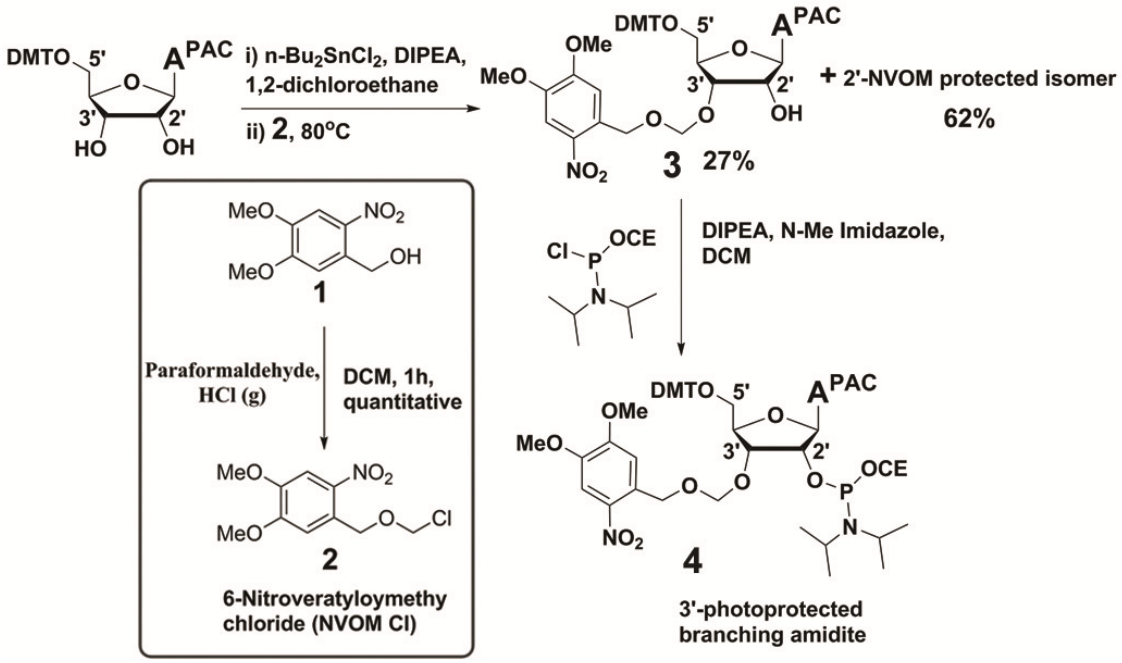


Scheme 2.1. Scheme for synthesis of bbRNA using a 3'-photoprotected phosophoramidite.

The advantages of using a 3'-photoprotected phosphoramidite is three fold: i) The photoprotecting group allows reagent-less deprotection of the 3'-OH of the branchpoint nucleotide which in turn allows a diverse set of modifications to be present in the RNA. ii) Because the 2'-branch is synthesized using standard amidites, now we can incorporate modified nucleotides in the branch. iii) Using this method we can incorporate terminal as well as internal modifications into the bbRNA required for biochemical and biophysical studies (Scheme 2.1). The following section describes the synthesis of the 3'-photoprotected phosphoramidite required for the bbRNA synthesis.

2.3.1. Synthesis of the 3'-photoprotected phosphoramidite

We chose to use *o*-nitrobenzyl (*o*NBn) based 3'-photoprotecting group for our bbRNA synthesis. The *o*NBn group has been extensively used as a 2'-OH protecting group for RNA synthesis as well as photocaging group for functional RNAs.[28,37–40] However one of the disadvantages of the *o*NBn group is that its absorption spectra overlap with that of the nucleobases. Therefore the UV light (~260nm) that is used to excite the *o*NBn group can also photodamage the nucleobases. We sought to synthesize an appropriately protected branching phosphoramidite with the 2'-OH protected by a 4,5-dimethoxy-*O*-nitrobenzyloxymethyl or 6-nitoveratrlyloxymethyl (NVOM) group. Addition of the two methoxy groups red shifts the absorption spectra of the NVOM group with absorption maxima around 365nm. This allows us to use 365nm UV light for the photodeprotection which does not affect the nucleobases.



Scheme 2.2. Synthesis of the 3'-photoprotected phosphoramidite (**4**).

The NVOM group was installed in the commercially available N-PAC (phenoxyacetyl) protected adenosine nucleoside in a single step following a known procedure (Scheme 2.2).[41] First the addition of the dibutyltin chloride and the *N*,*N*-diisopropylethylamine (DIPEA) to a solution of the base protected nucleoside promotes the formation of the stanylene acetal which activates the nucleophilic 2'- and 3'-OH groups. Then addition of the 6-nitoveratrlyloxymethyl chloride (NVOM Cl, **2**) reagent at 80˚C gives rise to a mixture of both 2'- and 3'-*O*-NVOM protected ribonucleoside. The 3'-*O*-NVOM protected isomer (**3**) can be isolated form this mixture using column chromatography in about 27% yield. The photoprotected alkylating reagent **2** used for this step was synthesized from the commercially available 6-nitroveratryl alcohol (**1**) (by Dr. Debasish Grahacharya) in a single step with almost quantitative yield using reported method (Scheme 2.2, inset)[42]. Following the installation of the 3'-*O*-NVOM group in **3**,

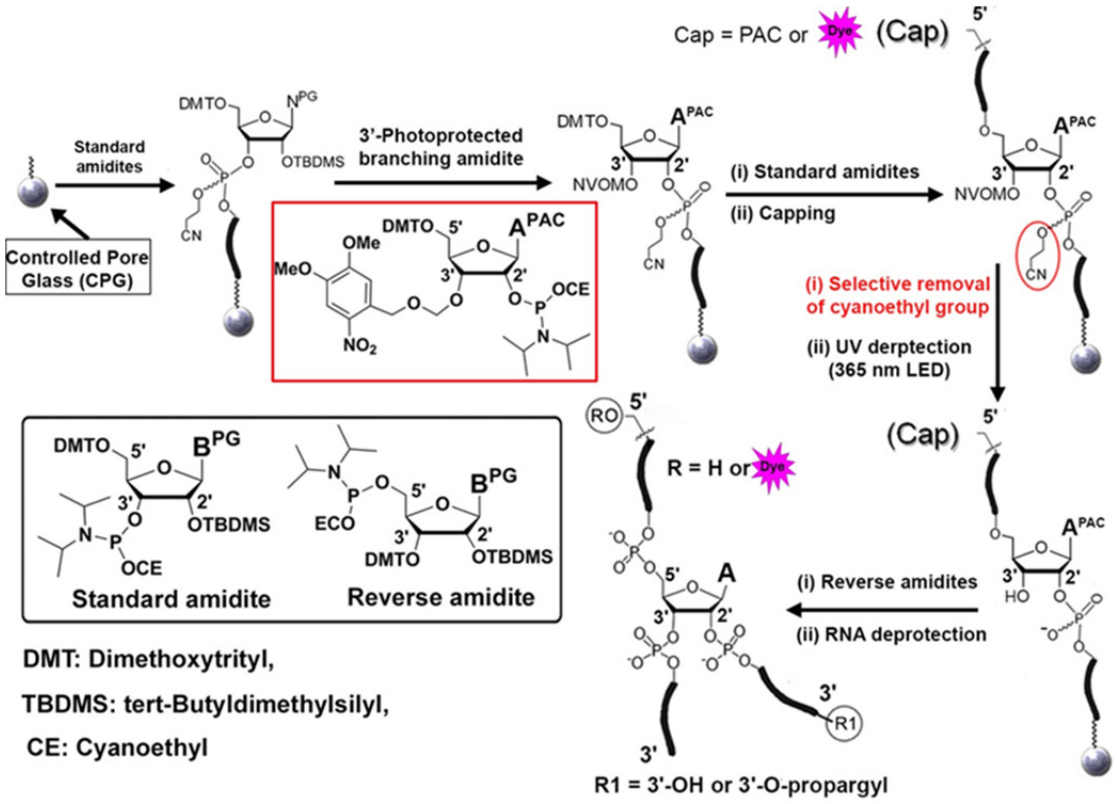
the 2'-OH was converted into phosphoramidite using standard procedure giving excellent yield of the 3'-photoprotected amidite **4** (Scheme 2.2).[41] The amidite **4** was then directly used for the solid phase synthesis of bbRNAs. Along with the 3'-*O*-NVOM protected amidite, we have also used the 2'-*O*-NVOM protected monomer for solid phase synthesis of bbRNAs using a protocol similar to described by Damha *et al.*[5]

Compared to the *o*NBn protecting group, incorporation of the NVOM group using this method enjoys the benefit of using fully protected (both nucleobase and 5'-OH) nucleoside which are commercially available. For synthesis of the 2'-*O*-*o*NBn protected ribonucleosides, the base and 5'-protecting groups are installed at a later stage requiring multiple step synthesis.[40] The use of the oxymethyl linker in the 3'-*O*-NVOM protected ribonucleoside also makes it less sterically hindered which promotes better coupling efficiency of the corresponding phosphoramidite. The following section describes the solid phase synthesis of 2'-5' phosphodiester linked bbRNAs in detail using the 3'-*O*-NVOM protected phosphoramidite.

2.3.2. Solid phase synthesis of bbRNAs using 3'-*O*-NVOM protected amidite

The solid phase synthesis of bbRNA described in Scheme 2.1 was performed in a Mermade-4 oligonucleotide synthesizer. The synthesis starts in 3' to 5' direction from the solid support using standard 2'-*O*-TBDMS protected phosphoramidites (Scheme 2.3). Then the 3'-*O*-NVOM protected amidite **4** is introduced at the branchpoint using similar coupling conditions as standard amidites (see experimental). Following this, the RNA is grown in the 3' to 5' direction again using standard amidites and then the 5' terminal is capped either using phenoxyacetyl (PAC) group or a florescent dye to prevent further

extension in this direction. Then the cyanoethyl (CE) groups are selectively removed from the phosphate backbone using a mixture of triethylamine and acetonitrile. The removal of the cyanoethyl group is necessary for the bbRNA synthesis to prevent any 2'- to 3'-branch migration or strand cleavage under neutral or basic condition.[31,33]



Scheme 2.3. Solid phase synthesis of native bbRNAs using 3'-*O*-NVOM protected phosphoramidite.

Following the cyanoethyl removal, the 3'-*O*-NVOM group is removed from the RNA strand on solid support using UV LED (365nm) to generate a free 3'-OH group. The removal of the 3'-*O*-NVOM group by UV light could be performed both outside the synthesizer and inside the synthesizer as well. Following photolysis of the 3'-*O*-NVOM group from the RNA strand, the rest of the 3'-section of the stem is synthesized using

commercially available reverse RNA phosphoramidites. The reverse amidites contains a 5'-phosphoramidite group and a 3'-*O*-DMT protecting group and therefore the RNA strand can be synthesized in 5' to 3' (reverse) direction using these amidites. Because the 2'-oxygen of the branchpoint adenosine already contains a phosphodiester bond at this point (Scheme 2.3), the reverse RNA amidites were coupled at higher concentration and for longer time to compensate for the steric hindrance to amidite coupling (see experimental). Once the bbRNA synthesis has been completed, the RNAs were cleaved from the solid support and the different protecting groups (nucleobase, 2'-*O*-TBDMS and cyanoethyl) were removed using standard condition. Although the final yields of the bbRNAs were lower compared to linear RNAs, they were comparable or better than other reported methods. For example, bbRNA1 (Table 2.1) which contains six nucleotides in each of the arms gave a final yield of 32% compared to 19% yield reported by Damha *et al.* for a bbRNA containing five nucleotides in each of the arms. Using this method, we have obtained variable branching efficiencies for the bbRNA synthesis which lies between 25% to 40% (Table 2.4) depending on the RNA synthesized.

As described earlier, this bbRNA synthesis strategy uses standard amidities for the synthesis of the 2'-branch. Therefore any commercially available modified nucleotide can be readily incorporated in the 2'-branch using this method. Using this strategy it is also possible to incorporate terminal modification to any of the three ends of this bbRNA which include fluorescent dye, biotin, amine, alkyne modifications etc. (Scheme 2.3). The following section describes the synthesis of several modified bbRNAs which are used in different Dbr1p related projects to highlight the versatile nature of this synthesis.

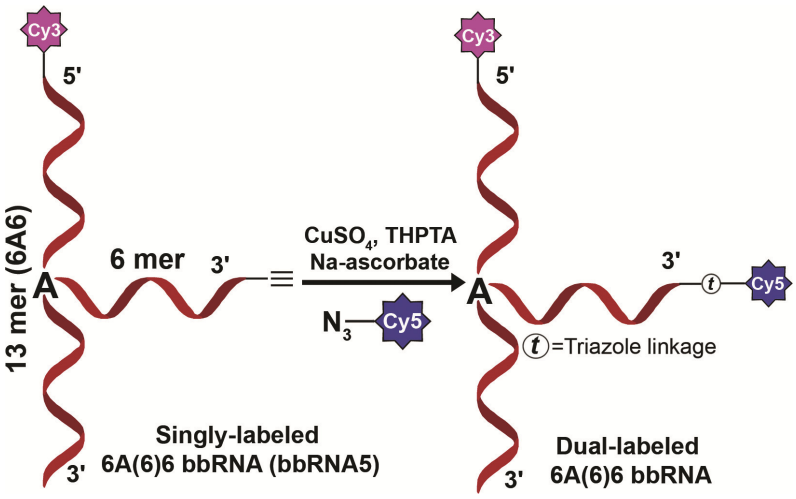
2.3.3. Synthesis of bbRNAs for biochemical and kinetics studies

Previous biochemical studies by Das lab members (Dr. Eduardo Paredes and Molly Evans) suggested that the first nucleotide in the 2'-branch may be important in the debranching reaction by Dbr1p. Therefore using this strategy, I have synthesized several bbRNAs where the first nucleotide in the 2'-branch is modified (bbRNA1-4). Table 2.1 shows the sequence of the different bbRNAs synthesized and their characterization by mass spectrometry. The sequences of the bbRNAs were chosen based on the yeast consensus sequence around the 5'-splice site and the branchpoint nucleotide.[43] These bbRNAs also contains a 5'-fluorescent dye; Dylight 547 (Dy547, a Cy3 analogue), which was incorporated using commercially available phosphoramidite during the capping step (Scheme 2.3). The fluorescent dye label at the 5'-terminus of the bbRNA was included to aid in visualization. This also helps us to avoid hazardous radioactive labeling of the bbRNAs for detection.

Name	Sequence	Mass calculated	Mass Found
bbRNA1	**5'-Dy547-GUACUA A-(2'-5'-GUAUGA)-CAA GUU–3'**	**6559.1**	**6557.1**
bbRNA2	**5'-Dy547-GUACUAA-(2'-5'-gUAUGA)-CAAGUU-3'**	**6544.0**	**6543.6**
bbRNA3	**5'-Dy547-GUACUAA-(2'-5'-(2'F-G)UAUGA) -CAAGUU-3'**	**6561.3**	**6564.8**
bbRNA4	**5'-Dy547-GUACUAA-(2'-5'-(2'-OMe-G)UAUGA)-CAA GUU-3'**	**6571.6**	**6571.8**
bbRNA5	**5'-Dy547-GUACUAA-(2'-5'-GUAUGA-*3'-O-Propargyl*) CAA GUU-3'**	**6597.1**	**6598.8**

Table 2.1 Sequence and MALDI mass of the bbRNAs for biochemical and kinetics studies. The g indicates deoxyguanosine.

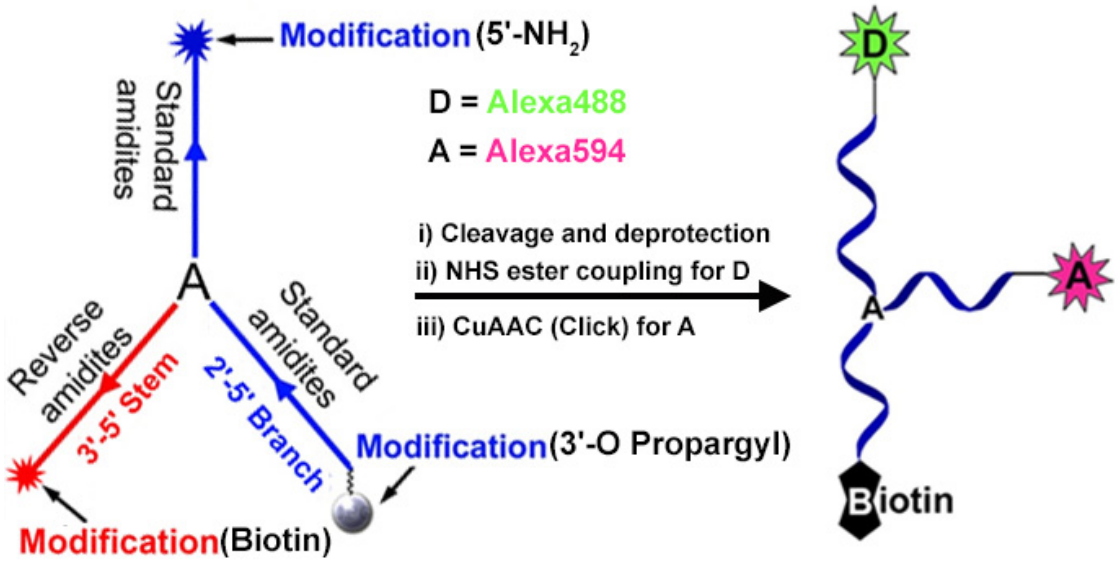
Although, some Dbr1p substrates have been biochemically studied, there has been no report of the reaction kinetics for the Dbr1p cleavage reaction. Therefore I have developed a FRET based kinetic assay for Dbr1p cleavage using a dual fluorescently (Cy3 and Cy5) labeled bbRNA substrate. The dual-labeled bbRNA was synthesized from a singly labeled bbRNA precursor (bbRNA5) which contained a 3'-*O*-Propargyl group at the 3'-terminal of the branch. The synthesis of bbRNA5 started from a 3'-*O*-Propargyl modified adenosine CPG and then the 5'-Cy3 dye was incorporated during the capping step (Scheme 2.4). Following deprotection and purification of bbRNA5, the Cy5 dye was incorporated using Cu catalyzed azide alkyne cycloaddition reaction (CuAAC) which is almost synonymous with click chemistry. Our previous efforts to install the Cy5 dye during solid phase synthesis failed because of the degradation of Cy5 dye during the 2'-*O*-TBDMS deprotection step. Following the CuAAC reaction, the dual fluorescently labeled bbRNA was purified by HPLC and characterized by mass spectrometry. We have used this bbRNA for kinetic analyses of the Dbr1p debranching reaction which will be described in chapter 4.



Scheme 2.4. Synthesis of dual fluorescently labeled bbRNA for kinetics studies using CuAAC.

2.3.4. Synthesis of bbRNAs for single molecule studies

We have also investigated the conformation and dynamics of bbRNAs using single molecule FRET (smFRET) methods. For the smFRET studies, we have synthesized dual-fluorescently (Alexa488 and Alexa594) labeled bbRNAs which also contained a 3'-biotin modification. The dual labeled bbRNA contains Alexa488 dye as the donor (D) fluorophore and Alexa594 dye as acceptor (A). Scheme 2.5 describes the synthesis of the dual fluorescently labeled bbRNA for the single molecule studies.



Scheme 2.5. Synthesis of dual fluorescently labeled bbRNA for single molecule studies.

As the Alexa dyes are not commercially available as phosphoramidite reagents, we chose to use a different strategy for incorporation of the fluorescent dyes. Similar to the synthesis of bbRNA5 described before, we used 3'-*O*-Propargyl modified adenosine CPG for the synthesis (Scheme 2.5) for post synthetic modification of the branch terminal. Then the 3'-*O*-NVOM protected amidite **4** is introduced at the branchpoint. Following this, the 5'-termial was capped using a C6 amino modifier phosphoramidite.

Then following the synthesis of the 3'-section of the stem using reverse amidites (red), the biotin was incorporated into the bbRNA using Biotin-TEG phosphoramidite. This biotin moiety will be used for surface immobilization of these bbRNAs for single molecule studies. Both the fluorescent dyes were incorporated into the bbRNA using post synthetic methods. First the donor dye (Alexa488) was conjugated to the 5'-amine using NHS ester chemistry (see experimental). Then the acceptor dye (Alexa594) was incorporated into the Alexa488 labeled bbRNA using CuAAC (see experimental). It is important to mention here that both the donor and acceptor dyes incorporated into this bbRNA are isomerically pure. Commercially available Alexa dyes based on rhodamine backbone (both Alexa488 and Alexa594) always have two regioisomers (see experimental).[44] However to avoid any complication during the single molecule studies, we decided to use single isomers of each of the donor and acceptor dyes. I have synthesized the Alexa488 dye following reported method[45]. Then I was able to purify one isomer from the other using reverse phase chromatography. Following this the NHS ester of the isomerically pure Alexa488 dye was synthesized using standard protocol. We have used azide modified Alexa594 dye to incorporate the acceptor dye into the bbRNA. The commercially available Alexa594 dye also comes as a mixture of two isomers. When it is conjugated to the Alexa594 labeled bbRNA, the two isomeric products can be separated by reverse phase HPLC. To understand the effect of branch length on the conformation and dynamics of bbRNAs, we have synthesized two different dual labeled bbRNAs with different 2'-branch length (6mer and 12 mer). Both the dual labeled bbRNAs were used for single molecule conformational studies which are described in chapter 7.

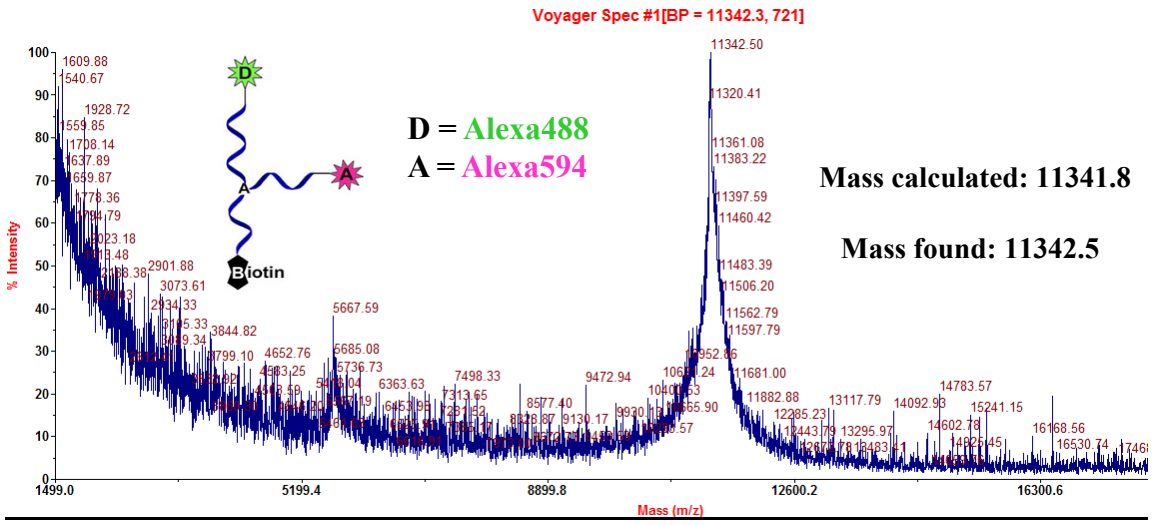


Figure 2.4. MALDI mass spectra of a dual fluorescently labeled bbRNA used for single molecule studies. Sequence of the bbRNA: 5'-Alexa488-ACU ACU GUA CUA A-(2'-5' GUAUGA-Alexa594)-CAA GUU ACU U-3'Biotin TEG.

Figure 2.4 shows the mass spectra of a 29mer dual labeled bbRNA synthesized by this method. This bbRNA also contains three different modifiers: one incorporated during solid phase synthesis and two installed through post synthetic methods. Successful synthesis of this multiply modified bbRNA demonstrates the power and versatility of our solid phase synthesis strategy.

2.4. Synthesis of 'click' branched RNAs: substrate analogue of Dbr1p

Although we can use native 2'-5' phosphodiester linked bbRNAs for biochemical and kinetics studies, they are not useful for binding studies with Dbr1p. The native bbRNAs will bind with Dbr1p, get cleaved and then they will be released from the active site. Therefore to investigate binding, we need a substrate analogue which will bind with Dbr1p similar to the native substrate, but will not get cleaved. Also because of the important role of Dbr1p in different biological regulatory processes such as miRNA biogenesis and retrovirus propagation, it is important to search for molecules which can

inhibit Dbr1p function. Tago and Damha reported the synthesis of 2'-5' phosphoramidate linked bbRNAs as an inhibitor of Dbr1p.[46] The 2'-5' phosphodiester linkage is replaced by 2'-5' phosphoramidate linkage (a phosphoramidate is a phosphate that has an NR2 instead of an OH group) in this analogue which cannot be cleaved by Dbr1p.[46] Although this substrate analogue is produced by minimal possible perturbation to the native bbRNA substrate, it requires synthesis of complex building blocks. We have synthesized 2'-5' triazole linked branched RNAs for the same purpose using commercially available amidites and reagents (Fig. 2.5). The triazole linkage has been shown to be biocompatible when present in DNA and RNA and also it can be read through by DNA polymerase and reverse transcriptase.[47–50] Because these branched RNAs are synthesized using copper catalyzed azide alkyne cycloaddition (CuAAC) reaction which is one of the most widely used reactions under the term 'click' chemistry, we refer to these 2'-5' triazole linked branched RNAs as 'click' branched RNAs or cbRNAs. We have synthesized several cbRNAs for both kinetics studies and single molecule studies. The following sections describe the synthesis of the different cbRNAs required for different projects in this thesis.

2.4.1. Synthesis of cbRNA for kinetics studies

The cbRNA is a non-cleavable substrate analogue of Dbr1p that contains a 2'-5' triazole linkage instead of a phosphodiester linkage and therefore cannot be cleaved by Dbr1p. To test its efficacy as Dbr1p inhibitor, we have synthesized a cbRNA with the same sequence as the dual-labeled (Cy3-Cy5) bbRNA for kinetics studies but without any dyes. The cbRNA was synthesized by conjugating a linear strand containing an internal 2'-*O*-propargyl with the branch sequence containing a 5'-azide using CuAAC reaction

(Fig. 2.5a). The stem strand containing a 2'-*O*-propargyl group was synthesized by regular solid phase synthesis using the commercially available 2'-*O*-propargyl adenosine (N-Bz) phosphoramidite at the branch point. The 5'-azide modification was introduced to the branch strand using post synthetic modification on the solid support. First the 5'-OH of the branch strand was modified to 5'-iodo using methyltriphenoxyphosphonium iodide[51] and then the 5'-iodo was converted to 5'-azido by treatment with sodium azide.[52] Following the synthesis of the stem and branch strand, they were conjugated using CuAAC using THPTA as Cu(I) stabilizing ligand.[53,54] Following the CuAAC reaction, the product was separated from the starting materials using gel electrophoresis. UV shadowing of the gel shows near-complete conversion of the stem strand into the cbRNA demonstrating the easy synthetic accessibility to these cbRNAs (Fig. 2.5b). Then the cbRNA was characterized by mass spectrometry and HPLC. I have used this unlabeled 6A(6)6 cbRNA for the kinetics studies which is described in chapter 4.

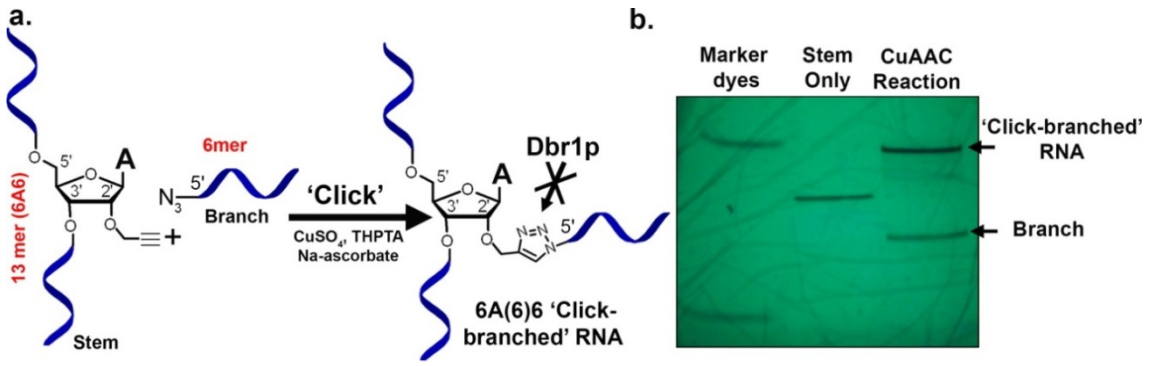
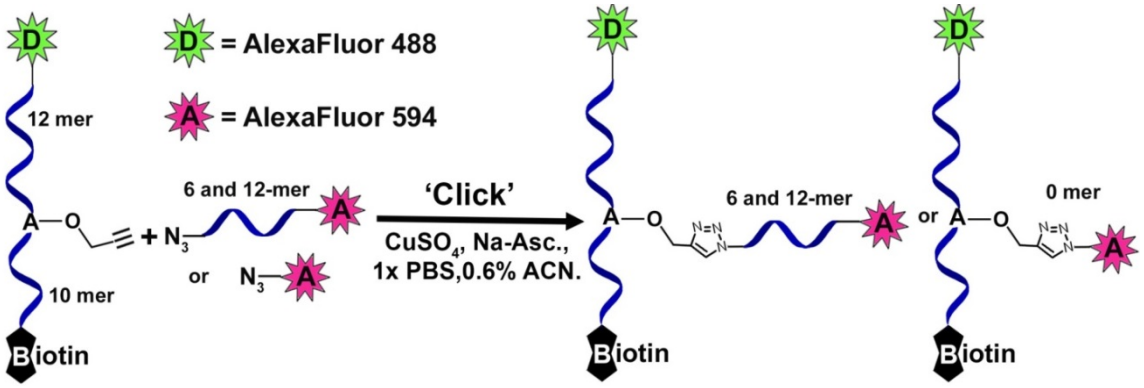


Figure 2.5. Synthesis of 'click' branched RNAs as a potential inhibitor of Dbr1p **a**. The 'click' branched RNA was synthesized by conjugating a stem strand containing an internal 2'-*O*-propargyl with the branch containing a 5'-azide using CuAAC reaction. Sequence of the cbRNA: 5'-GUA CUA A-(2'-*t*-5' GUA UGA)-CAA GUU where *t* signifies the triazole linkage. **b.** UV shadowing of the gel after the CuAAC reaction and separation shows complete conversion of the stem into the product.

2.4.2. Synthesis of dual-labeled cbRNAs for single molecule studies

Through kinetic analysis of the Dbr1p cleavage reaction it was discovered that the cbRNA is a competitive inhibitor of Dbr1p, suggesting that they bind Dbr1p in a similar fashion as the native substrate (Chapter 4, section 4.2). This prompted us to synthesize cbRNAs which could be used to study their conformational changes upon binding with Dbr1p. We have synthesized several dual fluorescently (Alexa488-Alexa594) labeled cbRNAs for single molecules studies of their conformation and dynamics before and after binding with Dbr1p (Scheme 2.6). I have used the same Alexa488 and Alexa594 donor acceptor pairs in this cbRNAs and they were also isomerically pure as well. Scheme 2.6 describes the synthesis of the Alexa488 and Alexa594 labeled cbRNAs.



Scheme 2.6. Synthesis of dual-labeled cbRNAs for single molecule studies.

Similar to the non-labeled cbRNA, the dual-labeled cbRNAs were also synthesized by conjugating a stem strand containing an internal 2'-*O*-propargyl with the branch containing a 5'-azide using CuAAC reaction. However the stem and branch strands were labeled with Alexa488 and Alexa594 dyes respectively before they were conjugated together to make the cbRNA. The stem strand containing an internal 2'-*O*-

propargyl group was labeled at the 5'-terminal with Alexa488 dye using NHS ester chemistry (see experimental, section 2.6). The branch strand containing a 5'-azide was labeled at the 3'-terminal also using NHS ester chemistry. Following the purification of the two strands, the cbRNA was synthesized by joining them through a trizole linkage using CuAAC (Scheme 2.6). I have synthesized two different cbRNAs with 6 and 12 nucleotides in the branch. I have also synthesized another dual-labeled bbRNA where the acceptor dye is directly conjugated to the Alexa488 conjugated stem strand (Scheme 2.6). Because there is no nucleotide in the branch, it is referred as '0 mer' cbRNA (Scheme 2.6). All the dual-labeled cbRNAs synthesized using this method also contained a 3'-biotin in the stem which is useful for surface immobilization during single molecule experiments. All these cbRNAs were used in single molecule experiments which are described in chapter 7.

2.5. Conclusion

This chapter describes the synthesis of several bbRNAs which are useful for biochemical and single molecule studies of debranching and relative processes. The solid-phase synthesis of bbRNAs was achieved using a 3'-photoprotected branching amidite. The 3'-*O*-NVOM protected amidite **4** could be readily synthesized from commercially available starting materials in three steps with high yields. Then this amidite **4** was used for solid phase synthesis of bbRNAs. These bbRNAs contain both terminal and internal modification in the stem and in the branch as well. Our synthetic strategy for access to bbRNAs is efficient, readily accommodates modifications and the use of photo-labile protecting group makes the method amenable to complete automation. We have also used CuAAC to make triazole linked cbRNAs as a substrate analogue of Dbr1p and we show

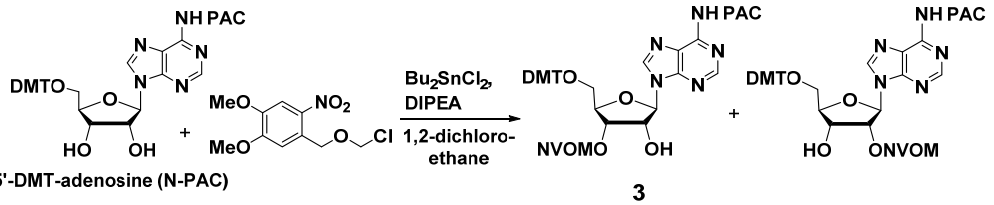
later that these cbRNAs are competitive inhibitor of Dbr1p. Overall this chapter describes important access to the various RNAs required for the different projects later in this thesis.

2.6. Experimental

Materials

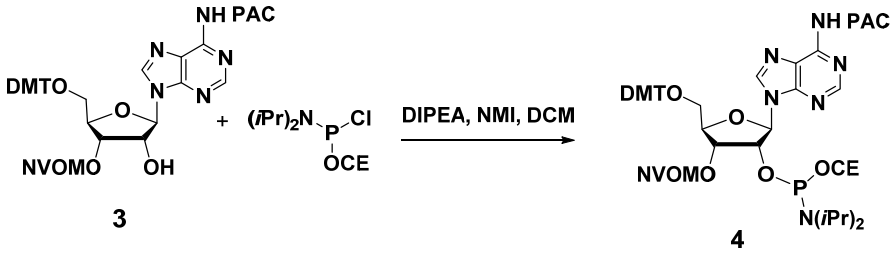
The 5'-DMT-adenosine (*N*-PAC) nucleoside, standard RNA phosphoramidites with 2'-*O*-TBDMS and ultramild protecting groups (A-PAC, C-PAC and G-PAC), CPG solid supports for RNA synthesis, 2'-*O*-Propargyl adenosine (N-Bz) phosphoramidite and 2-cyanoethyl-*N,N*-diisopropyl-chloro-phosphoramidite were purchased from Chemgenes (Wilmington, MA, USA). Appropriate reagents for solid phase RNA synthesis (deblock, activator, ultramild CapA, CapB and oxidation reagent), GlenPak cartridges, 5-amino modifier C6 PDA phosphoramidite and Dylight547 (an analogue of Cy3) phoshoramidte were purchased from Glen Research (Sterling, VA, USA). The Cy5 azide dye used for the dual-labeled bbRNA synthesis was provided by Dr. Brigette Schmidt. The Alexa594 azide dye was purchased from Life Technologies. All the dry solvents and other reagents were purchased for Fisher or Sigma Aldrich.

Synthesis of 5'-O-(Dimethoxytrityl)-3'-*O*-(6-Nitoveratrlyloxymethyl)-N^6-phenoxyacetyl-adenosine (3):



A solution of 5'-DMT-adenosine (*N*-PAC) (0.8 g, 1.14 mmol) and DIPEA (0.99 mL, 5.68 mmol) in anhydrous 1,2-dichloroethane (10 mL) was treated with n-Bu_2SnCl_2 (414 mg, 1.36 mmol) for 90 mins at room temperature under argon atmosphere. Then NVOM-Cl (0.39 g, 1.19 mmol with 2% $CaCl_2$ w/w) was added and the mixture was stirred for 20mins at 80 °C. The resulting mixture was subjected to flash chromatography. The column was started using EtOAc/hexane 3:2 (with 1% TEA) and the 2'-*O*-NVOM labeled isomer was eluted with 100% EtOAc (with 1% TEA). Then the 3'-*O*-NVOM labeled isomer was eluted using 1% MeoH in DCM (with 1% TEA) as the mobile phase. We finally afforded the 3'-*O*-NVOM labeled isomer (287 mg, 27%) as pale yellow solids. ^{1}H NMR (300 MHz, $CDCl_3$): δ 3.44 (dd, *J* = 3.9, 10.6 Hz, 1H), 3.54 (dd, *J* = 3.4, 10.4 Hz, 1H), 3.79 (s, 6H), 3.83 (s, 3H), 3.91 (s, 3H), 4.30 (app q, *J* = 3.5 Hz, 1H), 4.54-4.58 (m, 1H), 4.78-4.90 (m, 4H), 5.03 (s, 2H), 5.06 (t, *J* = 5.3 Hz, 1H), 6.27 (d, *J* = 5.7 Hz, 1H), 6.80-6.83 (m, 4H), 7.05-7.10 (m, 3H), 7.30-7.45 (m, 8H), 7.66 (s, 1H), 8.16 (s, 1H), 8.59 (s, 1H), 9.25 (s, 1H). Mass calculated: 928, mass obtained 929 (M^++H), 951 (M^++Na).

Synthesis of 5'-O-(Dimethoxytrityl)-3'-O-(6-Nitroveratrlyloxymethyl)-N^6-phenoxyacetyl-adenosine 3'-N,N-diisopropyl(cyanoethyl) phosphoramidite



To a solution of **3** (300mg, 0.322mmol) in dry CH_2Cl_2 (10ml), DIPEA (0.74 mL, 1.61 mmol), 2-cyanoethyl-*N,N*-diisopropyl-chloro-phosphoramidite (108 µL, 0.484 mmol) and 1-methyl-imidazole (12µL, 0.16mmol) were added under argon atmosphere. The

mixture was stirred for 30mins at 0 °C and 1.5 hour at r.t. Work up was done with $NaHCO_3$ (saturated)/ EtOAc. Column chromatography (EtOAc/Hexane, 4:1 with 2% TEA) afforded **4** (305mg) in 83.6% isolated yield. A single peak (doublet, δ: 150.95, 150.86) in the ^{31}P NMR spectra of the photoprotected phosphoramidite **4** conforms to its purity.

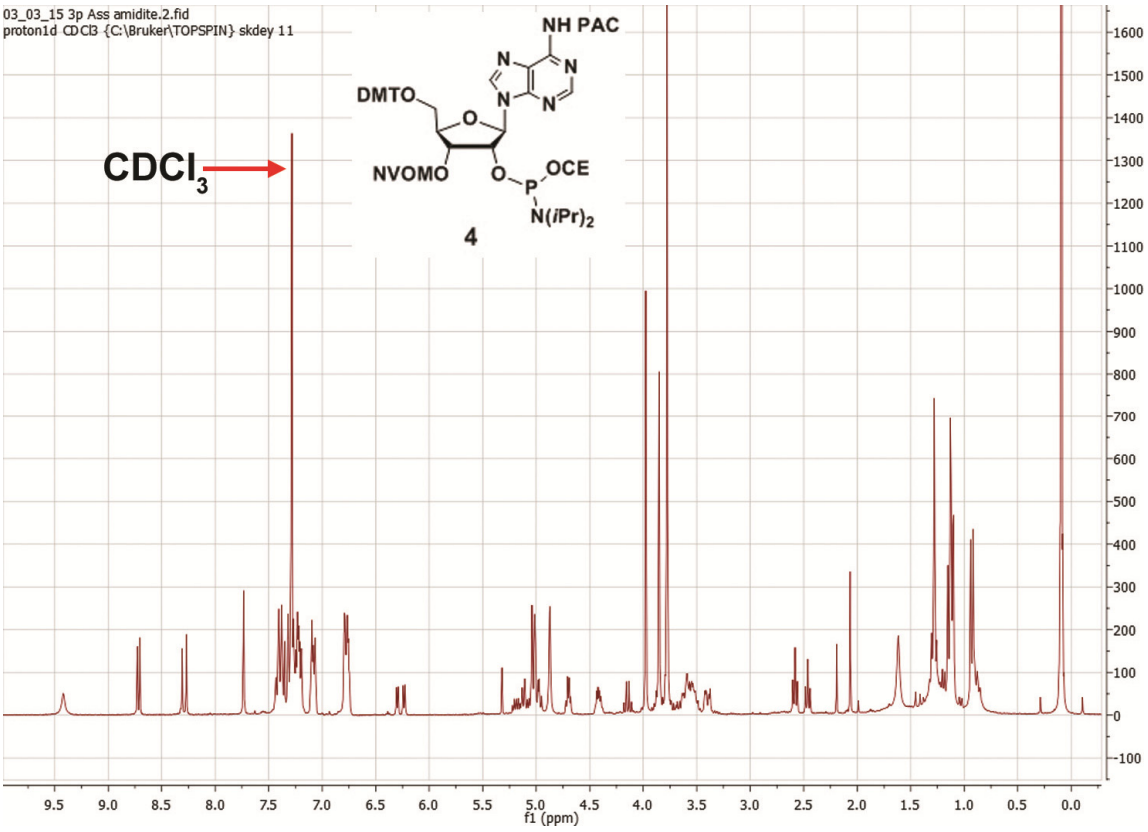


Figure 2.6. ^{1}H NMR spectrum of the 3'-*O*-NVOM protected phosphoramidite (**4**).

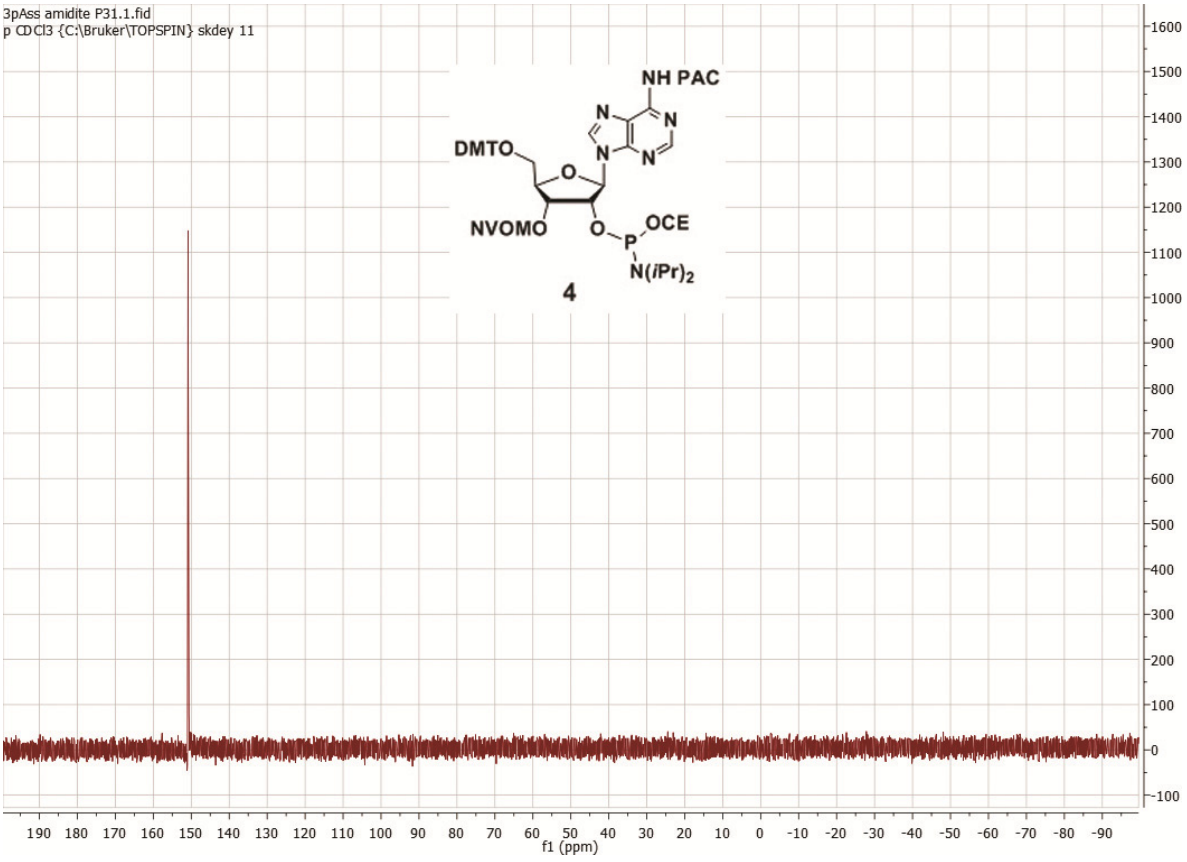


Figure 2.7. ^{31}P NMR spectrum of the 3'-*O*-NVOM protected phosphoramidite (**4**).

Solid phase synthesis of native bbRNAs

Solid phase synthesis of the bbRNAs was performed on a Mermade-4 (Bioautomation, Plano, TX, USA) automated synthesizer. The reagents and the conditions for solid phase synthesis of the linear portion of the bbRNA are given in Table 2.2. This include coupling of the standard 2'-*O*-TBDMS protected amidites as well as the 3'-*O*-NVOM protected amidites.

Synthesis step	**Reagent**	**Volume per injection**	**Reaction time per injection (s)**	**Total no of injections**	**Total reaction time (s)**
Deblock	3% Trichloroacetic acid	120 µL	50	2	100
Coupling	0.1M RNA (standard and photoprotected) amidite + 0.25M 5-(Ethylthio)-tetrazole activator	60 µL of amidite + 60 µL of activator	150 (for RNA amidites); 200 (for photo-protected amidite)	3	450 (for standard RNA amidites); 600 (for photo-protected) amidite
Capping	CapA: THF/Phenoxyacetic anhydride/Pyridine CapB: 16% 1-methylimidazole in THF	60 µL of CapA + 60 µL of CapB	60	1	60
Oxidation	0.02M I_2 in THF/Pyridine/H_2O	120 µL	50	2	100

Table 2.2. Reagents and synthesis conditions for linear portion of the bbRNA (1 µmole scale).

Following the synthesis of the linear strand, the cyanoethyl groups from the phosphate backbone were removed by treating the CPG beads with 2:3 trietylamine/acetonitrile mixture (10 mL) for 90 minutes[31]. The solution was passed

through the CPG beads using a syringe and pushed down every 15 minutes so that the CPG beads come in contact with fresh deprotection solution. Then the CPG beads were washed with acetonitrile (10 mL), THF (10 mL) and then again with acetonitrile (10 mL) and dried for 30mins under house vacuum. Following the removal of the cyanoethyl groups, the 1μmole CPG beads were divided into ~3x300nmoles by transferring them to 3 different 2 mL glass vials. Then 1 mL of HPLC grade acetonitrile was added to the vial and the photodeprotection was performed by using UV LED (Prizamatix, Israel) for 10 minutes with shaking the CPG beads after 5 minutes of photodeprotection. Then the CPG beads were transferred to a new column, washed with ACN and dried under house vacuum for 10 mins. After this, the column was transferred to the synthesizer to grow the branch using the reverse amidites. The reagents and the conditions for solid phase synthesis of the 3'-section of the bbRNA stem are given in Table 2.3.

Following solid phase synthesis of the bbRNA, the CPG beads were transferred to a glass vial and treated with 1.5 mL of 1:1 ammonia/40% aqueous methylamine mixture for 10 minutes at 65˚C to remove the base and cyanoethyl protecting groups. It is important to mention here that for the bbRNAs synthesized using a 5'-amino modifier; before the base deprotection, the CNOEt groups were removed using 10% solution of diethylamine in acetonitrile for 10 mins. This prevents inactivation of the 5'-amine due to its reaction with the acrylonitrile produced during base deprotection. The RNA solution after base deprotection is filtered through Pierce column to remove the CPG beads and then it is lyophilized. The 2'-*O*-TBDMS groups are deprotected from the RNA by treating this lyophilized powder with a mixture of DMSO (115 μL), trimethylamine (60 μL) and triethylamine trihydrofluoride (TEA.3HF, 75 μL). Then the bbRNAs are desalted using

cartridges following manufacturer's protocol. Following this the bbRNAs are purified either using HPLC or using gel electrophoresis.

Synthesis step	Reagent	Volume per injection	Reaction time per injection (s)	Total no of injections	Total reaction time (s)
Deblock	3% Trichloroacetic acid	120 µL	75	2	150
Coupling	0.25M RNA reverse amidite + 0.25M 5-(Ethylthio)tetrazole activator (For the first residue in the branch) 0.2M RNA reverse amidite + 0.25M 5-(Ethylthio)tetrazole activator (For rest of the residues)	60 µL of amidite + 60 µL of activator	360	4 (For the first residue in the branch). 2 (For the rest of the residues in the branch)	1440 (For the first residue in the branch). 720 (For the rest of the residues in the branch)
Capping	CapA: THF/Phenoxyacetic anhydride/Pyridine CapB: 16% 1-methyl-imidazole in THF	60 µL of CapA + 60 µL of CapB	60	1	60
Oxidation	0.02M I_2 in THF/Pyridine/H_2O	120 µL	50	2	100

Table 2.3. Reagents and synthesis conditions for the 3'-section of the bbRNA stem using reverse amidites (300 nmoles scale).

HPLC purification of native bbRNAs

Following the deprotection of the nucleobases and 2'protecting groups, the bbRNAs were purified by reverse phase HPLC using a Waters Breeze 2 HPLC system equipped with Waters 2996 phtotodiode array (PDA) detectors. Figure 2.8 shows the HPLC chromatogram of the crude bbRNA2 as an example. The branching efficiency was calculated based on the percentage of the bbRNA peak compared to all the other peaks in

the chromatogram. Figure 2.9 shows the HPLC chromatogram of the bbRNA2 after HPLC purification. Following the HPLC purification, the bbRNAs were also characterized by MALDI mass spectrometry using 3-hydoxy picolinic acid (3-HPA) as matrix. Table 2.4 shows the characterization of the different bbRNAs synthesized for the biochemical and kinetics studies.

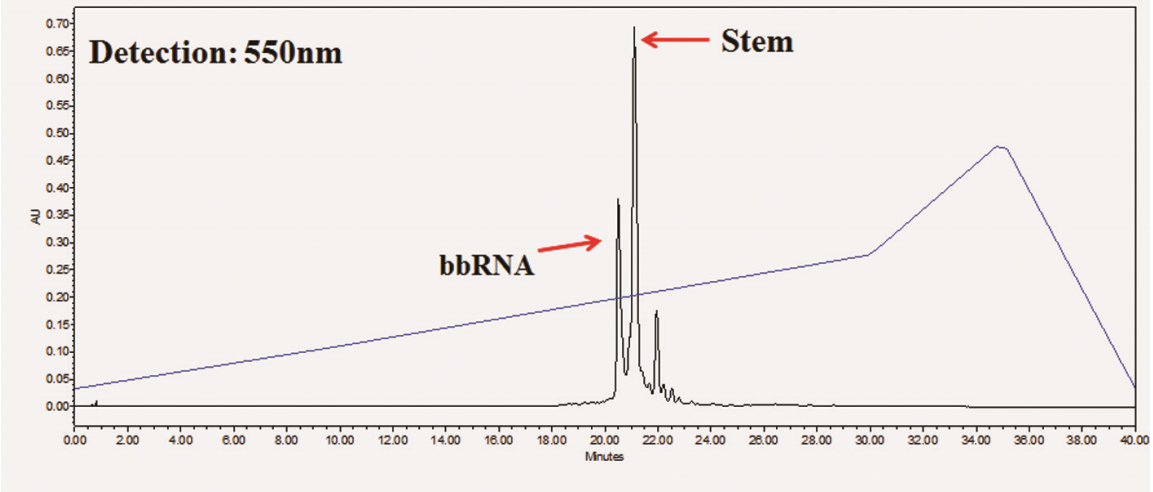


Figure 2.8. HPLC chromatogram of the crude bbRNA2. **HPLC condition**: Waters Xbridge C18 column (4.6x150 mm, 5 µm particle), Temp: 25˚C. Solvent A: 0.1M Triethylammonium acetate (TEAA) Buffer (pH 7), Solvent B: 80% Acetonitrile, 20% water, 0.1M TEAA buffer (pH 7), Gradient: 2.5% to 25% B in 10 mins, then 25% to 50% B in next 20 mins.

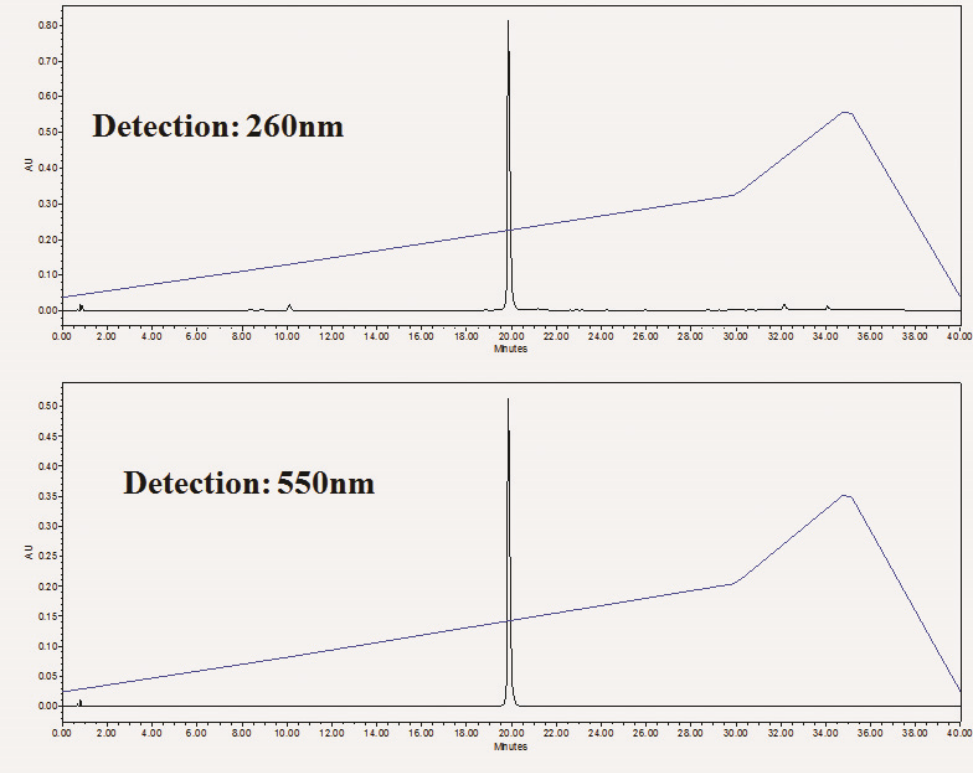


Figure 2.9. HPLC chromatogram of the purified bbRNA2. The HPLC conditions are exactly the same as the previous chromatogram used for purification.

Name	Sequence	Mass calculated	Mass Found	Branching efficiency (%)
bbRNA1	**5'-Dy547-GUACUAA-(2'-5'-GUAUGA)-CAAGUU-3'**	**6559.1**	**6557.1**	**24.8**
bbRNA2	**5'-Dy547-GUACUAA-(2'-5'-gUAUGA)-CAAGUU-3'**	**6544.0**	**6543.6**	**31.7**
bbRNA3	**5'-Dy547-GUACUAA-(2'-5'-(2'F-G)UAUGA) -CAAGUU-3'**	**6561.3**	**6564.8**	**27.6**
bbRNA4	**5'-Dy547-GUACUAA-(2'-5'-(2'-OMe-G)UAUGA)-CAA GUU-3'**	**6571.6**	**6571.8**	**24.6**
bbRNA5	**5'-Dy547-GUACUAA-(2'-5'-GUAUGA-*3'O-Propargyl*) CAA GUU-3'**	**6597.1**	**6598.8**	**40.5**

Table 2.4. Sequence and characterization of the different bbRNAs synthesized for biochemical and kinetics studies. The branching efficiencies of each of the bbRNA were calculated by integrating individual HPLC chromatograms.

Synthesis of the dual-labeled (Cy3-Cy5) bbRNA: conjugation of Cy5 using CuAAC

The Cy3 labeled bbRNA containing 3'-*O*-Propargyl group (bbRNA5) was conjugated to Cy5 azide using CuAAC to afford the dual labeled bbRNA for biochemical studies. The CuAAC reaction was performed by using bbRNA5 (containing both 5'-Dylight547 and 3'-alkyne, 100 μM final), Cy5 azide (250 μM final), 1X PBS (pH 7.4), $CuSO_4$ (1 mM final), THPTA (5 mM final), sodium ascorbate (5 mM final) and amino guanidinium chloride (4 mM final) in 50 μL total volume. The $CuSO_4$ and THPTA solution was premixed together before starting the reaction. Then all the components except sodium ascorbate and $CuSO_4$/THPTA were mixed together and degassed for 5 minutes by blowing argon. The sodium ascorbate and $CuSO_4$/THPTA solutions were also degassed separately by blowing argon for 5 minutes. Finally the reaction was started by adding the required amount of degassed sodium ascorbate and $CuSO_4$/THPTA solutions to the reaction mixture. The reactions were continued for 2 hours. Then the dual-labeled bbRNA was purified by HPLC and characterized by mass spectrometry. Figure 2.10 shows the HPLC chromatogram of the dual-labeled bbRNA and Figure 2.11 shows the MALDI mass spectra of the dual-labeled bbRNA.

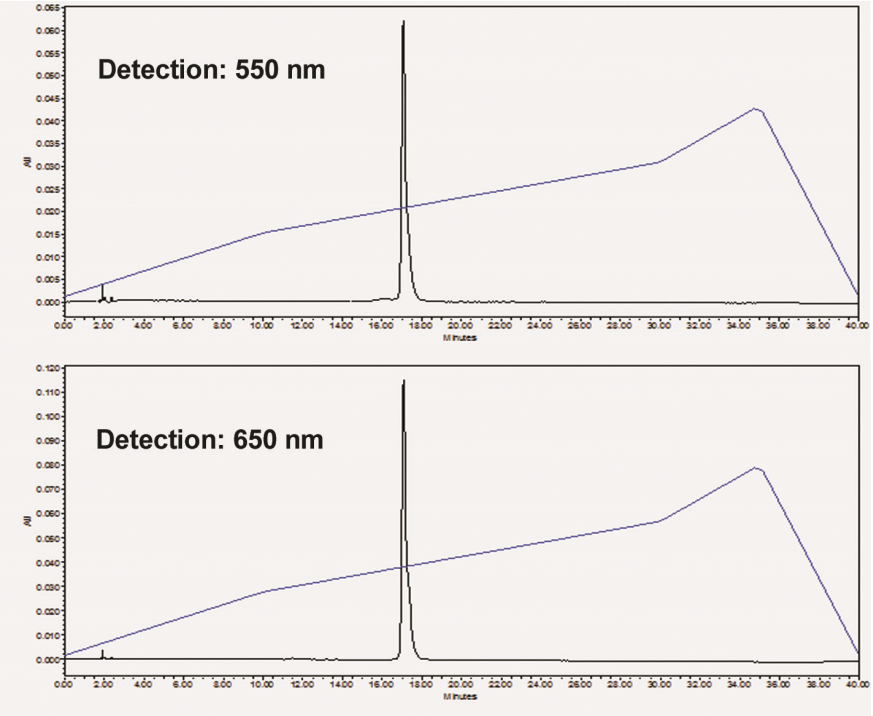


Figure 2.10. HPLC chromatogram of the purified dual-labeled bbRNA at 550nm (Cy3) and 650nm (Cy5). **HPLC condition**: Waters Xbridge C18 column (4.6x150 mm, 5 µm particle), Temp: 25˚C. Solvent A: 0.1M Triethylammonium acetate (TEAA) Buffer (pH 7), Solvent B: 80% Acetonitrile, 20% water, 0.1M TEAA buffer (pH 7), Gradient: 2.5% to 25% B in 10 mins, then 25% to 50% B in next 20 mins.

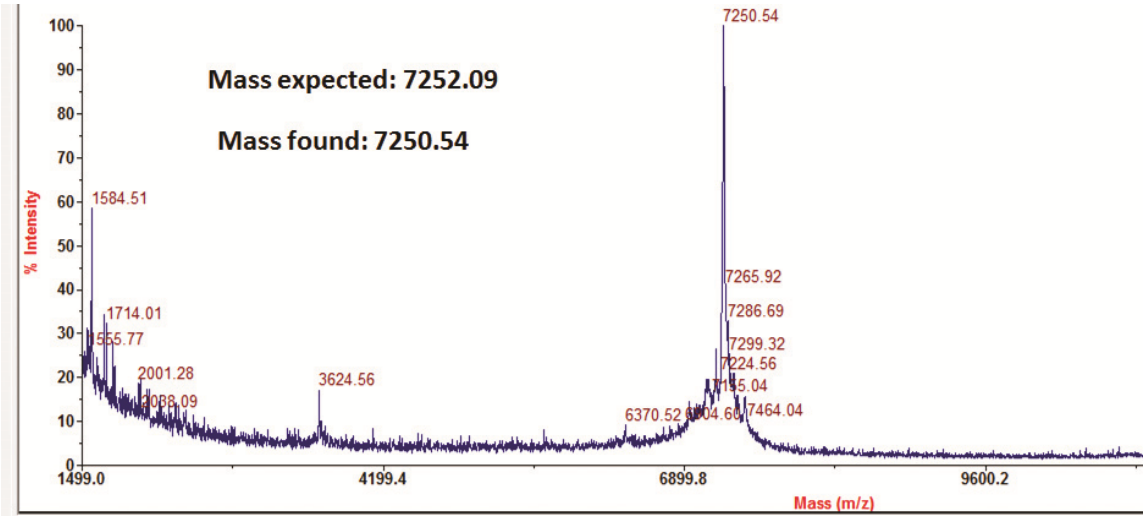


Figure 2.11. MALDI mass spectra of the dual-labeled (Cy3-Cy5) bbRNA.

Synthesis of the dual-labeled (Alexa488-Alexa594) bbRNAs for single molecule studies

Alexa488 labelling of the bbRNA

Following the synthesis of bbRNA containing 5'-amino modifier and 3'-Biotin-TEG in the stem and 3'-*O*-Propargyl group in the branch (Scheme 2.5), it was purified by HPLC. Then it was conjugated with isomerically pure Alexa488 NHS ester following protocol described by Life Technologies. First 1mg of the Alexa488 NHS ester was dissolved in 56 μL of dry DMSO. Then 7 μL of this solution was mixed 5 μL of 2mM HPLC purified bbRNA and 38 μL of freshly prepared 0.1M sodium tetraborate buffer (pH 8.5). The reaction was continued overnight with shaking. Then 5 μL of 3M NaOAc and 125 μL of chilled EtOH was added to the reaction mixture and the tube was kept in dry ice for 1 hour to precipitate the labeled and unlabeled bbRNA. Following this the tube was centrifuged at 14000 rpm for 30 mins to pellet out the precipitate. After removing the supernatant, the pellet was washed with chilled 70% EtOH. Then the pellet was air dried and finally dissolved in 100 μL of water. Then the Alexa488 labeled bbRNA was purified by HPLC.

Alexa594 labeling of the Alexa 488 labeled bbRNA using CuAAC

The Alexa594 azide was conjugated to the Alexa488 labeled bbRNA using CuAAC (Scheme 2.5) to afford the dual labeled bbRNA. It is important to mention that commercially available Alexa594 azide (Life Technologies) have two regioisomers. However we were able to separate those two regioisomers during HPLC purification of this dual-labeled bbRNA. The CuAAC reaction was performed by using Alexa488 labeled bbRNA (containing 3'-*O*-propargyl, 100 μM final), Alexa594 azide (250 μM final), 1X PBS (pH 7.4), $CuSO_4$ (1 mM final), THPTA (5 mM final), sodium ascorbate (5

mM final) and amino guanidinium chloride (4 mM final) in 50 μL total volume. The $CuSO_4$ and THPTA solution was premixed together before starting the reaction. Then all the components except sodium ascorbate and $CuSO_4$/THPTA were mixed together and degassed for 5 minutes by blowing argon. The sodium ascorbate and $CuSO_4$/THPTA solutions were also degassed separately by blowing argon for 5 minutes. Finally the reaction was started by adding the required amount of degassed sodium ascorbate and $CuSO_4$/THPTA solutions to the reaction mixture. The reactions were continued for 2 hours. Then the dual-labeled bbRNA was purified by HPLC and characterized by mass spectrometry. As it can be seen in the HPLC chromatogram (Fig. 2.12), there were two peaks corresponding to the Alexa488 and Alexa594 labeled bbRNA denoted by isomer 1 and isomer 2. These two peaks arise because of the two regioisomers of the Alexa594 dye. The two isomers of the Alexa594 azide can also be seen in the chromatogram as the two very close peaks around 32 minutes.

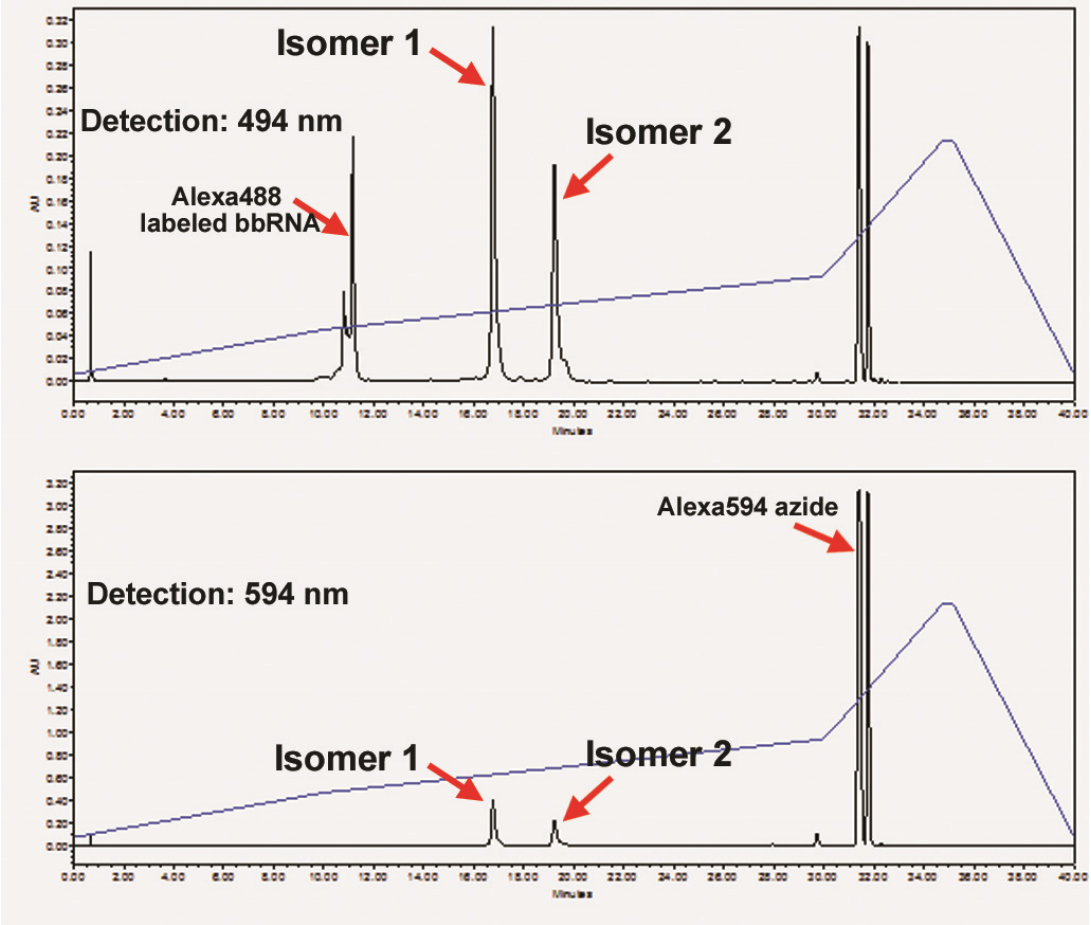


Figure 2.12. HPLC chromatogram of the CuAAC reaction between Alexa488 labeled bbRNA containing 3'-*O*-Propargyl and Alexa594 azide. **HPLC condition**: Waters Xbridge C18 column (4.6x150 mm, 5 µm particle), Temp: 25˚C. Solvent A: 0.1M Triethylammonium acetate (TEAA) Buffer (pH 7), Solvent B: 80% Acetonitrile, 20% water, 0.1M TEAA buffer (pH 7), Gradient: 2.5% to 15% B in 10 mins, then 15% to 30% B in next 20 mins.

Table 2.5 shows the sequence and MALDI mass spectroscopic characterization of the two dual labeled bbRNA synthesized for single molecule studies. The mass spectra of bbRNA6 is shown in Figure 2.4 to demonstrate that we can synthesize significantly large bbRNAs using this method.

Name	Sequence	Mass (calculated)	Mass Found
bbRNA6	**5'-Alexa488-ACU ACU GUA CUA A-(2'-5' GUAUGA-*t*-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**	**11341.8**	**11342.5**
bbRNA7	**5'-Alexa488-ACU ACU GUA CUA A-(2'-5' GUAUGA CUAUCC-*t*-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**	**13216.0**	**13218.9**

Table 2.5. Sequence and MALDI mass of the dual labeled bbRNAs for single molecule studies. The '*t'* signifies the triazole linkage formed between the 3'-*O*-Propargyl of the branch and Alexa594 azide.

Synthesis of 2'-5' triazole linked 'click' branched RNA for kinetics studies

The 'click' branched RNAs were synthesized by conjugating a stem strand containing an internal 2'-*O*-propargyl with the branch containing a 5'-azide using CuAAC reaction[55]. The stem strand (S1) containing a 2'-*O*-propargyl group was synthesized by regular solid phase synthesis using the commercially available 2'-*O*-propargyl adenosine (N-Bz) phosphoramidite at the branch point. The 5'-azide modification was introduced to the branch strand using post synthetic modification on the solid support following reported methods.[51,52]

Incorporation of 5'-azide on the branch strand

Following the synthesis of the linear branch strand (1 μmole), the 5'-*O*-DMT group was removed to generate a 5'-OH. Then the 5'-OH of the branch strand was modified to 5'-iodo by treating the CPG beads with 0.5 M methyltriphenoxyphosphonium iodide[51] solution in DMF for 15 minutes inside the synthesizer. Then the CPG beads were taken outside the synthesizer and treated with a saturated solution of sodium azide in

DMF for 90 minutes at 60˚C. This reaction converted the 5'-iodo group of the RNA to 5'-azido. Then the RNA was deprotected using standard protocol as described before.

Conjugation of the stem strand and branch strand to make cbRNA

The CuAAC reaction was performed by using the stem strand (containing internal 2'-*O*-Propargyl, 100 μM final), the branch strand (containing 5'-azide, 200 μM final), 1X PBS (pH 7.4), $CuSO_4$ (1 mM final), THPTA (5 mM final), sodium ascorbate (5 mM final) and amino guanidinium chloride (4 mM final) in 50 μL total volume. The $CuSO_4$ and THPTA solution was premixed together before starting the reaction. Then all the components except sodium ascorbate and $CuSO_4$/THPTA were mixed together and degassed for 5 minutes by blowing argon. The sodium ascorbate and $CuSO_4$/THPTA solutions were also degassed separately by blowing argon for 5 minutes. Finally the reaction was started by adding the required amount of degassed sodium ascorbate and $CuSO_4$/THPTA solutions to the reaction mixture. The reactions were continued for 2 hours and then 50 μL of STOP solution (90% formamide/10% 0.1M EDTA) was added to the reaction mixture. The reaction mixture was separated using 20% non-denaturing PAGE (8M Urea, 1X TBE). Following the loading of the reaction mixture, the gel was run (0.5X TBE running buffer) at 10W for 2 hrs to separate the starting materials and products (Fig. 2.5). Then the band corresponding to the cbRNA was cut and electroeluted at 250V using Elutrap electroelution device (GE Healthcare Life Sciences). Then the cbRNA was desalted using SepPak cartridges, lyophilized and finally dissolved in 20 μL of water. The HPLC and MALDI mass spectrum of the cbRNA is described in Figure 2.13.

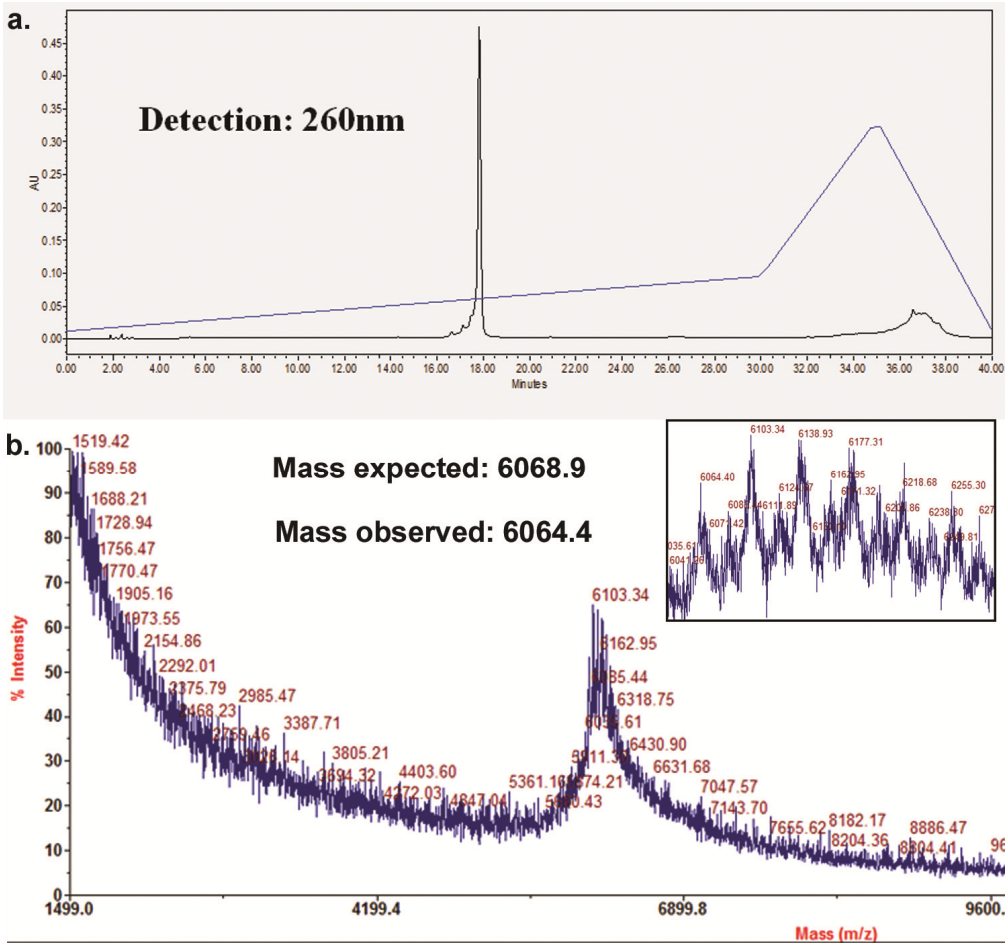


Figure 2.13. Characterization of the 2'-5' triazole linked cbRNA used for kinetics studies. **a.** HPLC chromatogram of the purified cbRNA **HPLC condition**: Waters Xbridge C18 column (4.6x150 mm, 5 µm particle), Temp: 25˚C. Solvent A: 0.1M Triethylammonium acetate (TEAA) Buffer (pH 7), Solvent B: 80% Acetonitrile, 20% water, 0.1M TEAA buffer (pH 7), Gradient: 2.5% to 20% B in 30 mins. **b.** MALDI mass spectrum of the cbRNA. Sequence of the cbRNA: 5'- GUA CUA A-(2'-*t*-5'-GUAUGA)-CAA GUU. The *t* signifies the triazole linkage between the stem and the branch.

Synthesis of dual-labeled 'click' branched RNAs for single molecule studies

The dual labeled cbRNA was synthesized by conjugating Alexa488 labeled stem strand containing internal 2'-*O*-Propargyl and Alexa594 labeled branch strand containing 5'-azide using CuAAC. Both the donor and acceptor dyes were incorporated into the stem strand and branch strand respectively using NHS ester chemistry.

Alexa488 labelling of the stem strand (S1)

Following the synthesis of the stem strand S1 containing 5'-amino modifier and internal 2'-*O*-Propargyl group at the branch point (Scheme 2.6), it conjugated with isomerically pure Alexa488 NHS ester using protocol described by Life Technologies. First 1mg of the Alexa488 NHS ester was dissolved in 56 µL of dry DMSO. Then 14 µL of this solution was mixed 10 µL of 2 mM stem strand and 76 µL of freshly prepared 0.1M sodium tetraborate buffer (pH 8.5). The reaction was continued overnight with shaking. Then 10 µL of 3M NaOAc and 250 µL of chilled EtOH were added to the reaction mixture and the tube was kept in dry ice for 1 hour to precipitate the labeled and unlabeled bbRNA. Following this the tube was centrifuged at 14000 rpm for 30 mins to pellet out the precipitate. After removing the supernatant, the pellet was washed with chilled 70% EtOH. Then the pellet was air dried and finally dissolved in 100 µL of water. Then the Alexa488 labeled S1 strand was purified by HPLC.

Alexa594 labelling of the branch strands (B1 and B2)

The branch strands were synthesized from a 3'-C6-phthalimidyl-amino modifier CPG. Following the synthesis of the RNA, the 5'-OH was converted into 5'-azide using the same protocol described in the previous section. Then the 5'-azido labeled branch was conjugated to Alexa594 dye at the 3' using Alexa594 NHS ester (Life Technologies) using the exact same protocol described before. Figure 2.14 shows the HPLC chromatogram of Alexa594 labeling of branch B1 (Sequence: 5'-N_3-GUAUGA-3'-C6 NH_2). The chromatogram shows two peaks designated as isomer 1 and isomer 2 which have absorption in both 260nm and 594nm channel. These two peaks were collected separately during HPLC and they were characterized using MALDI mass spectroscopy.

Figure 2.15 shows the mass spectra of the two different isomers of the Alexa594 labeled branch B1. The two isomers show almost the exact same mass indicating that the two peaks arises due to two regioisomers of the Alexa594 dye. I have used the isomer 1 for the synthesis of the dual-labeled cbRNA.

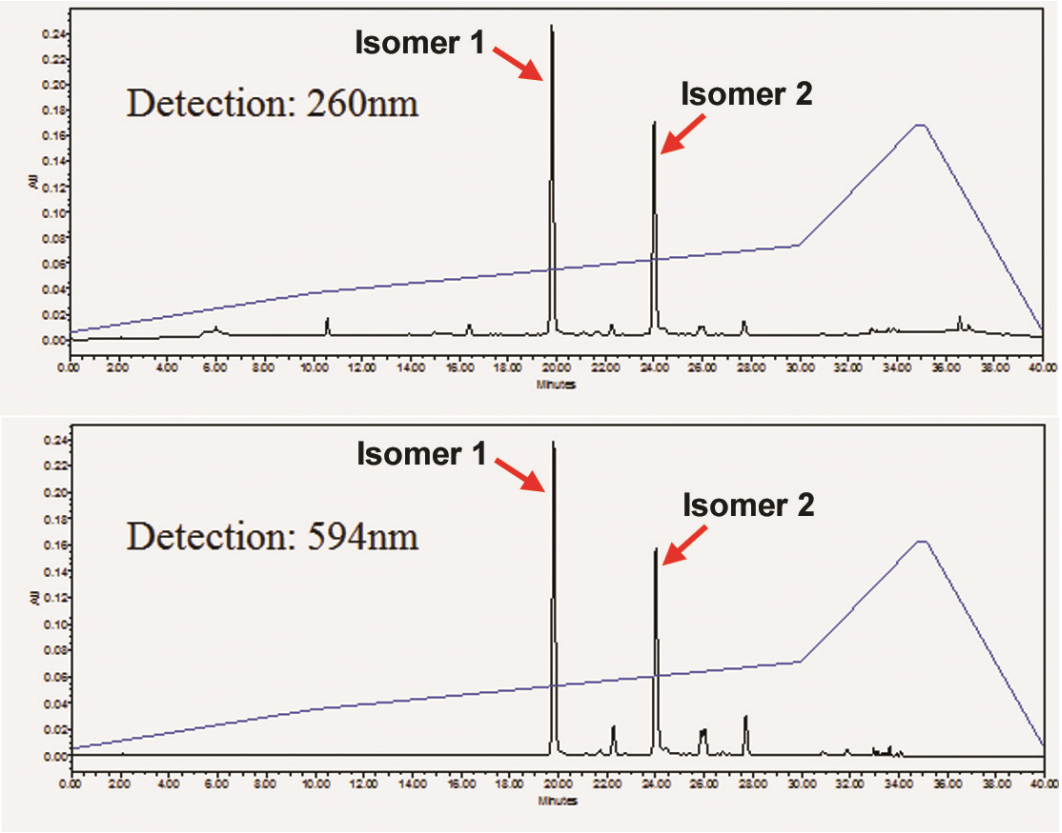


Figure 2.14. HPLC chromatogram of the Alexa594 labeling of branch strand B1 containing 5'-azide and 3'-C6-NH_2. **HPLC condition**: Waters Xbridge C18 column (4.6x150 mm, 5 µm particle), Temp: 25˚C. Solvent A: 0.1M Triethylammonium acetate (TEAA) Buffer (pH 7), Solvent B: 80% Acetonitrile, 20% water, 0.1M TEAA buffer (pH 7), Gradient: 2.5% to 15% B in 10 mins, then 15% to 30% B in next 20 mins.

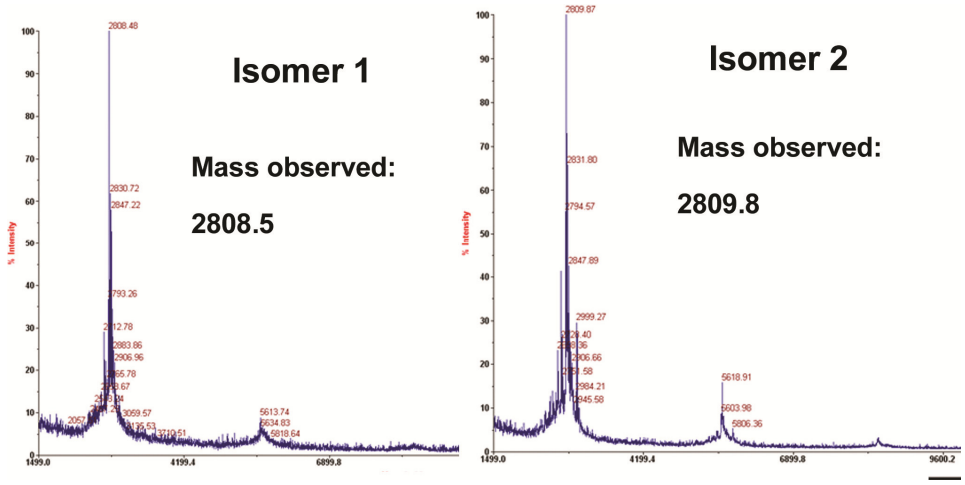


Figure 2.15. MALDI mass spectra of the two isomers of the Alexa594 labeled branch strand (B1). Mass expected: 2807.5.

Conjugation of Alexa488 labeled stem with A594 labeled branch to make dual-labeled cbRNAs

Following the purification of the Alexa488 labeled stem strand and Alexa594 labeled branch strands, they were conjugated using CuAAC. The CuAAC reaction was performed by using the stem strand (A488 S1 containing internal 2'-*O*-Propargyl, 100 µM final), the branch strand (B1-A594 or B2-A594 containing 5'-azide, 200 µM final), 1X PBS (pH 7.4), $CuSO_4$ (1 mM final), THPTA (5 mM final), sodium ascorbate (5 mM final) and amino guanidinium chloride (4 mM final) in 50 µL total volume. The $CuSO_4$ and THPTA solution was premixed together before starting the reaction. Then all the components except sodium ascorbate and $CuSO_4$/THPTA were mixed together and degassed for 5 minutes by blowing argon. The sodium ascorbate and $CuSO_4$/THPTA solutions were also degassed separately by blowing argon for 5 minutes. Finally the reaction was started by adding the required amount of degassed sodium ascorbate and $CuSO_4$/THPTA solutions to the reaction mixture. The reactions were continued for 2 hours and purified by HPLC. The HPLC chromatogram shows almost complete

conversion of the A488 labeled S1 stem (limiting reagent). The peak around 18 mins has both Alexa488 and Alexa594 absorption and therefore corresponds to the dual-labeled cbRNA. The peak at 16 mins is the unreacted A594 labeled branch B1 and therefore only appears in the 594nm channel (Fig. 2.16). As described before, I have synthesized three different dual-labeled cbRNAs (cbRNA 1-3) for single molecule studies having different branch length. They have respectively 0, 6 and 12 nucleotides in the 2'-branch. The sequences and MALDI mass characterization of the three cbRNAs and their precursors are provided in Table 2.6. All of these dual-labeled cbRNAs were used for single molecule studies which are described in chapter 7.

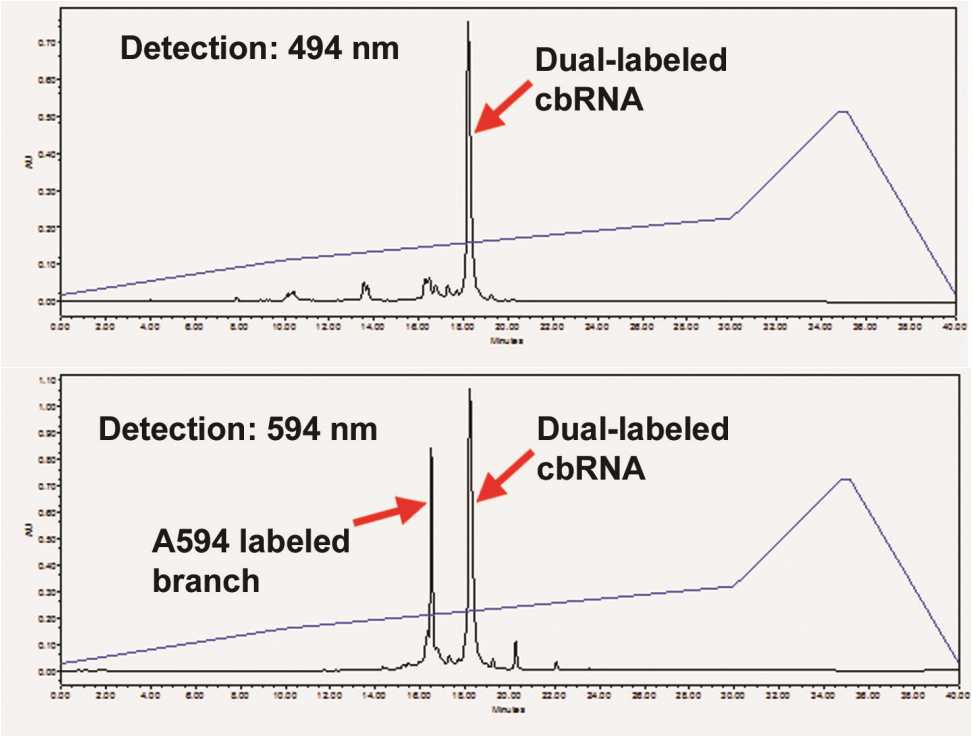


Figure 2.16. HPLC chromatogram of the CuAAC reaction between Alexa488 labeled stem containing 2'-O-Propargyl and Alexa594 labeled branch B1 containing 5'-azide. The peak around 18 mins have both Alexa488 and Alexa594 absorption and therefore corresponds to the dual-labeled cbRNA. **HPLC condition**: Waters Xbridge C18 column (4.6x150 mm, 5 µm particle), Temp: 25˚C. Solvent A: 0.1M Triethylammonium acetate (TEAA) Buffer (pH 7), Solvent B: 80% Acetonitrile, 20% water, 0.1M TEAA buffer (pH 7), Gradient: 2.5% to 15% B in 10 mins, then 15% to 30% B in next 20 mins.

Name	Sequence	Mass (calculated)	Mass Found
A488-S1	**5'-Alexa488-C6-ACU ACU GUA CUA A-(2'-*O*-Propargyl)-CAA GUU ACU U-3' Biotin-TEG**	**8536.3**	**8536.4**
B1-A594	**5'-N_3-GUAUGA-3'-C6-Alexa594**	**2807.5**	**2808.5**
B2-A594	**5'-N_3-GUAUGACUAUCC-3'-C6-Alexa594**	**4663.8**	**4665.9**
cbRNA1	**5'-Alexa488-C6-ACU ACU GUA CUA A-(2'-t--C6-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**	**9383.6**	**9386.2**
cbRNA2	**5'-Alexa488-C6-ACU ACU GUA CUA A-(2'-t-5' GUAUGA-C6-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**	**11343.8**	**11345.3**
cbRNA3	**5'-Alexa488-C6-ACU ACU GUA CUA A-(2'-t-5' GUAUGA CUAUCC-C6-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**	**13200.1**	**13202.9**

Table 2.6. Sequence and MALDI mass of the dual labeled cbRNAs and its precursors for single molecule studies. The *t* signifies the triazole linkage formed between the 2'O-propargylof the stem and 5'azide of the branch.

References

1. Wallace, J. C. & Edmonds, M. Polyadenylylated nuclear RNA contains branches. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 950–954 (1983).

2. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **38,** 317–331 (1984).

3. Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. Lariat RNAs as Intermediates and Products in the Splicing of Messenger RNA Precursors. *Science* **225,** 898–903 (1984).

4. Ruskins, B. & Green, M. R. An RNA Processing Activity that Debranches RNA Lariats. *Science* **229,** 135–140 (1985).

5. Katolik, A. *et al.* Regiospecific solid-phase synthesis of branched oligoribonucleotides that mimic intronic lariat RNA intermediates. *J. Org. Chem.* **79,** 963–975 (2014).

6. Nam, K. *et al.* Yeast Lariat Debranching Enzyme: Substrate and sequence specificity. *J. Biol. Chem.* **269,** 20613–20621 (1994).

7. Damha, M. J., Pon, R. T. & Ogilvie, K. K. Chemical Synthesis of branched RNA: novel trinucleoside diphosphate containing vicinal 2'-5' AND 3'-5' phosphodiester linkage. *Tetrahedron Lett.* **26,** 4839–4842 (1985).

8. Kierzek, R., Kopp, D. W., Edmonds, M. & Caruthers, M. H. Chemical synthesis of branched RNA. *Nucleic Acids Res.* **14,** 4751–4764 (1986).

9. Zhou, X. X., Nyilas, A., Remaud, G. & Chattopadhyaya, J. Regiospecific Synthesis of Branched Tetranucleotides - U3'p5'A2'p5'G3'p5'U, U3'p5'A2'p5'G3'p5'C, A3'p5'A2'5p'G3'p5'U and A3'p5'A2'p5'G3'p5'C. *Tetrahedron* **43,** 4685–4698 (1987).

10. Vial, J.-M., Balgobin, N., Remaud, G., Nyilas, A. & Chattopadhyaya, J. A

Convergent Regiospecific Synthesis of the Lariat-Trinucleotides and A2′p 5′G and A 2′p 5′G 3′p 5′C from A O 4 -(2-Nitrophenyl)-Uridine Building Block. *Nucleosides and Nucleotides* **6,** 209–226 (1987).

11. Fourrey, J. L., Varenne, J., Fontaine, C., Guittet, E. & Yang, Z. W. A new method for the synthesis of branched ribonucleotides. *Tetrahedron Lett.* **28,** 1769–1771 (1987).

12. Balgobin, N., Foldesi, A., Remand, G. & Chattopadhyaya, J. A new regiospecific synthesis of 'Branched' tetraribonucleotide and its three analogues to delineate the chemospecific role of the 'Branch-point' adenine nucleotide in splicing. *Tetrahedron* **44,** 6929–6939 (1988).

13. Foldesi, A., Balgobin, N. & Chattopadhyaya, J. Synthesis of 'Branched' trinucleotide using the H-phosphonate chemistry. *Tetrahedron Lett.* **30,** 647–650 (1989).

14. Damha, M. J. & Ogilvie, K. K. Synthesis and Spectroscopic Analysis of Branched RNA Fragments: Messenger RNA Splicing Intermediates. *J. Org. Chem.* **3722,** 3710–3722 (1988).

15. Zhou, X., Remaud, G. & Chattopadhyaya, J. New regiospecific synthesis of the 'branched' tri-, penta- & hepta-ribonucleic acids which are formed as the 'lariat' in the pre-mRNA processing reactions[Splicing]. *Tetrahedron* **44,** 6471–6489 (1988).

16. Zhou, X.-X., Nyilas, A., Remaud, G. & Chattopadhyaya, J. 270 MHz 1H-MNR Studies of four 'branched' tetraribonucleotides: A3'p5'A2'p5'G3'p5'U, A3'p5'A2'p5'G'3'p5'U, U3'p5'A2'p5'G3'p5'U & U3'p5'A2'p5'G3'p5'C which are formed as the lariat branch-point in the pre-mRNA processing reactions (splicing). *Tetrahedron* **44,** 571–589 (1988).

17. Sund, C., Agback, P. & Chattopadhyaya, J. Synthesis of Tetrameric cyclic branched-RNA (Lariat) modelling the Introns of Group II and Nuclear pre-mRNA Processing Reaction (Splicing). *Tetrahedron* **47,** 9659–9674 (1991).

18. Sund, C., Agback, P. & Chattopadhyaya, J. Synthesis of heptameric lariat-RNA modelling the lariat introns of group II and nuclear pre-mRNA processing reaction (splicing). *Tetrahedron* **49,** 649–668 (1993).

19. Reese, C. B. & Song, Q. A new approach to the synthesis of branched and branched cyclic oligoribonucleotides. *Nucleic Acids Res.* **27,** 2672–2681 (1999).

20. Wang, Y. & Silverman, S. K. Deoxyribozymes That Synthesize Branched and Lariat RNA. *J. Am. Chem. Soc.* **125,** 6880–6881 (2003).

21. Wang, Y. & Silverman, S. K. Characterization of Deoxyribozymes That Synthesize Branched RNA. *Biochemistry* **42,** 15252–15263 (2003).

22. Pratico, E. D., Wang, Y. & Silverman, S. K. A deoxyribozyme that synthesizes 2',5'-branched RNA with any branch-site nucleotide. *Nucleic Acids Res.* **33,** 3503–3512 (2005).

23. Wang, Y. & Silverman, S. K. Efficient one-step synthesis of biologically related lariat RNAs by a deoxyribozyme. *Angew. Chemie Int. Ed.* **44,** 5863–5866 (2005).

24. Lee, C. S., Mui, T. P. & Silverman, S. K. Improved deoxyribozymes for synthesis of covalently branched DNA and RNA. *Nucleic Acids Res.* **39,** 269–279 (2011).

25. Carriero, S. & Damha, M. J. Template-Mediated Synthesis of Lariat RNA and DNA introns. *J. Org. Chem.* **68,** 8328–8338 (2003).

26. Mitra, D. & Damha, M. J. A novel approach to the synthesis of DNA and RNA lariats. *J. Org. Chem.* **72,** 9491–9500 (2007).

27. Komatsu, Y. & Ohtsuka, E. *Chemical RNA Synthesis (Including RNA with Unusual Constituents)*. *RNA* (2001). doi:10.1016/B978-008043408-7/50027-7

28. Marshall, W. S. & Kaiser, R. J. Recent advances in the high-speed solid phase synthesis of RNA. *Curr. Opin. Chem. Biol.* **8,** 222–229 (2004).

29. Damha, M. J. & Zabarylo, S. Automated solid-phase synthesis of branched oligonucleotides. *Tetrahedron Lett.* **30,** 6295–6298 (1989).

30. Damha, M. J., Ganeshan, K., Hudson, R. H. & Zabarylo, S. V. Solid-phase synthesis of branched oligoribonucleotides related to messenger RNA splicing intermediates. *Nucleic Acids Res.* **20,** 6565–6573 (1992).

31. Carriero, S. & Damha, M. J. Solid-phase synthesis of branched oligonucleotides. *Curr. Protoc. Nucleic Acid Chem.* 4.14.1-4.14.32 (2002). doi:10.1002/0471142700.nc0414s09

32. Ganeshan, K. *et al.* Novel approaches to the synthesis and analysis of branched RNA. *Nucleosides Nucleotides* **14,** 1009–1013 (1995).

33. Braich, R. S. & Damha, M. J. Regiospecific solid-phase synthesis of branched oligonucleotides. Effect of vicinal 2',5'- (or 2',3'-) and 3',5'-phosphodiester linkages on the formation of hairpin DNA. *Bioconjug. Chem.* **8,** 370–377 (1997).

34. Gretli, M., Eritja, R. & Sproat, B. Solid-phase Synthesis of Branched RNA and Branched DNA / RNA Chimeras. *Tetrahedron* **53,** 11317–11346 (1997).

35. Damha, M. J. & Braich, R. S. Synthesis of a branched DNA/RNA chimera similar to the msDNA molecule of Myxococcus xanthus. *Tetrahedron Lett.* **39,** 3907–3910 (1998).

36. Srivastava, S. C., Pandey, D., Srivastava, N. P. & Bajpai, S. P. RNA synthesis by reverse direction Process: Phosphoramidites and high purity rnas and introduction of ligands, chromophores, and modifications at 3'-End. *Curr. Protoc. Nucleic Acid Chem.* 3.20.1-3.20.39 (2011). doi:10.1002/0471142700.nc0320s45

37. Ohtsuka, E. *et al.* Studies on tRNA and related compounds. XXXVII. Synthesis and physical properties of 2'- or 3'-O-(o-nitrobenzyl)nucleosides: the use of o-nitrophenyldiazomethane as a synthetic reagent. *Chem. Pharm. Bull. (Tokyo).* **29,** 318–324 (1981).

38. Das, S. R. & Piccirilli, J. a. General acid catalysis by the hepatitis delta virus ribozyme. *Nat. Chem. Biol.* **1,** 45–52 (2005).

39. Somoza, A. Protecting groups for RNA synthesis: an increasing need for selective preparative methods. *Chem. Soc. Rev.* **37,** 2668–2675 (2008).

40. Chaulk, S. G. & MacMillan, A. M. Synthesis of oligo-RNAs with photocaged adenosine 2'-hydroxyls. *Nat. Protoc.* **2,** 1052–8 (2007).

41. Pitsch, S. An Efficient Synthesis of Enantiomeric Ribonucleic Acids from D-Glucose. *Helvectica Chim. Acta* **80,** 2286–2314 (1997).

42. Connor, D. S., Klein, G. W., Taylor, G. N., Boeckman, R. K. J. & Medwid, J. B. Benzyl Chloromethyl Ether. *Org. Synth. Coll.* **6,** 101–104 (1988).

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

44. Panchuk-Voloshina, N. *et al.* Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. *J. Histochem. Cytochem.* **47,** 1179–1188 (1999).

45. Mao, F., Leung, W.-Y. & Haugland, R. P. Sulfonated Xanthene Derivatives. (2000).

46. Tago, N. *et al.* Design, Synthesis, and Properties of Phosphoramidate 2′,5′-Linked Branched RNA: Toward the Rational Design of Inhibitors of the RNA Lariat Debranching Enzyme. *J. Org. Chem.* **80,** 10108–10118 (2015).

47. El-Sagheer, A. H., Sanzone, A. P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli. *Proc. Natl. Acad. Sci.* **108,** 11338–11343 (2011).

48. El-Sagheer, A. H. & Brown, T. New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15329–15334 (2010).

49. El-Sagheer, A. H. & Brown, T. Synthesis and polymerase chain reaction amplification of DNA strands containing an unnatural triazole linkage. *J. Am.*

Chem. Soc. **131,** 3958–3964 (2009).

50. Chen, X., El-Sagheer, A. H. & Brown, T. Reverse transcription through a bulky triazole linkage in RNA: implications for RNA sequencing. *Chem. Commun. (Camb).* **50,** 7597–600 (2014).

51. Miller, G. P. & Kool, E. T. A simple method for electrophilic functionalization of DNA. *Org. Lett.* **4,** 3599–3601 (2002).

52. Miller, G. P. & Kool, E. T. Versatile 5'-Functionalization of Oligonucleotides on Solid Support: Amines, Azides, Thiols, and Thioethers via Phosphorus Chemistry. *J. Org. Chem.* **69,** 2404–2410 (2004).

53. Hong, V., Presolski, S. I., Ma, C. & Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed. Engl.* **48,** 9879–83 (2009).

54. Presolski, S. I., Hong, V. P. & Finn, M. G. Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Curr. Protoc. Chem. Biol.* **3,** 153–162 (2011).

55. Mack, S., Fouz, M. F., Dey, S. K. & Das, S. R. in *Current Protocols in Chemical Biology* 83–95 (John Wiley & Sons, Inc., 2016). doi:10.1002/cpch.1

Chapter 3

Biochemical studies of lariat debranching enzyme using backbone branched RNA substrates

3.1. Introduction

The roles of debranching in different cellular regulatory processes have been discussed in chapter 1. Although Dbr1p was initially thought to be only important for intron turnover, other regulatory roles of Dbr1p such as miRNA and snoRNA biogenesis have been discovered.[1,2] Dbr1p is a Mn^{2+} dependent phosphodiesterase that selectively cleaves the 2'-5' phosphodiester bond without affecting the vicinal 3'-5' phosphodiester bond in intronic lariats. A lariat is not essential and Dbr1p also hydrolyzes the 2′-5′ phosphodiester linkage found in Y shaped branched RNAs and multicopy single stranded DNAs (msDNAs) as well.[3,4] Hydrolysis of the 2'-5' phosphodiester bond by Dbr1p yields a 5'-phosphate end and a 2'-hydroxyl at the branch point residue. Dbr1p has sequence homology with other non-specific nucleases such as Mre11, which cleaves 3'-5' phosphodiester bonds in DNA non-specifically. Therefore it is important to understand how Dbr1p can selectively cleave only 2'-5' phosphodiester bond despite having structural similarities with non-specific nucleases. To recognize its DNA substrate, Mre11 depends only on the DNA phosphate backbone, as expected for a non-specific nuclease. However, for Dbr1p, specific cleavage of only the 2'-5'-phosphodiester bond in a branched RNA likely requires a different mechanism of recognition that it could use to distinguish the RNA specific 2'-5' linkage. To understand the specificity of Dbr1p, biochemical studies with different substrates and Dbr1p enzymes could be used. Although the DBR1 gene was isolated in 1991,[5] debranching activity was first observed

in 1985 from HeLa cell extract.[6] Since then many biochemical studies were performed on Dbr1p using either lariat RNAs or synthetic 2'-5' phosphodiester linked substrates.

Dbr1p was shown to be unable to cleave a substrate containing only 2'-5' phosphate without any nucleotides.[7] Additionally, a 3'-nucleotide adjacent to the 2'-branch point is also required for efficient cleavage; a 3'-hydroxyl does not undergo any reaction and a 3'-phosphate does not support efficient cleavage.[3,6,8] Requirements of the stem and 2'-branch nucleotide for efficient cleavage suggest that the RNA substrate either acts to stabilize the active site or may even act in catalysis.[7] It has also been shown that non-conserved branchpoints of C and G are not efficiently debranched by Dbr1p.[7,9] Synthetic branches and lariats with non-natural branch residues also show differing cleavage efficiencies.[10] This suggests that Dbr1p recognizes both the 2'- and 3'-residues at the branchpoint as well as the branchpoint nucleotide itself. Conversely, the identity of nucleotide 5' of the branchpoint residue contributes little to efficient cleavage; in fact, the nucleotide is not even required for efficient cleavage.[11]

Features of the 2'-branch sequence of the RNA substrate are essential for debranching activity. Initial indication of the essential nature of the 2'-branch sequence arises from studies of Dbr1p acting on a trinucleotide substrate but not a branched structure with only the 2'-phosphate.[7] Nam *et al.* showed that Dbr1p can tolerate the first residue off the 2'-branch to be even deoxy nucleotide.[3] However a purine rather than a pyrimidine nucleobase is better tolerated at the 2'-residue. Similar results were also obtained using synthetic bbRNAs containing ribonucleotides in the 2'-branch. Additionally, the pro-R_P oxygen of the 2'-5'-phosphate linkage is essential for debranching activity while the pro-S_P oxygen is not.[12] Thus, these residues of the bbRNA

substrate are required for efficient catalysis; however, the role each of the elements play (structural or catalytic) is yet to be determined. The crystal structure of Dbr1p from *E. histolytica* has also been reported recently by Montemeyer *et al.*[13] Although they were able to identify the important amino acid residues in the active site, they could only observe one metal ion in the active site and the exact mode of binding between the bbRNA and Dbr1p still remains elusive.

Although some of the substrate requirements or effects on Dbr1p are reported, the branched RNA substrates used for most of these studies were non-native sequences.[3,14] In some studies the substrates were RNA-DNA hybrid, whereas in others the substrates had same sequences in the 2'- and the 3'-arm of the branched RNAs. Additionally typically only relative reactivity of the substrates with Dbr1p at a specific time point is reported with little information about the cleavage kinetics. As the lariat RNA has conserved sequences around the branchpoint phosphodiester bond, some modifications of the nucleotides are not tolerated by the splicing machinery and thus hard to access. Therefore easy access to both native sequences and modified bbRNA substrates will allow us to study the Dbr1p mechanism in much greater details. In the previous chapter I have described a solid phase synthesis method using photoprotected amidite to produce different bbRNAs that allow for readily incorporating modifications to the 2'-branch. Our synthetic strategy for access to bbRNAs is efficient and readily accommodates modifications. In this chapter I will describe the debranching reaction of several modified bbRNA substrates. Additionally, I compare two different Dbr1p enzymes from - *Saccharomyces cerevisiae* (ScDbr1p) and *Entamoeba Histolytica* (EhDbr1p) in debranching. The ScDbr1p is the most biochemically characterized enzyme[3,14] and the

crystal structure of the EhDbr1p enzyme was published recently[13]. However no biochemical characterization is available for EhDbr1p. We first characterized these two Dbr1p variants in terms of their metal ion requirements. Then we compared the debranching reaction of these two Dbr1p enzymes using the modified substrates described in chapter 2. The following section discusses the metal ion requirement of the two different Dbr1p enzymes.

3.2. Metal ion requirement of different Dbr1p enzymes

Similar to most nucleases, Dbr1p also requires divalent metal ions for catalysis. During the early studies of ScDbr1p and hDbr1p (human Dbr1p), Mg^{2+} was used as the divalent metal ion for the debranching reaction.[3,7] However it was shown later that Mn^{2+} is a better cofactor for ScDbr1p[14] given the homology of its active site to MRE11 that is structurally well characterized.[15] Although the crystal structure of EhDbr1p has been recently described,[13] it is not well characterized in terms of its metal ion requirement. Therefore we compared the ScDbr1p and EhDbr1p in terms of their metal ion requirement for debranching activity. To compare the debranching activity under different metal ion conditions, previous reports have almost always relied upon a single time point measurements.[3,14] Therefore we have developed a debranching assay where the rate of the reaction could be measured and then compared under different conditions (Fig. 3.1). The progress of the debranching reaction was monitored by taking aliquot from a debranching reaction mixture at different time points. The debranching reaction at each time point was stopped by addition of a STOP solution (90% formamide and 10% 0.1 M EDTA) which inactivated Dbr1p. Then the reaction mixtures at different time points were analyzed by gel electrophoresis. We used bbRNA substrate with 5'-

Dylight547 dye for these studies. The 5'-Dylight547 dye helped in visualization and quantification of the product and substrate bands without the use of any radioactive labeling. Following this, the fractions of the substrate that remains uncleaved at each time point was plotted against the time. The plot was fitted with a single exponential decay to calculate the observed rate constant k_{obs}. It is important to mention here that we have used sub-saturating enzyme condition for the biochemical studies as single turnover reactions are too fast to study using manual methods.

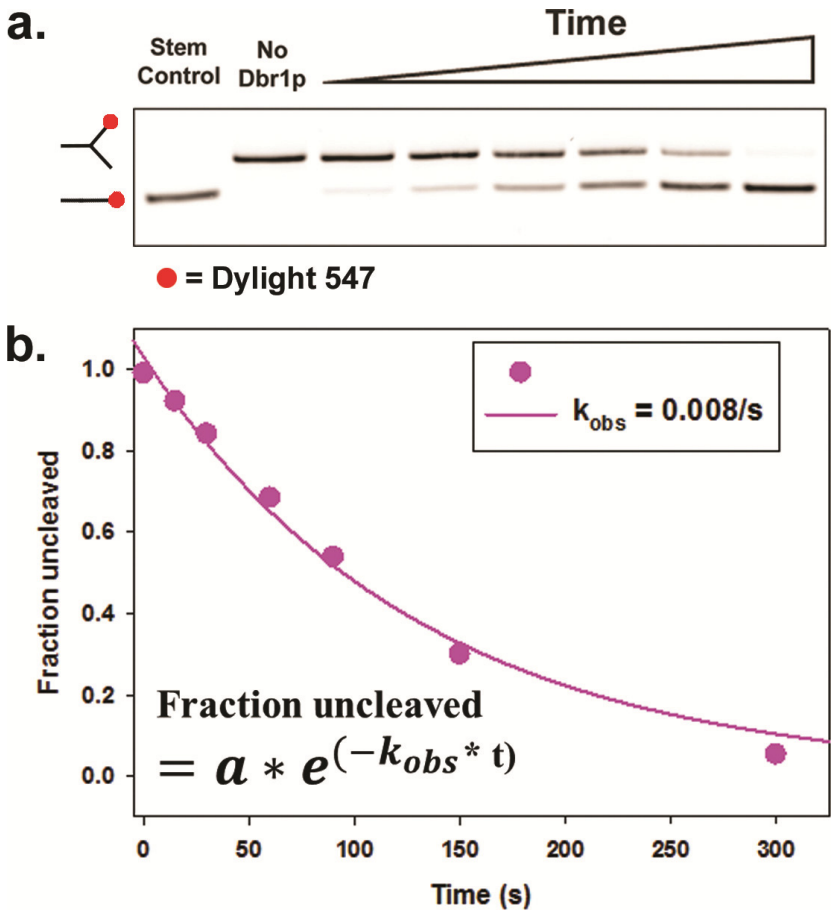


Figure 3.1. Quantitative approach for biochemical studies of the debranching reaction. **a.** Gel showing the progress of the debranching reaction by ScDbr1p at different time points. The first two lanes on the left are the control lane. The fraction uncleaved at time each time point was calculated by quantifying the intensity of the substrate and product bands for each reaction. **b.** The observed rate of the debranching (k_{obs}) reaction was calculated by fitting the fraction uncleaved vs time data with a single decay exponential.

We tested the effect of Mn^{2+} concentration on the debranching reaction of EhDbr1p and ScDbr1p using bbRNA1 substrate (experimental section, Table 3.1). The bbRNA1, used for these studies, is a 19 mer substrate containing six nucleotides in each of the arms and is the yeast consensus sequence. Figure 3.2 shows the plot of the k_{obs} values against the Mn^{2+} ion concentration for the two different enzymes. It can be concluded from this plot that the EhDbr1p and ScDbr1p have different Mn^{2+} ion requirement for debranching activity. Primarily the EhDbr1p has highest debranching activity at much lower concentration of Mn^{2+} (0.5 mM) compared to the ScDbr1p (4 mM). Also the activity of the enzymes followed a general trend of increase-maximum-decrease with increase in the Mn^{2+} concentration. Most interestingly, EhDbr1p has high activity (~60% of the maximum) without any added Mn^{2+} ion (0 mM) even in presence of 1 mM EDTA (Fig. 3.2). ScDbr1p however has almost no activity under exactly similar condition. This suggests that the Mn^{2+} ion is most probably bound very tightly to the active site of the EhDbr1p enzyme so that even EDTA could not chelate it. The ScDbr1p active site probably has lower binding affinity for the Mn^{2+} ion and therefore Mn^{2+} present in the active site could be easily chelated by EDTA rendering the enzyme inactive. This difference between the metal ion affinities of the two enzymes is an important factor to be considered for biochemical and structural studies. For all the further biochemical studies we have used 1 mM Mn^{2+} for EhDbr1p and 4 mM Mn^{2+} for ScDbr1p in the reaction buffer. The following section describes the biochemical studies of the different modified substrates for these two enzymes.

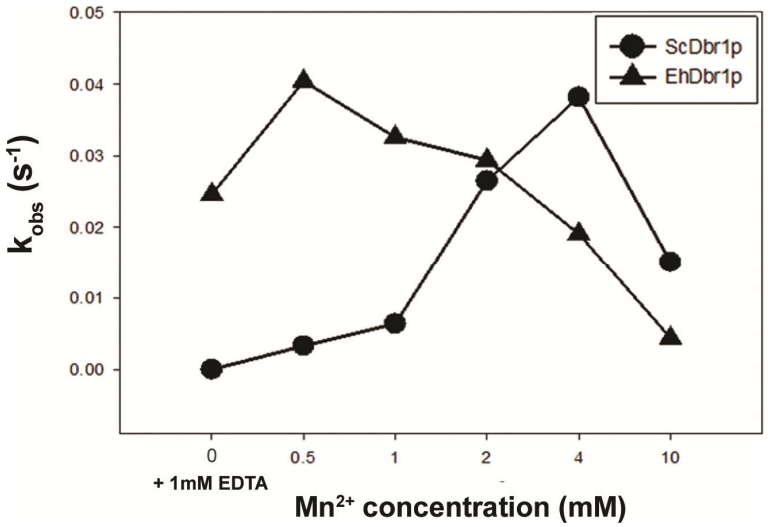


Figure 3.2. Mn^{2+} ion requirement for debranching activity of EhDbr1p and ScDbr1p. The k_{obs} values were calculated for each concentration using the method described in section 3.2. The 0 mM Mn^{2+} also contained 1 mM EDTA in the reaction mixture.

3.3. Biochemical studies of bbRNA substrates with modified 2'-branch

As discussed in section 3.1, the nucleotides in the 2'-branch of the bbRNA substrate is important for debranching reaction. Preliminary results from Das lab (Dr. Eduardo Paredes, Molly Evans and Dr Debasish Grahacharya, unpublished data) suggested that the 2'-OH of the first G nucleotide in the 2'-branch significantly affected for debranching activity by ScDbr1p. Although a substrate containing deoxyG as the first nucleotide in the 2'-branch has been shown to be cleaved by ScDbr1p, the rate of the debranching reaction has never been compared with the corresponding riboG containing substrate. Along with the native substrate, we have synthesized three more bbRNA substrates containing modified first nucleotide in the 2'-branch (Table 3.1). As shown in Figure 3.3a, the bbRNA substrates used for these studies are identical in very way except the 2'-subsituent X in the first nucleotide of the branch.

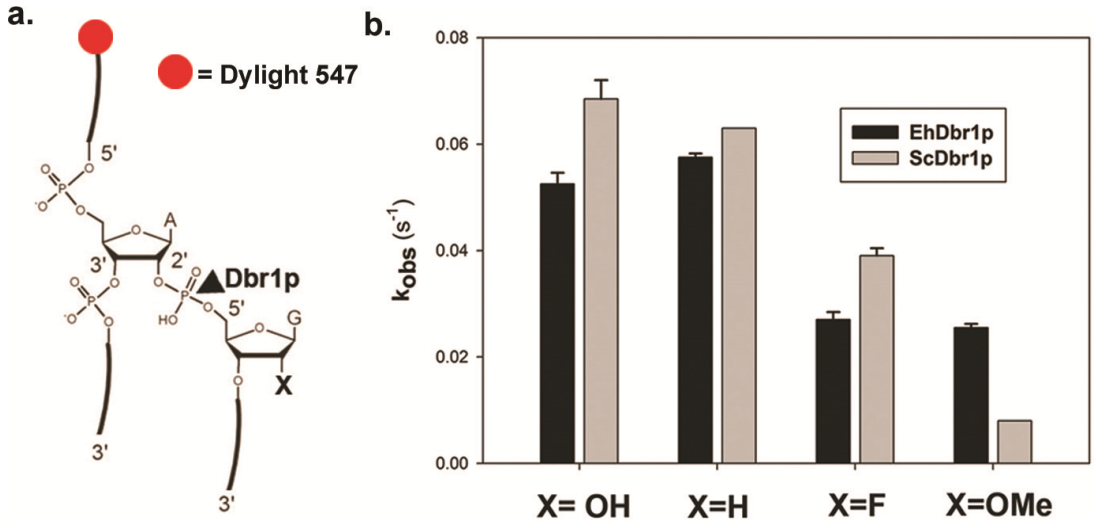


Figure 3.3. Debranching reaction of modified bbRNA substrates by EhDbr1p and ScDbr1p. **a.** The structure of the 5'-Dylight547 (red circle) modified bbRNA substrates used for the debranching studies. The arrow indicates the 2'-5' phosphodiester bond cleaved by Dbr1p. **b.** Bar diagram of the rate constants (k_{obs}) of debranching for the modified bbRNA substrates by ScDbr1p and EhDbr1p. The experiments were performed in triplicate and the error bars indicates standard deviation.

The observed reaction rate constants (k_{obs}) obtained for the different substrates with both enzymes are plotted in Figure 3.3b. Firstly we did not find any observable difference in reactivity between bbRNA1 and bbRNA2 (X=OH and H respectively) for both the enzymes. This suggests that a deoxyribonuclotide at the 2'-branch similar to that of msDNA is well tolerated by Dbr1p. For bbRNA3 (X=F) we observed that the rate of the debranching reaction reduced to almost half, indicating that the highly electronegative fluorine atom is not preferred in the 2'-position of the first nucleotide. The 2'-OMe modified substrate (bbRNA4) also showed slower debranching than the native substrate containing a 2'-OH (bbRNA1). Finally both the enzymes showed very similar activities for all the different substrates except bbRNA4 containing an X = OMe. The ScDbr1p shows almost three times lower k_{obs} values for bbRNA4 compared to EhDbr1p. This is probably because the bulky OMe group is interacting more closely with the active site

residues of the ScDbr1p compared to the EhDbr1p enzyme. Although we have seen some differences in the debranching activity of the different modified substrates, the exact reason for this difference – whether due to association of the substrate or due to a direct effect on the cleavage step – still remains to be uncovered.

3.4. Conclusion

In this chapter, I have described the biochemical studies of the debranching reaction of several modified bbRNA substrates. We have developed a debranching assay using bbRNAs containing 5'-Dylight547 dye. The Dylight547 dye has been used for detection and quantification of the substrate and cleavage product. The bbRNAs used for this assay were synthesized on the solid support without the need of any post synthetic modification. We have also avoided the use of any radioactive label for our debranching studies. Using this assay we first show that the two mostly studied lariat debranching enzyme (ScDbr1p and EhDbr1p) have striking difference in their metal ion (Mn^{2+}) requirement for debranching activity. This difference in activity arises probably because of the nature of binding of the metal ion in the active site of these two enzymes. Then we used this assay to show the difference in the debranching activity of several modified bbRNA substrates. These experiments suggest that an electronegative fluorine atom and a bulky methyl group is not well tolerated by both the Dbr1p. Although we have observed some difference in the reactivity of the bbRNA substrates, we still do not know the exact reason for those differences. One would require more detailed kinetics studies to uncover the true reason (binding or catalysis) for the difference in the reactivity obtained. A fluorescence based debranching kinetics assay is discussed in the next chapter.

3.5. Experimental

Name	X=	Sequence
bbRNA1	OH	5'-Dy547-GUACUAA-(2'-5'-GUAUGA)- CAAGUU-3'
bbRNA2	H	5'-Dy547-GUACUAA-(2'-5'-gUAUGA)-CAAGUU-3'
bbRNA3	F	5'-Dy547-GUACUAA-(2'-5'-(2'F-G)UAUGA) -CAAGUU-3'
bbRNA4	OMe	5'-Dy547-GUACUAA-(2'-5'-(2'-OMe-G)UAUGA)-CAA GUU-3'

Table 3.1. Sequence of the bbRNAs used for the biochemical studies.

Debranching reactions for Mn^{2+} dependence

The debranching reactions for the Mn^{2+} dependence were performed with bbRNA1 substrate in 1X reaction buffer (50 mM Tris-HCl (pH 7.5), 25 mM NaCl, Mn^{2+} (as required), 2.5 mM DTT, 0.1% TritonX-100). Table 3.2 shows the concentration and volume of the different components used for the reaction. First all the components of this reaction except the Dbr1p enzyme were mixed together and then the debranching reaction was started by addition of Dbr1p. At the desired time point a 20 µL solution of the reaction mixture was treated with 40 µL of STOP solution (90% formamide and 10% 0.1M EDTA) to inactivate the debranching reaction. The reaction mixtures were separated using analytical 20% denaturing PAGE (1X TBE buffer). 15 µL of the stopped reaction mixture was loaded in each of the wells and then the gel was run (0.5X TBE running buffer) at 250V for 3 hrs to separate the starting materials from the products. Then the gel was scanned in a Typhoon FLA900 (GE healthcare life sciences) fluorescence scanner using the settings for Cy3 dye. The bands were quantified using the

associated software to calculate the fraction cleaved. Then the fraction cleaved at each time point was plotted against the time and it was fitted with a single decay exponential using SigmaPlot. The decay constant signifies the k_{obs} values of the particular reaction.

Component	Stock concentration	Volume used (μL)	Final concentration
Dual-labeled bbRNA	1μM	20	100nM
Dbr1p (in1X Buffer)	10nM	20	1nM
2X reaction buffer	2X	90	1X
Water	-	70	-
Total		200	

Table 3.2. Reaction condition for the debranching of bbRNA1 by Dbr1p.

Debranching reactions for of the modified substrate

The debranching reactions for the modified substrates (bbRNA1-4) were performed using the exact same condition as described in the previous section except one change. The 1X debranching reaction buffer contained 4 mM Mn^{2+} for ScDbr1p and 1 mM Mn^{2+} for EhDbr1p. Figure 3.4 shows the debranching reaction rate of the different substrates for the two different enzymes.

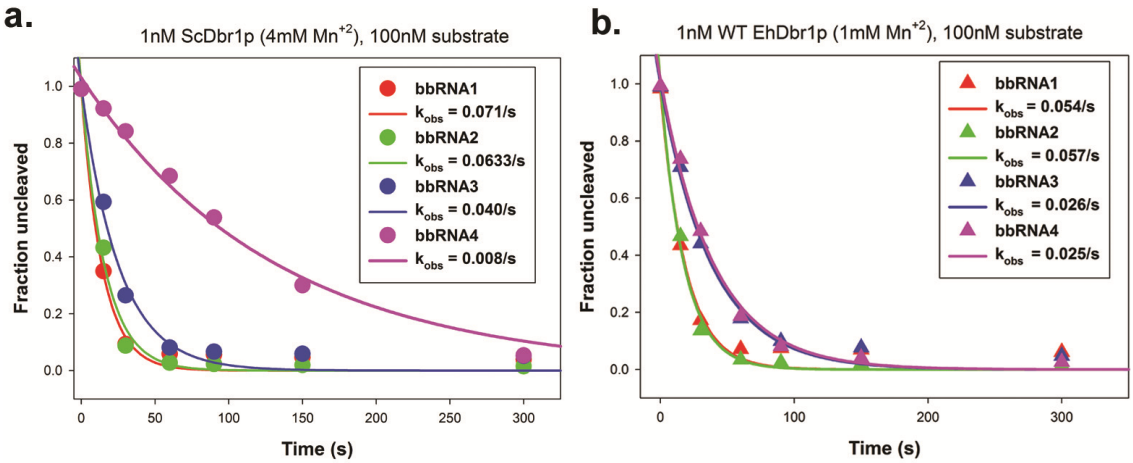


Figure 3.4. Debranching reactions of the modified bbRNA substrates for **a.** ScDbr1p and **b.** EhDbr1p. The k_{obs} values of each of the reaction was calculated by single exponential fitting of the fraction uncleaved vs time plots.

References

1. Dey, S. K., Paredes, E., Evans, M. & Das, S. R. in *From Nucleic Acids Sequences to Molecular Medicine* 475–501 (Springer Berlin Heidelberg, 2012). doi:10.1007/978-3-642-27426-8_19

2. Hesselberth, J. R. Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA* **4,** 677–691 (2013).

3. Nam, K. *et al.* Yeast Lariat Debranching Enzyme: Substrate and sequence specificity. *J. Biol. Chem.* **269,** 20613–20621 (1994).

4. Sutton, R. E. & Boothroyd, J. C. Evidence for Trans Splicing in Trypanosomes. *Cell* **47,** 527–535 (1986).

5. Chapman, K. B. & Boeke, J. D. Isolation and Characterization Encoding Yeast Debranching of the Gene Enzyme. *Cell* **65,** 483–492 (1991).

6. Ruskins, B. & Green, M. R. An RNA Processing Activity that Debranches RNA Lariats. *Science* **229,** 135–140 (1985).

7. Arenas, J. & Hurwitz, J. Purification of a RNA debranching activity from HeLa cells. *J. Biol. Chem.* **262,** 4274–4279 (1987).

8. Ooi, S. L. *et al. RNA Lariat Debranching Enzyme*. *Methods in Enzymology* **342,** (Elsevier Masson SAS, 2001).

9. Jacquier, A. & Rosbash, M. RNA Splicing and Intron Turnover are Greatly Diminished by a Mutant Yeast Branch Point. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 5835–5839 (1986).

10. Pratico, E. D., Wang, Y. & Silverman, S. K. A deoxyribozyme that synthesizes 2',5'-branched RNA with any branch-site nucleotide. *Nucleic Acids Res.* **33,** 3503–3512 (2005).

11. Ganeshan, K. *et al.* Novel approaches to the synthesis and analysis of branched RNA. *Nucleosides Nucleotides* **14,** 1009–1013 (1995).

12. Mourani, R. & Damha, M. J. Synthesis, characterization, and biological properties of small branched RNA fragments containing chiral (Rp and Sp) 2′,5′-phosphorothioate linkages. *Nucleosides, Nucleotides and Nucleic Acids* **25,** 203–229 (2006).

13. Montemayor, E. J. *et al.* Structural basis of lariat RNA recognition by the intron debranching enzyme Dbr1. *Nucleic Acids Res.* **42,** 10845–10855 (2014).

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

15. Hopfner, K. P. *et al.* Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105,** 473–485 (2001).

Chapter 4

Kinetics studies of the lariat debranching enzyme using dual-labeled bbRNA substrates

4.1. Introduction

The previous chapter described the biochemical studies with the different bbRNA substrates to investigate the Dbr1p enzyme cleavage activity. Most of the biochemical results obtained for Dbr1p enzyme thus far have been based on end point cleavage.[1–3] In those studies, the bbRNA substrates were incubated with Dbr1p for a specified amount of time and then separated from the cleavage product using gel electrophoresis. Different substrates gave differing amounts of cleavage products in a specified amount of time and suitability of bbRNAs as substrates were judged based on the amount of product formation. Although this approach is useful to initially screen a large number of substrate analogues, it has two major flaws. First, it does not give a quantitative measure of the reactivity of different substrates. Second, it does not inform as to why one substrate is cleaved faster than another. In the previous chapter, I described biochemical studies to compare a variety of different substrates based on their observed cleavage rate (k_{obs}), values of which provide quantitative information about their reactivities. These k_{obs} values represent a rate that reflects both the association of the Dbr1p as well as the cleavage rate for the substrate. However a complete kinetics analysis is required to quantitatively understand the origin of the different reactivities observed for different substrates.[4]

The two most important kinetic parameters that characterize any enzyme are the turnover number or k_{cat} and the Michaelis constant or K_m. The k_{cat} represents the number

of substrate molecules converted into product by an enzyme per unit time when the enzyme is fully saturated with substrate. Theoretically the Michaelis constant K_m represents the substrate concentration at which the rate of the enzymatic reaction is half the maximum rate. However, under certain circumstances, it is synonymous with the dissociation constant (K_d) of the enzyme-substrate complex.[4] Thus far there have been no reported kinetics studies of the Dbr1p enzyme. However to gain insight into how different substrate or enzyme modifications affect the debranching reaction, kinetics studies of Dbr1p are essential. Therefore we have developed a real time kinetics assay for the debranching reaction based on Förster Resonance Energy Transfer (FRET). For this purpose, we have synthesized a dual fluorescent dye (Cy3 and Cy5) labeled bbRNA which is described in details in chapter 2 (experimentals, section 2.6). The following section describes the use of the dual-labeled bbRNA substrate for kinetics studies of the debranching reaction.

4.2. Dual fluorescently labeled bbRNA for debranching kinetics

We chose to use a FRET-based method for the kinetics studies of the debranching reaction because the progress of the reaction can be monitored in real time. In addition, with the advancement of plate-reader-based fluorescence technologies, it is possible to perform multiple kinetics studies in a short amount of time using this fluorescence-based strategy. There have been several reports of FRET-based kinetics studies for restriction enzyme cleavage of DNA, protease cleavage, and translational kinetics of the ribosome.[5–9] For example, Ghosh and Miller have used a dual dye labeled (fluorescein and rhodamine) double stranded DNA (dsDNA) substrate for FRET- based kinetics studies of the PaeR7 restriction endonuclease.[5] The kinetics of the reaction was followed by loss of

FRET with substrate cleavage and the kinetics parameters obtained matched those obtained with gel-electrophoresis-based methods. Kettling and Eigen used dual-color cross correlation microscopy along with FRET to study the kinetics of the EcoRI cleavage of dsDNA substrates.[6] This is a very sensitive method which can detect enzymatic activity down to low picomolar concentration of the enzyme. Inspired by these studies based on DNA cleavage, we also devised a FRET-based method for kinetics studies of the debranching reaction.

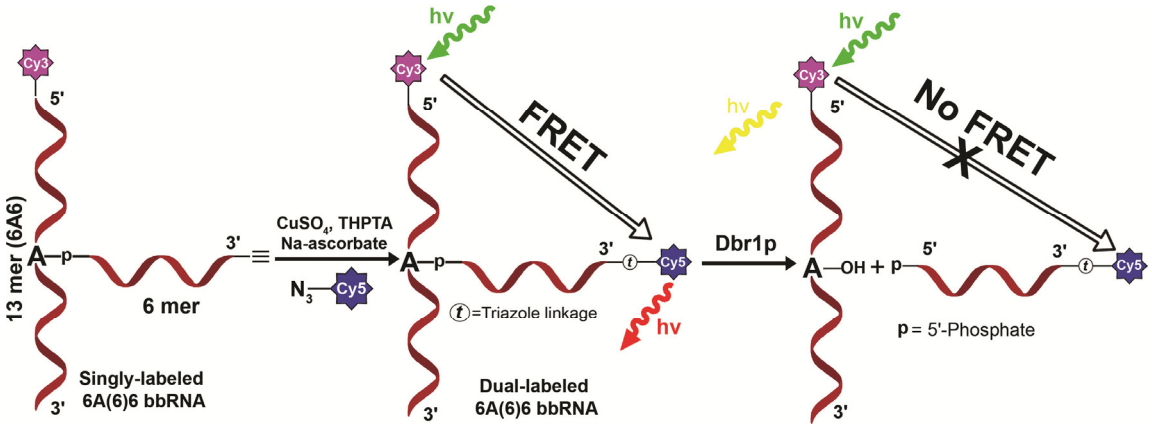


Figure 4.1. Kinetics of the debranching reaction using dual-labeled bbRNA substrate. The Cy3 labeled bbRNA containing 3'-O-propargyl was conjugated with Cy5-azide using copper catalyzed azide alkyne cycloaddition (CuAAC) reaction to make the dual labeled bbRNA. Upon debranching by Dbr1p, the donor dye (Cy3) will be separated from the acceptor dye (Cy5) giving rise to loss in FRET signal. The kinetics of the debranching reaction can be followed by following the loss in FRET.

One of the major challenges in using FRET for debranching kinetics is to design a proper dual-labeled bbRNA substrate. However the versatile nature of our method of bbRNA synthesis using photo-protected amidite allows incorporation of modifiers in both the terminus of the stem and in the branch. Our approach involves incorporating a donor dye (Cy3) into the bbRNA during the solid-phase synthesis and conjugating the acceptor

dye (Cy5) to the donor-labeled bbRNA (Fig. 4.1). The Cy3-labeled bbRNA containing the 3'-*O*-Propargyl group in the branch was conjugated to the Cy5 azide using a copper catalyzed azide alkyne cycloaddition (CuAAC) reaction (Fig. 4.1) to generate the dual-labeled bbRNA. Because of the close proximity of the donor (Cy3) and acceptor (Cy5) dyes in the dual-labeled substrate, energy transfer (FRET) from the donor dye to the acceptor dye is expected when the donor is excited. However, upon cleavage of the 2'-5' phosphodiester bond of the bbRNA by Dbr1p, the acceptor dye in the branch will be separated from the donor dye in the stem which will cause loss of FRET (Fig. 4.1). The kinetics of the debranching reaction can be monitored in real time by monitoring this loss of FRET with time.

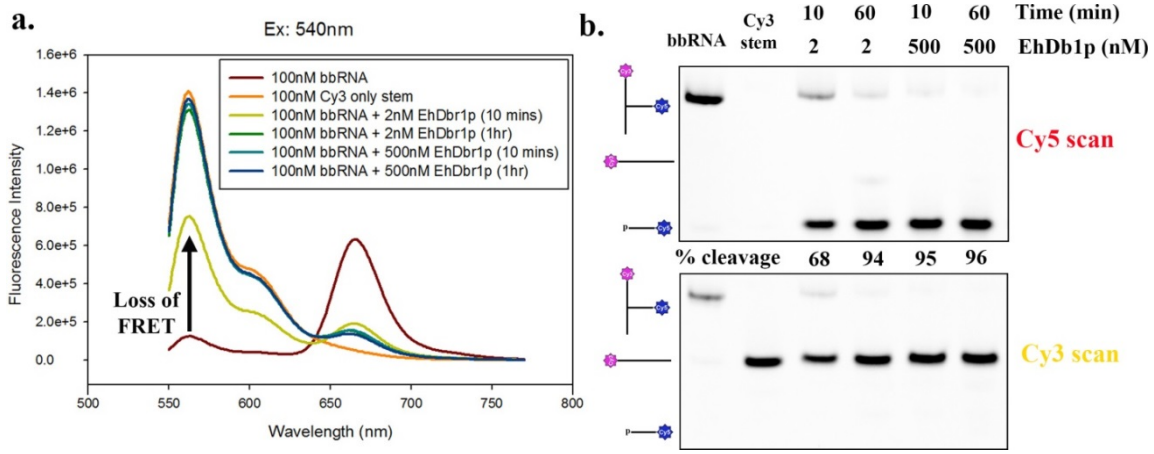


Figure 4.2. Debranching of the dual-labeled bbRNA substrate. **a.** Upon debranching of the dual-labeled bbRNA by EhDbr1p, the donor (Cy3) fluorescence intensity increases indicating loss of FRET. Using both sub-saturating and excess enzyme compared to the substrate, the reaction goes to almost completion in 60 and 10 minutes respectively. **Reaction condition**: 100 nM substrate, 2 nM/500 nM EhDbr1p, 50 mM Tris-HCl (pH 7.5), 25 mM NaCl, 1 mM Mn^{2+}, 2.5 mM DTT, 0.1% TritonX-100, 10 mins/1hr reaction **b.** The same reactions as in a. were analyzed by gel electrophoresis followed by fluorescence scanning. The scan for Cy5 can be used to calculate the percentage of product formed. The pink sphere indicates the Cy3 dye and the blue sphere indicates the Cy5 dye.

To check the suitability of the dual-labeled bbRNA substrate for debranching studies, we first measured its fluorescence spectra under different enzymatic condition. The dual-labeled bbRNA substrate contains 6 nucleotides in each arm. This puts the distance between the donor and acceptor dye at ~42Å assuming a 90˚ angle between the stem and the 2'-branch. As the reported R_0 value for the Cy3-Cy5 pair is 56Å, ~90% energy transfer is expected in the dual-labeled bbRNA. The dual-labeled bbRNA substrate shows high energy transfer (~95%) as evidenced by the drop in donor signal and increase in acceptor signal (red line, Fig. 4.2a) compared to the donor only (Cy3) labeled stem RNA (orange line, Fig. 4.2a). Upon complete debranching, the dual-labeled bbRNA will be completely converted into a Cy3-labeled stem and a Cy5-labeled branch (Fig. 4.1). In other words, the donor signal should go back to the Cy3 only stem level upon completion of the debranching reaction.

The Dbr1p enzyme from *Entamoeba histolytica* (EhDbr1p) was used for this FRET based debranching studies because the crystal structure of this enzyme is already reported.[10] Different reaction conditions were tested using either excess or sub-saturating concentration of the EhDbr1p compared to the substrate. Under the excess enzyme condition (100 nM substrate, 500 nM EhDbr1p), the donor signal returns almost completely to Cy3 only stem level in 10 minutes (blue line, Fig. 4.2a) as well as in 1hr (cyan line, Fig. 4.2a). However, under sub-saturating condition (100 nM substrate, 2 nM EhDbr1p), the donor signal does not revert to the Cy3-only stem level within 10 minutes indicating that the reaction is only partially complete. (yellow line, Fig. 4.2a). However, after 1 hour, the donor signal increases to nearly the Cy3 stem only level indicating a nearly complete reaction (green line, Fig. 4.2a). As can be seen from the fluorescence

spectra, the donor signal increases almost 10 times upon complete debranching of the bbRNA. This gives a large dynamic range within which the progress of the debranching reaction can be followed using the increase in donor fluorescence over time.

The extent of the debranching reaction was also measured by first stopping the reactions using formamide/EDTA solution then separating the reaction mixture using gel electrophoresis. Figure 4.2b shows the Cy5 and Cy3 scans of the gel for the debranching reactions described before. A fluorescence scan of the gel for Cy5 was used to calculate the percentage of the product formation (Fig. 4.2b, top panel). As the Cy3 excitation will give fluorescence at both the Cy3 and Cy5 wavelengths for the substrate due to FRET, it cannot be used for calculation of relative amount of product formation. Therefore the intensity of the Cy5 signal in the product was compared to that of the substrate by exciting at the Cy5 absorption maximum. As it can be seen in the Cy5 scans, the dual-labeled bbRNA undergoes complete cleavage by EhDbr1p under both excess and sub-saturating enzyme conditions in 10 and 60 minutes, respectively. Before moving on to kinetics studies, we investigated whether the increase in the donor fluorescence corresponds well with the amount of product formed. Table 4.1 represents the amount of product formed at different time points throughout the debranching reaction calculated using either the donor fluorescence or by gel scanning (experimental, section 4.5). As can be seen in Table 4.1, the amount of the product formed at different time points calculated using these two methods are in good agreement with some minor differences. This indicates that the increase in the donor signal is linear with the amount of product formation and we can use this increase in the donor signal to follow the extent of the debranching reaction with time. Thus we have successfully demonstrated the usefulness

of this dual-labeled bbRNA for FRET-based kinetic studies of the debranching reaction. The next section discusses the actual kinetics studies of Dbr1p using this bbRNA.

Time (s)	Trial 1		Trial 2	
	% Product formation (using fluorescence)	% Product formation (using gel scanning)	% Product formation (using fluorescence)	% Product formation (using gel scanning)
30	**7**	**9**	**5**	**6**
90	**22**	**21**	**15**	**16**
150	**33**	**33**	**23**	**25**
300	**55**	**61**	**40**	**45**
600	**80**	**83**	**64**	**72**
3600	**95.0**	**94**	**93**	**93**

Table 4.1. Calculation of the amount of product formation for the debranching reaction using both donor signal and gel scanning for two different trials. **Reaction condition**: 100 nM substrate, 2 nM EhDbr1p, 50 mM Tris-HCl (pH 7.5), 25 mM NaCl, 1 mM Mn^{2+}, 2.5 mM DTT, 0.1% TritonX-100.

4.3. Kinetics studies of the debranching reaction

To study the kinetics of the debranching reaction, we first used the Δ7 EhDBr1p (first 7 amino acid removed from the N-terminus) variant of the *E. histolytica* Dbr1p. This variant was chosen for kinetics analysis to match the form for which crystallization is being pursued. The donor dye (Cy3) of the dual labeled bbRNA was excited at 540 nm and the donor emission was monitored with time at 570nm at 0.5s intervals using a Fluoromax-2 fluorimeter. We used 1 nM Δ7 EhDBr1p enzyme and a variety of different substrate concentration (20 nM, 50 nM, 70 nM, 100 nM and 120 nM) for the kinetics studies. First the dual-labeled bbRNA substrate in the reaction buffer (50 mM Tris-HCl (pH 7.5), 25 mM NaCl, 1 mM Mn^{+2}, 2.5 mM DTT, 0.1% TritonX-100) was placed in a cuvette inside the fluorimeter and the time-based fluorescence measurement was started.

Then after ~15 seconds, the reaction was initiated by addition of the enzyme solution using a pipette while still continuing to monitor the fluorescence. The fluorescence traces obtained with different concentration of the bbRNA substrate are shown in Figure 4.3a. When the enzyme is added to the cuvette, there is a sharp rise in the donor fluorescence. This artifact arises because of the scattering of the excitation light due to insertion of the pipette tip inside the cuvette for proper mixing (Fig. 4.3a). Following this sudden increase, the donor fluorescence drops to a lower value very quickly and then increases steadily. This steady increase indicates the loss of FRET due to the debranching reaction.

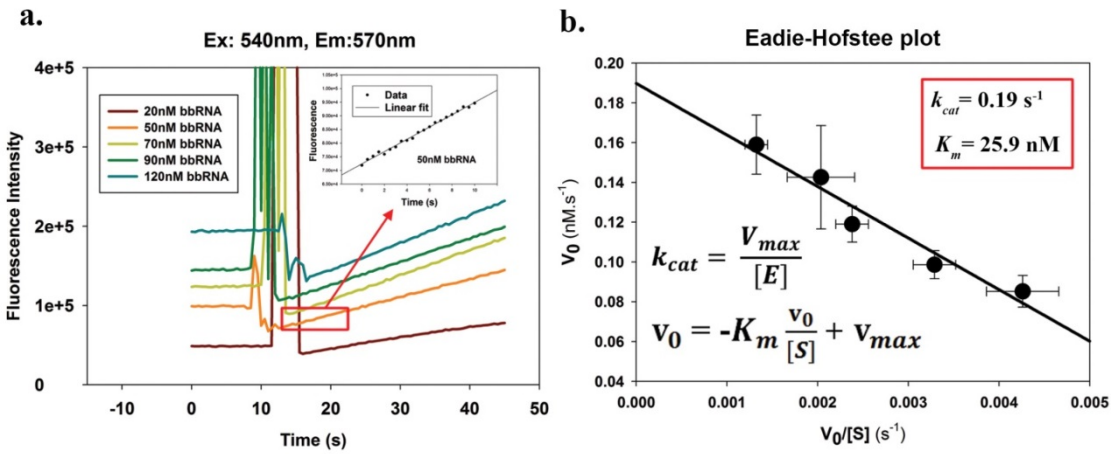


Figure 4.3. Debranching kinetics using dual-labeled bbRNA. **a.** Kinetic traces of the debranching reaction of Δ7 EhDBr1p for different substrate concentrations. The kinetics of the reaction can be monitored in real time by monitoring the increase in donor signal over time. The inset shows the linear fit of the donor fluorescence signal with time for calculation of the initial rate. **b.** Eadie-Hofstee plot for the debranching reaction. The initial rates were calculated by using the first 10s time points (inset).

The first 10s time points during the steady increase of the donor fluorescence were used to calculate the initial rates (v_0) of the debranching reaction for each substrate concentration (Fig. 4.3a, inset). Then the rates expressed in number of fluorescence counts units were converted into concentration using the difference between the product

and substrate fluorescence for the different concentrations of bbRNA (experimental, section 4.5). Once the initial rates were calculated for each concentration of the bbRNA substrate, an Eadie-Hofstee plot was constructed to determine the kinetics parameters of Dbr1p. We chose to use the Eadie-Hofstee plot for the debranching kinetics instead of Lineweaver-Burk (LB) plot, because the LB plot suffers from a highly biased weighing of errors due to its double reciprocal nature.[4,11] The data shows a good linear fit with the Michaelis–Menten equation (Fig. 4.3b and equation therein) and the k_{cat} and K_m values of the Δ7 EhDBr1p enzyme were determined to be $0.19s^{-1}$ and 25.9 nM respectively (Fig. 4.3b). To our knowledge this is the first report of the kinetics parameters of the Dbr1p.

The same kinetics studies were performed with two additional lariat debranching enzyme variants, the full length EhDbr1p and yeast Dbr1p (ScDBr1p). It is important to mention here that, to measure the ScDbr1p kinetics, 4 mM Mn^{2+} was used whereas 1 mM was used for the EhDbr1p and Δ7 EhDBr1p reactions. Our previous biochemical studies (Chapter 3) showed that 4 mM Mn^{2+} is the optimum condition for the ScDbr1p enzyme. The kinetics parameters of the three different lariat debranching enzymes are reported in Table 4.2. The three enzymes have similar K_m values indicating similar binding mechanism. However their k_{cat} values are significantly different. The Δ7 EhDBr1p form exhibits an almost four times slower turnover rate compared to the full length EhDbr1p, indicating that the first 7 amino acids may be important for catalytic activity. In addition, the turnover rate of the ScDbr1p is almost 10 times slower than that of EhDbr1p. This result is quite distinctive as these two enzymes were determined to have similar k_{obs} (Chapter 3). However the biochemical studies were performed at 37˚C and the kinetics studies were performed at room temperature. Therefore any temperature dependent

increase in k_{cat} values for ScDbr1p could still account for this rate difference at room temperature.[12]

Enzyme	K_m (nM)	k_{cat} (s^{-1})
Δ7 EhDBr1p	**25.9 ± 3.5**	**0.19 ± 0.01**
EhDBr1p	**24.7 ± 5.0**	**0.80 ± 0.06**
ScDBr1p	**27.5 ± 13.3**	**0.08 ± 0.02**

Table 4.2. Kinetics parameters of the different lariat debranching enzyme determined using FRET based kinetics.

In summary, this section demonstrates the use of the dual-labeled bbRNA substrate for the kinetics studies of Dbr1p. To understand the structure function relationship of Dbr1p, several modified bbRNA substrates (first base in the 2'-branch as deoxyG, 2'F-G, 2'OMe-G) were synthesized for dual labeling. The kinetics parameters for Dbr1p obtained for these various substrates will shine light on the actual reason (binding or catalysis) for the differences in rate. Going forward, it is also possible to compare the kinetics parameters of modified Dbr1p enzyme species in which various residues have been mutated in order to dissect their roles in the debranching reaction.

4.4. 'Click' branched RNA as a competitive inhibitor of Dbr1p

Easy access to the dual-labeled bbRNAs has allowed us to study the kinetics of the debranching reaction. However to study the binding of different bbRNA substrates to Dbr1p we require a non-cleavable substrate analogue. The substrate analogue will ideally bind Dbr1p in a similar fashion as the native substrate but will not be cleaved by Dbr1p. Tago and Damha previously reported the synthesis of 2'-5' phosphoramidate linked

bbRNAs as an inhibitor of Dbr1p.[13] The 2'-5' phosphodiester linkage is replaced by 2'-5' phosphoramidate linkage in this analogue which cannot be cleaved by Dbr1p. Although this substrate analog is produced by minimally perturbing the native bbRNA substrate, it requires the synthesis of complex building blocks. As an alternative, we set out to determine if 2'-5' triazole linked 'click' branched RNAs (cbRNAs) would act as an inhibitor of Dbr1p. The triazole linkage has been shown to be biocompatible when present in DNA and RNA and can be read through by DNA polymerase and reverse transcriptase.[14–17] The cbRNAs are a non-cleavable substrate analogue of Dbr1p that contains a 2'-5' triazole linkage that cannot be cleaved by Dbr1p (Fig. 4.4a). We synthesized the cbRNA by conjugating a stem strand containing an internal 2'-*O*-propargyl with the branch containing a 5'-azide using the CuAAC reaction (Fig. 4.4a). Both the stem and the branch strand can be synthesized using commercially available materials following reported methods (Section 2.6). Following completion of the CuAAC reaction, the product was separated from the starting materials using gel electrophoresis. UV shadowing of the gel shows near-complete conversion of the stem strand into the cbRNA, demonstrating the easy accessibility of these cbRNAs (Fig. 4.4b). The cbRNA synthesized for the inhibition studies contained 6 nucleotides in each of its three arms (Fig. 4.4b) and was not dye-labeled.

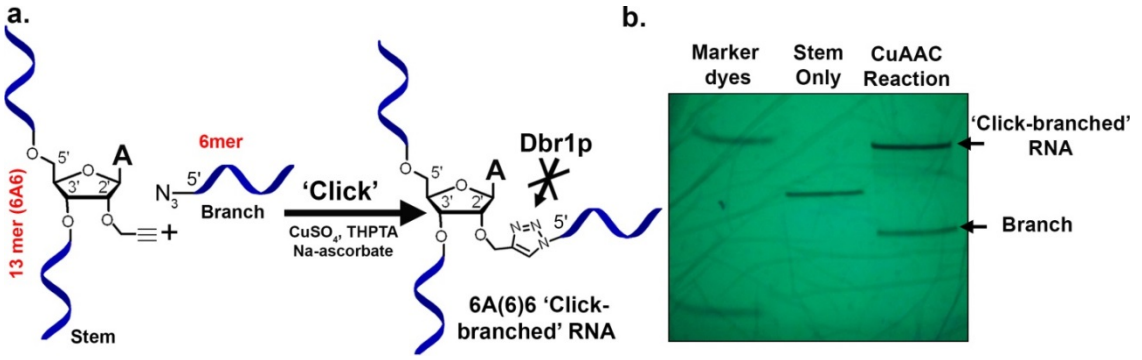


Figure 4.4. Synthesis of 'click' branched RNAs as a potential inhibitor of Dbr1p. **a**. The 'click' branched RNAs was synthesized by conjugating a stem strand containing an internal 2'-*O*-propargyl with the branch containing a 5'-azide using CuAAC reaction. **b.** UV shadowing of the gel after the CuAAC reaction and separation shows complete conversion of the stem into the product.

The FRET-based kinetics assay for debranching reaction can be used to screen for potential inhibitors of Dbr1p. To test whether the cbRNA is an inhibitor of Dbr1p, the same kinetics experiments described above were performed with Δ7 EhDBr1p in presence of three different concentrations of cbRNA: 30 nM, 45 nM and 60 nM. The Eadie-Hofstee plots for the debranching kinetics in presence of the cbRNA shows behavior typical of competitive inhibition (Fig. 4.5a). The plots for different inhibitor concentrations intercept at the Y axis almost same point but have different slopes (Fig. 4.5a). This indicates that, in presence of the inhibitor, the k_{cat} value of the enzyme remains the same but the K_m values change with varying inhibitor concentrations. This is a typical behavior of competitive inhibition in which the substrate competes with the inhibitor for binding with the enzyme hence changing K_m.[18–20] In case of non-competitive inhibition the plots for different concentrations will be parallel lines. A plot of the apparent K_m values versus inhibitor concentration gives a value of the inhibition constant K_i of 64.6 nM. This indicates strong binding between the inhibitor and Dbr1p (Fig. 4.5b). Damha *et al.* reported the IC_{50} value of the phosphoramidate based inhibitor to be ~40

nM meaning it is a slightly better inhibitor than cbRNA.[13] However the ease of synthesis of the cbRNAs makes them a very attractive choice as an inhibitor of Dbr1p.

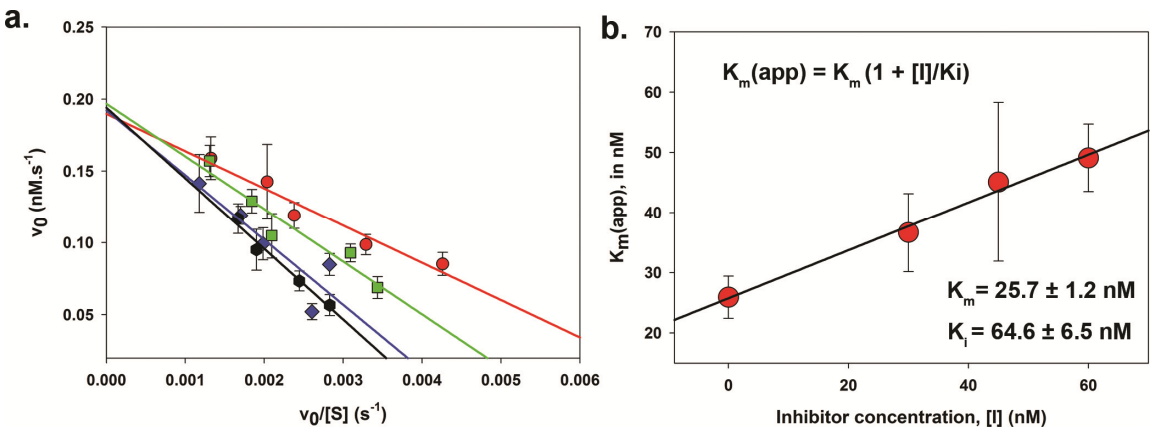


Figure 4.5. Competitive inhibition of Dbr1p by 'click' branched RNA. **a.** Eadie-Hofstee plot for the debranching kinetics of Δ7 EhDBr1p (1 nM) in presence of varying concentrations of cbRNA shows competitive inhibition. Under the different inhibitor concentrations, the k_{cat} value remains almost the same whereas the K_m values change. **b.** Plot of Km (app) versus the inhibitor concentration gives a Ki of 82.4 nM indicating strong binding of the inhibitor with Δ7 EhDBr1p.

The competitive inhibition of Dbr1p by cbRNA suggests that the enzyme most probably binds with this inhibitor at the same site as does the native substrate. Therefore cbRNA can be used to study substrate binding with Dbr1p. We synthesized several cbRNAs with various stem and branch lengths. These cbRNAs will be used for the inhibition studies to investigate the effect of the branch and stem length on binding with Dbr1p. We are also trying to co-crystalize cbRNAs with Dbr1p to investigate the different substrate enzyme contacts in order to elucidate the cleavage mechanism of Dbr1p. Because of the similar nature of binding between the native and the cbRNA substrate, we are also using it to monitor conformational changes upon binding with Dbr1p at the single molecule level.

4.5. Conclusion

This chapter describes the kinetics studies of the debranching reaction of the Dbr1p enzyme. We synthesized a novel dual fluorescently labeled bbRNA substrate using both solid phase synthesis and post-synthetic modification and showed that we can use it to follow the real time debranching kinetics using FRET. Using this FRET-based assay we reported the kinetics parameters for three different variant of the Dbr1p. Additionally, this kinetics assay was also used to investigate a potential inhibitor of Dbr1p: the 'click' branched RNAs (cbRNA). The cbRNA substrate demonstrated competitive inhibition for debranching activity suggesting that it binds Dbr1p in a similar fashion as the native substrate. This chapter is a very important portion of my thesis work, as it not only demonstrates a new method to study Dbr1p kinetics, but also opens up several avenues for biochemical and single molecule studies of bbRNAs and Dbr1p. For example, we are now trying to use the cbRNAs as co-crystal substrates for Dbr1p. Because of the non-cleavable nature of the cbRNAs, we are also using them for binding studies with Dbr1p at single molecule level. We envision that this kinetics assay will aid the molecular level understanding of the interaction between bbRNA substrates and Dbr1p.

4.6. Experimental

Fluorescence studies of the dual-labeled RNA

The fluorescence spectra (Fig. 4.2a) of the different RNAs were recorded at 100 nM concentration using a Fluoromax-2 (Horiba Scientific, New Jersey) fluorimeter. I used 540nm excitation wavelength and emission was monitored from 550-780 nm using 4 nm slit width for both the excitation and emission channel. The debranching reactions were performed in 1X reaction buffer (50 mM Tris-HCl (pH 7.5), 25 mM NaCl, 1 mM Mn^{2+}, 2.5 mM DTT, 0.1% TritonX-100) which we have also used for biochemical studies. Table 4.3 shows the concentration and volume of the different components used for the reaction.

Component	Stock concentration	Volume used (µL)	Final concentration
Dual-labeled bbRNA	1 µM	10	100 nM
EhDbr1p (in1X Buffer)	10 nM/2.5 µM	20	2 nM/500 nM
2X reaction buffer	2X	40	1X
Water	-	30	-
Total		100	

Table 4.3. Reaction condition for the debranching reaction of the dual-labeled bbRNA by Dbr1p.

First, all the components of this reaction except EhDbr1p were mixed together and then the debranching reaction was started by addition of EhDbr1p. After the desired time, the fluorescence spectra of the reaction mixture was collected by transferring 50 µL of the solution to a microcuvette (Starna Cells Inc, CA) and then placing the cuvette in

the fluorimeter. At the same time a 20 µL solution of the reaction mixture was treated with 40 µL of STOP solution (90% formamide and 10% 0.1M EDTA) to inactivate the debranching reaction. The reaction mixtures were separated using analytical 20% denaturing PAGE (1X TBE buffer). 15 µL of the stopped reaction mixture was loaded in each of the wells and then the gel was run (0.5X TBE running buffer) at 250V for 3 hrs to separate the starting materials from the products. Then the gel was scanned in a Typhoon FLA900 (GE healthcare life sciences) fluorescence scanner using the settings for Cy3 and Cy5 dyes. The bands were quantified using the associated software to calculate the percentage cleavage.

Kinetics studies of the debranching reaction

The same fluorimeter and the same cuvette were used for the kinetics studies as well. The table below describes the concentration and volume of the different components used for the reaction. Appropriate concentration of the dual-labeled bbRNA in 1x reaction buffer was introduced in the cuvette and then placed in the fluorimeter. Then the fluorescence measurement was initiated. I used 540 nm excitation wavelength and the emission was monitored at 570 nm using a 4 nm slit width for both excitation and emission. The fluorescence measurements were performed using 0.5 seconds time intervals. After about 15 seconds, 20 µL of the Δ7 EhDbr1p solution was mixed with the bbRNA in the cuvette while the fluorescence measurement is still going on. Then the measurements were continued for a total of 120 seconds. All experiments were performed in triplicate for each of the different concentration of the dual labeled bbRNA. The cuvette was washed vigorously with water and methanol (3 times each) in between each of the experiments to avoid contamination of Dbr1p from the previous experiments.

It is important to mention here that it is possible to go down to 1ms time intervals using this fluorimeter. Therefore we can also study much faster kinetics under this setup.

Component	Stock concentration	Volume used (µL)	Final concentration
Dual-labeled bbRNA	200/300/500/700/ 1200 nM	8	20/30/50/70/120 nM
Δ7 EhDbr1p (in1X Buffer)	4 nM	20	1 nM
2X reaction buffer	2X	30	1X
Water	-	22	-
Total		80	

Table 4.4. Reaction condition for the kinetics studies of Δ7 EhDbr1p using dual-labeled bbRNA.

The initial rates were calculated using the first 10 seconds time points (21 points total) once the steady increase of the fluorescence started. The rates were calculated from the slope of the fluorescence vs time plots. Figure 4.6 shows the fitting of the fluorescence data for calculation of the initial rates for 20 nM and 70 nM concentration of the bbRNA. The fitting of the data was performed using SigmaPlot software. However these initial rates are based on fluorescence unit and therefore needed to be converted into concentration unit.

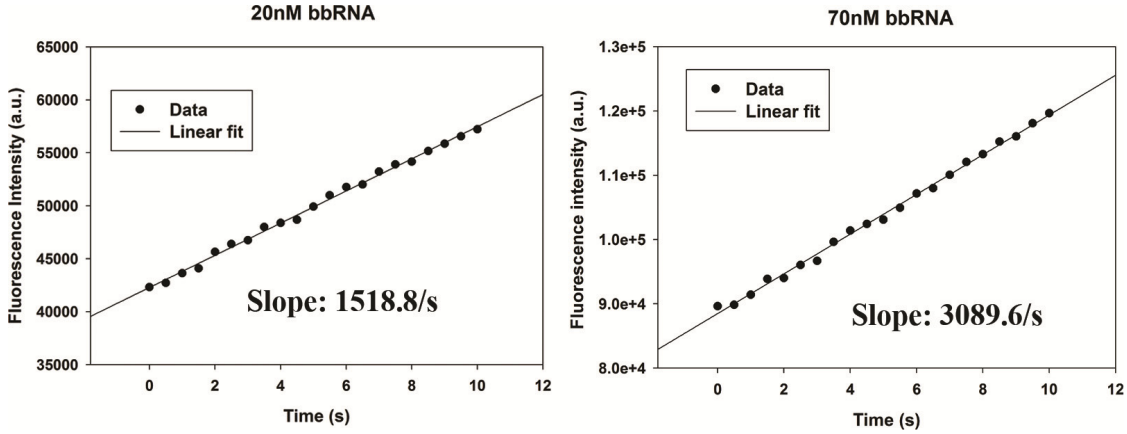


Figure 4.6. Calculation of the initial rates from fluorescence data.

To convert the initial rates from fluorescence unit to concentration unit we need to know the change in donor fluorescence signal after the complete cleavage of the dual labeled bbRNA by Dbr1p. If I_0 is the donor signal of the intact dual labeled bbRNA and I_f is the donor signal of the Cy3 only stem, then (I_f - I_0) is the total change in donor signal after 100% completion of the debranching reaction. I have calculated the quantity (I_f - I_0) for all the five different concentration of the bbRNA used for the debranching study. Now if the initial rate of the reaction is **a** in fluorescence unit for the concentration **c** of the bbRNA, then the rate of the reaction in terms of concentration will be **a*c**/(I_f - I_0). Table 4.5 shows the calculation of the initial rates of the debranching reaction in terms of the concentration units. Once the initial rates (v_0) were obtained then the values of the (v_0/[S]) were also calculated for each of the different concertation and plotted to determine the values of k_{cat} and K_m (Fig. 4.3a).

Conc (nM) of the bbRNA	Rate (Fl unit/s)	I_f	I_0	$I_f - I_0$	Rate ($nM.s^{-1}$)
20	**944.7**	**263514.5**	**41776**	**221738.5**	**0.085**
50	**1093.7**	**633409.4**	**73739.5**	**559669.9**	**0.099**
70	**1332.1**	**868769.3**	**96794.2**	**771975.1**	**0.119**
90	**1572.7**	**1118218.5**	**113563.9**	**1004654.6**	**0.143**
120	**1777.0**	**1495080**	**153657.4**	**1341422.6**	**0.159**

Table 4.5. Calculation of the initial rates in concentration unit from the fluorescence data.

Inhibition kinetics of Dbr1p using 'click' branched RNA

The inhibition kinetics of Dbr1p was performed following the exact same protocol as described before except one change. The dual-labeled bbRNA solution in the cuvette also contained the required concentration of cbRNA for each kinetics experiment. Therefore when the Δ7 EhDbr1p was introduced in the cuvette either the bbRNA or the cbRNA could bind with Dbr1p. The k_{cat} and K_m values were calculated for each concentration of the inhibitor cbRNA following the same protocol. Table 4.6 shows the different k_{cat} and K_m values obtained for the different inhibitor concentration. The apparent K_m values obtained for the different inhibitor concentration was used to calculate the value of inhibitor constant, K_i (Fig. 4.5b).

Inhibitor concentration (nM)	K_m (nM)	k_{cat} (s^{-1})
0	**25.9 ± 3.5**	**0.19 ± 0.01**
30	**36.6 ± 6.5**	**0.20 ± 0.01**
45	**45.1 ± 13.2**	**0.19 ± 0.03**
60	**49.1 ± 5.6**	**0.19 ± 0.01**

Table 4.6. Kinetics parameters of Dbr1p in presence of different inhibitor concentration.

References

1. Arenas, J. & Hurwitz, J. Purification of a RNA debranching activity from HeLa cells. *J. Biol. Chem.* **262,** 4274–4279 (1987).

2. Nam, K. *et al.* Yeast Lariat Debranching Enzyme: Substrate and sequence specificity. *J. Biol. Chem.* **269,** 20613–20621 (1994).

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

4. Johnson, K. A. A century of enzyme kinetic analysis, 1913 to 2013. *FEBS Lett.* **587,** 2753–2766 (2013).

5. Ghosh, S. S., Eis, P. S., Blumeyer, K., Fearon, K. & Millar, D. P. Real time kinetics of restriction endonuclease cleavage monitored by fluorescence resonance energy transfer. *Nucleic Acids Res.* **22,** 3155–9 (1994).

6. Kettling, U., Koltermann, a, Schwille, P. & Eigen, M. Real-time enzyme kinetics monitored by dual-color fluorescence cross-correlation spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 1416–1420 (1998).

7. Dittrich, P. S., Mŭller, B. & Schwille, P. Studying reaction kinetics by simultaneous FRET and cross-correlation analysis in a miniaturized continuous flow reactor. *Phys. Chem. Chem. Phys.* **6,** 4416 (2004).

8. Taliani, M. *et al.* A continuous assay of hepatitis C virus protease based on resonance energy transfer depsipeptide substrates. *Anal. Biochem.* **240,** 60–7 (1996).

9. Milon, P. *et al.* Transient Kinetics, Fluorescence, and FRET in Studies of Initiation of Translation in Bacteria. *Methods Enzymol.* **430,** 1–30 (2007).

10. Montemayor, E. J. *et al.* Structural basis of lariat RNA recognition by the intron debranching enzyme Dbr1. *Nucleic Acids Res.* **42,** 10845–10855 (2014).

11. Martin, R. B. Disadvantages of Double Reciprocal Plots. *J. Chem. Educ.* **74,** 1238 (1997).

12. Daniel, R. M. & Danson, M. J. A new understanding of how temperature affects the catalytic activity of enzymes. *Trends Biochem. Sci.* **35,** 584–591 (2010).

13. Tago, N. *et al.* Design, Synthesis, and Properties of Phosphoramidate 2′,5′-Linked Branched RNA: Toward the Rational Design of Inhibitors of the RNA Lariat Debranching Enzyme. *J. Org. Chem.* **80,** 10108–10118 (2015).

14. El-Sagheer, A. H., Sanzone, A. P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11338–11343 (2011).

15. El-Sagheer, A. H. & Brown, T. New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15329–15334 (2010).

16. El-Sagheer, A. H. & Brown, T. Synthesis and polymerase chain reaction amplification of DNA strands containing an unnatural triazole linkage. *J. Am. Chem. Soc.* **131,** 3958–3964 (2009).

17. Chen, X., El-Sagheer, A. H. & Brown, T. Reverse transcription through a bulky triazole linkage in RNA: implications for RNA sequencing. *Chem. Commun.* **50,** 7597–600 (2014).

18. Chen, N. & Justice, J. B. Cocaine acts as an apparent competitive inhibitor at the outward-facing conformation of the human norepinephrine transporter: kinetic analysis of inward and outward transport. *J. Neurosci.* **18,** 10257–10268 (1998).

19. Simm, A. M. *et al.* Bulgecin A: a novel inhibitor of binuclear metallo-beta-lactamases. *Biochem. J.* **387,** 585–590 (2005).

20. Michael D. Pluth, Robert G. Bergman, K. N. R. Acid Catalysis in Basic Solution. *Science* 85–88 (2007).

Chapter 5

Site specific labelling of Dbr1p using native chemical ligation and 'click' chemistry

5.1. Introduction

Site-specific labeling of proteins with chemical tags is a powerful technique that is used to study structure, dynamics and interactions of proteins both *in vitro* and *in vivo*.[1–3] Such studies can elucidate cellular localization of proteins and their functions.[4,5] In biophysical studies, having the label on a specific residue of the protein eliminates the heterogeneity due to inherent differences in the local environments in different regions of the protein.[6–10] Several biophysical studies of proteins, including high-throughput surface-based analyses for diagnostics purposes, also require immobilization of proteins on various surfaces.[11–13] Site-specifically labeled proteins can be immobilized on surfaces in the correct orientation and functionally-active form whereas non-specific labeling can lead to reduced activity.[14–16]

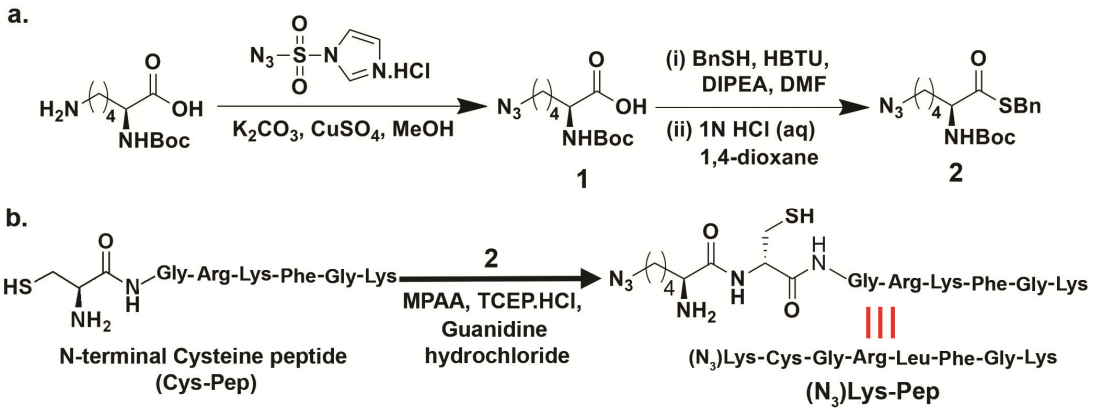
A protein of interest can be site-specifically labeled by genetic fusion to either a fluorescent protein[17,18] or a receptor proteins, such as SNAP, CLIP, ACP/MCP, Halo, TMP tag, which subsequently can be either covalently or non-covalently conjugated with probe molecules.[19–23] Relatively shorter peptide tags can also be used to label proteins terminally or internally using selective binding with fluorogenic dyes[24–28] or by selective incorporation of probe molecules through enzymatic reactions.[29–34] Another method for site-specific labeling involves incorporation of unnatural amino acids with bio-orthogonal functional groups at a specific residue.[35] These bio-orthogonal functional groups can be

used for conjugation with probe molecules using biocompatible reactions such as Staudinger ligation, copper-catalyzed or strain-promoted azide-alkyne cycloaddition reactions.[36,37] In a recent study, MacDonald *et al.* reported the site-specific modification of the N-terminus of native proteins using modified 2-pyridinecarboxyaldehyde reagents without need for genetic engineering.[38] This chapter discussed a method for site-specific N-terminal labeling of expressed proteins using an adapter molecule containing both thioester and azide moieties for native chemical ligation (NCL) and Cu(I)-catalyzed azide alkyne cycloaddition (CuAAC) reactions.

Site-specific labeling of proteins using NCL has been reported predominantly with stable proteins that can be lyophilized. Very few examples are reported with functional enzymes.[39–42] Recently, the NCL reaction was used in combination with the CuAAC reaction to label peptides and to dimerize and trimerize proteins.[43–45] In this chapter, we combined NCL and CuAAC using an adapter molecule to introduce site-specific labels at N-terminus of expressed proteins. We first tested the reactivity of this adapter molecule in the NCL reaction using a short-peptide-containing N-terminal cysteine. Following the NCL reaction with the adapter, the azido-labeled peptide was conjugated to either DNA or RNA using CuAAC to make peptide oligonucleotide conjugates. Such conjugates are useful for cellular and tissue-specific delivery of oligonucleotides.[46–54] We then demonstrated the usefulness of this adapter for N-terminal labeling of enzyme, namely the lariat debranching enzyme (Dbr1p), with a fluorescent dye and biotin. We show that the enzyme remains functional after both the NCL and CuAAC reactions, suggesting that this is a mild method for labeling proteins at the N-terminus with chemical tags. This site-specifically-labeled Dbr1p will be used for single

molecule FRET experiments where the enzyme carries one of the labels to probe its interaction with bbRNA substrate or to immobilize the enzyme on the surface.

5.2. Use of native chemical ligation and CuAAC to synthesize peptide oligo conjugates



Scheme 5.1. Synthesis of the adapter **2** and its native chemical ligation reaction with peptide. **a.** The adapter **2** was synthesized in three steps starting with α-NHBoc-protected lysine. **b.** Following the synthesis of **2**, it was conjugated with a peptide containing N-terminal cysteine (Cys-Pep) using native chemical ligation (NCL). The reaction was carried out in presence of 4-mercaptophenylacetic acid (MPAA) as a thiol exchange reagent and Tris-(2-carboxyethyl) phosphine hydrochloride (TCEP.HCl) as a reducing agent in 6M guanidine hydrochloride solution (pH 7). The resulting azido modified peptide, (N_3)Lys-Pep was purified by HPLC and analyzed by ESI-MS (see Fig. 5.5).

To combine the power of the NCL and CuAAC reactions for site-specific labeling, we required an adapter molecule that includes both azide and thioester functional groups. Though synthesis of an adapter molecule containing both azide and thioester groups has been previously reported, it requires complex synthesis methods.[55] Here we synthesize our adapter molecule **2** (synthesized by Dr. Debasish Grahacharya)**,** an azido-lysine thioester starting from α-amino Boc-protected lysine, in three simple steps (Scheme 5.1a). First the side chain amine group of Boc-protected lysine is

converted into azide using a previously reported azido transfer reagent[56] (Scheme 5.1a) to make azido-lysine **1**. The carboxylic acid group of **1** is then converted into a benzyl thioester and the Boc protecting group is removed from **1** with aqueous acid to generate the adapter **2** (Scheme 5.1a). This method produces the adapter **2** in three simple high yielding steps from inexpensive starting materials.

To test the effectiveness of the adapter molecule **2** in the NCL reaction, a peptide (Cys-Pep) containing an N-terminal cysteine (Scheme 5.1b) was used.[57,58] In this NCL reaction, 4-mercaptophenylacetic acid (MPAA) is used as a thiol transfer reagent and Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) is used as a reducing reagent to prevent disulfide bond formation. The reaction was complete within 15 minutes and we could obtain the azido labeled peptide (N_3)Lys-Pep in near quantitative yield as confirmed by HPLC and ESI mass spectrometry (Fig. 5.5).

Next we sought to evaluate the reactivity of the (N_3)Lys-Pep obtained in a subsequent click reaction to make peptide oligonucleotide conjugates (POCs). The (N_3)Lys-Pep could be used in a CuAAC reaction to an alkyne modified DNA1 or RNA1 both of which contained a 5'-fluorescent dye (Fig. 5.1a and experimental). The fluorescent dye (Dylight547, a Cy3 analogue) at the 5'-terminus of DNA1 or RNA1 was included to aid in visualization. For the CuAAC reaction, we used 0.6% acetonitrile as a co-solvent, a pseudo-ligandless condition developed in our lab.[59,60] Under these conditions along with the desired POCs, we observed disulfide bridged adduct formation between the unreacted peptide and the conjugated peptide (Fig. 5.6 and Fig. 5.7a). However this dithiol adduct can be easily eliminated by treating the reaction mixture with an excess of dithiothreitol (DTT) before gel electrophoresis (Fig. 5.7). The integrity of

the desired POCs following the CuAAC reaction was tested by a trypsin digest assay and gel electrophoresis (Fig. 5.1a). We observed that for both DNA1 and RNA1 conjugates, the trypsin digested product runs faster than the respective conjugates but slower than the DNA1 or RNA1 itself (Fig. 5.1b). This suggests that the oligonucleotide remains intact in the POCs after the CuAAC reaction and trypsin digestion as trypsin only cleaves the peptide part of the conjugate at the carboxyl side of arginine and lysine reside.

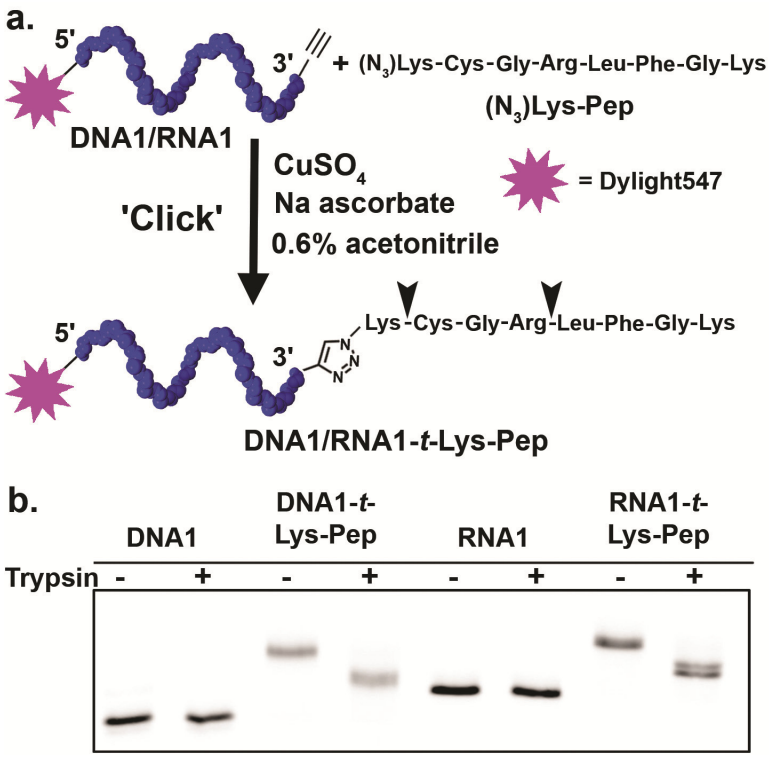


Figure 5.1. Synthesis of peptide oligonucleotide conjugates (POCs) of DNA1 and RNA1 with (N_3)Lys-Pep and their characterization. **a.** The azido modified peptide, (N_3)Lys-Pep was conjugated with DNA1 and RNA1 both containing a terminal alkyne (3'-*O-Propargyl*) and a 5'-fluorescent dye (Dylight547, a Cy3 equivalent) using CuAAC. A pseudo-ligandless CuAAC condition using 0.6% acetonitrile as a co-solvent was utilized to make the POCs (see Fig. 5.6). **b.** To test the integrity of the POCs after the NCL and CuAAC reactions, they were digested with trypsin and separated using gel electrophoresis. The fluorescence scan for Dylight547 showed complete cleavage of both the POCs of DNA1 and RNA1 by trypsin indicating that the oligonucleotide remains intact in the POCs after the CuAAC reaction and trypsin digestion.

5.3. Labeling of Dbr1p using native chemical ligation and CuAAC

With the success in labeling and conjugation of the Cys-Pep with the adapter **2**, we evaluated applicability of the adapter to label a large protein. We chose the lariat debranching enzyme Dbr1p that selectively cleaves the 2'-5' phosphodiester bonds of the lariat introns and is involved in important biological pathways such as miRNA and snoRNA biogenesis.[61] To conjugate Dbr1p to adapter **2** using NCL we required a Dbr1p that contains an N-terminal cysteine (Cys-Dbr1p). Methods to generate proteins containing an N-terminal cysteine for site specific labeling using NCL or other methods are well known.[39–41,62–64] To generate Cys-Dbr1p, we simply introduced a cysteine residue instead of glycine at the C-terminus of the TEV protease cleavage site using site directed mutagenesis (Fig. 5.2a). Subsequent cleavage of the expressed protein with TEV protease that is typically used in the protocol removed the histidine tag and generated Cys-Dbr1p (by Elizabeth Ransey and Dr. Mark MacBeth) (Fig. 5.2a).

Initial attempts of the NCL reaction of Cys-Dbr1p with adapter **2** were not successful in aqueous buffer at pH 7, likely due the lower solubility of **2**. Earlier with Cys-Pep, 6M guanidinium chloride was used to solubilize both the peptide and the dual adapter **2.**[58] However with Cys-Dbr1p, 6M guanidinium chloride would denature the enzyme. Therefore to solubilize the dual adapter **2**, we reacted it with an excess of MPAA before the NCL reaction. The thioester group of **2** undergoes a trans-thioesterification with MPAA and generates a thioester that is soluble in aqueous buffer at pH 7 (Scheme 5.2). The Cys-Dbr1p immediately reacted with this *in situ* generated soluble thioester to produce N-terminal azido-Dbr1p; N_3-Dbr1p (Fig. 5.2b).

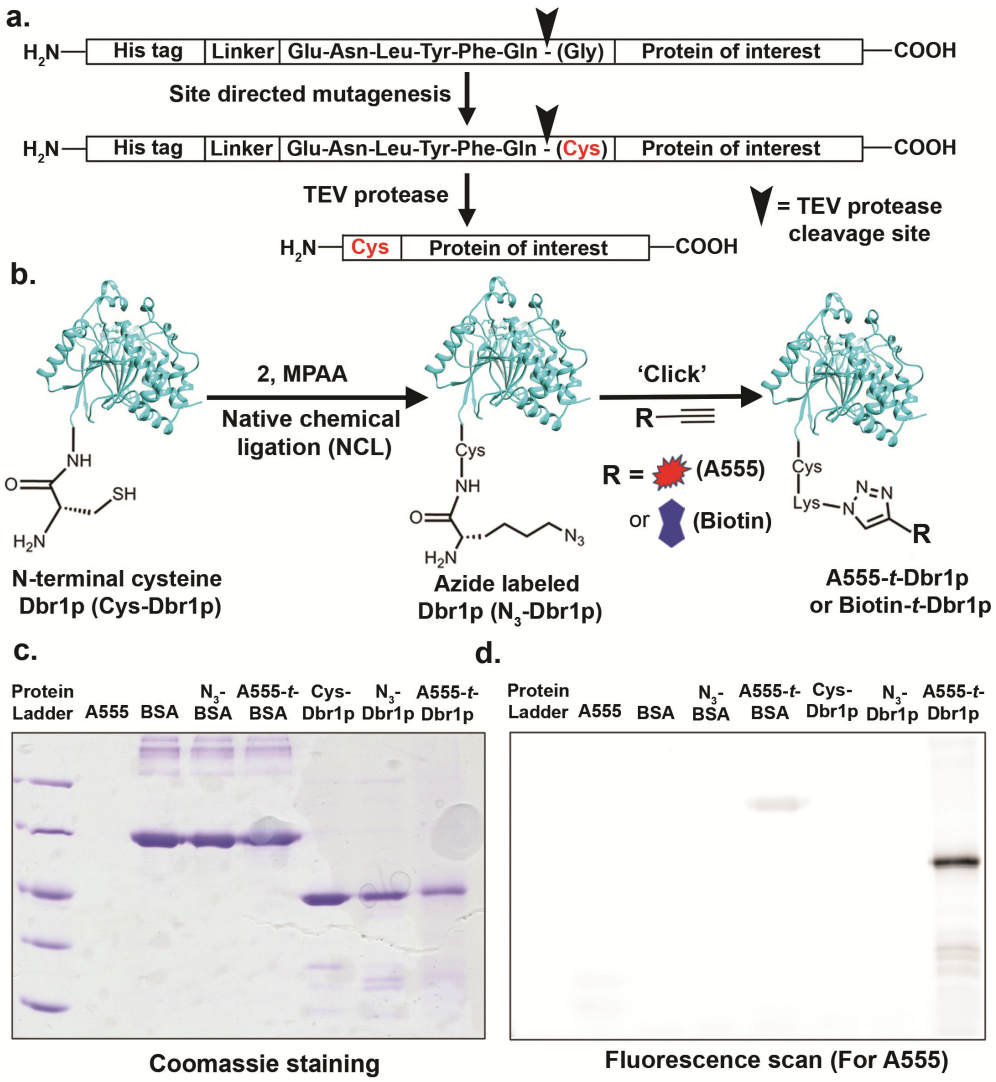


Figure 5.2. N-terminal labeling of Dbr1p using NCL and CuAAC reaction. **a.** To generate Dbr1p with N-terminal cysteine, the TEV protease cleavage site residue was modified from Gly to Cys (in red) using site directed mutagenesis. Following this, cleavage by TEV protease generated Dbr1p containing N-terminal cysteine (Cys-Dbr1p). **b.** The Cys-Dbr1p was reacted with the adapter **2** in presence of MPAA (See SI for exact procedure) to generate an azido modified Dbr1p (N_3-Dbr1p). Following this, the N_3-Dbr1p was labeled with alkyne derivatives of Alexa Fluor 555 (A555) or biotin using CuAAC reaction to generate A555-*t*-Dbr1p or biotin-*t*-Dbr1p respectively (See Experimentals for exact procedure). **c.** The labeling reaction of N_3-Dbr1p with A555 was analyzed by SDS-PAGE. We used BSA without N-terminal cysteine as a control to check non-specific labeling. After coomassie staining, both Dbr1p and BSA showed no degradation under either the NCL or CuAAC reaction condition. D) Fluorescence scan of the same gel from **c.** for A555 showed labeling of Dbr1p with A555 while BSA showed very little non-specific labeling.

Once N_3-Dbr1p has been accessed, it is possible to label it using the CuAAC reaction with a variety of commercially available alkyne tagged molecules. Using 0.6% acetonitrile pseudo-ligandless[59,60] conditions for CuAAC led to precipitation of N_3-Dbr1p, so we switched to the use of Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) as a Cu(I) ligand as reported by Hong *et al.*[65] Using this condition, we observed successful labeling of N_3-Dbr1p with Alexa Fluor 555 (A555) alkyne dye without any precipitation and generated A555-*t*-Dbr1p (Fig. 5.2b). In the reaction, bovine serum albumin (BSA) that contains no N-terminal cysteine does not get labeled and serves as a negative control for the NCL and the CuAAC reaction.

The labeling of N_3-Dbr1p with A555 alkyne to generate A555-*t*-Dbr1p was analyzed using gel electrophoresis and visualized by both coomassie staining and fluorescence scanning (Fig. 5.2c). Both the Cys-Dbr1p and the BSA negative control are stained by the coomassie dye before and after the labeling reactions. However, the fluorescence scan for A555 shows only a band in the A555-*t*-Dbr1p lane indicating that N_3-Dbr1p is successfully labeled with A555 after the CuAAC reaction. We only observed slight non-specific labeling of BSA. BSA is a highly negatively charged protein at pH 7 (pI 4.7) and therefore it could react non-specifically with the *in situ* generated active thioester that is a strong electrophile. To minimize the non-specific labeling of BSA, we varied the time for both the NCL and CuAAC reactions. We observed that reducing both the NCL and CuAAC reaction time to 2 hours reduces the non-specific labeling of BSA without reduction of the Dbr1p labeling (Fig. 5.8).

5.4. Debranching activity of the labeled Dbr1p

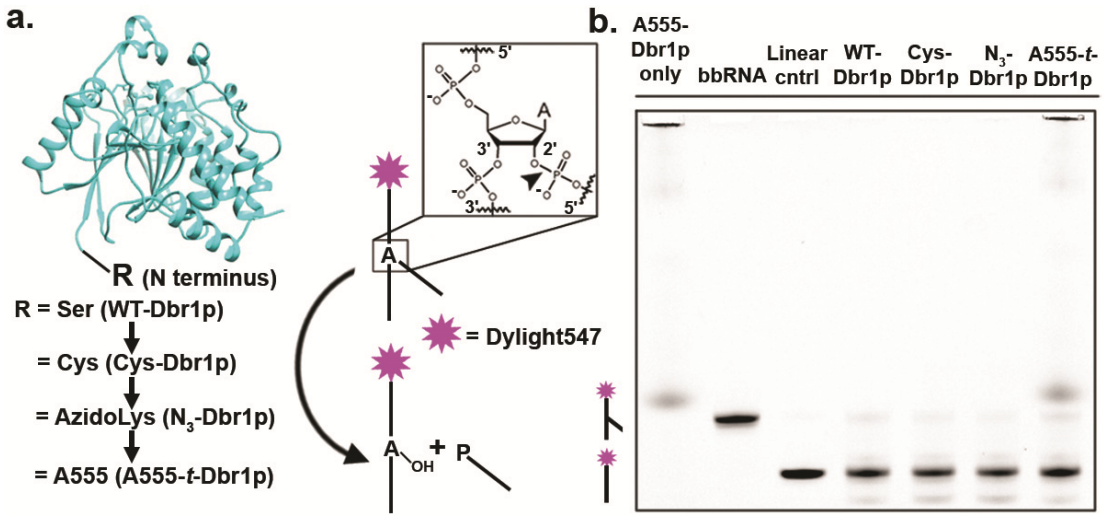


Figure 5.3. Dbr1p enzyme remains active after N-terminal labeling with A555. **a.** To check the debranching activity of Dbr1p after both the NCL and CuAAC reactions, we used a backbone branched RNA (bbRNA) substrate containing a 2'-5' phosphodiester bond (Inset) and a 5'-fluorescent dye (Dylight547, a Cy3 analogue). The bbRNA substrate and the cleaved product will run differently in gel electrophoresis and can be visualized using fluorescence scan. **b.** Fluorescence scan of the gel shows near complete cleavage of the substrate by WT-Dbr1p as well as by Cys-DBr1p, N_3-Dbr1p and A555-*t*-Dbr1p. This indicates that Dbr1p enzyme remains functional after both the NCL and CuAAC reaction. The A555-*t*-Dbr1p can be seen in the top wells of the gel (Lane 1 and 7), as it did not enter the gel due to much higher molecular weight compared to the substrate and the product.

To determine the integrity of the debranching activity of Dbr1p before and after the labeling reactions, we tested them with a backbone branched RNA (bbRNA) substrate (Fig. 5.3a and inset). The bbRNA substrate also contains a fluorescent label (Dylight 547) at the 5'-terminus. Following debranching by Dbr1p, the reaction mixture was separated by gel electrophoresis (Fig. 5.3a). The fluorescence scan of the gel for Dylight547 shows that along with the wild type Dbr1p (WT-Dbr1p), Cys-Dbr1p, N_3-Dbr1p and A555-*t*-Dbr1p also cleave the bbRNA substrate to near completion (Fig. 5.3b). Thus all the modified Dbr1p enzymes remain active after the NCL and CuAAC reactions indicating

the mild nature of the labeling using the adapter **2**. This suggests that we can use labeled Dbr1p produced in this method for biochemical and single molecule studies.

Proteins and enzymes labeled site-specifically with biotin are useful for biophysical studies that require surface immobilization.[66–68] Biotinylation of proteins can be achieved both non-specifically and site-specifically. However site-specific biotinylation can provide a single population of immobilized protein and is significant in single molecule studies of folding and dynamics.[8,10] Non-specific biotinylation can lead to multiple populations and heterogeneity of the immobilized proteins can generate ambiguous results. We therefore tested the use of adapter **2** for conjugation of biotin alkyne to Dbr1p. Using the previously used CuAAC reaction condition with THPTA, we generated biotin-*t*-Dbr1p from N_3-Dbr1p. To test the activity of the site-specifically labeled biotin-*t*-Dbr1p immobilized on solid surface, it was added to streptavidin beads and the beads were used in a debranching assay (Fig. 5.4a). We also tested the activity of the supernatant after separating the immobilized biotin-*t*-Dbr1p by centrifugation to check the amount of unbound Dbr1p. The scanned gel following the assay shows that compared to the supernatant, the immobilized biotin-*t*-Dbr1p has more debranching activity indicating that it remains active while bound to the beads (Fig. 5.4b). In contrast, N_3-Dbr1p shows significant activity in the supernatant compared to the beads indicating some non-specific binding of N_3-Dbr1p with streptavidin beads (Fig. 5.4b, lane 6). These results suggest that the biotinylation of an N-terminal cysteine containing enzyme using the NCL and the CuAAC reactions can readily provide functional immobilized enzyme.

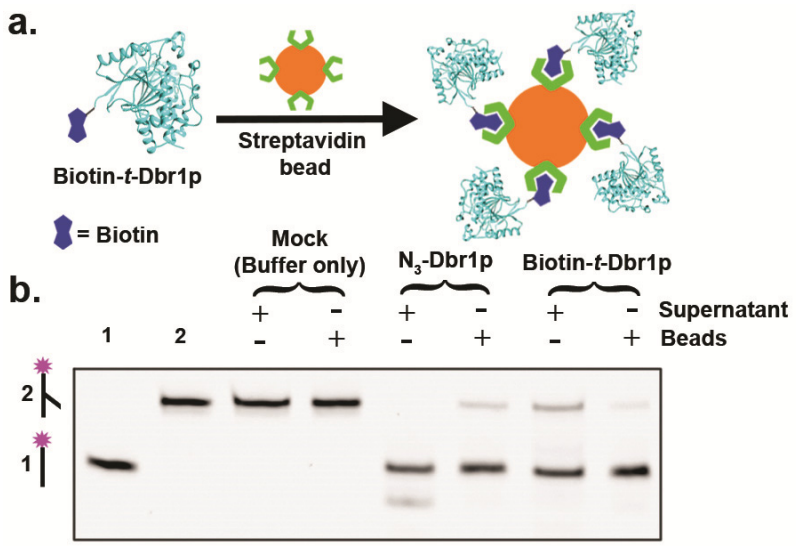


Figure 5.4. Activity of biotin-*t*-Dbr1p attached to solid beads. **a.** To evaluate the activity of biotin-*t*-Dbr1p on solid surface, it was immobilized on streptavidin beads. Following this, we used the same debranching assay as before to check Dbr1p activity of both the immobilized beads and the supernatant (See SI for reaction conditions). **b.** Fluorescence scan of the gel shows Dbr1p labeled with biotin is still active even when immobilized on solid surface. We used the N_3-Dbr1p as a control to evaluate non-specific binding with streptavidin beads. The mock sample without any Dbr1p showed no cleavage of the bbRNA, whereas the N_3-Dbr1p showed more cleavage in the supernatant compared to the beads indicating some non-specific binding to the streptavidin beads. However the biotin-*t*-Dbr1p showed more cleavage in the beads compared to the supernatant indicating that the biotin modified enzyme becomes immobilized on the streptavidin beads and remains functional even after immobilization.

5.5. Conclusion

We have developed a general and very mild method to label a protein of interest at the N-terminus using a two-step method that combines NCL and click chemistry. The adapter molecule for this can be readily obtained from commercially available starting materials in high yields. We use this adapter molecule to generate both peptide oligonucleotide conjugates and an N-terminally labeled enzyme. We demonstrate that the lariat debranching enzyme, Dbr1p remains active after both the NCL and CuAAC

reactions, indicating that our method of site specific labeling is mild, such that the enzyme retains function. The debranching assay following surface immobilization of biotin-t-Dbr1p shows that our method can be used to immobilize an enzyme on beads and surfaces for biophysical studies. This site-specific incorporation of the azide group into a protein using the adapter opens up avenue for attaching diverse array of biologically important tags using CuAAC reaction. We envision that an even milder copper free version of the conjugation reaction between an azide and strained alkyne could be used to label proteins using our adapter **2**[69]. Thus our approach enhances the toolset of methods to site-specifically label any protein the N-terminus with biophysical probes for the study of protein structure, dynamics and interactions. Labeled Dbr1p enzyme generated using this method would be useful for future single molecule experiments.

5.6 Experimental

Materials and methods

The adapter molecule **2** was synthesized by Dr. Debasish Grahacharya. The Cys-Pep was purchased from Lifetein. MPAA and TCEP.HCl were purchased from Alfa Aesar. Trypsin was purchased from New England Biolabs. Alexa Fluor 555 alkyne (A555 alkyne) was purchased from Invitorgen. biotin-TEG alkyne was purchased from Click Chemistry Tools. Standard DNA and RNA phosphoramidites with ultramild protecting groups (dA-PAC, dC-PAC and dG-^{t}Bu-PAC, A-PAC, G-PAC and C-PAC) were purchased from Chemgenes (Wilmington, MA, USA). Appropriate reagents for solid phase DNA synthesis (deblock, activator, ultramild CapA, CapB and oxidation reagent) and Dylight 547 phosphoramidite were purchased from Glen Research (Sterling,

VA, USA). CPG solid supports for DNA synthesis were purchased from Biosearch Technologies (Novato, CA, USA) or Chemgenes (Wilmington, MA, USA).

Fluorescence scan of the PAGE gels were performed using a Typhoon FLA-9000 fluorescence scanner. The synthesis of the DNA and RNA were performed in a MerMade4 automated synthesizer (Bioautomation, Plano, Texas) and then deprotection were performed using manufacturer's protocols.

NCL of the dual adapter 2 with Cys-Pep

An NCL reaction buffer containing 6M gunidium chloride (Gu.HCl) and 0.2M sodium phosphate dibasic (Na_2HPO_4) was prepared by dissolving 5.74 g (60 mmol) of Gu.HCl and 284 mg (2mmol) of Na_2HPO_4 in water and making the final volume to 10 mL. The final pH of the solution was adjusted to 7 using NaOH. The buffer was then degassed for 15 minutes by blowing argon. Following this, 2 mL of this NCL buffer was added to 11.5 mg (0.04 mmol) of tris-2-carboxyethylphosphine hydrochloride (TCEP.HCl) and 16.8 mg (0.1 mmol) of 4-mercaptophenlacetic acid (MPAA) to generate a final solution containing 20 mM TECP.HCl and 50 mM MPAA. The pH of this mixture was adjusted to 7 again using NaOH and then this solution was filtered using a 0.2 μm syringe filter. Then this solution was degassed for another 30 minutes. 2.8 mg (0.01 mmol) of the dual adapter (**2**) and 7 mg (0.009 mmol) of the Cys-Pep were weighed out in separate vials. Then 2 mL of the degassed buffer mixture (containing phosphate, 6M Gu.HCl, MPAA, TCEP.HCl) was added first to the Cys-Pep to dissolve it and then this solution was added to the dual adapter **2**, to initiate the reaction. The final concentration of Cys-Pep and **2** in the reaction mixture are 5 mM and 4.5 mM respectively. The reaction was continued for 15 minutes while blowing argon in the solution. After 15

minutes, the pH of the solution was adjusted to 2 using TFA in order to stop the reaction. Then the (N_3)Lys-Pep was purified by RP-HPLC and characterized by ESI-MS (Fig. 5.5).

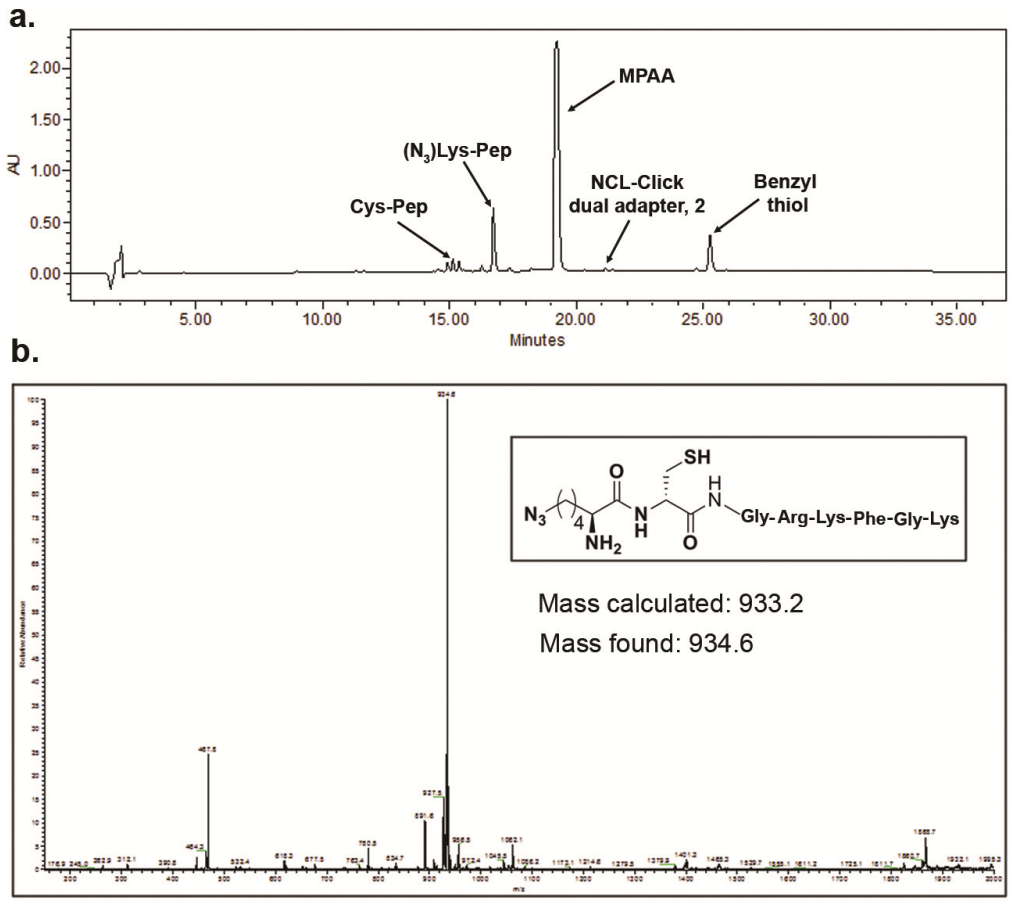


Figure 5.5. Synthesis and characterization of azido labeled peptide ((N_3)Lys-Pep). **a.** Following the NCL reaction of the dual adapter **2** with Cys-Pep, the reaction mixture was separated by HPLC. HPLC Condition: Column: Waters Xbridge 4.6x150 mm Column (5 µm particle), Column Temp: 40^0C. SolventA: 0.1% TFA in water, SolventB 0.1% TFA in Acetonitrile, Gradient: 0% to 80% B in 30 minutes. The chromatogram shows almost complete conversion of the dual adapter **2** into (N_3)Lys-Pep (The Cys-Pep was used in slight excess). **b.** ESI mass spectrometry of the (N_3)Lys-Pep after HPLC purification confirms the successful NCL reaction and conjugation of azido group to the peptide.

DNA/RNA synthesis

Solid phase synthesis of the DNA1 and RNA1 were performed on a Mermade-4 (Bioautomation, Plano, TX, USA) automated synthesizer. Following the synthesis the CPG beads were dried using a stream of argon. The deprotection of the DNA1 and RNA1

were performed using manufacturer's protocol. The DNA1 and the RNA1 used for the click reaction were purified using 15% denaturing polyacrylamide gel electrophoresis (with 8M urea, 1x TBE). The bands in the gel was excised and eluted overnight in $TE_{0.1}$ buffer (10 mM Tris.HCl, 0.1 mM EDTA, pH 7.5). The eluted DNA1 and RNA1 were desalted again using a WatersC18 Sep-Pak cartridge (Waters, Milford, MA, USA) and lyophilized. They were characterized by MALDI mass spectrometry using 3-hydroxypicolinic acid (3-HPA) as matrix.

Name	Sequence	5'-terminus	3'-terminus	Mass Calculated	Mass Found
DNA1	5'-TCT AAG GCG CGT TAG A-3'	Dylight547	3'-*O*-Propargyl	5464.12	5464.36
RNA1	5'-UCU AAG GCG CGU UAG A-3'	Dylight547	3'-*O*-Propargyl	5647.97	5646.40

Table 5.1. Sequence and MALDI mass of the DNA and RNA for peptide oligo conjugates.

Click reaction of (N_3)Lys-Pep with DNA1 and RNA1 for making DNA1/RNA1-t-Lys-Pep

Click reactions to make POCs were performed by using either DNA1 or RNA1 (both containing 3'-alkyne and 5'-Dylight547, 50 μM final), (N_3)Lys-Pep (100 μM final), Tris.HCl (pH 7.5, 20 mM final), $CuSO_4$ (500 μM final), sodium ascorbate (2.5 mM final), acetonitrile (0.6% v/v final) in 50 μL total volume. All the components except sodium ascorbate and $CuSO_4$ were mixed together and degassed for 5 minutes by blowing argon. The sodium ascorbate and $CuSO_4$ solutions were also degassed separately by blowing argon for 5 minutes. Finally the reaction was started by adding the required

amount of degassed sodium ascorbate and $CuSO_4$ solutions to the reaction mixture. The reactions were continued for 2 hours. After 2 hours the reaction mixture was quenched by 50 µL of gel loading solution (90% Formamide, 10% 0.1M EDTA) and purified by a 12% denaturing PAGE (0.5x TBE running buffer). The bands corresponding to the products were cut (see Fig. 5.6) and eluted from the gel pieces by passive elution using $TE_{0.1}$ (10 mM Tris.HCl, 0.1 M EDTA, pH 7.5) buffer and then desalted using Waters SepPak columns.

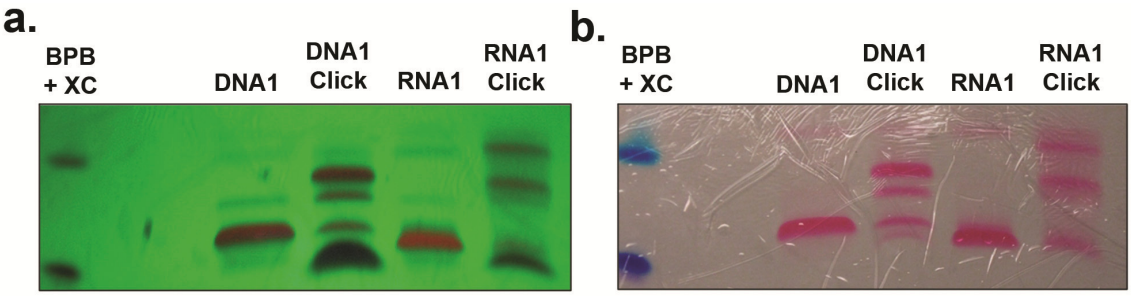


Figure 5.6. Click reaction of (N_3)Lys-Pep with DNA1 and RNA1 to make POCs. **a.** The (N_3)Lys-Pep was conjugated to DNA1 and RNA1 both containing 3'-O-Propargyl group and a 5'-fluorescent dye (DyLight547, a Cy3 equivalent) using a pseudo ligandless click reaction condition. The reaction was analyzed by PAGE. The UV shadowing of the gel shows successful conjugation of the DNA1 and RNA1 with the (N_3)Lys-Pep resulting in products which runs slower. The two product bands observed for both the DNA1 and the RNA1, are due to disulfide bond formation between the POCs with the excess peptide which can be easily eliminated. BPB: bromophenol blue and XC: xylene cyanol are marker dyes. **b.** Visual image of the same gel as in A) showing both the products and the starting materials contain the fluorescent dye.

Cleavage of the disulfide bridged product

As it can be seen from the PAGE for the click reaction, two product bands were observed for both the DNA1 and the RNA1. The bottom product band was the expected POC whereas the top band disulfide bridged adduct of the POC with the unreacted peptide (Fig. 5.7a) as predicted by mass spectrometry. We show here that the disulfide

bridged adduct can be eliminated for the DNA1 by treatment with dithiothreitol (DTT). Two click reactions for the DNA1 and the peptide using the condition described before were performed. One of the reactions was treated with DTT (2 mM final) at room temperature for 30 minutes whereas the other reaction was treated with water. Both the reaction mixture were mixed with the gel loading solution and run on a 12% denaturing PAGE (0.5x TBE running buffer). In the gel, the band corresponding to the disulfide adduct is almost below detection limit when the reaction mixture was treated with DTT (Fig. 5.7b), indicating that the DTT treatment is effective.

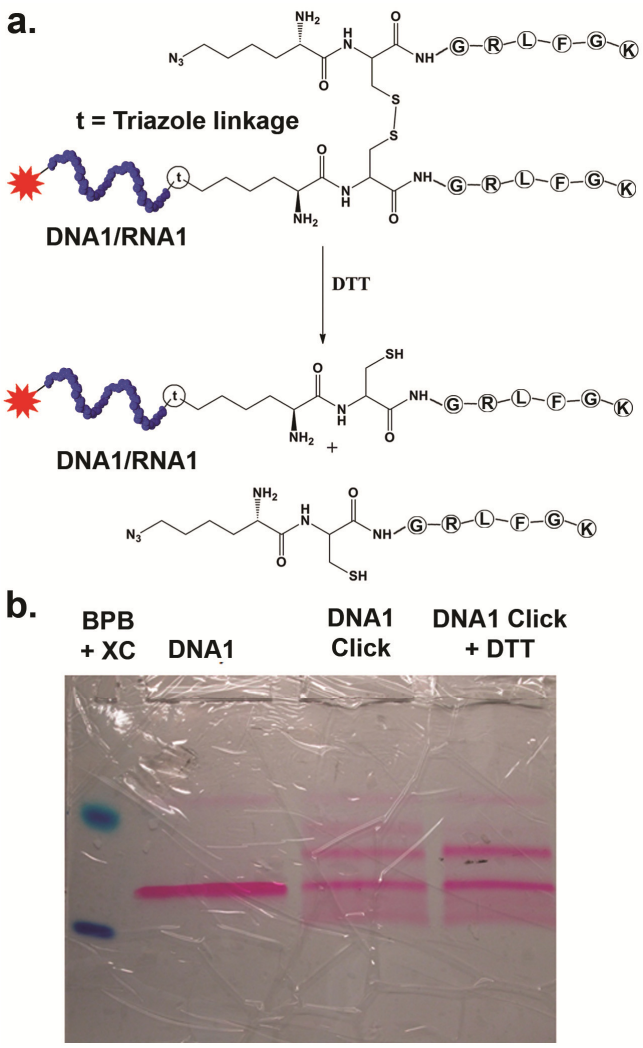


Figure 5.7. Disulfide bond formation between the POCs and the excess peptide under click reaction condition and its elimination. **a.** Under the click reaction condition, the POCs react with the excess peptide through the cysteine thiols to generate a disulfide linked product. This can be easily eliminated by treating the reaction mixture with an excess of dithiothreitol (DTT) before running the reaction in PAGE. **b.** Visual scan of the gel confirms elimination of the disulfide adduct after treating the reaction mixture with DTT. The DNA1 click lane shows a band above the usual product band which is the disulfide adduct. However when the reaction mixture was treated with DTT before loading it on gel (DNA1 click + DTT), that band disappears indicating cleavage of the disulfide adduct.

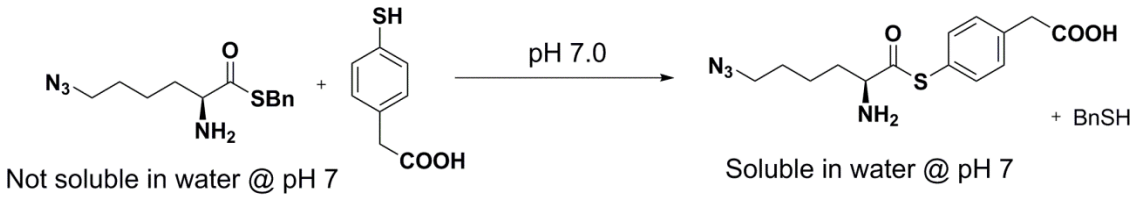
Trypsin digestion of the peptide-oligonucleotide conjugates (POCs)

The DNA1 and RNA1 POCs were treated with trypsin in order to evaluate the integrity of the conjugates after the click reaction. The DNA1 and RNA1 conjugates (40 pmoles each) were mixed with 0.8 μg of trypsin in 20 μL of Trypsin digestion buffer and incubated at 37°C for 1 hour. Following this, the reaction mixture was mixed with 20 μL of gel loading solution and run on a 20% denaturing PAGE (0.5x TBE running buffer) along with the unconjugated DNA1 and RNA1 controls. Then the gel was scanned in a Typhoon FLA900 fluorescence scanner using the settings for Cy3 dye. Complete digestion of the POCs by trypsin indicates that the conjugates did not degrade under the click reaction condition (Fig. 5.1b).

NCL reaction of 2 with Cys-Dbr1p

The Cys-Dbr1p enzyme was overexpressed in *E.Coli* and purified by Elizabeth Ransey and Dr. Mark MacBeth. Due to insolubility of **2** in aqueous buffer at pH 7, it was first reacted with MPAA to make a water soluble thioester (Scheme 5.2) and then reacted with Cys-Dbr1p. 5 mg (30 μmol) of MPAA was dissolved in 1.5 mL of 100 mM Tris.HCl buffer to a final concentration of 20 mM and the pH was adjusted to 7. 1 mL of

this MPAA solution was then added to 1.2 mg (4 μmol) of the dual adapter **2.** The solution was vortexed for 5mins and the white precipitate slowly disappeared. Then this solution was filtered through a 0.2 μm filter to remove any particulate matter. Then 100 μL of this solution was added to 100 μL of 50 μM Cys-Dbr1p and mixed well. The final NaCl concentration was adjusted to 200mM using 1M NaCl solution. The reaction was continued for 30 minutes at room temperature and then for either 1.5 or 14.5 hours at 4°C. Upon completion, the excess of MPAA and thioester were removed using a 10 KDa MWCO filter and also the buffer for the N_3-Dbr1p enzyme was exchanged to buffer D which was used for SEC (See expression and purification of Cys-Dbr1p). We have used bovine serum albumin (BSA) without any N-terminal cysteine as a negative control for the reaction and we followed the exact same protocol as Cys-Dbr1p described above for the NCL reaction.



Scheme 5.2. Thiol exchange reaction to convert insoluble dual adapter **2** into soluble thioester. The NCL click adapter **2** was not very soluble in aqueous buffer at pH 7. However when treated with MPAA, it readily undergoes trans-thioesterification reaction resulting in a much more soluble and reactive thioester which can be readily coupled to the Cys-Dbr1p containing N-terminal cysteine to label it at the N-terminus with azide groups.

Click reaction of N_3-Dbr1p with Alexa555 or bitoin-TEG alkyne

The N_3-Dbr1p was conjugated to either Alexa555 or biotin-TEG alkyne using CuAAC. A catalyst for the CuAAC reaction was prepared by mixing 50 μL of 1 mM $CuSO_4$ and 25 μL of 10 mM THPTA ligand and then degassing the solution for 5

minutes by blowing argon. Following this 50 µL of 50 µM N_3-Dbr1p was mixed with 7 µL of 1 mM A555 alkyne or biotin-TEG alkyne, 20 µL of 1 M NaCl (degassed) and 15 µL of the catalyst solution. Then the reaction was started by adding 5 µL of 50 mM degassed sodium ascorbate solution. The reaction was continued for 30 minutes at room temperature and then either for 1.5 or 14.5 hours at 4°C. Upon completion of the reaction, the excess dye and catalyst solution was separated from the enzyme using a zeba spin column (Pierce) following manufacturer's protocol. An exact similar reaction was done using the N_3-BSA to check the non-specific labeling of BSA by this method. After purification of the products exact same amount of the enzymes were run in SDS-PAGE gel along with the appropriate controls. The gels were first scanned using a Typhoon FLA9000 fluorescence scanner for Alexa555 (Fig. 5.2d). Following this the gels were stained with Coomassie stain to visualize the Dbr1p and BSA bands that are not fluorescently labeled (Fig. 2c). The amount of fluorescence signal was quantified using the inbuilt software in the instrument. Also another SDS-PAGE gel for comparison of the non-specific labeling under different conditions were run and scanned using the fluorescence scanner (Fig. 5.8).

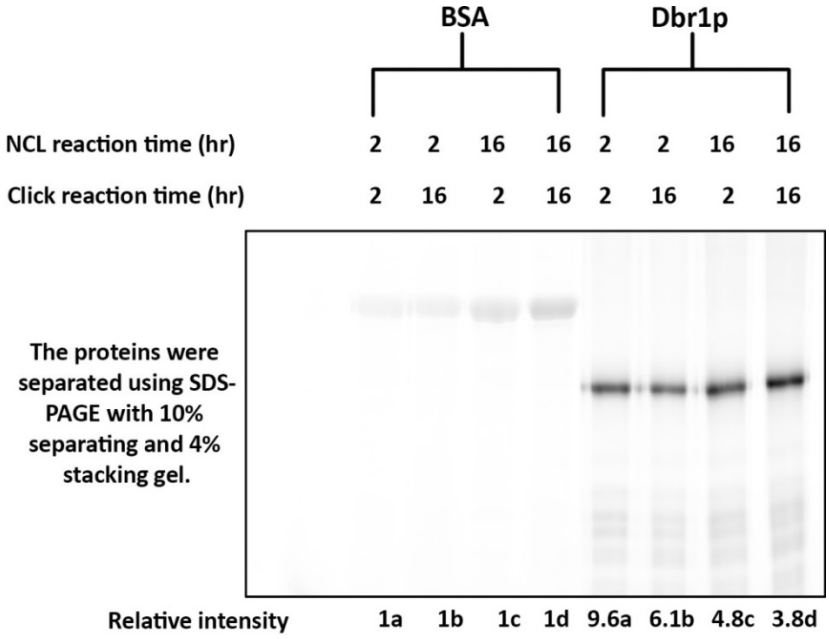


Figure 5.8. Optimization of the NCL and click reaction condition to reduce non-specific labeling of BSA compared to site specific labeling of Dbr1p. We have tried several different time conditions for the NCL and click reaction to make A555-*t*-Dbr1p along with the BSA control. Following the labeling reaction the proteins were run on SDS-PAGE and visualized using fluorescence scan for A555. 1a-d indicates the intensity of the BSA bands for each condition. The next four numbers indicates the relative intensity of the Dbr1p bands compared to the BSA bands. From the relative intensity of the bands it is clear that we get minimum non-specific labelling when both the NCL and the CuAAC reactions are carried out for 2 hours. The highest non-specific labeling was observed when both the reactions were carried out for 16 hours.

Debranching assay for checking activities of different Dbr1p enzymes

To check the activity of the different Dbr1p enzymes after the NCL and CuAAC reactions, a debranching assay using a fluorescently labeled bbRNA substrate was used. The sequence of the bbRNA used for the assay was [5'Dylight547-GUAUGA-A-(2'-5'-GUAUGA)-CAAGUU-3']. The final concentrations of the different components in 20 μL final volume are as follows: 500 nM Dbr1p (WT-Dbr1p, Cys-Dbr1p, N_3-Dbr1p and A555-*t*-Dbr1p), 100 nM bbRNA substrate, 50 mM Tris.HCl (pH 7.5), 4 mM $MnCl_2$, 25 mM NaCl, 2.5 mM DTT and 0.05% TritonX-100. The reactions were incubated at 30°C for 10 minutes and then stopped using 20 μL of the gel loading solution (90% formamide

and 10% 0.1 M EDTA). Then the reaction mixtures were run in 20% denaturing PAGE to separate the cleaved product from the bbRNA substrate. Following this the gels were scanned using the Typhoon FLA9000 fluorescence scanner and quantified.

Surface immobilization followed by debranching activity of biotin-t-Dbr1p

To check the activity of the Dbr1p on solid surface, the biotin-*t*-Dbr1p was immobilized on the streptavidin beads. 50 µL of the bead suspension (containing about 0.5 nmoles of streptavidin) was washed with buffer D two times and then incubated with either 30 µL of 10 µM (30 pmoles) biotin-*t*-Dbr1p or N_3-Dbr1p and 70 µL of buffer D. The incubation was continued overnight at 4°C. Then the beads were centrifuged at 5000 rpm for 2 minutes to form a pellet and the supernatant was taken out into a separate tube. The beads were washed with the buffer D three times to remove any unbound Dbr1p. Then both the beads and the supernatant (100 µL) were used in debranching assays with the same bbRNA1 substrate described earlier. The final concentrations of the different components in 200 µL final volume are as follows: 100 nM bbRNA substrate, 50 mM Tris.HCl (pH 7.5), 4 mM $MnCl_2$, 25 mM NaCl, 2.5 mM DTT and 0.05% TritonX-100. The reactions were incubated at 30°C for 1hour and then stopped using 200 µL of the gel loading solution (90% formamide and 10% 0.1M EDTA). For the reaction with the Dbr1p immobilized beads, it was centrifuged @ 6000rpm for 3 minutes to pellet out the beads and the supernatant was collected for running in the gel. 20 µL from each reaction mixture were run in 20% denaturing PAGE to separate the cleaved product from the bbRNA substrate. Following this the gels were scanned using the Typhoon FLA9000 fluorescence scanner and quantified.

References

1. O'Hare, H. M., Johnsson, K. & Gautier, A. Chemical probes shed light on protein function. *Curr. Opin. Struct. Biol.* **17,** 488–94 (2007).
2. Miyawaki, A. Proteins on the move: insights gained from fluorescent protein technologies. *Nat. Rev. Mol. Cell Biol.* **12,** 656–68 (2011).
3. Jing, C. & Cornish, V. W. Chemical Tags for Labeling Proteins Inside Living Cells. *Acc. Chem. Res.* **44,** 784–792 (2011).
4. Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312,** 217–24 (2006).
5. Marks, K. M. & Nolan, G. P. Chemical labeling strategies for cell biology. *Nat. Methods* **3,** 591–6 (2006).
6. Brustad, E. M., Lemke, E. a, Schultz, P. G. & Deniz, A. a. A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.* **130,** 17664–5 (2008).
7. Shi, X. *et al.* Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. *Nat. Methods* **9,** 499–503 (2012).
8. Borgia, A., Williams, P. M. & Clarke, J. Single-molecule studies of protein folding. *Annu. Rev. Biochem.* **77,** 101–25 (2008).
9. Hoskins, A. a *et al.* Ordered and dynamic assembly of single spliceosomes. *Science* **331,** 1289–95 (2011).
10. Cecconi, C., Shank, E. a, Bustamante, C. & Marqusee, S. Direct observation of the three-state folding of a single protein molecule. *Science* **309,** 2057–2060 (2005).
11. Pal, P., Lesoine, J. F., Lieb, M. A., Novotny, L. & Knauf, P. a. A novel immobilization method for single protein spFRET studies. *Biophys. J.* **89,** L11–L13 (2005).
12. Steen Redeker, E. *et al.* Protein engineering for directed immobilization. *Bioconjug. Chem.* **24,** 1761–1777 (2013).

13. Camarero, J. a. Recent developments in the site-specific immobilization of proteins onto solid supports. *Biopolym. - Pept. Sci. Sect.* **90,** 450–458 (2008).

14. Nemoto, N. *et al.* A versatile C-terminal specific biotinylation of proteins using both a puromycin-linker and a cell-free translation system for studying high-throughput protein-molecule interactions. *Anal. Chem.* **86,** 8535–8540 (2014).

15. Smith, J. J., Conrad, D. W., Cuneo, M. J. & Hellinga, H. W. Orthogonal site-specific protein modification by engineering reversible thiol protection mechanisms. *Protein Sci.* **14,** 64–73 (2005).

16. de Lorimier, R. M., Tian, Y. & Hellinga, H. W. Binding and signaling of surface-immobilized reagentless fluorescent biosensors derived from periplasmic binding proteins. *Protein Sci.* **15,** 1936–1944 (2006).

17. Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **67,** 509–544 (1998).

18. Wu, B., Piatkevich, K. D., Lionnet, T., Singer, R. H. & Verkhusha, V. V. Modern fluorescent proteins and imaging technologies to study gene expression, nuclear localization, and dynamics. *Curr. Opin. Cell Biol.* **23,** 310–7 (2011).

19. Miller, L. W., Sable, J., Goelet, P., Sheetz, M. P. & Cornish, V. W. Methotrexate conjugates: a molecular in vivo protein tag. *Angew. Chem. Int. Ed. Engl.* **43,** 1672–5 (2004).

20. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21,** 86–9 (2003).

21. Gautier, A. *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **15,** 128–36 (2008).

22. Los, G. V *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3,** 373–382 (2008).

23. Szent-Gyorgyi, C. *et al.* Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nat. Biotechnol.* **26,** 235–40 (2008).

24. Griffin, B. A., Adams, S. R. & Tsien, R. Y. Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells. *Science* **281,** 269–272 (1998).

25. Gaietta, G. *et al.* Multicolor and electron microscopic imaging of connexin trafficking. *Science* **296,** 503–7 (2002).

26. Hauser, C. T. & Tsien, R. Y. A hexahistidine-Zn2+-dye label reveals STIM1 surface exposure. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 3693–7 (2007).

27. Guignet, E. G., Hovius, R. & Vogel, H. Reversible site-selective labeling of membrane proteins in live cells. *Nat. Biotechnol.* **22,** 440–4 (2004).

28. Halo, T. L., Appelbaum, J., Hobert, E. M., Balkin, D. M. & Schepartz, A. Selective recognition of protein tetraserine motifs with a cell-permeable, pro-fluorescent bis-boronic acid. *J. Am. Chem. Soc.* **131,** 438–9 (2009).

29. Chen, I., Howarth, M., Lin, W. & Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2,** 99–104 (2005).

30. Zhou, Z., Koglin, A., Wang, Y., McMahon, A. P. & Walsh, C. T. An eight residue fragment of an acyl carrier protein suffices for post-translational introduction of fluorescent pantetheinyl arms in protein modification in vitro and in vivo. *J. Am. Chem. Soc.* **130,** 9925–30 (2008).

31. Zhou, Z. *et al.* Genetically Encoded Short Peptide Tags for Orthogonal Protein Labeling by Sfp and AcpS Phosphopantetheinyl Transferases. *ACS Chem. Biol.* **2,** 337–346 (2007).

32. Tanaka, T., Yamamoto, T., Tsukiji, S. & Nagamune, T. Site-specific protein modification on living cells catalyzed by Sortase. *Chembiochem* **9,** 802–7 (2008).

33. Rashidian, M., Dozier, J. K. & Distefano, M. D. Enzymatic labeling of proteins: techniques and approaches. *Bioconjug. Chem.* **24,** 1277–94 (2013).

34. Wu, P. *et al.* Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 3000–5 (2009).

35. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413–44 (2010).

36. Johnson, J. a, Lu, Y. Y., Van Deventer, J. a & Tirrell, D. a. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14,** 774–80 (2010).

37. Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **114,** 4764–806 (2014).

38. MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **11,** 326–331 (2015).

39. Tolbert, T. J. & Wong, C. New Methods for Proteomic Research: Preparation of Proteins with N-Terminal Cysteines for Labeling and Conjugation. *Angew. Chem. Int. Ed. Engl.* **41,** 2171–2174 (2002).

40. Gentle, I. E., De Souza, D. P. & Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug. Chem.* **15,** 658–63 (2004).

41. Busch, G. K. *et al.* Specific N-terminal protein labelling: use of FMDV 3C pro protease and native chemical ligation. *Chem. Commun. (Camb).* 3369–71 (2008). doi:10.1039/b806727a

42. Erlanson, D. A., Chytil, M. & Verdine, G. L. The leucine zipper domain controls the orientation NFATeAP-VDNA complex of AP-1 in the. *Chem. Biol.* (1996).

43. Musiol, H.-J. *et al.* Toward semisynthetic lipoproteins by convergent strategies based on click and ligation chemistry. *Chembiochem* **6,** 625–8 (2005).

44. Lee, D. J., Mandal, K., Harris, P. W. R., Brimble, M. a & Kent, S. B. H. A one-pot approach to neoglycopeptides using orthogonal native chemical ligation and click chemistry. *Org. Lett.* **11,** 5270–3 (2009).

45. Xiao, J. & Tolbert, T. J. Synthesis of N-terminally linked protein dimers and trimers by a combined native chemical ligation-CuAAC click chemistry strategy. *Org. Lett.* **11,** 4144–7 (2009).

46. Lu, K., Duan, Q.-P., Ma, L. & Zhao, D.-X. Chemical strategies for the synthesis of peptide-oligonucleotide conjugates. *Bioconjug. Chem.* **21,** 187–202 (2010).

47. Diezmann, F., Eberhard, H. & Seitz, O. Native chemical ligation in the synthesis of internally modified oligonucleotide-peptide conjugates. *Biopolymers* **94,** 397–404 (2010).

48. Juliano, R. L., Ming, X. & Nakagawa, O. The chemistry and biology of oligonucleotide conjugates. *Acc. Chem. Res.* **45,** 1067–76 (2012).

49. Tung, C. H. & Stein, S. Preparation and applications of peptide-oligonucleotide conjugates. *Bioconjug. Chem.* **11,** 605–18 (2000).

50. Järver, P. *et al.* Peptide-mediated Cell and In Vivo Delivery of Antisense Oligonucleotides and siRNA. *Mol. Ther. Nucleic Acids* **1,** e27 (2012).

51. Stetsenko, D. a & Gait, M. J. Efficient conjugation of peptides to oligonucleotides by 'native ligation'. *J. Org. Chem.* **65,** 4900–8 (2000).

52. Gramlich, P. M. E., Wirges, C. T., Manetto, A. & Carell, T. Postsynthetic DNA modification through the copper-catalyzed azide-alkyne cycloaddition reaction. *Angew. Chem. Int. Ed. Engl.* **47,** 8350–8 (2008).

53. Fonvielle, M. *et al.* Decoding the logic of the tRNA regiospecificity of nonribosomal FemX(Wv) aminoacyl transferase. *Angew. Chem. Int. Ed. Engl.* **49,** 5115–9 (2010).

54. Chemama, M. *et al.* Stable analogues of aminoacyl-tRNA for inhibition of an essential step of bacterial cell-wall synthesis. *J. Am. Chem. Soc.* **129,** 12642–3 (2007).

55. Xiao, J. & Tolbert, T. J. Supporting Information Synthesis of N-Terminally Linked Protein Dimers and Trimers by a Combined Native Chemical Ligation-CuAAC Click Chemistry Strategy. 1–15

56. Goddard-Borger, E. D. & Stick, R. V. An efficient, inexpensive, and shelf-stable diazotransfer reagent: imidazole-1-sulfonyl azide hydrochloride. *Org. Lett.* **9,** 3797–800 (2007).

57. Dawson, P. E., Churchill, M. J., Ghadiri, M. R. & Kent, S. B. H. Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives. *J. Am. Chem. Soc.* **119,** 4325–4329 (1997).

58. Johnson, E. C. B. & Kent, S. B. H. Insights into the mechanism and catalysis of the native chemical ligation reaction. *J. Am. Chem. Soc.* **128,** 6640–6 (2006).

59. Paredes, E. & Das, S. R. Click chemistry for rapid labeling and ligation of RNA. *Chembiochem* **12,** 125–31 (2011).

60. Paredes, E. & Das, S. R. Optimization of acetonitrile co-solvent and copper stoichiometry for pseudo-ligandless click chemistry with nucleic acids. *Bioorg. Med. Chem. Lett.* **22,** 5313–6 (2012).

61. Dey, S. K., Paredes, E., Evans, M. & Das, S. R. in (eds. Erdmann, V. A. & Barciszewski, J.) 475–501 (Springer Berlin Heidelberg, 2012). doi:10.1007/978-3-642-27426-8

62. Erlanson, D. A., Chytil, M. & Verdine, G. L. The leucine zipper domain controls the orientation of AP-1 in the NFAT AP-1 DNA complex. *Chem. Biol.* **3,** 981–991 (1996).

63. Liu, D., Xu, R., Dutta, K. & Cowburn, D. N-terminal cysteinyl proteins can be prepared using thrombin cleavage. *FEBS Lett.* **582,** 1163–7 (2008).

64. Ren, H. *et al.* A Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chemie* **121,** 9838–9842 (2009).

65. Hong, V., Presolski, S. I., Ma, C. & Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed. Engl.* **48,** 9879–83 (2009).

66. Helms, B., van Baal, I., Merkx, M. & Meijer, E. W. Site-specific protein and peptide immobilization on a biosensor surface by pulsed native chemical ligation. *Chembiochem* **8,** 1790–4 (2007).

67. Govindaraju, T. *et al.* Surface immobilization of biomolecules by click sulfonamide reaction. *Chem. Commun. (Camb).* 3723–5 (2008). doi:10.1039/b806764c

68. Szychowski, J. *et al.* Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J. Am. Chem. Soc.* **132,** 18351–60 (2010).

69. Jewett, J. C. & Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **39,** 1272 (2010).

Chapter 6

Mitigating unwanted photophysical processes in single-molecule FRET experiments

6.1. Introduction to single-molecule Förster Resonance Energy Transfer

All the biochemical and kinetics results related to Dbr1p obtained in the previous chapters are based on bulk experiments. In bulk experiments, the value of any property measured (e.g. the FRET efficiency of a dual-labeled RNA) is the weighted average of the different components present in the system. Therefore, for a multicomponent system, the bulk measurements might produce an experimental value which does not properly represent the system because of the ensemble averaging.[1,2] Studies of single molecules, however, are very well suited for revealing the different components in a multicomponent system. Another advantage of single molecule measurements is that, in a dynamically complex system in which it is not possible to synchronize different components, single molecule measurements can be used to follow the detailed kinetics of individual molecules.[1] With advances in technology, it has been possible to observe fluorescence from single dye molecules in biologically relevant conditions and use them to study the structure and dynamics of biomolecules.[3,4] Of all the single-molecule fluorescence-based methods available for studies of biomolecules, the most widely used method is single molecule Förster Resonance Energy Transfer or smFRET.[5,6] The following section describes applications of smFRET.

FRET between a donor and acceptor dye is a very powerful spectroscopic technique to measure distances on the order of 20-70Å.[7,8] When a donor dye is excited, the excitation energy is transferred to an acceptor dye *via* a dipole-induced-dipole interaction. Due to the dipolar nature of the interaction between the donor and acceptor fluorophore, the efficiency of this energy transfer, E_{FRET}, is highly sensitive to the distance (R) between the donor and acceptor dye (Eq1) where R_0 is the Förster radius, or the distance at which 50% energy transfer occurs.

$$E_{FRET} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \qquad \text{Equation 1}$$

The parameter R_0 depends on many factors such as the quantum yield of the donor dye (φ_D), the overlap between the donor emission and acceptor absorption spectra (J), orientation factor (κ^2) between the donor and acceptor dipole and the refractive index of the medium (η). The orientation factor κ^2 can vary between 0 and 4 depending on the dipole alignment of the fluorophores. It is generally assumed to be 2/3 for free rotation of the fluorophores on a time scale much faster than the excited state lifetime. However it has been shown in some cases that this assumption is too general and the value of κ^2 deviates significantly from the expected value.[9–11] The value of R_0 can be tuned by choosing different donor acceptor pairs. Ideal FRET pair dyes should have large overlap between the donor fluorescence and the acceptor absorption and their emission spectra should be well separated to avoid cross talk between the donor and acceptor channels. Figure 6.1a shows the absorption and fluorescence spectra of both the donor (Alexa488) and acceptor (Alexa594) dyes used for our studies. These show favorable overlap between the donor fluorescence and the acceptor absorption. The donor and acceptor

emission maxima are also separated by ~100 nm which makes them a very good FRET pair.

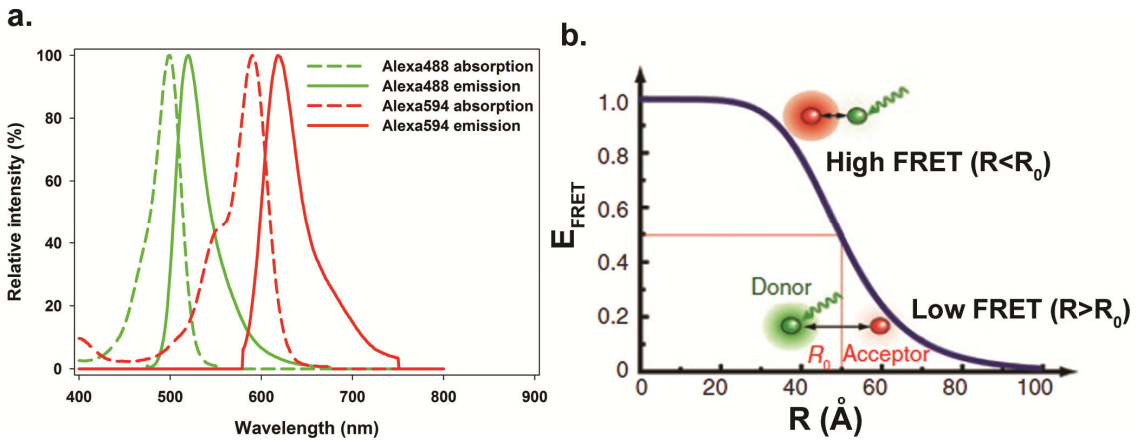


Figure 6.1. FRET and its distance dependence. **a.** The absorption (dotted) and fluorescence spectra (solid) of the donor (Alexa488) and acceptor (Alexa594) dyes used for our single molecule experiments. The large overlap between the donor fluorescence and acceptor absorption and also large (~100 nm) separation between the donor and acceptor fluorescence spectra suggests that these form a good FRET pair. **b.** The energy transfer efficiency, E_{FRET}, as a function of the inter dye distance R. R_0 indicates the distance at which $E_{FRET} = 0.5$. At distances near R_0 the change in E_{FRET} with R is linear. This figure is modified from ref. 6.

Energy transfer from donor dye to the acceptor dye results in decrease of the donor fluorescence yield and lifetime. The value of E_{FRET} can be calculated by quantifying these changes. Then distance between the donor and acceptor dye can then be calculated using form E_{FRET} using Eq.1. Figure 6.1b shows the changes in E_{FRET} with the distance (R) between the donor and acceptor dyes. When the donor and acceptor dyes are close ($R<R_0$), E_{FRET} is large as indicated by low donor fluorescence and high acceptor fluorescence. However when the donor and acceptor dyes are further apart ($R>R_0$), E_{FRET} decreases. Near R_0, the increase in E_{FRET} with R is approximately linear. Because of the strong distance dependence, FRET can be used as a spectroscopic ruler. For example, two different sites of a biomolecule can be labeled with the donor and acceptor dyes, respectively. Any conformational change resulting in a change in the distance between

the two sites will also result in change in the FRET values. Therefore, by monitoring the change in FRET, we can monitor the conformational change of that biomolecule. SmFRET measurements can also be used to observe the time evolution of different conformations of a dual-labeled molecule directly.

$$E_{FRET} = \frac{I_A}{I_A + \gamma I_D} = (1 - \frac{\tau_{DA}}{\tau_D}) \qquad \text{Equation 2}$$

The energy transfer efficiency, E_{FRET} of a donor acceptor pair could be expressed in terms of detected donor (I_D) and acceptor (I_A) fluorescence intensities using a ratiometric formula (Eq.2). The term γ is defined as $\gamma = \varphi_A n_A / \varphi_D n_D$ where φ and n are respectively the quantum yield and the detector efficiencies for the donor (D) and acceptor (A) dyes. Because energy transfer also causes lowering of the excited state lifetime of the donor, the energy transfer efficiency could also be calculated using the lifetime of the donor dye in the dual-labeled sample (τ_{DA}) and the donor-only (τ_D) sample (Eq.2). However the accurate determination of the lifetime of single molecule is very difficult and therefore this formula was mostly used for bulk measurements. The following section describes the setup used for smFRET experiments for this chapter.

6.1.1. Microscopy apparatus for smFRET

To measure E_{FRET} values of single biomolecules, the donor and acceptor intensity of individual dual-labeled molecules must be measured. The donor and acceptor intensities of fluorescently labeled biomolecules can be measured by two different configurations, either immobilized (on surface or in a matrix) or freely diffusing in solution.[2,12,13] With immobilized molecules, a particular molecule can be observed for an extended period of

time and the change in FRET over time can report the dynamics of the conformational change of biomolecules.[5,6,14] However, care must be taken to prevent perturbation of the structure of the biomolecule due to surface immobilization. Moreover, because the fluorophores are continuously excited in such an experiment, undesirable photophysical processes such as photobleaching and blinking must be minimized.[15] In solution-based smFRET, a freely diffusing single molecules traveling through a laser excitation volume generates fluorescence photon bursts.[12,16] Using appropriate filters, the donor and acceptor emission in these bursts can be measured and then E_{FRET} can be calculated. A histogram of the FRET efficiencies of a dual-labeled molecule can be constructed by performing the measurements on a large number of freely diffusing molecule over a period of time. Because these bursts typically last for short amount of time (~1 ms), any dynamical information slower than this timescale cannot be extracted using this method. However, invaluable information about the distributions of molecular properties of interest is obtained using this method. Moreover, the measured properties are unaffected by surface-induced effects and the impact of undesirable photophysical processes is reduced because each molecule stays in the laser focal volume for very short amount of time.

Confocal microscopy has been widely used for single molecule detection and characterization of fluorophores. In the confocal setup, a laser beam is focused to a diffraction-limited spot using a high numerical aperture (NA) objective (Fig. 6.2a and Fig. 6.11). The fluorescence emission is then collected using the same objective. A properly placed pinhole in the emission pathway reduces the out-of-focus fluorescence. The combination of a high NA objective and the pinhole creates a confocal detection

volume on the order of few femtoliters (See experimental section for detailed description). If a low enough concentration of a biomolecule is used (in the order of tens of femtomolar), there will be one or no molecule in the excitation volume at any given time. The probability of having two or more molecules in the excitation volume will be very low and therefore majority of the fluorescence bursts will be originating from single molecules (Fig. 6.2a). After passing through the pinhole, the donor and acceptor fluorescence from a dual-labeled molecule is then separated by using a dichroic mirror and focused into individual APD detectors (see experimental section and Fig. 6.11). After background subtraction, the value of E_{FRET} for each burst can be calculated using the donor and acceptor fluorescence intensity (Eq.2). Using fast single photon counting techniques and pulsed laser excitation, this confocal configuration is well used for single-molecule time-resolved fluorescence measurements.

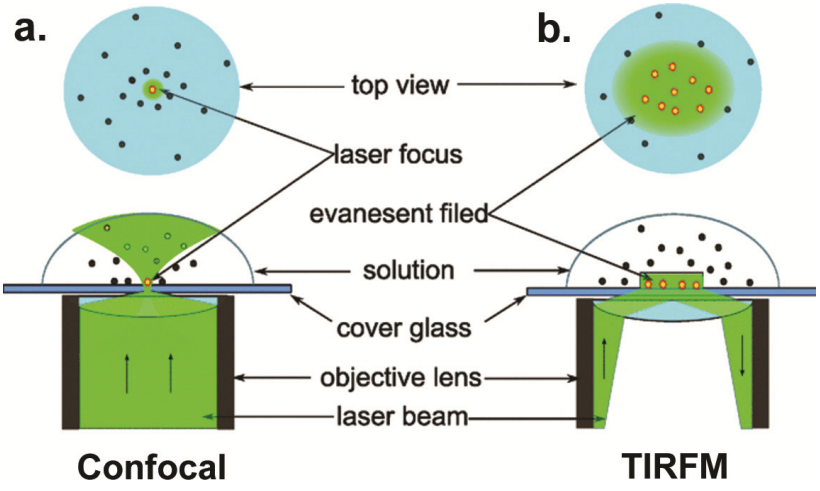


Figure 6.2. Schematic overview of basic characteristics of fluorescence confocal microscope and total internal reflection fluorescence microscope for smFRET measurements. **a.** Fluorescence confocal microscope with excitation beam waist ~250–300 nm (depends on the actual excitation wavelength) at the focal point. **b.** Total internal reflected fluorescence microscope (TIRFM) with wide excitation area, which is typically much larger than confocal spot in objective-type configuration. This figure is modified from ref. 13.

6.1.2. Unwanted photophysical processes in smFRET experiments

The use of confocal microscopy for smFRET measurements was first demonstrated in 1999 by Deniz *et al.* to determine conformation of dual-labeled DNA molecules.[16] Figure 6.3 shows the FRET histogram of three dual-labeled (tetramethylrhodamine, TMR and Cy5) DNAs with varying distance between the donor and acceptor dye. Because double stranded DNAs are conformationally restricted, they are expected to have only one population. However along with the expected FRET value, a second peak centered near zero FRET efficiency is also observed for all three DNAs. This zero FRET state is not predicted based on the known structures of the DNA strands. Therefore it is considered to be spurious or the result of dye photophysics rather than structural fluctuations.

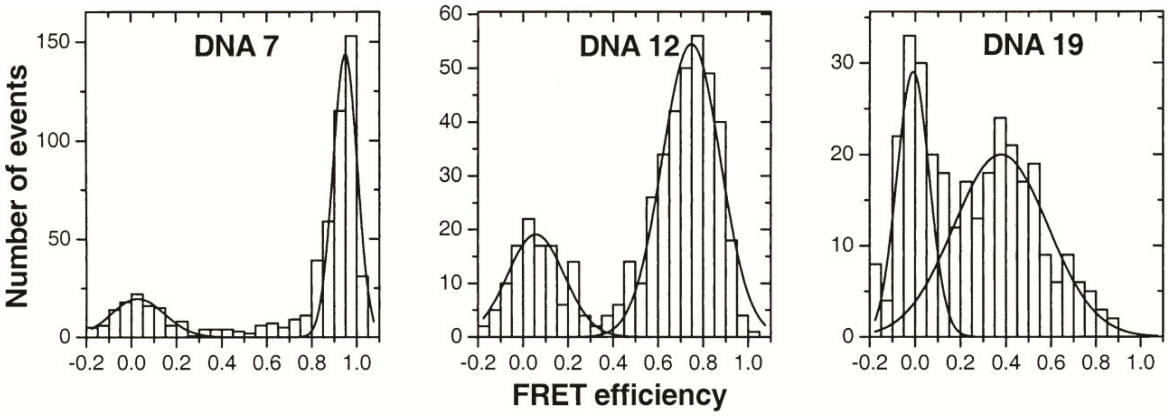


Figure 6.3. FRET histograms extracted from time traces for DNAs 7, 12, and 19 using a threshold of 20 photons. The number indicates the distance between the donor and acceptor dyes in terms of number of base pairs. Double gaussian fits extract numbers for the mean (width) of the higher efficiency peak of 0.95 (0.05), 0.75 (0.13), and 0.38 (0.21) respectively. This figure is modified from ref. 16.

The presence of a zero FRET state does not affect the FRET histogram of DNA strands that have high FRET values (DNA7 and 12) as the two histogram peaks does not obverlap. However, for DNA19, the high FRET and zero FRET population overlap considerably which complicates the analysis of the FRET distribution. The authors

attributed the zero FRET states to molecules with only donor dye resulting due to accelerated photobleaching of the acceptor or incomplete labeling of the DNA. They also suggested that a dark state of the acceptor dye will also result in this zero FRET population. Similar results were also obtained for other dual labeled DNA molecules with different dye pairs.[17,18]

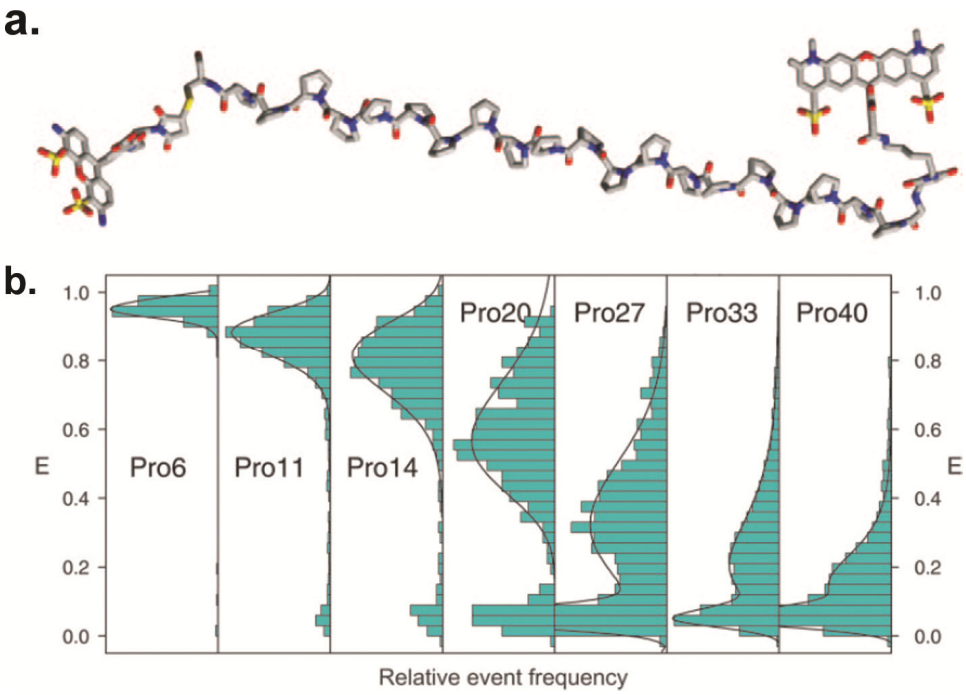


Figure 6.4. Single-molecule FRET efficiency measurements on polyproline peptides. **a.** Molecular model of a polyproline peptide used in this study. The acceptor and donor chromophore are linked to the chain by amino- and carboxyl-terminal glycine and cysteine residues, respectively. The conformation of the proline 20-mer is based on the crystal structure; the linkers and dyes were placed in arbitrary orientations. **b.** Transfer efficiency histograms obtained from confocal single-molecule measurements on polyproline peptides of various lengths. This figure is modified from ref. 21.

Spurious zero FRET states are not limited to DNA samples. Schuler *et al.* used dual fluorescently labeled (Alexa488 and Alexa594) polyproline molecules of various length to determine whether smFRET measurement could be used as a spectroscopic ruler (Fig. 6.4a).[19] In this paper, the authors reported the appearance of the zero FRET state in the smFRET experiments for the dual labeled polyproline molecules (Fig. 6.4).

Moreover the peak near zero efficiency was found to increase with increasing laser intensity. Based on these observations, the authors suggested that the acceptor dye undergoes a photophysical change into a dark state which does not accept excitation energy from the donor. The zero FRET peak has also been observed in smFRET experiments with dual-labeled proteins.[20–22]

To eliminate the spurious zero FRET population in smFRET experiments, several techniques were developed. For example, Achilles *et al.* developed an alternating laser excitation (ALEX) technique in which the donor and acceptor dyes are excited alternatively while in the confocal volume by rapidly switching between lasers that can excite the donor and acceptor, respectively.[23–25] The FRET histograms are then calculated from molecules that show both the donor and acceptor peaks in a burst confirming the presence of the fluorescent acceptor dye. Seidel *et al.* developed a muliparameter based smFRET measurement in which several other properties of single diffusing molecules, such as donor lifetime and anisotropy, are also calculated.[26–28] The zero FRET states can then be corrected for using the lifetime information as these states will have a donor lifetime similar to donor-only labeled sample. Although these techniques are very useful for smFRET studies, their implementation require added instrument cost.

Interestingly, the ratio of zero FRET to high FRET emitters is higher in molecular structures having a lower predicted FRET efficiencies (Fig. 6.4b). This was also observed for dual-labeled DNA molecules (Fig. 6.3). In a later report Grunwell *et al.* showed that the spurious zero FRET states could be partly eliminated by using an oxygen scavenger system. This suggested the involvement of molecular oxygen in the photophysical processes giving rise to the zero FRET state.[19] In an smFRET measurement using

immobilized molecules, Sabanayagam *et al.* showed that the blinking rate of the acceptor in dual-labeled DNA (TMR or Cy3 and Cy5) depends on the donor and acceptor dye separation.[29] When the donor and acceptor dyes are placed in close proximity (high FRET value), the blinking of the acceptor dye is much less frequent compared to that for a DNA having a large donor-acceptor separation. That a proximal donor dye can decrease the blinking observed in the acceptor dye emission suggests that a contact-induced reaction between the excited Cy3 and the dark Cy5 is important to the process. Later Rasnik *et al.* demonstrated that the long lasting dark states of the acceptor dye are actually non-radiative triplet states that could be easily depopulated by using a powerful triplet state quencher, Trolox.[30] It was also demonstrated that the antiblinking and antibleaching effects of Trolox are due to a combination of its reducing properties and a quinone form which is produced during its dissolution in aqueous buffers.[31] Trolox is now frequently used for immobilized smFRET measurements to prevent dye blinking and photobleaching.[15,30]

The dependence between the blinking characteristics of the immobilized molecules and appearance of the zero FRET state in freely diffusing molecules suggest a common mechanism for their appearance. As the proportion of the zero FRET population increases with decreasing predicted E_{FRET} values and increasing laser power, we postulated that the zero FRET state appears because the acceptor dye is going into a triplet state immediately after being excited by transfer of energy from the donor. We performed smFRET studies on several dual-labeled (Alexa488 and Alexa594) DNAs using the confocal setup and showed that the zero FRET population indeed appears because the acceptor dye goes into a dark triplet state. We also demonstrated that the zero

FRET state could be completely eliminated by using Trolox as a triplet state quencher during the smFRET experiments. Thereby we could perform smFRET experiments with freely diffusing molecules using the simple confocal setup.

6.2. Dual-labeled DNAs for smFRET studies using confocal microscopy

We synthesized several dual-labeled DNAs with varying donor-acceptor distances for smFRET studies using the confocal setup. In most of the previous studies by other groups, the dual-labeled DNAs were prepared by hybridizing a donor-labeled strand with another acceptor-labeled strand.[16,29,32] Thus there remains a possibility of having a sub-population of donor-only labeled strands if the hybridization is not 100% efficient. To avoid the possibility of having donor-only labeled strand due to inefficient hybridization, we conjugated both the donor (Alexa488) and acceptor Alexa594) dyes to a single stranded DNA (ssDNA). Following this the dual-labeled ssDNA was hybridized to an unlabeled complementary strand to generate a more rigid dual-labeled duplex DNA. These duplex DNAs were then used for the smFRET studies.

Figure 6.5a shows the structure of the different dual-labeled DNAs used for the smFRET studies. Both terminally (DNA NA) and internally (DNA NB and NC) labeled DNAs with varying donor acceptor distances were synthesized for the smFRET studies. For each of the dual-labeled DNAs constructs showed in Figure 6.5a, we prepared three dual-labeled DNAs with 7, 14 and 21 base pairs between the donor and acceptor dye. The predicted FRET efficiencies therefore ranged from 90% to 20%. To synthesize the dual-labeled ssDNA, the donor dye (Alexa488) was first conjugated to the DNA (see experimental section). The acceptor dye (Alexa594) was then conjugated the donor-only

labeled DNA to synthesize the dual-labeled ssDNA. Because of the highly hydrophobic nature of the acceptor dye, a large shift in the HPLC chromatogram was observed when the acceptor dye was conjugated to the donor-only labeled ssDNA (Fig. 6.5b). This ensured the absence of any donor-only labeled DNA in the purified dual-labeled sample due to insufficient purification.

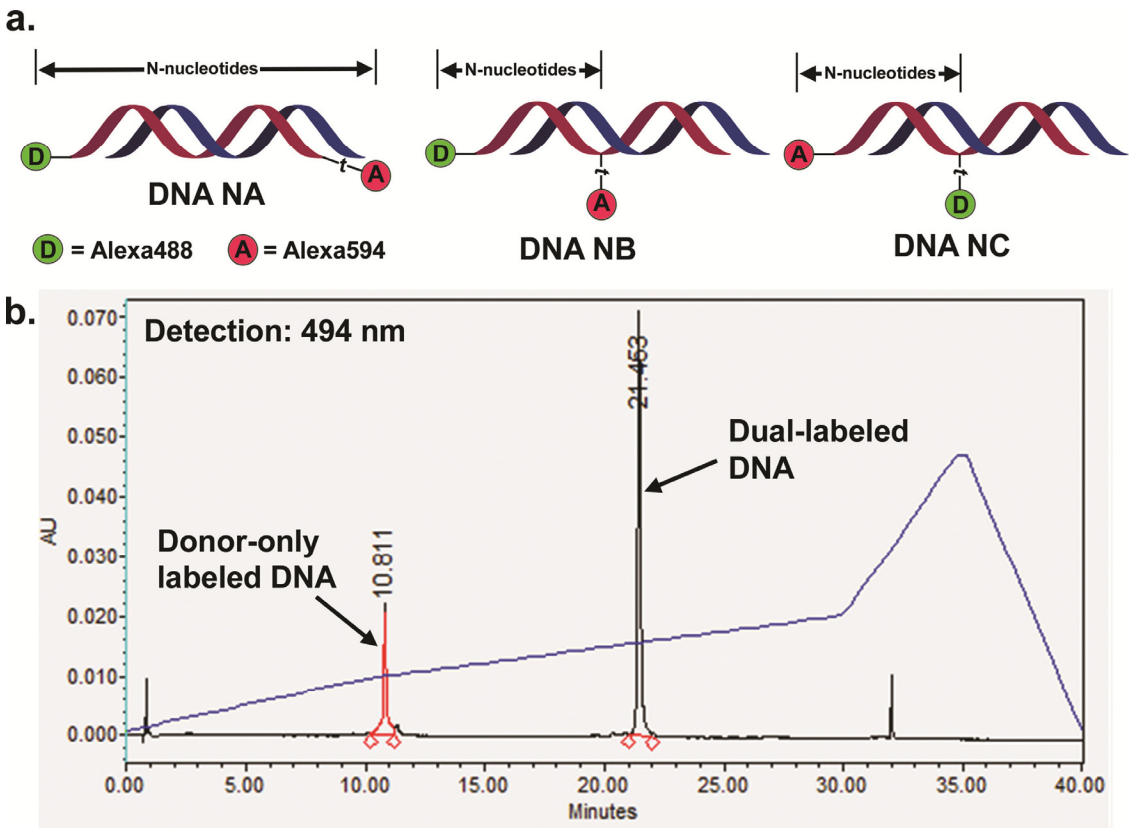


Figure 6.5. Dual-labeled DNAs for smFRET studies. **a.** The structure of the dual-labeled DNAs synthesized for smFRET studies using confocal setup. We have synthesized both terminally (DNA NA) and internally (DNA NB and NC) labeled duplex DNAs with varying donor acceptor distance. The number N indicates the distance between the donor and acceptor dyes in terms of DNA base pairs. The '*t*' between the DNA and the donor or acceptor dye indicates a trizole linkage due to the CuAAC reaction (See experimental) which was used for the conjugation reaction. **b.** HPLC chromatogram for conjugation of the acceptor dye to a donor-only labeled ssDNA which is a precursor to DNA 21A. Large separation between the donor-only and dual labeled peaks ensures high purity of the dual-labeled ssDNA without any donor-only impurity.

6.3. smFRET experiments with the dual-labeled DNA

In the confocal experiments, the donor dye (Alexa488) was excited using a pulsed 485nm diode laser (40MHz), and the donor and acceptor (Alexa594) emissions were recorded simultaneously using two different APD detectors (see Scheme 6.1 and experimental for details). The donor and acceptor channels were binned with 1 ms integration time. Fluorescent bursts containing at least 30 photons in the combined donor and acceptor channels were used for calculation of FRET efficiencies. Following background subtraction, the donor (I_D) and acceptor (I_A) intensities of each burst were used to calculate E_{FRET} using the ratiometric formula (Eq.2). The smFRET experiments were performed on dual-labeled duplex DNAs freely diffusing in buffer solution (50 mM Tris.HCl (pH 7.5), 50 mM NaCl and 0.005% TritonX-100). Figure 6.6a shows the fluorescence intensities of the donor and acceptor channels for 1s interval measured for dual-labeled DNA 7B. Because of the low concentration (50 pM) of the DNA used, most of the points consist of background signal, resulting predominantly from Raman and Rayleigh scattering. However the fluorescent bursts that are occasionally observed with signal much higher than the background corresponds to dye-labeled DNAs. As can be seen in Figure 6.6a, some of the fluorescent bursts have donor and acceptor intensities appearing at the same time. However some fluorescent bursts show only donor signal while the acceptor signal is close to the background level (indicated by *). Because of the absence of acceptor intensity, such bursts will result in E_{FRET} values close to zero. Figure 6.6b shows the E_{FRET} histogram of the dual-labeled DNA 7B. As expected, the FRET histogram showed two dominant peaks, one centered around zero efficiency and another at much higher efficiency. Careful synthesis of the dual-labeled ssDNA ensured that there

is no donor-only labeled DNA present in the sample (Fig. 6.5b). Therefore it could be clearly stated from these data that the peak near zero efficiency appears because of a dark state of the acceptor dye which can no longer accept excitation energy from the donor. Similar results were also obtained for all the other dual-labeled DNAs synthesized for this study.

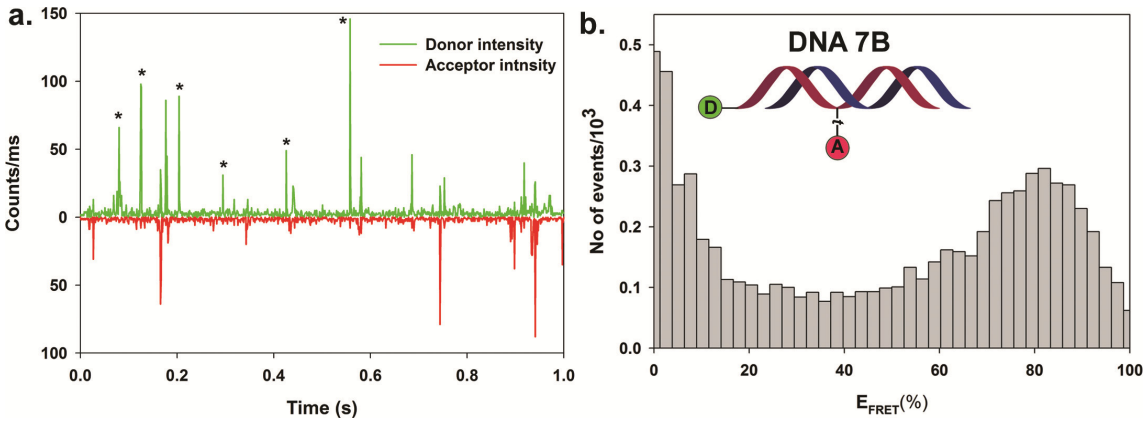


Figure 6.6. smFRET studies of the dual-labeled DNA 7B. **a.** Time traces of both the donor (green and acceptor (red) channels for dual-labeled DNA 7B freely diffusing in solution. The donor-only bursts with insignificant acceptor signals are marked by *. **b.** FRET efficiency (E_{FRET}) histogram of DNA 7B extracted from the bursts with a threshold of 30 photons.

To test our hypothesis that the zero FRET peak appears because of the acceptor dye goes to a dark triplet state, we measured the FRET efficiencies of the dual-labeled DNA at varying laser powers. With increasing laser power, the rate of excitation of the donor dye also increases, which, in turn, increases the number of excited acceptor molecules due to energy transfer. Therefore, the chances that an acceptor molecule will go to a dark triplet state also increases with increasing laser power. Previous experiments with dual-labeled DNA and peptides showed an increasing zero FRET population with increasing laser power.[16,21] We used three different laser powers (100 μW, 170 μW and

400 μW) for the smFRET measurements of the dual-labeled DNA. The optimal laser power for the smFRET experiments was determined to be 170 μW as the donor and acceptor signal no longer increases linearly with power above this value. Figure 6.7 shows the E_{FRET} histograms of the terminally dual-labeled DNA 14A for the three different laser powers. The relative enhancement of the zero FRET peak with increasing power suggests that a triplet state of the acceptor dye might be responsible for the zero FRET peak. We observed a similar power-dependent increase in the zero FRET population for all the dual-labeled DNAs synthesized. The appearance of the zero FRET population was not affected by labeling scheme (A, B and C) of the dual-labeled DNA (experimental, Fig. 6.12). The following section discusses the elimination of the zero FRET state using a triplet state quencher.

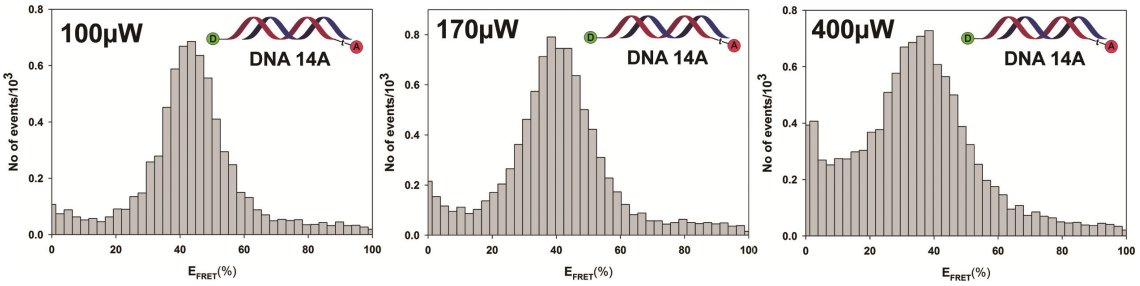


Figure 6.7. Representative FRET efficiency histograms of dual-labeled DNA 14A show power depended increase in the zero FRET population.

6.4. Elimination of the zero FRET population using a triplet state quencher

As the previously described experiments suggested that a dark triplet state of the acceptor dye might be responsible for the zero FRET population, we evaluated the effect of Trolox as a triplet state quencher on these smFRET experiments. We performed smFRET experiments with several dual-labeled DNAs in absence and presence of 1 mM Trolox.

Figure 6.8 shows the FRET histograms and time traces of terminally dual-labeled DNA in absence and presence of 1 mM Trolox using 400 µW power.

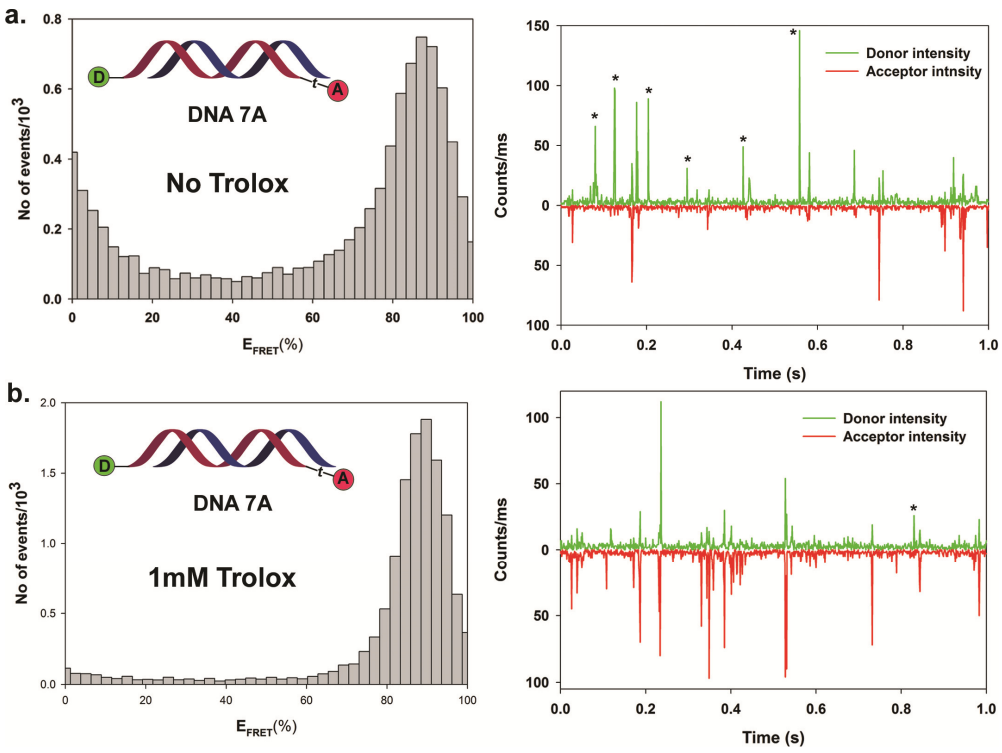


Figure 6.8. Representative FRET histogram and time traces of dual-labeled DNA 7A in presence and absence of 1 mM Trolox (400 µW laser power). **a.** The FRET histogram for DNA 7A in absence of any triplet state quencher shows a zero FRET population and the corresponding burst data show number of peaks with donor only fluorescence (indicated by *). **b.** Addition of 1 mM Trolox to the buffer solution almost completely eliminates the zero FRET population. The corresponding burst data also shows most of the bursts with both donor and acceptor intensity above the background level.

As it can be seen in Figure 6.8, the zero FRET population could be completely eliminated by addition Trolox to the buffer solution. Additionally, the burst data in presence of Trolox also shows almost no donor only peak (Fig. 6.8b). This indicates that the addition of the Trolox eliminates the dark state of the acceptor dye making it available for accepting energy form the donor dye. Therefore the chance of having a donor only peak is significantly reduced. The smFRET data shown in Figure 6.8 were obtained using a laser power of 400 µW which is the highest power used for these

experiments. As we have previously shown that the relative proportion of the zero FRET state increases with laser power (Fig. 6.7). The absence of a zero FRET peak even at 400 µW power means that Trolox is a very efficient triplet state quencher. We also tested another triplet state quencher, dithiothrietol (DTT) to see its effect on the zero FRET population. Similar to Trolox, 2.5 mM DTT also eliminated the zero FRET population which supports our hypothesis that slow de-population of the acceptor triplet state is responsible for the zero FRET population (Fig. 6.13). We observed the complete elimination of the zero FRET peak for all the different dual-labeled DNAs synthesized in presence of 1 mM Trolox.

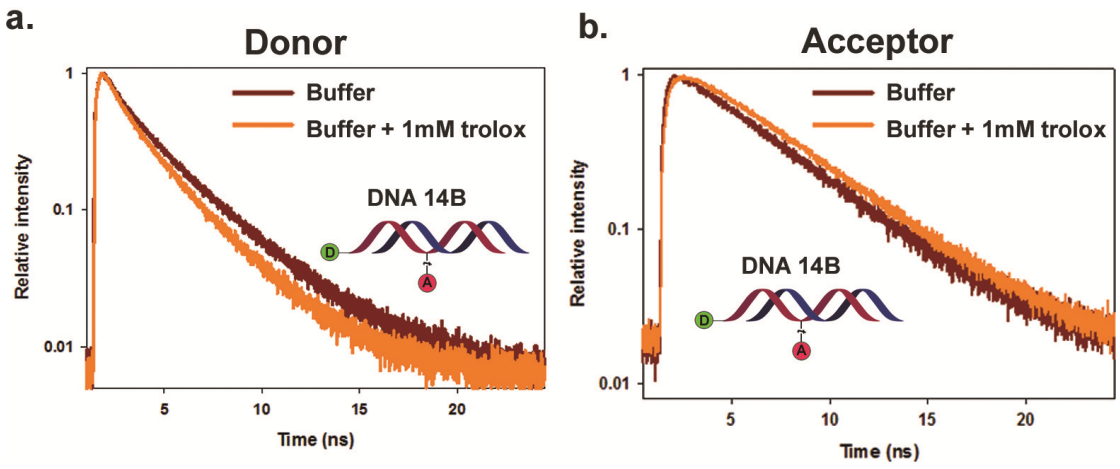


Figure 6.9. Bulk lifetime analysis of the dual-labeled DNA 14B in absence and presence of Trolox. **a.** Donor and **b.** acceptor fluorescence decay of the dual-labeled DNA in presence and absence of 1 mM Trolox. Buffer: 50 mM Tris.HCl (pH 7.5), 50 mM NaCl and 0.005% TritonX-100.

To test the effect of Trolox on the fluorescence lifetime of the donor and acceptor dyes, their fluorescence decay was measured in the bulk using time correlated single photo counting (TCSPC) using the identical setup. The only difference between the single molecule experiments and the bulk lifetime measurements is that a higher concentration (2 nM) of the dual-labeled DNA was used for the lifetime measurements.

As it can be seen in Figure 6.9a, the donor dye shows faster fluorescence decay in presence of the Trolox. The decrease in the donor fluorescence lifetime in presence of Trolox could be explained by increased energy transfer to the acceptor dye due to elimination of the dark triplet state. Similarly, the acceptor dye also shows a slight increase in the fluorescence lifetime in presence of Trolox indicating an increase in the number of bright fluorescent states.

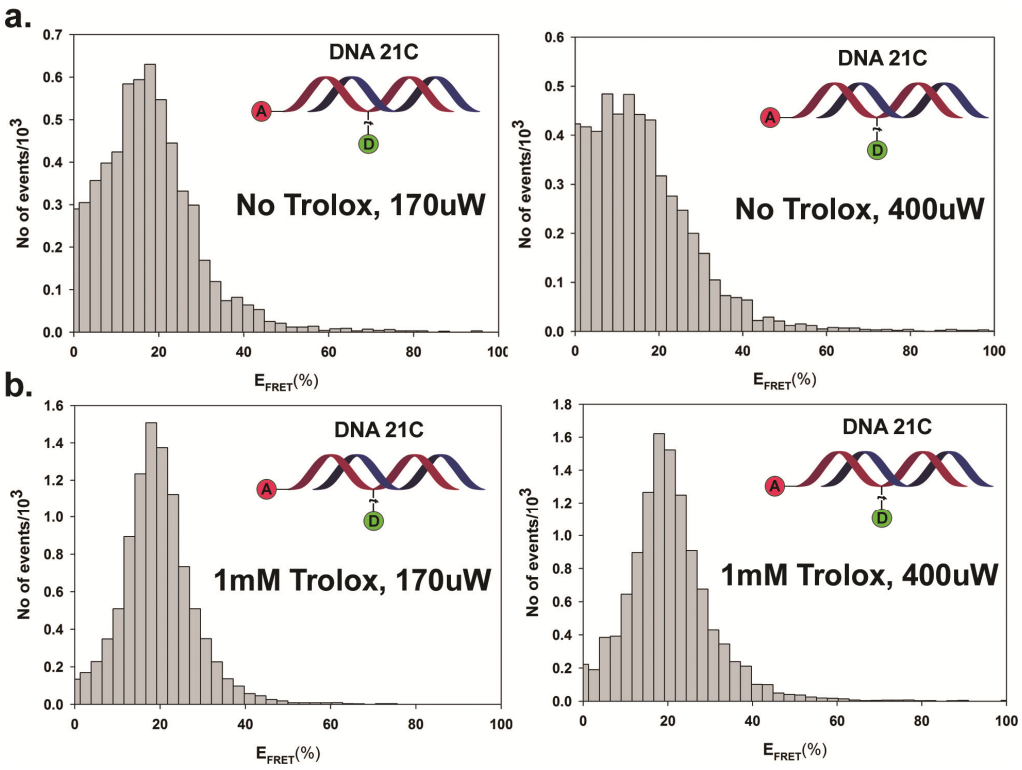


Figure 6.10. Representative FRET histogram and time traces of dual-labeled DNA 21C in presence and absence of Trolox at different laser power. **a.** FRET efficiency histogram of dual labeled DNA 21C without any added Trolox shows zero FRET state and obscures the histogram especially at higher power. **b.** The zero FRET state can be completely eliminated by using 1 mM Trolox even at high laser power.

As discussed before, for a molecule with closely positioned donor-acceptor pairs (high FRET) the zero FRET state does not distort the FRET histogram. However, for molecules with intrinsically low FRET values, overlap of the histogram with that arising

from zero FRET population can render the data more difficult to interpret. Figure 6.10a shows the FRET efficiency of donor labeled DNA 21C with an expected E_{FRET} of ~20%. In this case, the zero FRET population obscures the FRET histogram of DNA 21C especially at higher power. However, use of Trolox as a triplet state quencher successfully eliminates the zero FRET population (Fig. 6.10b) even at higher laser power. The FRET histogram of DNA 21C shows a clear single Gaussian peak in presence of Trolox at both 170 μW and 400 μW laser power. These data clearly show the effectiveness of Trolox as a triplet state quencher for smFRET experiments in solution phase. Therefore any change in biomolecular conformation which leads to a low FRET state could still be correctly identified as long as a triplet state quencher is used for the smFRET experiments.

6.5. Conclusion

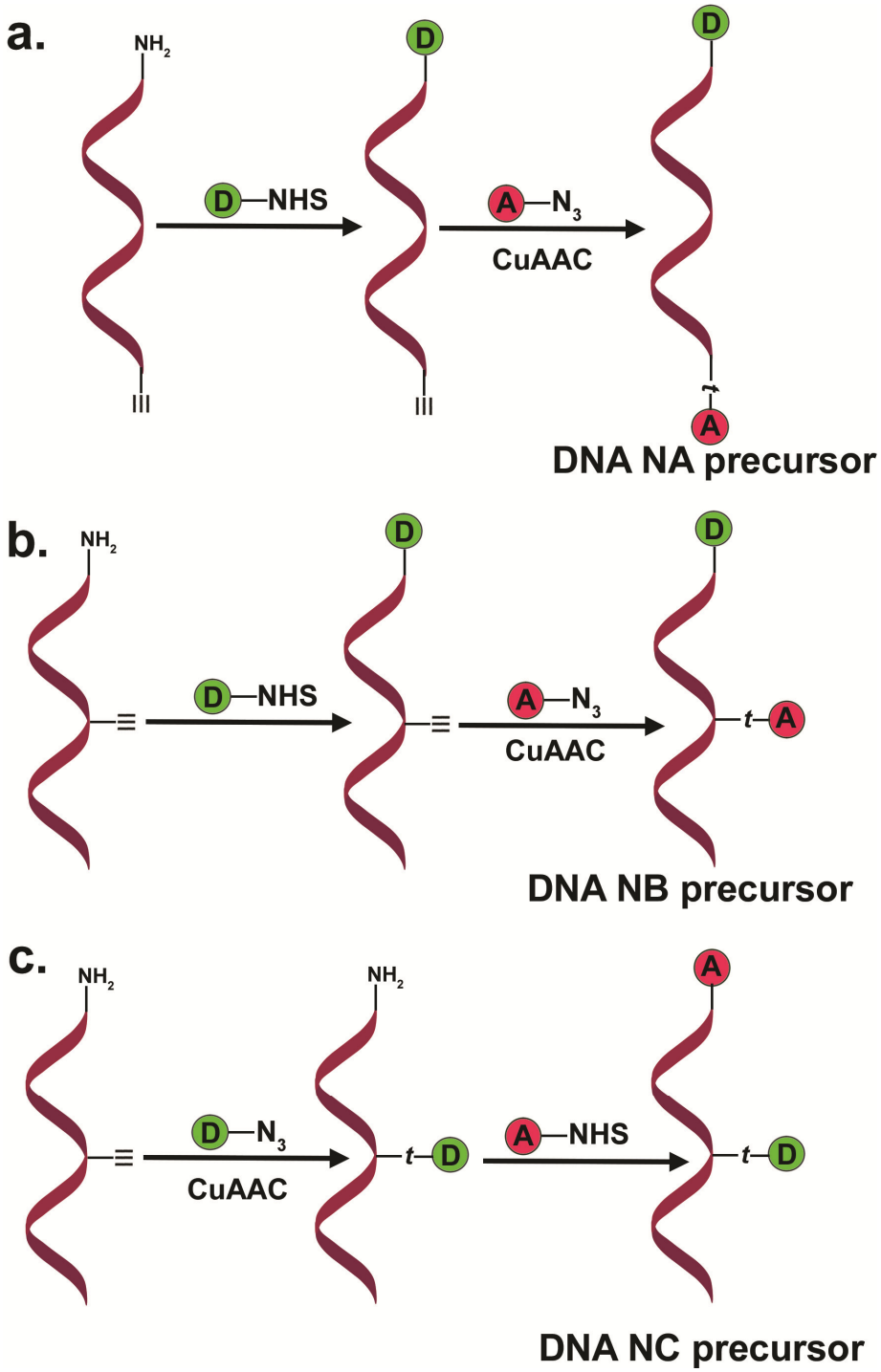
The use of confocal microscopy for smFRET measurement of freely diffusing molecules is a simple and powerful technique to detect subpopulations in a multicomponent system. However such measurements are complicated by some unwanted photophysical processed which results in a peak around zero FRET efficiency. The zero FRET peaks are especially problematic when the intrinsic FRET values of the biomolecules are low. Although previous studies have suggested different origins for this zero FRET state, it had not previously been analyzed in detail. In this chapter I showed that the zero FRET state appears because of a dark triplet state of the acceptor dye. The laser power dependent increase in the zero FRET peak suggested a mechanism involving the triplet state of the acceptor. When a triplet quencher such as Trolox or DTT is added to the buffer, the zero FRET state was completely eliminated, even when using high laser

power. Using a dual-labeled DNA with large donor-acceptor separation and low E_{FRET} elimination of the zero FRET state resulted in E_{FRET} values that formed a clear gaussian distribution. Therefore by simply adding Trolox to the buffer mixture we could eliminate the zero FRET peak without the need of more sophisticated instruments or analysis. In the next chapter I have used these findings for smFRET studies of conformations of dual-labeled bbRNAs.

6.6. Experimental

Synthesis of dual-labeled duplex DNAs for smFRET studies

The precursor DNA strands to the dual-labeled ssDNA contained a 5'-C6-NH_2 modification and either a terminal 3'-*O*-Propergyl modifications (for DNA NA) or an internal 2'-*O*-Propargyl modification (DNA NB and NC). These were synthesized using regular phosphoramidite chemistry and then deprotected following standard protocol. Scheme 6.1 shows the synthesis scheme to access the three different dual-labeled DNAs (pre-DNA NA, NB and NC) which are precursor to the three classes of dual-labeled duplex DNA. Following the synthesis and deprotection of the precursor DNAs, the donor and acceptor dyes were conjugated as indicated in Scheme 6.1. The conjugation of the donor (Alexa488) and acceptor (Alexa594) dyes were performed using the same method as described in chapter 2. Following each conjugation reaction the labeled DNA was purified by HPLC and then finally the dual-labeled ssDNAs were characterized by mass spectrometry. Table 6.1 shows the sequence and mass spectroscopic characterization of the dual-labeled ssDNAs which are precursor to the duplex-DNAs used for smFRET studies.



Scheme 6.1. Synthetic access to dual-labeled ssDNAs. **a.** Synthesis of dual-labeled ssDNA which are precursor to DNA NA. **b.** Synthesis of dual-labeled ssDNA which are precursor to DNA NB. **c.** Synthesis of dual-labeled ssDNA which are precursor to DNA NC. N indicates the separation between the donor and acceptor dyes in terms of number of base pairs. The symbol '*t*' indicates the triazole linkage formed between the azide and alkyne groups during the CuAAC reaction.

Name	Sequence	Mass (calculated)	Mass Found
Pre-DNA 7A	**5'-Alexa488-C6-TCC CAT A-3'-t-C6 Alexa594**	**3237.8**	**3234.2**
Pre-DNA 14A	**5'-Alexa488-C6-A GTT GGA TCC CAT A-3'-t-C6 Alexa594**	**5859.1**	**5857.3**
Pre-DNA 21A	**5'-Alexa488-C6-GCA CTG CAG TTG GAT CCC ATA-3'-t-C6 Alexa594**	**8001.5**	**8000.1**
Pre-DNA 7B	**5'-Alexa488-C6- AGT TGG A(-2'-t-C6 Alexa594)- TCC CAT A-3'**	**5859.1**	**5857.5**
Pre-DNA 14B	**5'-Alexa488-C6-GCA CTA CAG TTG G A(-2'-t-C6 Alexa594)-TCC CAT A-3'**	**7986.6**	**7982.4**
Pre-DNA 21B	**5'-Alexa488-C6- TAC AAT CGC ACT ACA GTT GG A(-2'-t-C6 Alexa594)- TCC CAT A-3'**	**10113.0**	**10111.8**
Pre-DNA 7C	**5'- Alexa594-C6- AGT TGG A(-2'-t-C6 Alexa488)- TCC CAT A-3'**	**5859.1**	**5850.4**
Pre-DNA 14C	**5'- Alexa594-C6-GCA CTA CAG TTG G A(-2'-t-C6 Alexa488)-TCC CAT A-3'**	**7986.6**	**7980.6**
Pre-DNA 21C	**5'- Alexa594-C6- TAC AAT CGC ACT ACA GTT GG A(-2'-t-C6 Alexa488)- TCC CAT A-3'**	**10113.0**	**10113.2**

Table 6.1. Sequence and MALDI mass data of the three different types of dual-labeled ssDNAs (pre DNA NA, NB and NC) which are precursor to the dual-labeled duplex DNA. The number N indicates the separation between the donor and acceptor dye in term of the base pairs.

Following purification of the dual-labeled ssDNAs, they were hybridized with the respective complementary strand to make duplex DNAs. The hybridization reaction mixture contained 100 μL of 1 μM dual-labeled DNA and 1.5 μM of the complementary strand in 1X hybridization buffer (50 mM Tris.HCl (pH 7.5), 50 mM NaCl and 0.005% TritonX-100). Because the complementary strand is not labeled, an excess could be used to ensure complete duplex formation. The hybridization was performed by heating the reaction tube at 90˚C for two minutes and then at 65˚C for ten minutes. Following this, the tube was slowly cooled to room temperature for 1 hr. The dual-labeled DNA was then

used for smFRET experiments. Freshly-prepared samples of the duplex DNAs were used whenever possible. However if necessary the sample was stored at 4˚C for few days.

Confocal microscope for smFRET experiments

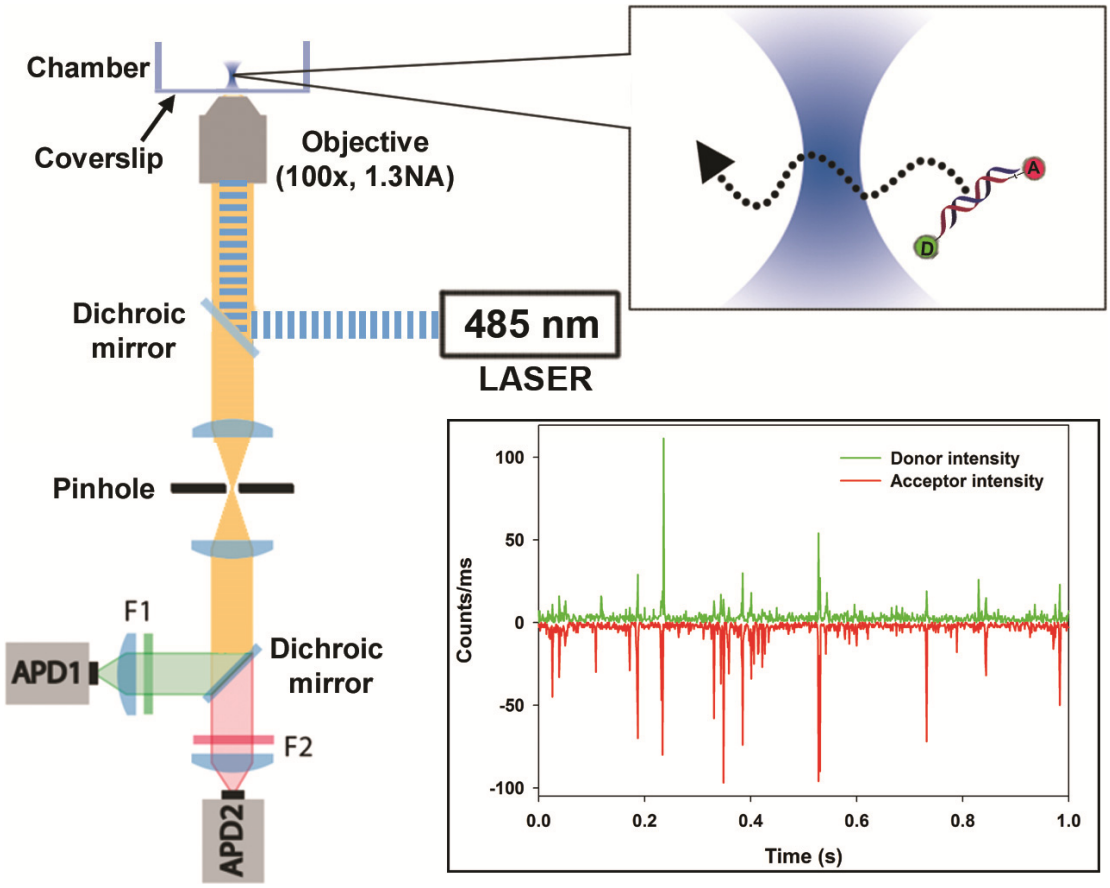


Figure 6.11. Schematics of the confocal microscope for smFRET measurements. The top inset shows the focal point of the excitation laser beam. When a dual-labeled molecule freely diffuses through the confocal volume it generates fluorescent bursts in both the donor and acceptor channels shown in the bottom inset. The out of focus light is eliminated using a pinhole.

The confocal microscope for the smFRET detection is built around an inverted microscope base (Olympus IX-71) using epi-illumination with the addition of a confocal pinhole, and two-color time-resolved detection (Fig. 6.11). A 485 nm pulsed diode laser (40MHz) (Picoquant, model P-C-485) is used as an excitation source for fluorescent

molecules. The laser beam is attenuated with neutral density filters to achieve the required power. The power of the laser beam is measured just before it entered through the back port of the microscope. The laser beam is directed to the objective (Olympus oil immersion, 100x, NA1.3) using a 500 nm dichroic mirror (Semrock). The fluorescence from the excited dual-labeled molecules is also collected using the same objective which is then directed towards the emission pathway through a 500 nm long pass filter (Semrock). The fluorescence emission collected by the objective is focused by the tube lens at a distance of 108 mm away from the left side port of the microscope.

To eliminate the out of focus light from the confocal excitation volume a 75 μm (Thorlabs) pinhole is placed at the same place where the emission light is focused. After passing the fluorescence emission through the pinhole the fluorescence image is focused at infinity using a 50 mm convex lens making the rays parallel. Following this the donor and acceptor emissions are separated using a 580 nm dichroic mirror (Semrock). This dichroic mirror reflects the donor (Alexa488) emission and transmits the acceptor (Alexa594) emission. Both the donor and acceptor emissions are then passed through bandpass filters to further eliminate the background noise. For the donor dye, a 531/46 nm bandpass filter (Semrock) is used and for the acceptor dye a 641/75 nm bandpass filter is used. Finally the both the donor and acceptor emissions are focused onto the active area of single photon avalanche diode (SPAD detectors, PDM series, Picoquant) using two separate micro lenses (NewPort). The single photon detectors are mounted onto xyz stage for easy movement during the alignment procedures. The signal generated in the SPAD detectors due to the fluorescent photons is first directed into a 4 channel router (PHR 800, Picoquant) and then the output is fed into a TCSPC board (PicoHarp

300, Picoquant). The smFRET data collected using this confocal microscope is analyzed using commercially available software (Symphotime 64, Picoquant) to generate lifetime data and FRET histogram.

For the lifetime measurement, the dual-labeled duplex DNAs were diluted to 2 nM concentration in 1X buffer (50 mM Tris.HCl (pH 7.5), 50 mM NaCl and 0.005% TritonX-100). Then 400 µL of the sample is placed in a home-build chamber consisting of a glass coverslip (22x22 mm, #1.5) at the bottom**.** The laser beam is then focused onto the solution and the bulk data is collected for 1 min. Following this, the solution of dual-labeled DNA in the chamber is diluted 40 times to reach a final concentration of 50 pM at which the smFRET measurements were performed. A typical smFRET measurement is performed for 10 mins which yielded 5000-10000 single bursts. The burst data is then analyzed using the Symphotime 64 software to calculate the FRET histograms.

To prepare the Trolox solution for smFRET experiments, 25 mg of Trolox is added to 1mL of 10X buffer and 8mL of water. The solution was vortexed vigorously for 10 mins to dissolve the Trolox. The pH of the solution was then adjusted to 7.5 using NaOH. The addition of NaOH also ensured the dissolution of the remaining solid Trolox. The final volume of the solution is then adjusted to 10 mL and finally the solution is filtered through a 0.2 µm filter. This yields 10 mL of 10 mM Trolox solution which was aliquoted into several tubes and stored at -20˚C. The solution of the Trolox was thawed before the smFRET experiments and it was mixed with the 1x buffer to a final concentration of 1 mM for the diffusing single molecule experiments.

smFRET histograms for different dual-labeled DNAs under different conditions

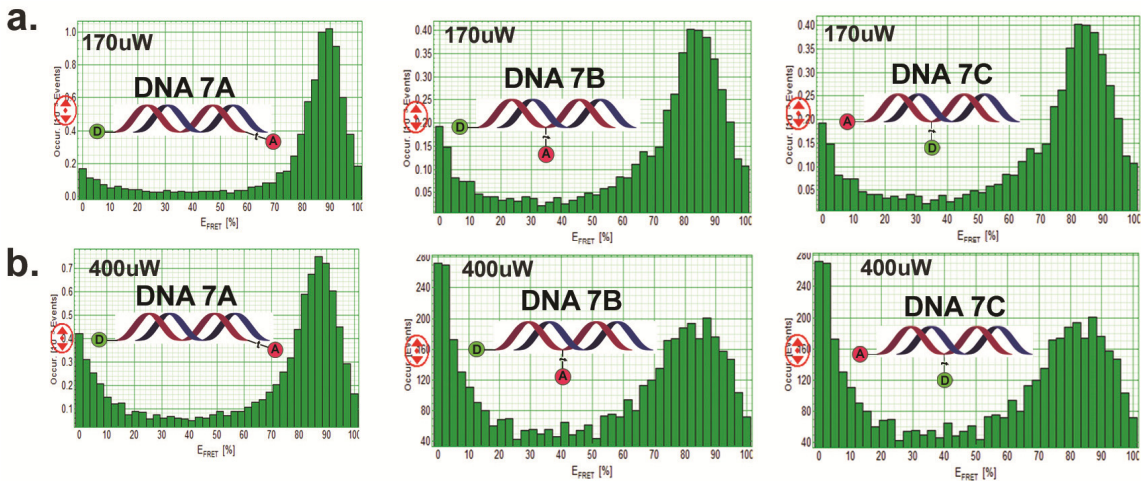


Figure 6.12. The smFRET histograms for three different types of dual-labeled DNAs with similar donor acceptor distance in absence of Trolox. **a.** The appearance of the zero FRET peak does not depend on the labeling scheme of the dual-labeled DNAs. However the terminally labeled DNA 7A shows slightly lower amount of the zero FRET state compared to the internally labeled DNAs (DNA 7B and 7C). **b.** The smFRET histograms at higher laser power shows more of the zero FRET peak as expected for a triplet dark state of the acceptor dye.

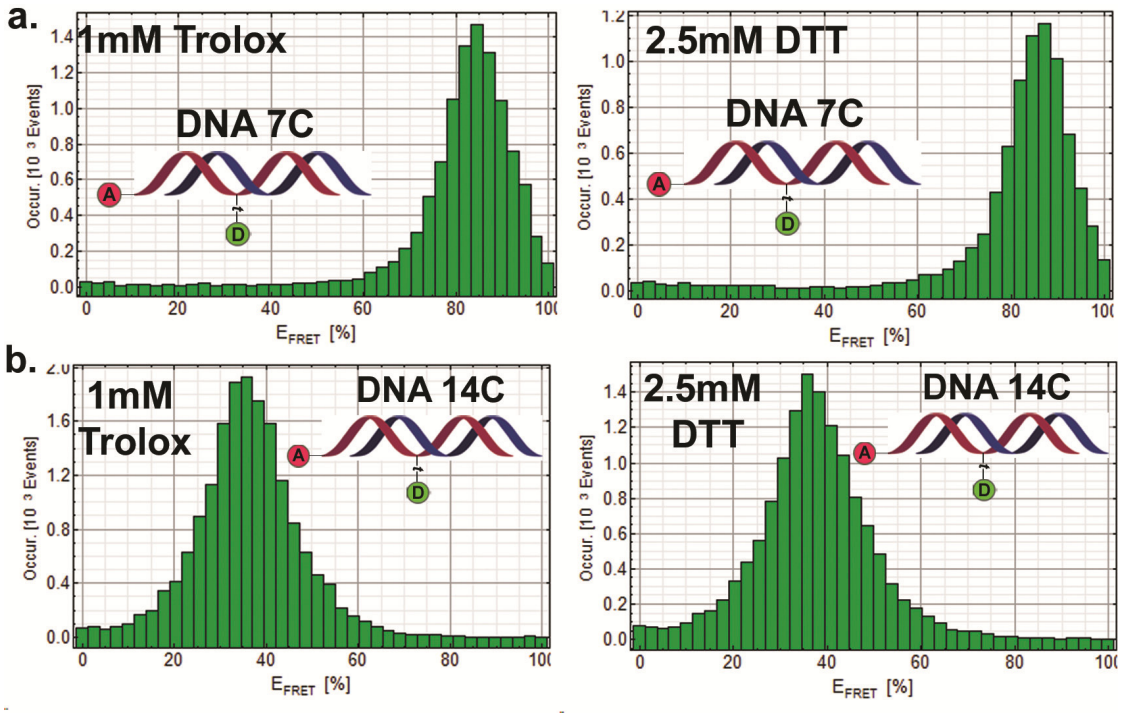


Figure 6.13. Use of dithiothreitol (DTT) as a triplet state quencher for smFRET experiments. **a.** The zero FRET population for DNA 7C could be completely eliminated using either 1 mM Trolox (right) or 2.5 mM DTT (left). **b.** The zero FRET population for DNA 14C could be completely eliminated using either 1 mM Trolox (right) or 2.5 mM DTT. 170 μW of laser power was used for these smFRET experiments.

References

1. Ha, T. Single-molecule fluorescence resonance energy transfer. *Methods* **25,** 78–86 (2001).

2. Moerner, W. E. & Fromm, D. P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Rev. Sci. Instrum.* **74,** 3597–3619 (2003).

3. Weiss, S. Fluorescence Spectroscopy of Single Biomolecules. *Science* **283,** 1676–1683 (1999).

4. Deniz, A. a, Mukhopadhyay, S. & Lemke, E. a. Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface* **5,** 15–45 (2008).

5. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. & Ha, T. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **77,** 51–76 (2008).

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

7. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 719–726 (1967).

8. Selvin, P. R. The renaissance of fluorescence resonance energy transfer. *Nat. Struct. Biol.* **7,** 730–734 (2000).

9. VanBeek, D. B., Zwier, M. C., Shorb, J. M. & Krueger, B. P. Fretting about FRET: correlation between kappa and R. *Biophys. J.* **92,** 4168–78 (2007).

10. Iqbal, A. *et al.* Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11176–81 (2008).

11. Ouellet, J., Schorr, S., Iqbal, A., Wilson, T. J. & Lilley, D. M. J. Orientation of cyanine fluorophores terminally attached to DNA via long, flexible tethers. *Biophys. J.* **101,** 1148–1154 (2011).

12. Deniz, A. A. *et al.* Ratiometric Single-Molecule Studies of Freely Diffusing Biomolecules. *Annu. Rev. Phys. Chem.* **53,** 233–53 (2001).

13. Zheng, D., Kaldaras, L. & Lu, H. P. Total internal reflection fluorescence microscopy imaging-guided confocal single-molecule fluorescence spectroscopy. *Rev. Sci. Instrum.* **83,** (2012).

14. Ha, T. *et al.* Probing the interaction between two single molecule s: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 6264–6268 (1996).

15. Ha, T. & Tinnefeld, P. Photophysics of Fluorescent Probes for Single-Molecule Biophysics and Super-Resolution Imaging. *Annu. Rev. Phys. Chem.* **63,** 595–617 (2012).

16. Deniz, A. A. *et al.* Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Förster distance dependence and subpopulations. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 3670–5 (1999).

17. Grunwell, J. R. *et al.* Monitoring the conformational fluctuations of DNA hairpins using single-pair fluorescence resonance energy transfer. *J. Am. Chem. Soc.* **123,** 4295–4303 (2001).

18. Dietrich, A., Buschmann, V., Meller, C. & Sauer, M. Fluorescence resonance energy transfer (FRET) and competing processes in donor-acceptor substituted DNA strands: A comparative study of ensemble and single-molecule data. *Rev. Mol. Biotechnol.* **82,** 211–231 (2002).

19. Schuler, B., Lipman, E. a, Steinbach, P. J., Kumke, M. & Eaton, W. a. Polyproline and the 'spectroscopic ruler' revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2754–2759 (2005).

20. Merchant, K. a, Best, R. B., Louis, J. M., Gopich, I. V & Eaton, W. a. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 1528–1533 (2007).

21. Mukhopadhyay, S., Krishnan, R., Lemke, E. A., Lindquist, S. & Deniz, A. A. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 2649–54 (2007).

22. Ferreon, A. C. M., Gambin, Y., Lemke, E. a & Deniz, A. a. Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5645–5650 (2009).

23. Kapanidis, A. N. *et al.* Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 8936–41 (2004).

24. Lee, N. K. *et al.* Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.* **88,** 2939–2953 (2005).

25. Kapanidis, A. N. *et al.* Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38,** 523–533 (2005).

26. Widengren, J. *et al.* Single-molecule detection and identification of multiple species by multiparameter fluorescence detection 1. *Anal.Chem.* **78,** 2039–2050 (2006).

27. Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P. J. & Seidel, C. A. M. *Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection*. *Methods in Enzymology* **475,** (Elsevier Inc., 2010).

28. Sindbert, S. *et al.* Accurate distance determination of nucleic acids via Förster resonance energy transfer: Implications of dye Linker length and rigidity. *J. Am. Chem. Soc.* **133,** 2463–2480 (2011).

29. Sabanayagam, C. R., Eid, J. S. & Meller, A. Long time scale blinking kinetics of cyanine fluorophores conjugated to DNA and its effect on Förster resonance energy transfer. *J. Chem. Phys.* **123,** 1–7 (2005).

30. Rasnik, I., McKinney, S. a & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods* **3,** 891–893 (2006).

31. Cordes, T., Vogelsang, J. & Tinnefeld, P. On the mechanism of Trolox as antiblinking and antibleaching reagent. *J. Am. Chem. Soc.* **131,** 5018–5019 (2009).

32. Di Fiori, N. & Meller, A. The Effect of dye-dye interactions on the spatial resolution of single-molecule FRET measurements in nucleic acids. *Biophys. J.* **98,** 2265–2272 (2010).

Chapter 7

Single molecule FRET studies of conformation and dynamics of backbone branched RNAs

7.1. Introduction

In the previous chapters of this thesis, I have discussed the biochemical and kinetics studies of Dbr1p with various bbRNA substrates. Such studies were possible because of our synthetic access to bbRNA substrates. However, these studies do not give any information about the structure and dynamics of the bbRNA in solution – either in the presence or absence of Dbr1p. To understand how Dbr1p selectively cleaves bbRNA substrates, it is important to understand the conformations of the bbRNA and how Dbr1p may bind it and if any conformational change of the substrate is induced upon binding. The crystal structure of EhDbr1p was recently published with a bbRNA substrate present in the active site.[1] In that structure, the bbRNA substrate mimic positions itself such that the 2'-5' phosphodiester bond is away from the active site. Therefore the binding and conformation of the bbRNA substrate with Dbr1p remains poorly understood. Although various bbRNAs have been synthesized for almost thirty years, there have been few detailed studies of their conformation in solution. Some early NMR-based studies on branched tri- and tetra-nucleotides suggested the existence of base stacking between the nucleobases of the branchpoint residue and the first nucleotide in the 2'-branch.[2–4] However, no other information is available about the conformation of bbRNAs in solution. As discussed in chapter 6, single molecule experiments are very well suited to identifying conformations of biomolecules in a multicomponent system and monitoring

the dynamics of conformational changes. In this chapter we have used single molecule Förster Resonance Energy Transfer (smFRET) to investigate the conformation and dynamics of bbRNA substrates and substrate mimics.

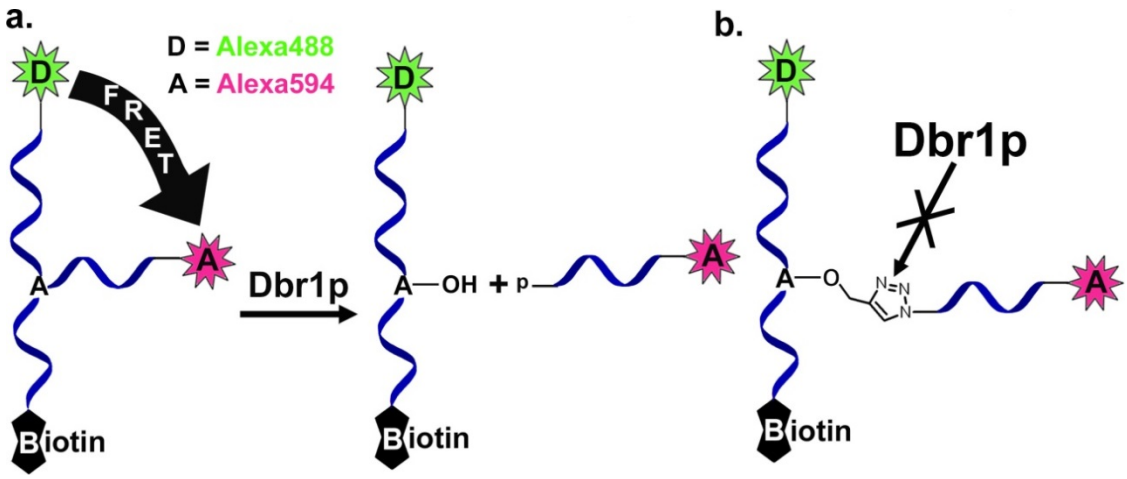


Figure 7.1. Dual-labeled branched RNAs substrates for smFRET studies. **a.** Upon binding**,** the native bbRNA is cleaved by Dbr1p and therefore cannot be used for binding studies. **b.** The 2'-5' triazole linked 'click' branched RNAs are resistant to Dbr1p cleavage and therefore can be used for single molecule binding studies. Both the native bbRNA and cbRNAs are labeled with Alexa488 as the donor and Alexa594 as the acceptor dyes.

To study the binding of bbRNA substrates with Dbr1p, a native substrate cannot be used as it will be readily cleaved by Dbr1p (Fig. 7.1a). We have previously shown that a 2'-5' triazole linked 'click' branched RNA (cbRNAs) is a competitive inhibitor of Dbr1p (chapter 4). This suggests that the cbRNA binds Dbr1p in a similar fashion as the native substrate and is not cleaved by Dbr1p. Therefore it is a good candidate to evaluate the conformational changes of bbRNA upon binding with Dbr1p (Fig. 7.1a). In this chapter, our investigations into the conformation and dynamics of dual-labeled (Alexa488-Alexa594) cbRNAs using smFRET under various physiological conditions are

described. The conformation and dynamics of native bbRNA substrates were investigated as well. Confocal-microscopy-based smFRET experiments were used to investigate the conformation of the cbRNAs and the native bbRNAs while TIRFM was used to investigate their dynamics. These studies will help us to understand the conformation of Dbr1p substrates in solution.

7.2. Bulk fluorescence studies of dual-labeled cbRNAs

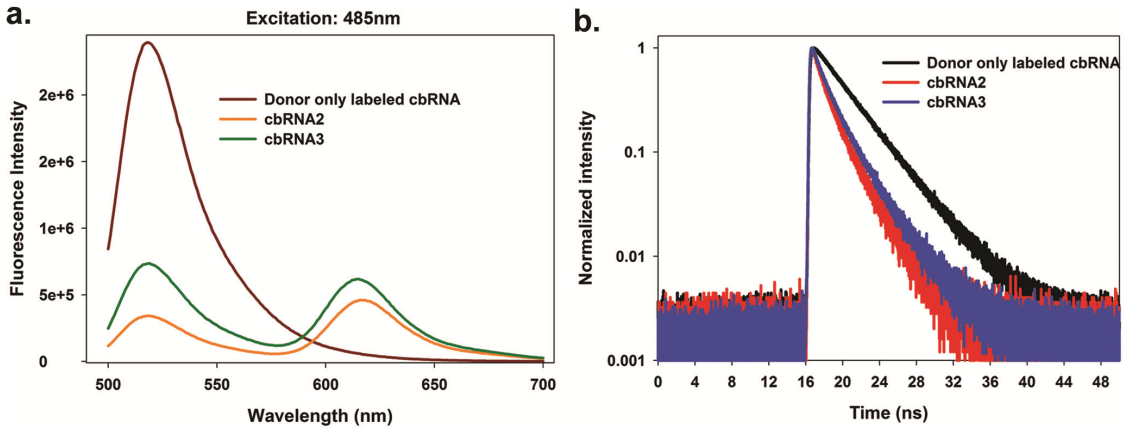


Figure 7.2. Ensemble fluorescence characterization of the dual-labeled cbRNAs. **a.** Fluorescence spectra of the dual-labeled cbRNAs along with donor-only control. **b.** Fluorescence decay of the dual-labeled cbRNAs along with donor-only control.

Before the smFRET studies, steady-state and time-resolved fluorescence measurements were performed with the dual-labeled cbRNAs to assess their ensemble behavior. Two different cbRNAs with 6mer (cbRNA2) and 12mer (cbRNA3) branches were used for the bulk measurements in addition to a donor-only labeled cbRNA (Table 7.3). The synthesis and characterization of these cbRNAs are described in chapter 2. The fluorescence spectra of the dual-labeled cbRNAs were measured in the 1x debranching reaction buffer at 100 nM concentration. Both dual-labeled cbRNAs show decreased donor emission and increased acceptor emission compared to the donor-only labeled

RNA (Fig. 7.2a). This indicates energy transfer from the donor dye to the acceptor dye. As expected, cbRNA2 with 6mer branch shows higher drop in donor intensity compared to cbRNA3 with 12 mer branch indicating higher energy transfer for cbRNA2. Surprisingly higher acceptor emission is also observed cbRNA3 compared to cbRNA2. This is unusual as higher energy transfer for cbRNA2 should result in higher acceptor emission as well.

Following the steady-state fluorescence measurements of the dual-labeled cbRNAs, the fluorescence decay of the dual-labeled sample was also measured using time correlated single photon counting (TCSPC). Figure 7.2b shows the fluorescence spectra of the dual-labeled cbRNAs along with the donor-only labeled cbRNA. The cbRNAs showed faster fluorescence decay compared to the donor-only sample, also indicating energy transfer. As expected, the 6mer cbRNA2 showed slightly faster decay than cbRNA3. The lifetime data could be fit using the tailfit method to recover the lifetime of the donor dye in the different cbRNAs (Fig. 7.3). As can be seen in Figure 7.3a, the data for the donor-only labeled cbRNA can be fit well using a single exponential with lifetime of 3.7 ns. However, the fluorescence decays of the dual-labeled cbRNAs cannot be fit with a single exponential; a two exponential decay provides a superior fit. Figure 7.3b shows the two-exponential fit of the fluorescence decay for the cbRNA2. The two different lifetimes obtained from the bi-exponential fit are 2.9 ns and 1.3 ns. The 2.9 ns lifetime corresponds to ~33% of the total population and represents a low-FRET state. This might be a result of large separation between the donor and acceptor dyes or it could be due to some photo-physical process which affects the energy transfer between the donor and acceptor. The lifetimes of 1.2 ns correspond to almost 67% of the total

population and represents a high-FRET state (E~66%) which is close to the expected FRET efficiency of cbRNA2 (54%) assuming a 90˚ angle between the stem and branch**.** Bi-exponential fluorescence decay curve was observed for cbRNA3 as well. Similar multicomponent fluorescence decays have also been reported for other dyes pairs.[5,6]

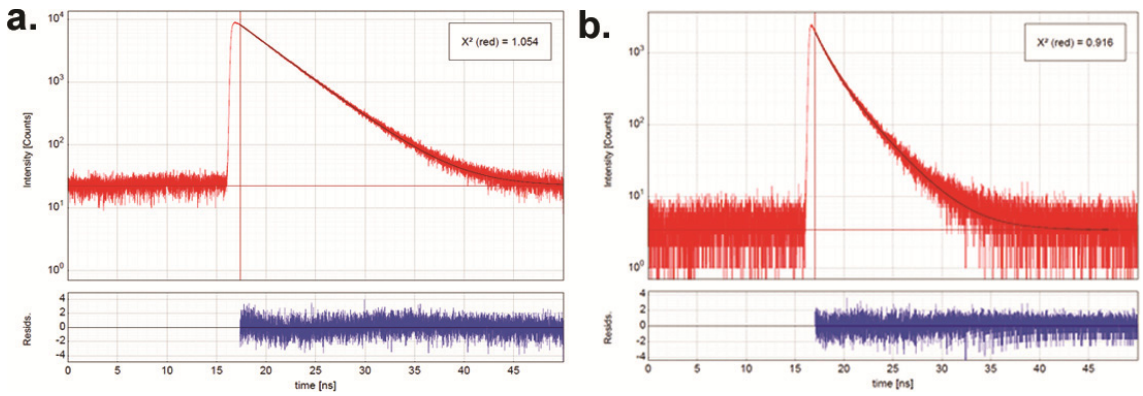


Figure 7.3. Ensemble fluorescence decay curves and their exponential fitting for cbRNAs **a.** the donor only cbRNA and **b.** cbRNA2.

Using the bulk fluorescence characterization of the dual-labeled cbRNAs, their FRET efficiencies were calculated. There are several ways by which the efficiency of the energy transfer can be measured for the dual-labeled samples. Firstly the drop in donor intensity in the dual labeled samples compared to the donor only labeled samples is a measure of FRET (Method 1). In another ratiometric method, FRET can be measured using the ratio of the acceptor emission and the sum of donor and acceptor emission in the dual-labeled sample (Method 2). In the second method, no reference from the donor-only labeled sample is required and therefore it should be independent of any error due to concentration measurements. Method 2 is used for single-molecule measurements because the donor-only-labeled single molecules will exhibit fluctuating intensities. The FRET efficiencies of the dual-labeled samples can also be measured by comparing the

lifetime of the donor dye in the dual-labeled samples to that of the donor-only sample (Method3). Method 3 is also independent of sample concentration. Table 7.1 shows the FRET efficiency of the various dual-labeled cbRNAs measured using the three methods described here. The expected values of the FRET efficiencies were calculated based on R_0 value of 56.7 Å and a 90° angle between the stem and the branch.

RNA	E_{FRET} (expected)*	E_{FRET} (Method 1)	E_{FRET} (Method 2)	Lifetime of donor (ns)	E_{FRET} (Method 3)	Angle (α)
cbRNA2	**54**	**86±3**	**57±3**	**1.3±0.1**	**67±2**	**79°**
cbRNA3	**28**	**69±2**	**45±2**	**1.6±0.2**	**59±10**	**69°**

Table 7.1. FRET efficiency of the various dual labeled cbRNAs using three different methods. Method1 compares the donor fluorescence in the donor only labeled sample vs the dual labeled samples. E_{FRET} **(Method1) = 100*(1-** $\frac{I_{DA}}{I_D}$**)** where I_{DA} is the intensity of the donor in the dual-labeled sample and I_D is the intensity of the donor in donor only labeled sample. Method 2 is a ratiometric method that is also used for single molecule measurements. E_{FRET} **(Method2)** = $100*(\frac{I_A}{I_D+I_A})$ where I_D and I_A are the intensities of the donor and acceptor respectively in the dual-labeled sample. The FRET efficiencies were calculated using the formula E_{FRET} **(Method3) =100 * (1-** $\frac{\tau_{DA}}{\tau_D}$**)** where τ_{DA} is the lifetime of the donor dye in the dual-labeled sample and τ_D is the lifetime of the donor dye in donor only labeled sample. τ_D is measured to be 3.7±0.1 ns. * The expected values of the FRET efficiencies were calculated based on R_0 value of 56.7 Å and a 90° angle between the stem and the branch. The stem and the branch lengths were calculated based on the duplex RNAs.

The three methods resulted in significantly different FRET efficiencies (Table7.1). However the general trend was same in all three cases which was that cbRNA2 has a higher FRET efficiency than does cbRNA3. The highest FRET efficiencies were produced by Method 1. Method 2 produced FRET efficiencies that are closest to the expected value. The expected values are based on a 90° angle between the

stem and the branch. Therefore any deviation from that structure will result in a different FRET value. Finally the FRET efficiencies calculated using Method 3 also differed from the expected values. For calculation of the FRET efficiencies, only the faster decay component was used.

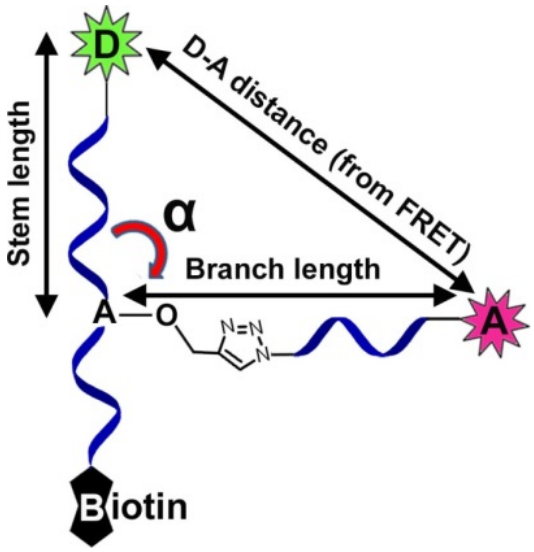


Figure 7.4. Calculation of the angle (α) between the stem and the branch for dual-labeled cbRNAs. Stem length and the branch lengths are fixed for a given cbRNA. The D-A distance could be calculated from the value of the FRET efficiencies. Therefore knowing the three sides of a triangle should give us the value of the angle α.

Knowing the FRET efficiencies of the cbRNAs, it is possible to calculate the distance between the donor and acceptor dye. Once the distance between the two dyes is known, the angle between the stem and the branch can be calculated given that the length of the stem and the branch are known (Fig. 7.4). Calculation of the angle α gives us a rough idea of the conformation of the cbRNAs in solution (Table 7.1). The value of the angle α is shown in Table 7.1 for the two different cbRNAs. The FRET efficiencies

obtained from method 3 was used for the calculation of the angle α. Both the cbRNAs give a value for α that is lower than 90 degrees as calculated using all three methods.

7.3. smFRET studies of dual-labeled cbRNA conformation(s)

The ensemble FRET efficiencies of the cbRNAs determined using the three different methods described yielded somewhat different values making it difficult to obtain unambiguous information regarding the structure of cbRNAs. Moreover, the emission lifetimes measured for the cbRNAs show bi-exponential decays. The two components of the fluorescence decay could arise from two different pairs populations of the cbRNA or it could be a result of unwanted photophysical processes as well. To avoid these ambiguities arising from the ensemble measurements, we initiated single molecule measurements to study the conformation of cbRNAs. The FRET efficiencies of the dual-labeled cbRNAs were determined using the same confocal microscope described in chapter 6 and following the same protocol. The smFRET measurements were performed for cbRNA2 and cbRNA3 in 1x debranching reaction buffer without the Mn^{2+} ion (see experimental). The Mn^{2+} ions were avoided for the initial experiments to probe conformations of cbRNAs in the absence of metal ion. Similar to the dual-labeled DNAs described in the previous chapter, a significant zero FRET population is observed for the cbRNAs (Fig. 7.5a). However addition of 1 mM Trolox to the buffer almost completely eliminated the zero FRET population (Fig. 7.5b). The cbRNA3 showed similar reduction of the zero FRET population in the presence of Trolox. For all subsequent smFRET experiments 1 mM of Trolox was added to the buffer used for the experiments.

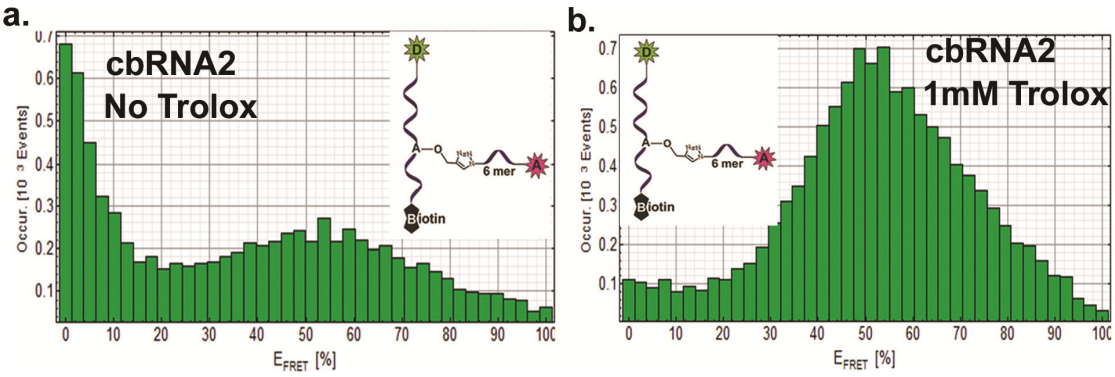


Figure 7.5. Representative FRET efficiency histograms of cbRNA2 with and without Trolox. **a.** The dual-labeled cbRNA2 shows large amount of zero FRET population when Trolox is not used. **b.** In presence of 1 mM Trolox the zero FRET state is significantly diminished.

The FRET efficiency histogram of cbRNA2 shows a broad distribution that is slightly skewed to high values of E_{FRET}. This suggests that a broad range of conformations can be adopted by the cbRNAs. Because both the stem and branch of the cbRNAs are single stranded, they are expected to be flexible. To test whether the conformational flexibility arises because of the single-stranded nature of the cbRNAs, the stem strand of the cbRNA (see experimental) was hybridized to a complementary strand resulting in more rigid duplex stem. Figure 7.6 shows the conformation of the different cbRNAs with and without duplex stem. A cbRNA1 which does not have a 2'-branch nucleotide (the acceptor dye is directly conjugated at the 2'-position of the branchpoint adenosine) was used as a control. As expected, little difference is seen between the single stranded and duplex stem forms of cbRNA1 (Fig. 7.6a and 7.6b, first column). Therefore cbRNA1 has a rigid conformation regardless of the nature of the stem. However a large difference in the conformational distribution of both cbRNA2 and cbRNA3 was observed when in duplex form. In the duplex stem form, the FRET efficiency distribution of these two cbRNAs becomes narrower (Fig. 7.6a and 7.6b, second and third column) indicating

that a more rigid conformation is adopted. In both cases the branch of the cbRNA is still single stranded and should have conformational flexibility. Nonetheless, these results suggest that in the duplex stem form, the cbRNA adopts a conformation in which the branch has decreased flexibility. Another possibility is that the cbRNA is so flexible that it moves very quickly between different populations than we can detect experimentally. It is also interesting that upon making the stem duplex, the cbRNAs loses its skewed high FRET population and adopts a more centered medium FRET values which are close to the expected values. This suggests that cbRNA conformations in which the stem and branch are closer than 90 degrees (lower α angle) are less likely when the stem is duplex.

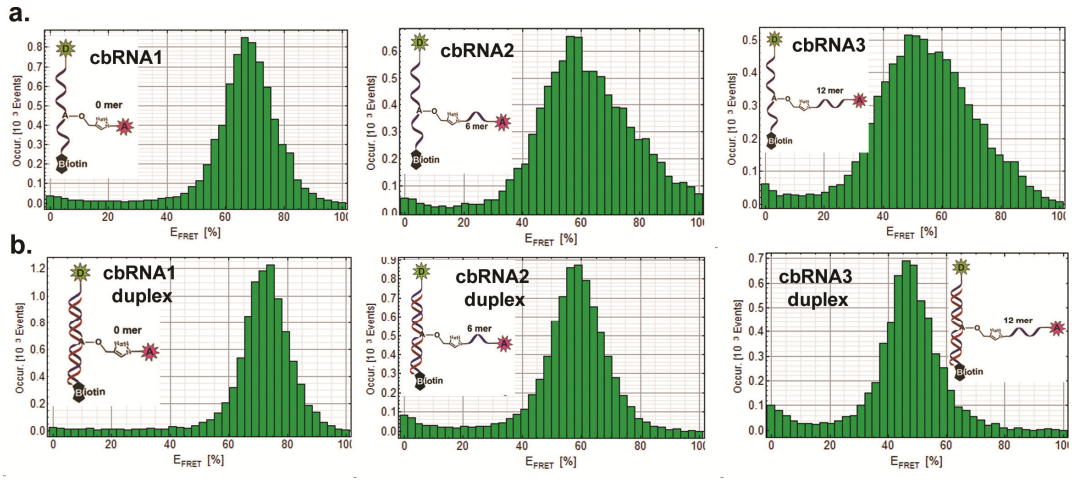


Figure 7.6. Representative FRET efficiency histograms of three different cbRNAs with and without duplex stem. **a.** Except cbRNA1, both cbRNA2 and cbRNA3 show broad FRET efficiency when the stem is single stranded. **b.** In the duplex stem form, all three cbRNAs show a narrow FRET distribution.

In previous chapters the importance of divalent metal ions for the debranching reaction was discussed.[7] Specifically, the crystal structure of the EhDbr1p shows a Mn^{2+} ion in the active site.[1] Although the role of Mn^{2+} in the Dbr1p active site is well understood, it is not known if Mn^{2+} affects the structures of the substrate bbRNAs. This

motivated us to determine whether the cbRNAs change their conformations in presence of Mn^{2+}. Upon addition of 1 mM Mn^{2+} to the buffer, a new population appears in the histogram with very high FRET efficiency (~90%). Thus, in presence of Mn^{2+}, cbRNA2 can adopt two different conformations, with medium (~57%) and high (~93%) FRET efficiencies, respectively. Although at this point it is not possible to propose a role for the high FRET population in the debranching reaction, it is interesting to observe the two different populations of cbRNA2 in presence of Mn^{2+}. A slightly increased contribution due to the zero FRET state is also observed for the experiments with 1 mM Mn^{2+} ion. This is probably because Mn^{2+} ions are known to be capable of inducing fluorescent dyes to relax to their triplet state upon excitation.[8,9] However as the FRET values for both sub-populations are much larger than zero, the spurious zero FRET peak is not expected to distort the distributions.

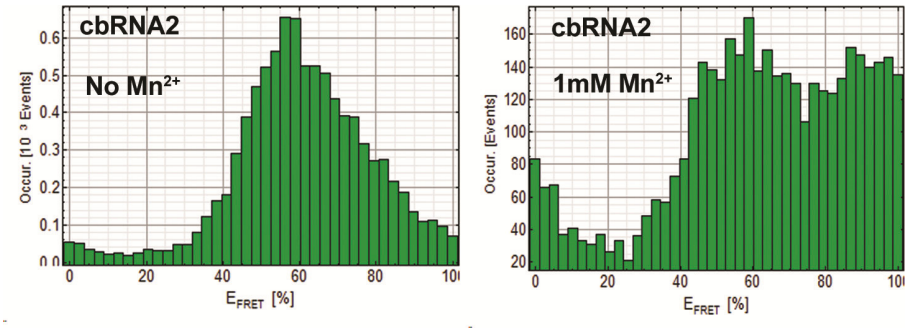


Figure 7.7. Representative FRET efficiency histograms of cbRNA2 with and without Mn^{2+} ion. In presence of 1 mM Mn^{2+}, cbRNA2 adopts a second conformation with high FRET efficiency.

7.4. smFRET studies of dual-labeled native bbRNA conformation(s)

In the previous section the conformational changes of cbRNAs monitored by smFRET were discussed. As the debranching reaction requires a 2'-5' phosphodiester

linked bbRNA, it is also important to study the conformation of the native bbRNAs under various physiological conditions. The synthesis of the native dual-labeled (Alexa488-Alexa594) bbRNA6 used for the single molecule studies is described in chapter 2. The native bbRNA6 has same sequence as the cbRNA2 except that the 2'-5' phosphodiester linkage replaces the triazole linkage (Table 7.3). The bbRNA6 was characterized by steady state and time-resolved fluorescence methods (Fig. 7.8).

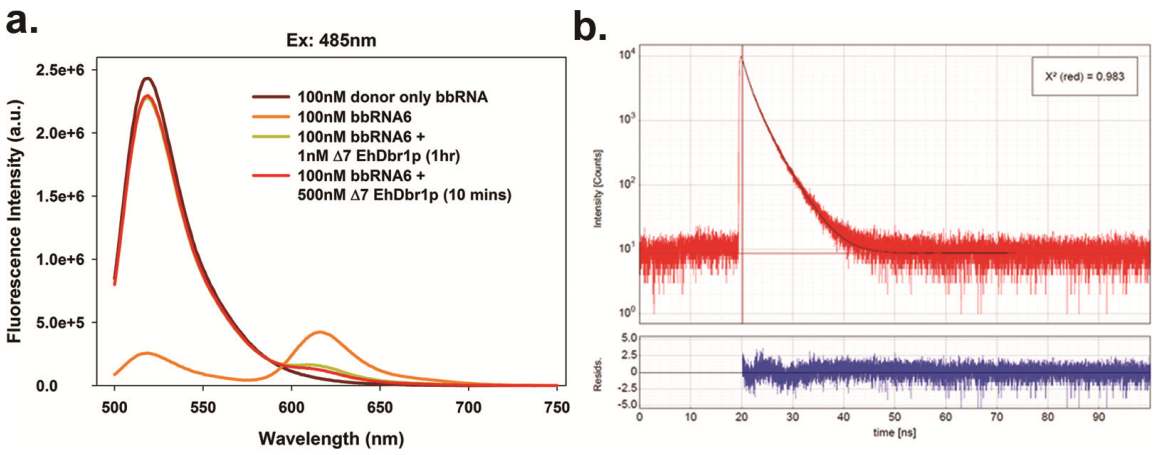


Figure 7.8. Ensemble fluorescence characterization of the bbRNA6. **a.** Fluorescence spectra of the bbRNA6 with and without Dbr1p. Upon debranching of the bbRNA6 by Δ7 EhDbr1p, the donor (Alexa488) fluorescence intensity increases indicating loss of FRET. Using both sub-saturating and excess enzyme compared to the substrate, the reaction goes to almost completion in 60 and 10 minutes respectively. **b.** Fluorescence decay of the dual-labeled bbRNA6 showing biexponential decay.

The dual-labeled bbRNA6 substrate shows high energy transfer as evident by the drop in donor signal and increase in acceptor signal (orange line, Fig. 7.8a) compared to the donor only labeled bbRNA (brown line, Fig. 7.8a). Under the excess enzyme condition (100 nM bbRNA6, 500 nM Δ7 EhDbr1p), the donor signal falls to nearly the donor-only level in 10 minutes (red line, Fig. 7.8a) indicating that the debranching reaction is complete. However, under sub-saturating condition (100 nM substrate, 1 nM

Δ7 EhDbr1p), it takes 1 hour to complete the debranching reaction (green line, Fig. 7.8a). Complete cleavage of the dual-labeled bbRNA substrate indicated by donor fluorescence increase suggests that the inclusion of the fluorescent dyes does not affect the debranching reaction. Therefore bbRNA6 is suitable for studying the conformation of bbRNAs in solution using smFRET**.** Figure 7.8b shows the two-exponential fit of the fluorescence decay for the bbRNA6. Similar to cbRNA2, bbRNA6 also shows two different lifetime components, 3.2 ns and 1.3 ns. The 3.2 ns lifetime corresponds to ~36% of the total population and represents a low-FRET state. The lifetimes of 1.3 ns corresponds to almost 2/3rd of the total population and represents a high-FRET state (E~65%) which is close to the expected FRET efficiency for bbRNA6 (~54%)**.**

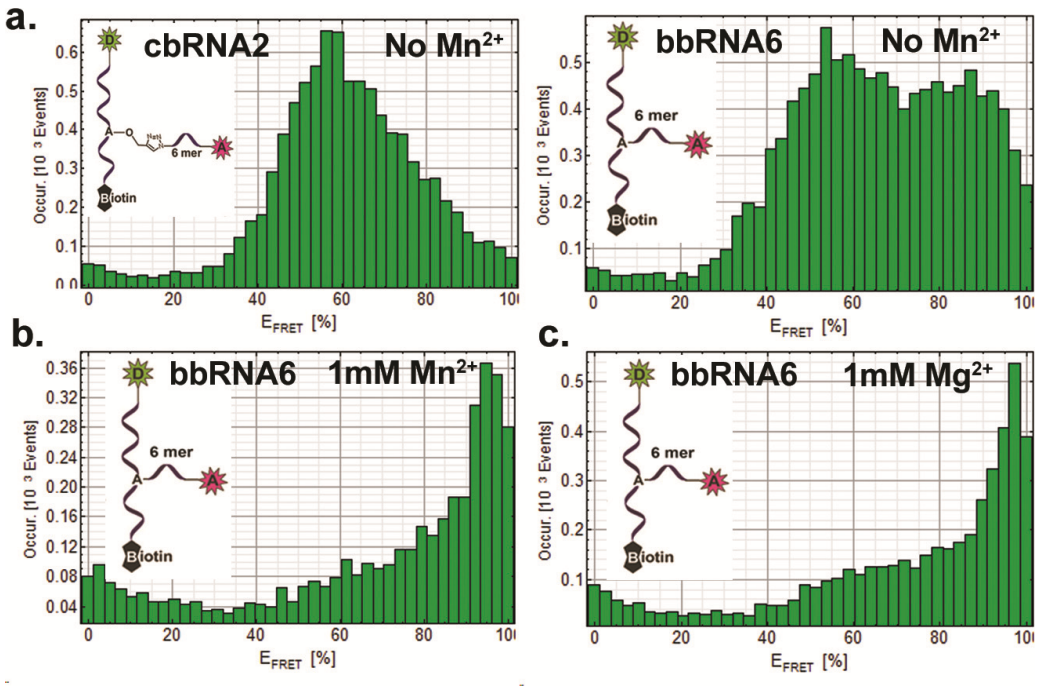


Figure 7.9. Representative FRET efficiency histograms of native bbRNA6 under different physiological conditions. **a.** Comparison of FRET histograms of cbRNA2 and native bbRNA6. The bbRNA6 shows two different conformations even without any added metal ion. **b.** FRET efficiency histogram of bbRNA6 in presence of 1 mM Mn^{2+}. Addition of the Mn^{2+} ion results in increase of the high FRET population. **c.** FRET efficiency histogram of bbRNA6 in presence of 1 mM Mg^{2+} shows a similar increase in the high FRET population.

The dual-labeled native bbRNA6 was used for smFRET studies as described before. Figure 7.9a shows a comparison of the FRET efficiency histogram of the cbRNA2 and bbRNA6. Unlike cbRNA2, bbRNA6 shows two distinct peaks in the FRET efficiency histogram, even in the absence of metal ion. This indicates that the native bbRNA can exist in two different conformations; one with medium FRET value (~58%) and another with high FRET value (~93%). The medium FRET population in bbRNA6 shows a similar maximum as in cbRNA6 (~57%). However, the high-FRET state may be due to base stacking between the nucleobases of the branchpoint residue and the first nucleotide in the 2'-branch as suggested by previous NMR-based conformational studies.[2–4] This interaction would bring the stem and the branch closer and therefore increase the FRET efficiency. The absence of a high-FRET state in cbRNA2 under comparable condition suggests that this conformation cannot be accessed in the absence of the 2'-5' phosphodiester linkage. To get an idea about the relative conformations of the two FRET states, the angle α is calculated for these two populations of the bbRNA6 (Fig. 7.4 and Table 7.2). As expected, the high FRET state shows an angle of 53° between the stem and the branch indicating close proximity of the donor and acceptor dyes. The medium FRET state shows almost 90° angle between the stem and branch which matches well with our initial assumption. It is important to mention here that cbRNA2 also adapts similar conformations in presence of 1 mM Mn^{2+}.

Population	E_{FRET} (%)	D-A distance (Å)	Angle (α)
Medium FRET	**58**	**53.7**	**88°**
High FRET	**93**	**36.8**	**53°**

Table 7.2. FRET efficiencies and angle between the stem and branch for the two different populations of the bbRNA6. The angle calculations were performed similar to that of Table 7.1.

When smFRET experiments were performed in the presence of 1 mM Mn^{2+}, bbRNA6 adopts predominantly a high FRET conformation (Fig. 7.9b). This conformational change in presence of Mn^{2+} is similar to that seen for cbRNA2. However the relative population of the high FRET conformation is much higher for bbRNA6 than for cbRNA2. To test whether this conformational change is specific for the Mn^{2+} ion, smFRET measurements of bbRNA6 were performed in presence of 1 mM Mg^{2+}. Under these conditions, a high FRET conformation is observed indicating that this conformational change is due to the presence of a divalent metal ion and not specific to Mn^{2+}.

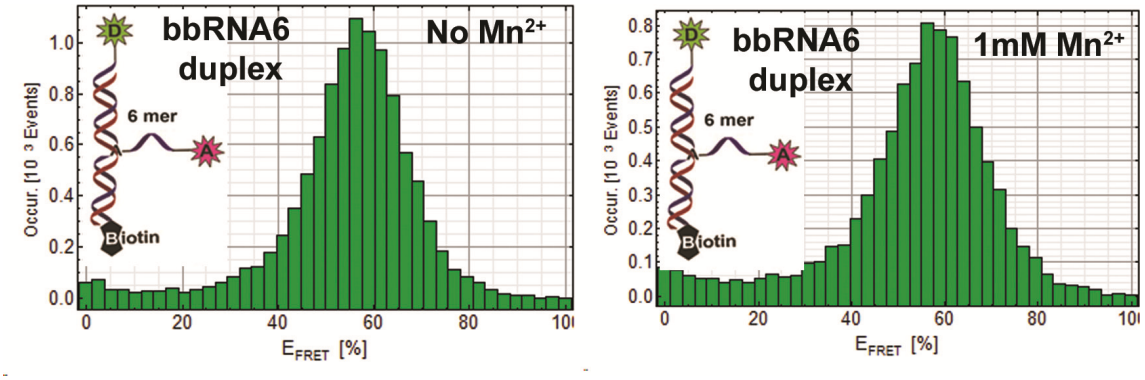


Figure 7.10. Representative FRET efficiency histograms of bbRNA6 with duplex stem in absence and presence of Mn^{2+}. **a.** Upon making the stem duplex the bbRNA6 shows a narrow FRET distribution indicating a more rigid structure. The high FRET population of the bbRNA6 is completely absent when the stem is in a duplex form. **b.** Addition of Mn^{2+} does not affect the FRET efficiency of the bbRNA6 with a duplex stem.

Finally, the effect of a duplex stem on the conformation of bbRNA6 was examined (Fig. 7.10). Interestingly, making the stem duplex eliminated the high FRET population of bbRNA6 (Fig. 7.10a). Instead, bbRNA6 shows a single narrow FRET efficiency distribution that is very similar to that observed for duplex cbRNA2 (Fig. 7.6b). Even addition of 1 mM Mn^{2+} did not affect this FRET distribution (Fig. 7.10b).

Therefore it can be concluded from these data that the conformational changes responsible for the high FRET state of both the native bbRNAs and cbRNAs requires the stem to be single-stranded. When the stem is in the duplex form, the nucleobases of the stem strand participate in hydrogen bonding and stacking interactions. Therefore any stacking interaction between the stem nucleobases with the nucleobases in the branch strand will be unfavorable. Therefore these studies support previous NMR-based structural model of the bbRNAs.[2,3]

7.5. smFRET studies of cbRNA dynamics

The confocal-based smFRET studies discussed in the previous sections are very well suited for probing different sub-populations of a multicomponent system. However the diffusing single molecules stay in the confocal volume for a short period of time (~1ms). Therefore any dynamical information slower than this timescale, cannot not be extracted using this technique. SmFRET time trajectories are most commonly recorded by imaging surface immobilized samples using total internal reflected fluorescence microscopy (TIRFM).[10–13] TIRFM takes advantage of the evanescent field which is created during total internal reflection of light (see experimental and Fig. 7.13). The intensity of the evanescent field decreases exponentially from the surface and therefore it almost diminishes to zero above ~200 nm from the surface. Therefore fluorescent molecules which are close to the surface are excited and any molecule above ~200 nm from the surface is not excited reducing the background suitable for single molecule detection. For measurement of FRET efficiencies of individual molecules using TIRFM, they must be immobilized on the surface and therefore they can be observed for an extended period of time.

SmFRET measurements of cbRNAs were performed using an objective-type TIRF microscope equipped with a dual-channel detection system for donor and acceptor (see experimental for details). The cbRNA molecules were immobilized on glass coverslip surface modified with PEG and biotin-PEG (~10%). The 3'-biotin of the cbRNAs was used to immobilize them on the modified surface using the biotin-streptavidin interaction. To minimize the blinking and photobleaching of cbRNAs, a pyranose oxidase, catalase and gluocose oxygen scavenger system was used (see experimental).[14] This oxygen removal system is superior to the commonly used glucose oxidase based buffer because the pH does not decrease with time. Because a decrease in the pH could result in a conformational change in cbRNAs, this buffer is more suited for our smFRET experiments. Following the imaging of cbRNAs, the data was analyzed using an open source smFRET software (iSMS).[15]

The cbRNA1 with a duplex stem was imaged as a control for the smFRET experiments. Figure 7.11a shows superimposed images of both the donor and acceptor channel for cbRNA1 rendered in false color (red and green). Most of the molecules in the image are orange, indicating that they have both high acceptor (red) and low donor (intensity) as expected for a molecule with a high FRET efficiency. Figure 7.11b shows the FRET efficiency histogram of cbRNA1 calculated from the all the molecules observed in Figure 7.11a. Similar to what was seen using burst analysis, cbRNA1 shows a single peak and narrow distribution of FRET efficiency in TIRFM. The two-color smFRET data of a representative molecule is shown in Figure 7.11c. The high intensity of the acceptor channel compared to the donor channel indicates a high FRET value (~0.8) of this cbRNA1 molecule. Figure 7.11d shows the dynamics of a single cbRNA1

over time. As expected for a rigid duplex RNA, a steady FRET value around 0.8 was observed for cbRNA1 without any large fluctuations.

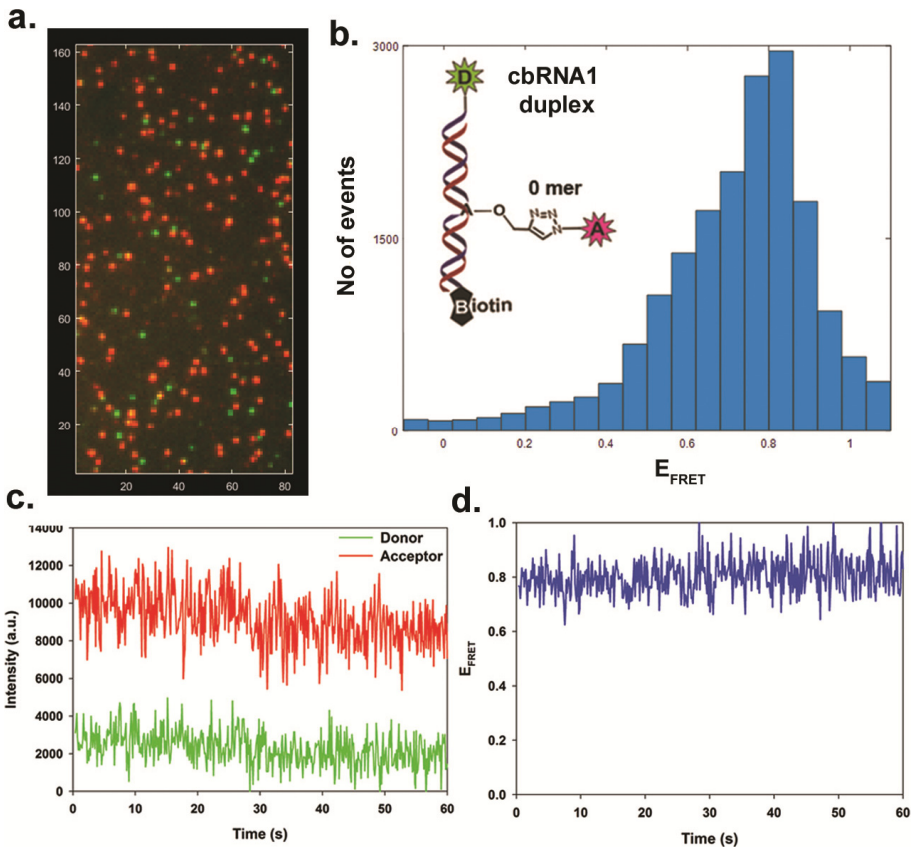


Figure 7.11. smFRET studies of dual-labeled cbRNA1 using TIRFM. **a.** Superimposed image of false color coded donor and acceptor channels. **b.** Representative FRET histogram of cbRNA1 measured from all the molecules shown in a. **c.** Two color smFRET data of a single cbRNA1 molecule showing the donor and acceptor intensity with time. **d.** FRET efficiency of the cbRNA1 molecule with time calculated from the smFRET data in c.

Next, the conformation and dynamics of the cbRNA2 (6mer branch) were investigated using smFRET (Fig. 7.12). As was seen using burst analysis, cbRNA2 shows a broad distribution of FRET efficiencies with a maximum of ~0.6. The donor and acceptor intensity of a single cbRNA2 molecule is shown in Figure 7.12c. The anti-

correlated nature of the donor and acceptor intensity suggests that these fluctuations occur due to a change in the efficiency of energy transfer rather than to other photophysical processes. The FRET values at different time points were calculated using the donor and acceptor intensity. cbRNA2 shows a dynamic switching behavior between a medium FRET (~0.6) and high FRET (~0.95) state (Fig. 7.12d). This conformational switch between a high and low FRET state suggest that cbRNA molecules continuously transit from one specific conformation to another. The role of this conformational switching on binding of cbRNAs with Dbr1p will be investigated in the future.

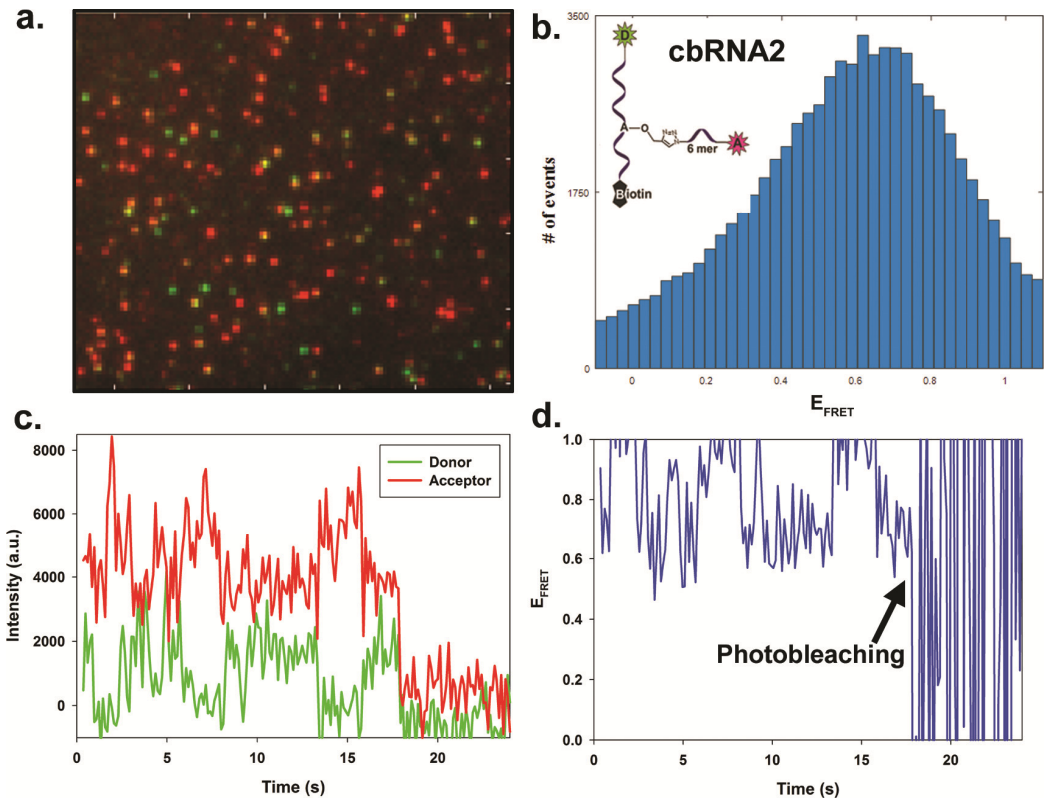


Figure 7.12. smFRET studies of dual-labeled cbRNA2 using TIRFM. **a.** Superimposed image of false color coded donor and acceptor channels for cbRNA2. **b.** Representative FRET histogram of cbRNA1 measured from all the molecules shown in a. **c.** Two color smFRET data of a single cbRNA1 molecule showing the donor and acceptor intensity with time. The anti-corelated nature of the donor and acceptor signal suggests that these changes are due to conformational fluctuations. **d.** FRET efficiency of the cbRNA1 molecule with time calculated from the smFRET data in c. The photobleaching of the donor dye is shown by an arrow.

7.6. Conclusion

Although 2'-5' phosphodiester linked synthetic bbRNAs have been accessible for long, their conformation(s) in solution is not well studied. Such information is useful in understanding how Dbr1p can selectively recognize bbRNA substrates. In this chapter, studies of the conformation and dynamics of both native and non-native 'click' branched RNAs using smFRET are described. The 2'-5' triazole linked cbRNAs exhibited a broad range of conformations in solution as indicated by a broad FRET efficiency histogram with a single maxima. However, upon making the stem of the cbRNA duplex, a more restricted conformation is adopted by the cbRNAs. In presence of Mn^{2+} ions, an important cofactor for Dbr1p, cbRNAs show a new high FRET population indicating stabilization of a high FRET state by the divalent metal ion. The conformations of native 2'-5' phosphodiester linked bbRNAs were compared with that of the cbRNAs. Unlike the cbRNAs, bbRNA6 showed two different FRET populations even in the absence of added Mn^{2+} suggesting that the native bbRNA can exist in two possible conformations. The observation of a high FRET state in bbRNA6 supports a previous prediction of base stacking between the branchpoint adenosine nucleobase and nucleobases in the 2'-branch based on NMR. Upon addition of 1 mM Mn^{2+}, the majority of the bbRNA molecules adopts a high FRET conformation. Finally the dynamics of cbRNAs in solution were studied using smFRET combined with the TIRF microscopy. These studies showed that the cbRNAs undergoes a conformational switching between a high FRET and medium FRET state. This high FRET state could be the same state that is induced by Mn^{2+} ions. Although the data obtained for dynamics studies of the cbRNAs are preliminary in nature, it provides a framework for future smFRET experiments with different bbRNA

substrates in presence of Dbr1p. Such studies will allow us to monitor any conformational change of the substrate RNA upon binding with Dbr1p.

7.7. Experimental

Dual-labeled RNAs for smFRET experiments

The sequence of the different dual-labeled RNAs used for the smFRET experiments are shown in Table 7.3. The syntheses of these RNAs are reported in chapter 2.

Name	Sequence
cbRNA1	**5'-Alexa488-C6-ACU ACU GUA CUA A-(2'-t--C6-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**
cbRNA2	**5'-Alexa488-C6-ACU ACU GUA CUA A-(2'-t-5' GUAUGA-C6-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**
cbRNA3	**5'-Alexa488-C6-ACU ACU GUA CUA A-(2'-t-5' GUAUGA CUAUCC-C6-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**
bbRNA6	**5'-Alexa488-ACU ACU GUA CUA A-(2'-5' GUAUGA-*t*-Alexa594)-CAA GUU ACU U-3' Biotin-TEG**

Table 7.3. Sequences of dual-labeled RNAs used for smFRET experiments.

The dual-labeled RNAs were hybridized with the respective complementary strand to make duplex DNAs. The hybridization reaction mixture contained 100 µL of 1 µM dual-labeled RNA and 1.5 µM of the complementary RNA strand in 1X hybridization buffer (50 mM Tris.HCl (pH 7.5), 50 mM NaCl and 0.005% TritonX-100). Because the complementary strand is not labeled, an excess could be used to ensure complete duplex formation. The hybridization was performed by heating the reaction

tube at 85˚C for two minutes and then at 65˚C for ten minutes. Following this, the tube was slowly cooled to room temperature for 1 hr resulting in duplex RNAs.

smFRET experiments in confocal setup

The smFRET experiments for the dual labeled RNAs using the confocal microscope were performed as described in the previous chapter except one modification. 1x debranching reaction buffer (50 mM Tris.HCl (pH 7.5), 25 mM NaCl and 0.005% TritonX-100) with added 1 mM Trolox was used for the smFRET measurements. Then 1 mM Mn^{2+} or 1 mM Mg^{2} was added to the reaction buffer as required.

smFRET experiments in TIRF setup

The TIRF microscope for smFRET imaging is built around an inverted microscope base (Olympus IX-71) and it uses evanescent field generated during total internal reflection for exciting surface immobilized biomolecules. The 488 nm laser line from a mixed Ar-Kr gas laser (SpectraPhysics) is used as an excitation source for fluorescent molecules. The laser beam is attenuated with neutral density (ND) filters to achieve the required power. Then the diameter of the laser beam is increased by using a 5x beam expander. Following this the laser beam is introduced into the microscope through the back port of the microscope. A 200 mm focusing lens on a xyz stage is placed in the laser path just before the back port of the microscope. This lens focusses the laser beam on the back focal plane of the TIRF objective (Olympus oil immersion, 60X, 1.4NA) so that the beam that comes out of the objective have parallel rays. The power of the laser beam is measured just before it entered through the back port of the microscope. To achieve total internal reflection, a coverslip (#1.5) is placed on the microscope with the immersion oil and then

the 200 mm lens is moved using the XYZ stage such that parallel beam comes out of the objective at an acute angle. The XYZ stage is moved until the laser beam hits the critical angle and undergoes total internal reflection.

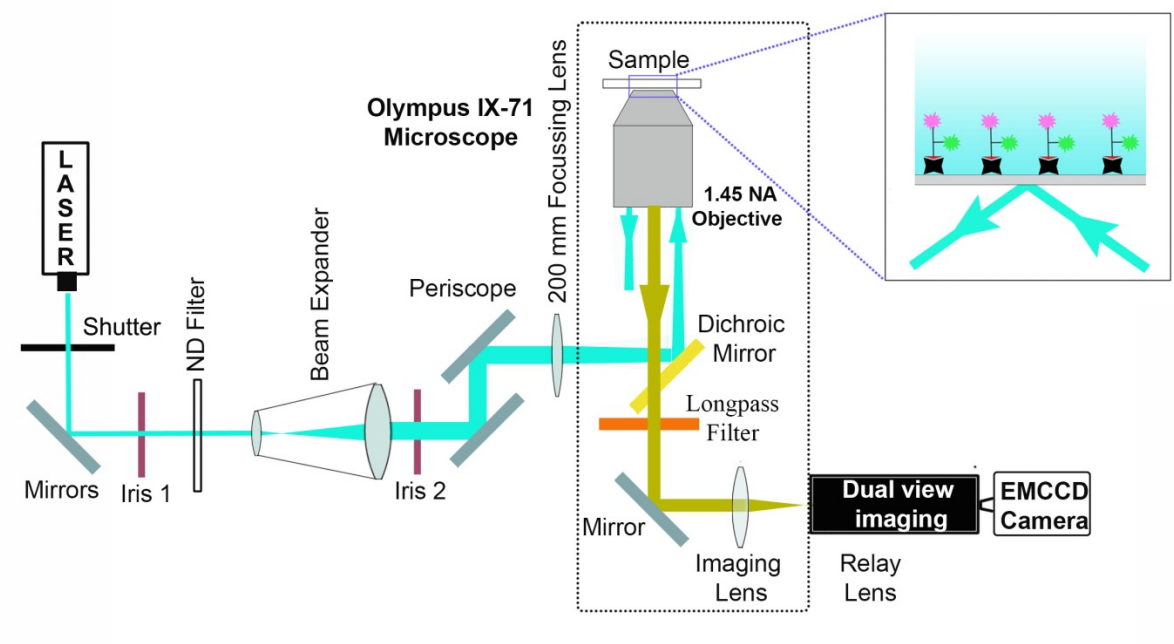


Figure 7.13. Schematics of the TIRF microscope for smFRET imaging. The inset shows the evancescent field created due to total internal reflection of the laser beam. Because the evanescent field rapidly decays as it move away from the surface, molecules which are only close to the surface are excited.

The fluorescence emission from the dual-labeled RNA molecules is also collected using the same objective which is then directed towards the emission pathway through a 500 nm long pass filter (Semrock). The fluorescence emission collected by the objective is focused using the imaging lens. The fluorescence emission is then separated into the donor and acceptor channel using two channel simultaneous imaging system DV2 (Photometrics) which was attached to the C mount of the microscope side port. A 565nm dichroic mirror was utilized in the DV2 system to separate the donor and acceptor

emission. Both the donor and acceptor emissions are then passed through bandpass filters to further eliminate the background noise. For the donor dye, a 531/46 nm bandpass filter (Semrock) is used and for the acceptor dye a 641/75nm bandpass filter is used. Finally the both the channels were focused on to the active imaging area of an Evolve 512 EMCCD camera (Photometrics) to image the immobilized single molecules.

Preparation of flow cells for smFRET imaging of RNAs

The flow cell for single molecule imaging was prepared by sandwiching a glass (25x75 mm) slide and a coverslip (#1.5, 22x25 mm) using double-sided sticky tape. Both the slide and the coverslip surfaces were passivated with PEG (MW5000) and a small amount of biotin-PEG (~10%). The PEG prevents non-specific attachment of labeled biomolecules to glass surface. The biotin PEG is used to immobilize the biomolecules through biotin-streptavidin interaction. The cleaning and surface passivation of the slide and coverslips were performed following reported protocols.[12,13] A 200 μL pipette tip and a 1 mm wide tubing was introduced in the flow cell through two holes drilled into the slide. Then all the connections in the flow cell were sealed using Epoxy (Devcon). To introduce buffer or any other solution in the flow cell, it was first placed into the pipette tip. Then the solution was pulled into the flow cell by using a syringe which was attached to the tubing through a needle.

smFRET imaging of dual-labeled RNAs using TIRF microscopy

For the smFRET experiments first the flow cells were washed with the 1x imaging buffer (50 mM Tris.HCl (pH 7.5), 50 mM NaCl and 0.005% TritonX-100) three times (100 μL each). Then 50 μL of 0.1 mg/mL streptavidin solution was introduced into the chamber

and incubated for 2 mins. Then the flow cells were washed again with the imaging buffer. Then the dual-labeled RNAs were introduced in the chamber which also contained a biotin-TEG modification for surface attachment. A 50 pM sample of the RNAs were used for the surface immobilization. Following this, the oxygen scavenger system (OSS) was introduced in the flow cell to remove dissolved oxygen from the imaging buffer and also prevent blinking of the dyes. The OSS solution contained pyranose oxidase (3 U/mL), catalase (100 U/mL), 4 mM glucose and 2 mM Trolox in 1x imaging buffer. After introducing the OSS solution, the immobilized molecules were imaged using TIRF microscope. I used 1mW laser power before the microscope for the imaging of the dual-labeled molecules. 100 ms integration time was used for the smFRET imaging which resulted in a frame rate of ~8frames/s. Following the data collection, the images were analyzed using the iSMS software which is freely accessible.[15]

References

1. Montemayor, E. J. *et al.* Structural basis of lariat RNA recognition by the intron debranching enzyme Dbr1. *Nucleic Acids Res.* **42,** 10845–10855 (2014).

2. Damha, M. J. & Ogilvie, K. K. Synthesis and Spectroscopic Analysis of Branched RNA Fragments: Messenger RNA Splicing Intermediates. *J. Org. Chem.* **3722,** 3710–3722 (1988).

3. Zhou, X.-X., Nyilas, A., Remaud, G. & Chattopadhyaya, J. 270 MHz 1H-MNR Studies of four 'branched' tetraribonucleotides: A3'p5'A2'p5'G3'p5'U, A3'p5'A2'p5'G'3'p5'U, U3'p5'A2'p5'G3'p5'U & U3'p5'A2'p5'G3'p5'C which are formed as the lariat branch-point in the pre-mRNA processing reactions (splicing). *Tetrahedron* **44,** 571–589 (1988).

4. Foldesi, A., Agback, P., Glemarec, C. & Jyoti, C. Synthesis of tetrameric branched RNA-DNA conjugate & branched-RNA analogue & their comparative conformational studies by 500 MHz NMR spectroscopy. *Tetrahedron* **47,** 7135–7156 (1991).

5. Sabanayagam, C. R., Eid, J. S. & Meller, A. Using fluorescence resonance energy transfer to measure distances along individual DNA molecules: Corrections due to nonideal transfer. *J. Chem. Phys.* **122,** 1–6 (2005).

6. Dietrich, A., Buschmann, V., Meller, C. & Sauer, M. Fluorescence resonance energy transfer (FRET) and competing processes in donor-acceptor substituted DNA strands: A comparative study of ensemble and single-molecule data. *Rev. Mol. Biotechnol.* **82,** 211–231 (2002).

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

8. Stennett, E. M. S., Kodis, G. & Levitus, M. Photobleaching and blinking of TAMRA induced by Mn 2+. *ChemPhysChem* **13,** 909–913 (2012).

9. Ciuba, M. A. & Levitus, M. Manganese-induced triplet blinking and photobleaching of single molecule cyanine dyes. *ChemPhysChem* **14,** 3495–3502 (2013).

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

11. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. & Ha, T. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **77,** 51–76 (2008).

12. Zhao, R. & Rueda, D. RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods* **49,** 112–117 (2009).

13. Lamichhane, R., Solem, A., Black, W. & Rueda, D. Single-molecule FRET of protein-nucleic acid and protein-protein complexes: Surface passivation and immobilization. *Methods* **52,** 192–200 (2010).

14. Swoboda, M. *et al.* Enzymatic oxygen scavenging for photostability without ph drop in single-molecule experiments. *ACS Nano* **6,** 6364–6369 (2012).

15. Preus, S., Noer, S. L., Hildebrandt, L. L., Gudnason, D. & Birkedal, V. iSMS: single-molecule FRET microscopy software. *Nat. Methods* **12,** 593–594 (2015).

Chapter 8

Conclusion and future directions

The lariat RNA generated during pre-mRNA splicing includes a branch-point adenosine residue that is linked through the 2'-*O* position to the 5'-end of the RNA sequence. This 2'-5'-phosphodiester linkage in the lariat RNA backbone is debranched by the lariat debranching enzyme (Dbr1p). Following debranching, some introns can participate in highly important biological processes like snoRNA biogenesis, microRNA pathways etc.[1] Although Dbr1p activity was initially identified more than thirty years ago as being necessary for efficient intron turnover,[2] the key roles Dbr1p plays in a variety of regulatory processes were identified over the years.[3] To study debranching and related processes easy synthetic access to backbone branched RNA (bbRNA) substrates is necessary.

We developed a solid phase synthetic method for easy access to bbRNAs utilizing a photoprotected phosphoramidite reagent as the key component. The photoprotected phosphoramidite can be synthesized in three easy steps starting from commercially available materials in high yields. Use of the photoprotected amidite at the branchpoint provides a reagent-less deprotection condition for branch synthesis which can accommodate a multitude of functional group in the bbRNAs. Our new method of bbRNA synthesis is versatile – not only does it allow us to incorporate chemical modifications such as fluorescent dyes, biotin, azide or alkyne functionalities into the bbRNAs, but it also for the first time permits incorporation of modified nucleotides in the branch. Such modifications are critical for biochemical and biophysical analysis of bbRNAs and Dbr1p. The true power of our strategy was demonstrated by synthesis of a

29 mer bbRNA substrate containing three different modifications (two fluorescent dyes and biotin). The later part of my thesis work was built based on our facile access to bbRNA substrates.

Utilizing a solid-phase synthesis strategy, we have synthesized several different modified bbRNAs for biochemical and kinetics studies. The bbRNA substrates synthesized for the biochemical studies contain a 5'-fluorescent dye (Dylight547) for detection and quantification. This eliminated the need for the use of hazardous radioactive elements. The biochemical studies showed for the first time how two different Dbr1p variants (ScDbr1p and EhDbr1p) can have different Mn^{2+} ion requirements for debranching activity. Next, bbRNA substrates with modified nucleotides in the 2'-branch were studied. The different reactivity of those modified substrates indicates important role of the 2'-branch residues in substrate binding and cleavage by Dbr1p. To better understand the Dbr1p binding and cleavage mechanism, we require more detailed kinetics analysis. For that purpose, we developed a fluorescence-based assay for the debranching reaction kinetics. A dual-labeled (Cy3 and Cy5) bbRNA substrate was synthesized which, upon cleavage by Dbr1p, causes separation of the donor (Cy3) and acceptor (Cy5) dyes resulting in a decrease in FRET. The loss of FRET can be monitored in real time to follow the progress of the reaction. Using this method, the kinetics parameters (k_{cat} and K_m) of three different Dbr1p variants were measured for the very first time. This kinetics analysis will be a benchmark for studying the Dbr1p mechanism and will be used in the future to study debranching reactions using modified substrates and enzymes. To study the binding of bbRNA substrates with Dbr1p, we synthesized a non-cleavable substrate analogue of Dbr1p: ‘click’ branched RNA (cbRNA), in which the

2'-5' phosphodiester bond is replaced by a triazole ('click') linkage. Using the kinetics assay described earlier, this 'click' branched RNA was shown to be a competitive inhibitor of Dbr1p that binds Dbr1p in a similar fashion to the native substrate, permitting its use in binding studies.

While synthetic access to bbRNAs has enabled biochemical studies of the debranching reaction and the substrate requirements for Dbr1p, these studies do not give any information about the structure and dynamics of the bbRNAs in solution in presence and absence of Dbr1p. Single molecule experiments are ideal for studying different sub-populations in a macroscopic system and their time evolution. A confocal microscope setup was used to study the conformation of the different branched RNAs in solution under different physiological conditions. While developing the conditions for these experiments, we discovered an unexpected result of using Trolox as an efficient triplet state quencher in these experiments. Specifically, Trolox resulted in elimination of the zero FRET population without the need for acceptor-only excitation.[4,5] While the native bbRNA substrate could be used for studying cleavage by Dbr1p, the cbRNAs could be used to study binding with Dbr1p. Dual-fluorescently labeled native bbRNAs and non-native cbRNAs were synthesized for single molecule FRET (smFRET) studies. While the native bbRNA shows a bi-modal FRET distribution in solution, the cbRNA exhibits a broad distribution of FRET values with a single peak. Both the bbRNAs and cbRNAs underwent a conformational change to a high FRET state in presence of the Mn^{2+} ion which is an important cofactor for Dbr1p activity. Interestingly, the conformational changes in these branched RNAs are dependent on the nature of the stem strand. In presence of a duplex stem strand, both the bbRNAs and cbRNAs adopt a more rigid

conformation which does not change even in presence of Mn^{2+}. These results support previous NMR based findings which predicted base stacking interaction between the branchpoint nucleobase and the 2'-branch nucleobase.[6,7]

As a complementary technique to confocal microscopy, an objective based total internal reflection fluorescence (TIRF) microscope was used to study the dynamics of the branched RNAs immobilized on the surface. Unlike the confocal microscope, TIRF allows observation of single molecules over much longer time which is useful to study the dynamics of the branched RNAs both bound to and free of protein. The initial results obtained with the dual-labeled cbRNA showed a conformational switch between a medium FRET and high FRET state associated with protein binding. Such conformational switch could be an important element for the recognition of bbRNA substrates by Dbr1p.

The work presented in this thesis has set up several directions for future studies of the Dbr1p cleavage mechanism. The versatile nature of our solid-phase synthesis strategy developed in this work gives us access to a variety of modified bbRNAs. Debranching kinetics studies with branch-modified bbRNAs will provide us a more detailed explanation (binding or catalysis) of the difference in the debranching activity. To isolate the role of important amino acid residues in the Dbr1p active site, kinetics studies will also be performed with several mutants of the Dbr1p enzyme. The FRET-based debranching kinetics assay developed in this work will be very useful for such kinetics studies. To understand the nature of the interactions between the Dbr1p enzyme and the bbRNA substrates, we are trying to use the cbRNAs as a substrate analogue in crystallography experiments (collaboration with Prof. Andrew VanDemark at University

of Pittsburgh). Because cbRNAs are competitive inhibitor of Dbr1p, they are expected to bind with the enzyme in a similar fashion as the native bbRNA substrate thereby showing the important contacts between Dbr1p and its substrate.

Finally the single molecule studies have revealed novel conformation of the branched RNA substrates in different environments. However the relevance of these conformations to the actual debranching reaction remains to be discovered. The presence of Dbr1p in the single molecule experiments will allow us to study whether Dbr1p prefers any particular conformation of the branched RNA substrates. Using smFRET, we should be also able to monitor any conformational change of the substrate upon binding with Dbr1p. These experiments are currently been performed by Dr. John Pettersson, a postdoc in the lab. Future studies of debranching reaction kinetics at single molecule level will also be performed using the native bbRNA substrates synthesized.

Although the studies presented in the thesis were performed *in-vitro* with purified enzyme, we could use the knowledge gained from these studies to control Dbr1p function *in vivo*. For example the dual-labeled bbRNA substrate can be transfected into a cell and used to localize Dbr1p based on the cleavage of the substrate.[8] Because of the important role of Dbr1p in different cellular processes, inhibition of Dbr1p activity could also have therapeutic potential.[3,9] We are also trying to use the non-cleavable cbRNAs to prevent proliferation of HIV virus by deactivating Dbr1p. I envision that the work presented in this thesis will further our knowledge of the elusive lariat debranching enzyme which will improve our fundamental understanding of enzymatic reactions as well as give us potential therapeutic targets.

References

1. Dey, S. K., Paredes, E., Evans, M. & Das, S. R. in *From Nucleic Acids Sequences to Molecular Medicine* (eds. Erdmann, A. V. & Barciszewski, J.) 475–501 (Springer Berlin Heidelberg, 2012). doi:10.1007/978-3-642-27426-8_19

2. Ruskins, B. & Green, M. R. An RNA Processing Activity that Debranches RNA Lariats. *Science* **229,** 135–140 (1985).

3. Hesselberth, J. R. Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA* **4,** 677–691 (2013).

4. Kapanidis, A. N. *et al.* Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38,** 523–533 (2005).

5. Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P. J. & Seidel, C. A. M. Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection. *Methods Enzymol.* **475,** 455–514 (2010).

6. Damha, M. J. & Ogilvie, K. K. Synthesis and Spectroscopic Analysis of Branched RNA Fragments: Messenger RNA Splicing Intermediates. *J. Org. Chem.* **3722,** 3710–3722 (1988).

7. Zhou, X.-X., Nyilas, A., Remaud, G. & Chattopadhyaya, J. 270 MHz 1H-MNR Studies of four 'branched' tetraribonucleotides: A3'p5'A2'p5'G3'p5'U, A3'p5'A2'p5'G'3'p5'U, U3'p5'A2'p5'G3'p5'U & U3'p5'A2'p5'G3'p5'C which are formed as the lariat branch-point in the pre-mRNA processing reactions (splicing). *Tetrahedron* **44,** 571–589 (1988).

8. Kataoka, N., Dobashi, I., Hagiwara, M. & Ohno, M. hDbr1 is a nucleocytoplasmic shuttling protein with a protein phosphatase-like motif essential for debranching activity. *Sci. Rep.* **3,** 1–7 (2013).

9. Armakola, M. *et al.* Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat. Genet.* **44,** 1302–1309 (2012).