Structure and Dynamics of Non-coding RNAs by Single-Molecule Microscopy and Complementary Methods

by

Kalika R. Pai

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Chemistry

Dissertation Committee:

James Prell, Chair

Julia R. Widom, Advisor

Jeffery Cina, Core Member

Diane Hawley, Core Member

Alice Barkan, Institutional Representative

University of Oregon

Spring 2026

© 2026 Kalika R. Pai
This work is openly licensed via BY-NC 4.0.

DISSERTATION ABSTRACT

Kalika R. Pai

Doctor of Philosophy in Chemistry

Title: Structure and Dynamics of Non-coding RNAs by Single-Molecule Microscopy and Complementary Methods

Non-coding RNAs span a huge range of size, complexity, and biological functions. This variety means that a variety of techniques are necessary to study them. In this dissertation, I describe the use of different methods, used in tandem, to study different non-coding RNAs. I present a study which aims to expand the use of a single-molecule microscopy technique to long non-coding RNAs, by improving the design of probes used to investigate the structural accessibility of RNA local structure. I describe the utilization of the probe design principles to study a heart-specific long non-coding RNA, paired with chemical probing experiments to further refine probe design. Finally, I discuss the use of single-molecule fluorescence resonance energy transfer to study a class of riboswitches, along with bulk fluorescence spectroscopy to further survey ligand binding. These experiments show that pairing single-molecule microscopy with various complementary methods can yield a greater understanding of the systems of interest than any individual method.

This dissertation contains previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Kalika R. Pai

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
New Mexico Institute of Mining and Technology

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2026, University of Oregon
Bachelor of Science, Chemistry, 2020, New Mexico Institute of Mining and Technology

AREAS OF SPECIAL INTEREST:

Structure and Dynamics of Nucleic Acids
RNA Biophysics
Single-Molecule Microscopy

PROFESSIONAL EXPERIENCE:

Graduate Research Assistant, University of Oregon, 2021-2026

Graduate Teaching Assistant, University of Oregon, 2020-2021

Undergraduate Teaching Assistant, New Mexico Institute of Mining and Technology, 2017-2020

GRANTS, AWARDS, AND HONORS:

NSF GRFP, University of Oregon, 2022

Oregon ARCS Foundation Scholar Award, University of Oregon, 2020

Melvin Hatch Chemistry Pentathlon Prize, New Mexico Institute of Mining and Technology, 2020

Carl J. Popp Senior Thesis Award, New Mexico Institute of Mining and Technology, 2020

Morris F. Stubb Award, New Mexico Institute of Mining and Technology, 2020

Outstanding Physical Chemistry Award, New Mexico Institute of Mining and Technology, 2019

Outstanding Analytical Chemistry Award, New Mexico Institute of Mining and Technology, 2018

PUBLICATIONS:

Pai, K. R.; Martin, A. M.; Kadrmas, M.; Widom, J. R. Alternative Probe Chemistries for Single-Molecule Analysis of Long Non-Coding RNA. *Journal of Biological Chemistry* **2026**, *0* (0). https://doi.org/10.1016/j.jbc.2026.111441.

Hossain, S.; Pai, K. R.; Piyasena, M. E. Fluorescent Lipo-Beads for the Sensitive Detection of Phospholipase A2 and Its Inhibitors. *ACS Biomater. Sci. Eng.* **2020**, *6* (4), 1989–1997. https://doi.org/10.1021/acsbiomaterials.9b01720.

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Dr. Julia Widom for all of her support and mentorship relating to both research and my career. The feedback and resources she has provided to me have been extremely valuable, and I have grown significantly as a scientist during my time in her lab.

I would also like to thank those who mentored me prior to graduate school. I thank Dr. Menake Piyasena for welcoming me into his lab at New Mexico Tech as a sophomore undergraduate and mentoring me in my first research experience. I thank Dr. Jeffery Altig and Dr. Sally Pias for the teaching opportunities and mentorship I received, as well as the entire New Mexico Tech chemistry department for their teaching and support. I also thank Dr. Reza Foudazi for the opportunity to gain new research skills and build confidence in working independently and adapting to different labs.

I thank all members of the Widom lab. I thank former lab member Dr. Nico Conoan Nieves for building our single-molecule instrument and for training me in single-molecule microscopy. I thank Madison Kadrmas for their contributions to maintaining and improving the microscope, and for contributions to data collection and analysis. I thank the undergraduates I had the privilege of mentoring- Megan Adamec, Natalie Sande, and Aimee Martin, both for their contributions to my research, and for allowing me to grow as a mentor. I thank Taylor Coulson for always being available for brainstorming sessions. I also thank Madison Kadrmas and Taylor Coulson for their support beyond the lab space as friends.

Thank you to all the members of the UO community who have supported me along the way. I thank my committee members for their support and feedback over the years. I thank all faculty in the Physical Chemistry division for their teachings, advice, and insights. I thank the

administrative staff in Chemistry, IMB, and OMQ for all of their contributions that make our research possible. I thank Bryan Rebar for his advice and help with my GRFP proposal, and for mentorship support. I thank the past and present members of WGS for welcoming me into such a supportive organization.

I thank the Oregon ARCS Foundation for their scholarship award, and their continued support network. I would also like to acknowledge the National Science Foundation, National Institutes of Health, and American Heart Association for providing research funding.

I would like to thank all members of my family. To my husband and best friend, Jon, thank you for everything you do to make my life better. You are truly essential to the success I've had over the past 9 years. I thank his family for welcoming me and for supporting my decisions that led us a thousand miles away. I thank Denny for all the support only a crusty white dog can provide. Finally, I would like to thank my parents, Doug and Karen Pai, for their love and support. Thank you for raising me to believe not just that I can do anything, but that I should, and for instilling the values that have led to the person I am today.

This dissertation is dedicated to Jon and Denny for their endless support and love.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION 15

1. Overview 15

2. RNA Structure and Functions 16

3. Long Non-Coding RNAs 19

4. Riboswitches 21

II. METHODS 24

1. Methods to Study RNA Structure 24

2. Oligonucleotide Design for Single-Molecule Studies 25

3. Single-Molecule and Bulk FRET 26

4. Single-Molecule Kinetic Analysis of RNA Transient Structure and Chemical Probing 30

5. Bridge to Chapter III 34

III. ALTERNATIVE PROBE CHEMISTRIES FOR SINGLE-MOLECULE ANALYSIS OF LONG NON-CODING RNA 35

1. Overview 35

2. Introduction 35

3. Results 38

A. Experimental Design 38

B. SiM-KARTS Probe Chemistry Impacts RNA Binding Properties 42

C. Melting Temperatures Confirm Increased Binding Stability of LNA and MO Probes 48

D. LNA Probes Distinguish Structures on a Trace-by-Trace Basis..... 50

E. Divalent Cations Impact Probe Binding to Bvht H9:cs..... 54

4. Discussion..... 59

5. Experimental Procedures..... 62

A. Oligonucleotides..... 62

B. RNA Preparation for SiM-KARTS..... 63

C. SiM-KARTS Data Collection..... 63

D. SiM-KARTS Data Analysis..... 64

E. Melting Curves..... 66

F. Circular Dichroism Spectroscopy..... 68

6. Bridge to Chapter IV..... 68

IV. STRUCTURAL INVESTIGATIONS OF THE LNCRNA MHRT..... 69

1. Overview..... 69

2. Introduction..... 69

3. Structural Determination of Mhrt by Predicted Secondary Structure and SHAPE..... 72

4. Preliminary Results of Mhrt Structure and Dynamics via SiM-KARTS..... 80

5. Conclusions..... 84

6. Future Directions..... 85

7. Bridge to Chapter V..... 87

V. STUDIES OF LIGAND BINDING TO VARIANTS OF THE NAD^+-II RIBOSWITCH..... 88

1. Overview..... 88

2. Introduction..... 88

3. NAD^+-II Riboswitch 91

4. Mini-NAD^+-II Riboswitch 94

5. Conclusions 97

6. Future Directions 98

VI. CONCLUDING REMARKS 100

APPENDIX: SUPPORTING INFORMATION FOR CHAPTER III-
ALTERNATIVE PROBE CHEMISTRIES FOR SINGLE-MOLECULE
ANALYSIS OF LONG NON-CODING RNA 102

REFERENCES CITED 112

LIST OF FIGURES

Figure Page

1.1. RNA base pairing and stacking 18

1.2. Secondary structure of *Braveheart* 20

1.3. Riboswitch structure and function 21

2.1. Setup of single-molecule FRET 26

2.2. Setup of SiM-KARTS 30

2.3. SHAPE adduct formation and reverse transcription termination 33

3.1. Model target RNA for SiM-KARTS and probe backbone chemistry 39

3.2. Impact of probe chemistry on dwell times 43

3.3. Melting curves of each probe type 49

3.4. Clustering of traces according to average dwell times 51

3.5. Impact of ionic concentration on dwell times of LNA2 or MO binding 56

3.6. CD spectroscopy of RNA at varying $MgCl_2$ concentration 58

4.1. Proposed mechanism for function of *Mhrt* 70

4.2. Binding of each SHAPE primer to *Mhrt* 73

4.3. Gel exploring annealing conditions of primer 1 and primer 4 74

4.4. Gel optimizing SHAPE sequencing 75

4.5. Optimization of 1M7 concentration 77

4.6. Optimization of NAI concentration 78

4.7. Preliminary results of *Mhrt 779 t1* SHAPE 79

4.8. SiM-KARTS probe for poly-A region of *Mhrt 779 t1* 80

4.9. Binding of *Mhrt 779 t1* to Cy3-labeled t1 81

4.10. SiM-KARTS design and example traces of poly-A probe 81

4.11. Cumulative distributions of poly-A probe dwell times..... 83

4.12. Fano factor of unbound dwell time of poly-A probe 83

4.13. Cumulative counts of unbound dwells in- and out-of-burst 84

5.1. Secondary structures and labeling sites of NAD^+-II and mini-NAD^+-II 89

5.2. smFRET data from NAD^+-II 92

5.3. Histograms for NAD^+-II, fit with a sum of four Gaussians 93

5.4. Bulk FRET titration of NAD^+ 94

5.5. Mini-NAD^+-II bulk FRET titrations of NMN 96

5.6. Example smFRET data of mini-NAD^+-II 97

A1. Impact of probe concentration of dwell times of LNA2 102

A2. Overview of dwell time distribution analysis 102

A3. Representative single-molecule traces of control conditions..... 104

A4. SiM-KARTS data reproducibility 105

A5. Processed melting curves represented as fraction bound vs temperature 106

LIST OF TABLES

Table Page

3.1. Sequences of oligonucleotides 41

3.2. Characteristic dwell times of probes binding to Bvht H9 or Bvht H9:cs 45

3.3. Characteristic dwell times for LNA2 and MO binding to Bvht H9:bs 46

3.4. Melting temperature for each probe type 50

3.5. Quantification of overlap between the ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) distributions for Bvht H9 and Bvht H9:cs 53

3.6. Characteristic dwell times of LNA2 or MO binding in indicated salt concentrations 57

A1. Detailed fitting results for Bvht H9 data reported in Table 3.2 107

A2. Detailed fitting results for Bvht H9:cs data reported in Table 3.2 107

A3. Detailed fitting results for Bvht H9 data reported in Table 3.3 108

A4. Detailed fitting results for experimental replicates 108

A5. Detailed fitting results for bootstrapped data sets 109

A6. Detailed fitting results for Bvht H9 data reported in Table 3.6 (LNA2) 109

A7. Detailed fitting results for Bvht H9:cs data reported in Table 3.6 (LNA2) 110

A8. Detailed fitting results for Bvht H9 data reported in Table 3.6 (MO) 110

A9. Detailed fitting results for Bvht H9:cs data reported in Table 3.6 (MO) 111

CHAPTER I: INTRODUCTION

1. Overview

This dissertation will focus on the use of single-molecule microscopy alongside complementary techniques to study the structure and dynamics of non-coding RNAs. While single-molecule methods are advantageous for a variety of reasons, supporting data from biochemical or spectroscopic techniques can allow for a more comprehensive view of an RNA's structure and dynamics. The diversity of structural and dynamical behaviors exhibited by non-coding RNAs that make them interesting to study also mean that no single technique will give sufficient information for a full understanding of the RNA. I will discuss studies of two types of non-coding RNA, long non-coding RNA (lncRNA) and riboswitches, each using different sets of techniques.

This dissertation describes work done in the lab of Dr. Julia Widom, who advised on all experiments and written material. Chapter I serves as a general introduction to RNA and the specific systems that are discussed in this dissertation. Chapter II describes the theory and details of the methodology used, including single-molecule microscopy, bulk fluorescence spectroscopy, and chemical probing assays. Chapter III contains work recently published in *Journal of Biological Chemistry*, co-authored with Aimee Martin, Madison Kadrmas, and Julia Widom, and discusses the use of alternative probe chemistries for single-molecule kinetic analysis of RNA transient structure (SiM-KARTS). In this work, Aimee Martin and Madison Kadrmas contributed to the collection and analysis of SiM-KARTS data. Chapter IV describes further work done to utilize SiM-KARTS in tandem with chemical probing to study lncRNA. This work was done in collaboration with Natalie Sande, Megan Adamec, and Julia Widom.

Biochemical experiments were performed with assistance by Natalie Sande and Megan Adamec. Chapter V presents single-molecule and bulk Förster resonance energy transfer (FRET) studies of NAD^+-II riboswitches, done in collaboration with Nico Conoan Nieves, Madison Kadrmas, and Julia Widom. smFRET and bulk FRET data for the NAD^+-II riboswitch were collected by Nico Conoan Nieves, including preliminary data analysis. Madison Kadrmas contributed to the collection and analysis of mini-NAD^+-II smFRET data.

2. RNA Structure and Functions

RNA is often viewed within the Central Dogma of Molecular Biology, where RNA is the intermediate between the genetic information carried by DNA and proteins which contribute a variety of functions. Since the 1950s, it has been known that RNAs exist that perform functions beyond encoding the sequences of proteins, with the identification of the ribosome and later transfer RNA (tRNA).[1] How RNA was viewed shifted further with the discovery of ribozymes, which are catalytic RNAs.[2] Today, it is known that up to 99% of the human genome does not encode proteins, but most is transcribed into RNA.[3] Non-coding RNAs are diverse in function, playing roles in translation, catalysis, gene regulation, and processing.[4]

These functions are closely related to the structure of the RNA. While RNA is often depicted as unstructured and "floppy", this does not give an accurate image of RNA. Unlike DNA, which is generally double-stranded, RNA is initially synthesized in single-stranded form. RNA is able to fold into complex secondary and tertiary structures, which, for some RNAs, have been suggested to be conserved more heavily than their primary sequences.[5] RNA secondary structure is defined by hydrogen bonds comprising the canonical Franklin-Watson-Crick base pairs (Figure 1.1), as well as noncanonical pairing. These interactions stabilize and facilitate

RNA folding, forming structures which can impact the function of the RNA. An archetypical example is the 14-3-3ξ mRNA in *Drosophila melanogaster*, which is present in three different conformational states, each of which lead to different splicing pathways.[6] Specifically, differences in the exact nucleotides that are base-paired to produce stem-loops lead to the activation of a specific exon, while the others are suppressed due to their base-pairing.

While the structural details revealed in this dissertation focus primarily on secondary structure through canonical base-pairing, it is important to acknowledge that other features of RNA can impact structure. Like DNA, RNA is helical in nature, due to stacking of the planar aromatic rings that make up the bases (Figure 1.1). This base stacking is important for the overall three dimensional shape of an RNA molecule, as it impacts the orientation of different bases, and further stabilizes the molecule through nonspecific van der Waals interactions between bases.[7] Interestingly, it has been shown that RNA is nearly always compact, with long RNA molecules folding into structures where the 3' and 5' ends are in close proximity.[8] Single-molecule studies were used to show that the observed small end-to-end distances were largely due to canonical base-pairing, and that these end-to-end distances remain small even within different possible conformational states. While RNA is referred to as intrinsically structured, the distinction must be made that RNA is not inherently static.[9]

Some RNAs function by changing structure in response to the binding of proteins, small molecules, or ions, while many others sample multiple conformational states without external stimulus. The dynamic nature of RNA structure leads to complications in determining RNA structure, since most techniques will provide an ensemble average view of the RNA in question. Well-studied examples of the structure-function relationship include tRNA, which undergo a large amount of post-transcriptional modification, but have a highly conserved structure which is

necessary for correct translation of mRNA to proteins.[10] The major features of the canonical "cloverleaf" structure are a stem where the 3' and 5' ends exist and three stem-loops referred to as "arms". Together, these features are responsible for aspects of the RNA's function, such as correct positioning on the ribosome.[11] While RNA structure is closely related to function, it is often the case that function is determined without investigation into structural details. For example, the mechanism by which an RNA named BC069792 suppresses tumor growth was determined using overexpression and knockout assays, but structural investigations were not performed.[12] This is often the case for non-coding RNAs due to their diversity and abundance.

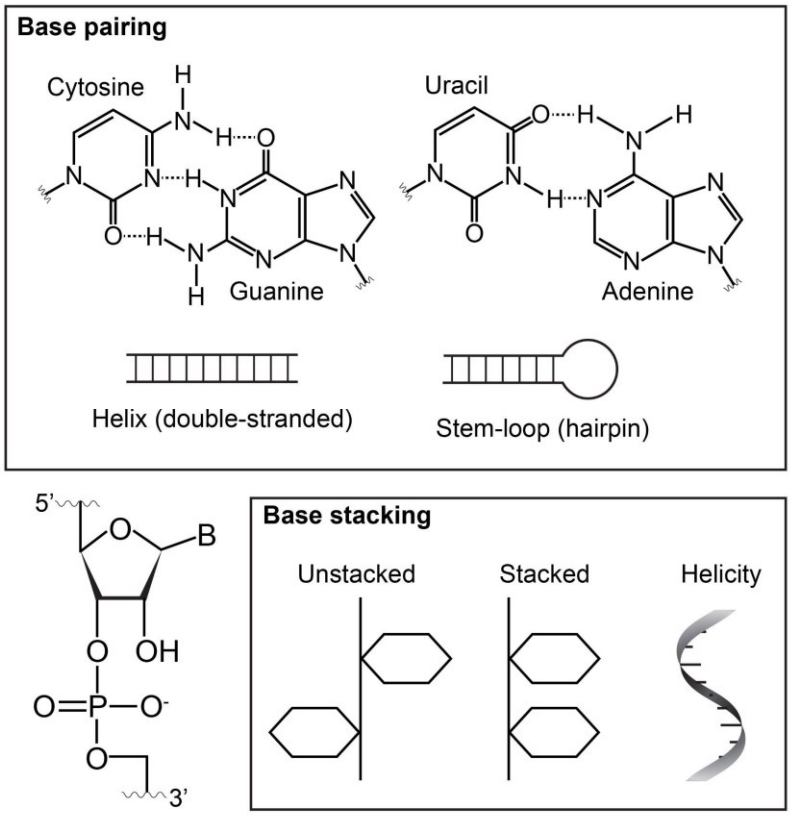


Figure 1.1. RNA base pairing and stacking. Upper panel: Base structures and hydrogen bond locations for pairs between cytosine (C) and guanine (G), and uracil (U) and adenine (A). Represented are two major features formed by base pairs in RNA- the helix and stem-loop. Lower left: Chemical structure of the backbone of RNA. B signifies any of the four RNA bases. Lower panel: Simplified representations of unstacked and stacked bases, and the helical form caused by stacking.

Within non-coding RNAs, there are many classes which contribute different functions. These include the previously mentioned tRNAs, ribosomal RNAs, small nuclear RNAs, and small nucleolar RNAs. Each of these have specific localizations and functions in the cell. This dissertation focuses on long non-coding RNAs (lncRNAs) and riboswitches, which are two distinctly different classes of non-coding RNAs.

3. Long Non-Coding RNAs

The two defining features of lncRNAs are straightforward: lncRNAs are 200 nucleotides or greater in length, and do not encode proteins. In recent years, functions of lncRNAs have become increasingly identified as relevant in human diseases.[13] In cardiac research alone, it has been found that many lncRNAs play roles in heart development and function, such as H19,[14] MALAT1,[15] *Braveheart*,[16] and *Myheart*.[17] Among these, the structures of many remain unrefined, but some have been thoroughly studied.

Braveheart, an lncRNA studied in Chapter III, was first identified to be a regulator of cardiovascular commitment, suggesting that lncRNAs can direct cell fate.[16] Notably, *Braveheart* was the first lncRNA found to have implications on cardiovascular development. Later, the structure of *Braveheart* was studied using multiple types of chemical secondary structure probing (Figure 1.2).[18] The structure was found to be highly modular, meaning shorter fragments of the full RNA were found to have the same structural features as the full length transcript. Interestingly, the three main domains of *Braveheart* correspond roughly with its exons. The same study also determined that the 5' asymmetric G-rich internal loop (AGIL) was a necessary feature for Braveheart to carry out the role of cellular differentiation. Specifically, it was discovered that the AGIL motif interacts with factors that are known to bind G-rich nucleic

acids, such as cellular nucleic acid binding protein (CNBP). The binding of *Braveheart* to CNBP is suggested to be a critical interaction in how *Braveheart* functions in cardiovascular lineage commitment.[18] Further studies have looked at CNBP binding and the conformational changes to Braveheart that accompany such binding.[19] It was found that upon binding of CNBP, there is an interesting change in structure of *Braveheart* that is more apparent for the full RNA than with individual fragments. This conformational change is observed as a compaction in structure, and it is suggested that at least two distinct sites on *Braveheart* form interactions with CNBP.[19] The work done on *Braveheart* has revealed that not only do lncRNAs have structures that can be well defined, but that interactions such as protein binding can induce conformational changes.

Further studies have revealed many cardiac-related lncRNAs. Once such group, termed myosin heavy-chain-associated transcripts (*Myheart* or *Mhrt*), consists of six splice isoforms and was identified as cardioprotective in function.[17] It was demonstrated that *Mhrt* repression was needed for myopathy to occur, which makes *Mhrt* an interesting basis for potential therapeutics. Further studies have primarily looked into expression levels of *Mhrt* in patients with various cardiac conditions, presenting *Mhrt* as a potential biomarker.[20–22] However, to date, no studies have uncovered the structure or dynamics of *Mhrt*.

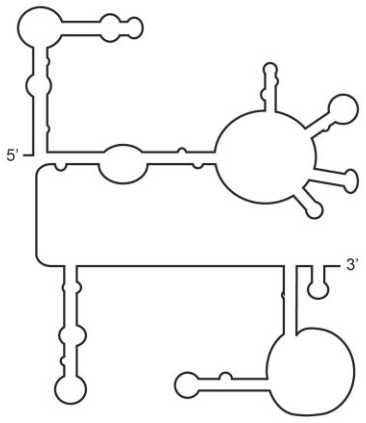


Figure 1.2. Simplified representation of the solved secondary structure of *Braveheart*.

4. Riboswitches

Riboswitches are cis-acting non-coding regions of RNA found in bacteria, which regulate gene expression in response to ligand binding through conformational changes (Figure 1.3). In the face of ever increasing levels of antibiotic resistance, targeting riboswitches is a promising avenue to combat bacterial infections.[23] To efficiently design drugs to target a riboswitch and impact cell health, it is important to understand how the riboswitch in question binds its native ligand, as well as the dynamics of this interaction and how this functionality can be disrupted.

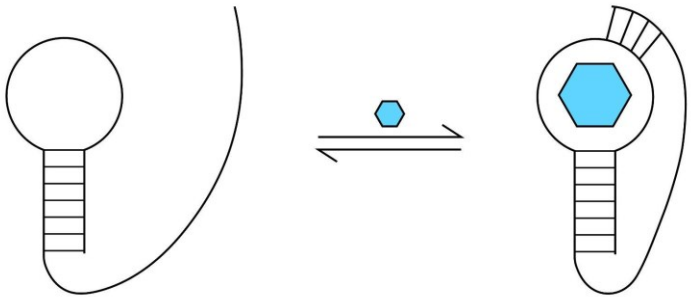


Figure 1.3. Cartoon of general riboswitch structure and function. Upon ligand binding, a conformational change is induced, enabling regulatory functions.

One interesting group of riboswitches bind the reduced form of nicotinamide adenine dinucleotide (NAD^+), which is a key cofactor in many biosynthetic and signaling pathways.[24] NAD^+ riboswitches are categorized into two classes: NAD^+-I[25] and NAD^+-II.[26] The identification of these riboswitches began with the finding that RNAs with the *nadA* motif function as NAD^+ riboswitches, specifically binding only the adenosine moiety of NAD^+.[25] Further studies revealed two distinct binding domains, one of which might aid in recognizing the nicotinamide moiety of NAD^+.[27] These riboswitches were later defined to belong to class I following the discovery of a class of NAD^+ sensing riboswitches that selectively bind nicotinamide riboside and derivatives

such as NAD^+ and nicotinamide mononucleotide (NMN) using the *pnuC* motif.[26] The structures for both the NAD^+-I and NAD^+-II riboswitch classes were studied, with specific interest in the differences between the free and ligand-bound states. For the NAD^+-II riboswitch, it was suggested that distinct binding pockets are responsible for binding nicotinamide, and that multiple RNA architectures that sense NAD^+ might exist.[26] Further structural investigations found that with NAD^+ or NMN bound, a unique RNA fold is present, with tiers of base tetrads and base triples forming the binding pocket around the ligand.[28] As is true of many riboswitches,[29,30] ligand binding facilitates the formation of a pseudoknot, in which the Shine-Dalgarno (SD) sequence is sequestered, switching off translation of the downstream open reading frame.[28]

More recently, new variants of the NAD^+-II riboswitch have been identified. These riboswitches, termed mini-NAD^+-II riboswitches, are found more abundantly than the initially reported NAD^+-II riboswitch from the *pnuC* motif.[31] The variants found of the mini-NAD^+-II riboswitch form a simpler pseudoknot scaffold, and lack the P1a stem found in the classic NAD^+-II riboswitch. Like NAD^+-II, mini-NAD^+-II binds both NAD^+ and NMN, with stronger binding to NMN than NAD^+. NAD^+ is bound more weakly by the mini-NAD^+-II riboswitch as compared to the full length version, likely due to the conserved adenines around the P1a helix that have been reported to interact with the adenosine moiety of NAD^+.[31] The mini-NAD^+-II riboswitch likely still functions by binding NAD^+ rather than NMN, due to the high relative concentration of NAD^+ in cells. While the structures of both mini- and full-length NAD^+-II riboswitches have been studied in the absence and presence of binding partners, the conformational dynamics of these RNAs in different conditions are not yet well described.

RNA structure comprises many different aspects, from primary and secondary structure to more complex conformational dynamics. Due to this complexity and the variety brought about by different types and sizes of RNA, there is no single best technique to study RNA structure. To gain a complete understanding of the structure and dynamics of an RNA, multiple complementary techniques should be used in tandem.

CHAPTER II: METHODS

1. Methods to Study RNA Structure

Many methods exist to study the structure of RNA, each with its advantages and limitations. Computational tools have been developed to predict RNA secondary structure, including methods based on free energy minimization calculations and, more recently, the incorporation of machine-learning algorithms.[32] While these calculations are useful for understanding RNA structure, they work best in tandem with experimental techniques, as accuracy is not guaranteed with the large diversity of structures known for RNAs. X-ray crystallography is a method which has been used successfully for decades to determine atomic-resolution structures.[33] However, the necessity to crystalize the RNA leaves questions about the biological relevance of results. Additionally, highly dynamic regions of RNA may be poorly resolved or may prevent the RNA from crystallizing entirely, again leaving unresolved questions. Small-angle x-ray scattering is done in solution to determine size and and global shape of RNA molecules, but can only give information on the average structure of the RNA.[34] Cryo-electron microscopy is another powerful technique,[35] as is nuclear magnetic resonance[36] for measuring RNA structure in solution. Various chemical probing methods are widely used, such as dimethyl sulfate probing and selective 2' hydroxyl acylation analyzed by primer extension (SHAPE).[37] All the techniques above provide information only on the ensemble average of the RNA, and cannot report on dynamics. Methods utilizing single-molecule microscopy can resolve the dynamics of individual RNA molecules, while also contributing structural information. These techniques are well suited to the dynamic nature of RNA. The following chapters will utilize combinations of these techniques to study the structure and dynamics of non-coding RNAs.

2. Oligonucleotide Design for Single-Molecule Studies

Single-molecule microscopy of RNA relies on labeling with specific fluorophores depending on the optics and lasers utilized. Our system is configured for Cyanine 3 (Cy3) and Cyanine 5 (Cy5), and alternative dyes with similar excitation and emission profiles. There are multiple routes to label RNA depending on the desired synthesis type, and labeling positions. Smaller RNAs can be obtained through solid phase synthesis from companies such as Horizon (Dharmacon), with the option for 3' or 5' labeling with Cy3 or Cy5. If a different labeling site is desired, an internal modification 5-aminohexylacrylamino-uridine (5-LC-N-U) can be used in the place of U, which then allows for the attachment of dyes with a succinimidyl ester reactive group. For the smFRET studies here, RNA was purchased with 3' Cy3 labeling and an internal 5-LC-N-U modification to which Cy5 was coupled. Additionally, the inclusion of a biotin label (5' labeling was used in these studies) on the RNA is necessary for immobilization in the single-molecule sample chamber and can be included by the manufacturer. Longer RNAs, such as *Mhrt*, are better obtained by purchasing DNA through companies such as Integrated DNA Technologies (IDT) and performing *in vitro* transcription. RNA obtained through transcription can be 5' or 3' end labeled if desired. For use of RNA obtained via *in vitro* transcription for single-molecule microscopy, a biotinylated DNA tether sequence which exhibits stable binding with the RNA is needed. This tether can be labeled with a fluorophore in order to localize molecules without labeling the RNA, assuming efficient hybridization of the tether to the RNA.

3. Single-Molecule and Bulk FRET

Single-molecule microscopy is powerful for the study of RNA structure and dynamics because information from individual molecules is retained, allowing for the measurement of conformational dynamics in a heterogeneously behaving sample. The single-molecule techniques discussed in this dissertation use prism type total internal reflection fluorescence (TIRF) microscopy, where a prism is used to direct laser light to a glass/water interface at a critical angle for the light to undergo total internal reflection. This setup creates an evanescent field, which is used to gently excite fluorescently labeled RNA molecules tethered to the glass surface (Figure 2.1). As a result of this excitation method, TIRF microscopy allows for generally dark backgrounds and high quality data. The fluorescence collected by the microscope is then directed through optics to split the image into two paths depending on wavelength. These two paths are then directed onto a camera chip side-by-side to allow for the simultaneous collection of two channels, one containing red fluorescence, the other green fluorescence. Movies are collected, and traces of time-dependent intensities of each channel are extracted for further analysis.

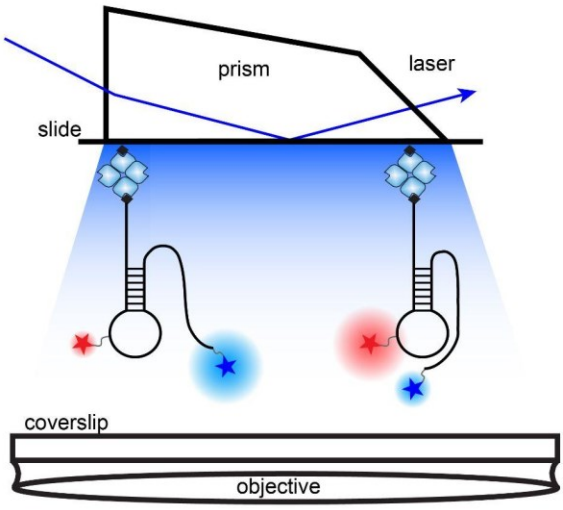


Figure 2.1. General setup of a single-molecule FRET experiment.

A popular use of TIRF microscopy is single-molecule Förster resonance energy transfer (smFRET), which utilizes two fluorophores at specific positions on the molecule.[38] Bulk FRET and smFRET operate on the same principle, where the efficiency of FRET reports on inter-dye distance following Equation 2.1, where R is the inter-dye distance and R_0 is the Förster radius where $E_{FRET} = 0.5$.

$$E_{FRET} = \frac{1}{1+\left(\frac{R}{R_0}\right)^6} \quad \text{Equation 2.1}$$

Experimentally, the apparent FRET efficiency, E_{app}, is determined by Equation 2.2, which relates the intensities of the acceptor and donor, I_A and I_D respectively.

$$E_{app} = \frac{I_A}{I_A+I_D} \quad \text{Equation 2.2}$$

Correction factors to convert E_{app} to an actual FRET efficiency are needed. One factor, γ, corrects for differences in quantum yield and detection efficiency between the two fluorophores, and is the ratio of the change in acceptor intensity to the change in donor intensity upon acceptor photobleaching.[39] The corrected FRET efficiency, E, is determined by Equation 2.3 where βI_D is an additional correction factor that accounts for bleed-through of donor emission into the acceptor channel.

$$E = \frac{I_A-\beta I_D}{(I_A-\beta I_D)+\gamma I_D} \quad \text{Equation 2.3}$$

E is calculated from the green and red channel traces for each molecule, resulting in E_{FRET} traces which can be further analyzed. Histograms of E_{FRET}, populated by the first 50 frames of each trace, show the distribution of FRET states for a particular set of data. They can be used to determine what FRET states exist in the underlying data, and the prevalence of the conformations that give rise to those states.

System kinetics can be measured using smFRET on a tens-of-milliseconds time scale. The length of time a molecule is in any given conformation is determined by the dwell time of the molecule at a particular E_{FRET} value, or "FRET state". A popular analysis technique for state and transition determination is Hidden Markov Modeling (HMM), which uses a manually inputted initial model to idealize FRET traces. The use of HMM for analysis of smFRET systems was first reported in 2006, and has become widespread since.[40] HMMs are used to describe stochastic processes, specifically to predict the evolution of observable events.[41,42] A model is used to assign states and determine transitions between states that might be obscured by noise. Available software such as QuB Online allow the user to input starting values for transition rates between each possible state, the signal amplitudes of each state, and the approximate standard deviation of signal observed within each state.[43] These inputs can be guided by the histogram for the data, considering the peak heights, E_{FRET} values, and peak widths respectively. The modeling results in what can be referred to as "idealized traces" in which transitions between states are clear and no longer obscured by noise.[40]

The number of states and how they interconvert can lead to various conclusions about RNA structure and conformational dynamics. Specific information about dynamics can be obtained from transition occupancy density plots (TODPs), which show the prevalence of transitions between different states. Which pairs of states interconvert directly and which require an intermediate state can be easily seen in these plots. From this information, rate constants can be extracted for each transition type. These rates can be compared for different conditions for an RNA of interest, such as a riboswitch in ligand free or ligand bound contexts, which will reveal potential differences in conformational dynamics and stability.

To gain a full understanding of a system studied by smFRET, it can be advantageous to also perform bulk FRET measurements using steady-state fluorescence spectroscopy. For studies of systems like riboswitches which bind ligands, bulk FRET titrations can be done to determine the dissociation equilibrium constant (K_D) much more efficiently than could be accomplished by smFRET while also confirming the general nature of the shift in structure. By measuring the change in FRET efficiency at different ligand concentrations, fractional saturation, θ, can be calculated using Equation 2.4, where I_i is the Cy5 intensity (normalized to the intensity of Cy3) at a specific concentration, I_{min} is the Cy5 intensity at the minimum ligand concentration, and I_{max} is the Cy5 intensity at the maximum ligand concentration.

$$\theta = \frac{I_i - I_{min}}{I_{max} - I_{min}} \qquad \text{Equation 2.4}$$

In the case of FRET using fluorophores such as Cy3 and Cy5, photobleaching occurs over the titration period, which convolutes changes in FRET efficiency with an overall reduction in fluorescence intensity. This can be corrected for by performing a "mock titration" where the photobleaching of the RNA is measured over a titration of water only, and subtracting the normalized Cy5 intensity of the mock titration, $I_{mock,i}$ from the matching titration data to obtain the contrast. The contrast value of each concentration of the titration can be used in the place of normalized intensity in Equation 2.4 to obtain the corrected fractional saturation. The corrected fractional saturation curve can be fit with the Hill equation, Equation 2.5, where *[L]* is the concentration of ligand, K_D is the dissociation constant, and *n* is the Hill coefficient.

$$\theta = \frac{[L]^n}{K_D + [L]^n} \qquad \text{Equation 2.5}$$

4. Single-Molecule Kinetic Analysis of RNA Transient Structure and Chemical Probing

For longer, more complex RNAs, it is not easy to label the RNA as is required for smFRET, but there are still other single-molecule techniques which can be used. One of these is single-molecule kinetic analysis of RNA transient structure (SiM-KARTS).[44] SiM-KARTS can be done without direct labeling of the RNA, or using a simple end-labeling reaction. Instead of using the intensities of each fluorophore to calculate the distance between two locations, one fluorophore is used for localization of the molecule, while the other is used to track binding and dissociation of a short, oligonucleotide probe (Figure 2.2). The binding behavior of this probe is used to determine the accessibility of the complementary site for binding, which gives information on conformation. Changes over time to the accessibility give information about the dynamics of the RNA.

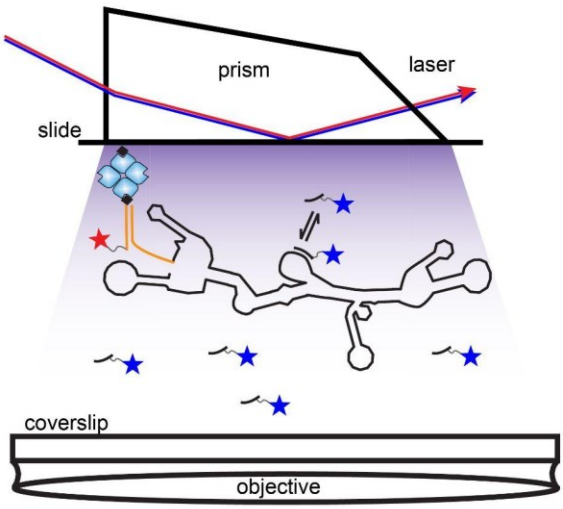


Figure 2.2. General setup of a SiM-KARTS experiment.

Similar to smFRET, HMM is useful to idealize the information embedded within SiM-KARTS traces. However, instead of idealizing the FRET efficiency trace, the fluorescence trace from the probe channel is idealized using HMM, and the model will always consist of two states,

bound and unbound. TODPs can be used to extract the dwell times for these two states, which can then be further analyzed. A convenient way to view the binding kinetics of SiM-KARTS probes is plotting the cumulative distributions of dwell times in each state. This produces a curve showing the frequency that a bound or unbound dwell is equal to or shorter than a given time. These cumulative distributions can be described by their representative dwell times obtained from single- or double-exponential fits, as shown in Equations 2.6 and 2.7.

$$y = a(1 - e^{-kt}) \qquad \text{Equation 2.6}$$

$$y = a_1(1 - e^{-k_1 t}) + a_2(1 - e^{-k_2 t}) \qquad \text{Equation 2.7}$$

To analyze if an RNA is undergoing conformational changes, and determine the time scale of those changes, the Fano factor for dwell times can be calculated. Simply, the Fano factor is a measure of dispersion for a counting process and is the variance (σ_t^2) to mean (μ_t) ratio (Equation 2.8).

$$F(t) = \frac{\sigma_t^2}{\mu_t} \qquad \text{Equation 2.8}$$

For a Poisson distribution, which is expected for a consistent, random binding process, the Fano factor is expected to be 1.[45] Variation away from 1 signifies nonrandom behavior, which, in the context of SiM-KARTS, shows a change in binding accessibility, presumably from a conformational change. The counting window over which the Fano factor strays away from ideal is indicative of the time scale of the molecule's dynamics.

For a closer view of the molecule's dynamics, a burst analysis based on spike train analysis from neuroscience is useful.[46] Specifically, the "rank surprise" statistic is useful for the type of data generated by SiM-KARTS. Instead of assuming that the probability of observing a number of spikes in an interval of certain length is a Poisson distribution, the rank surprise method orders interspike intervals by length to assign each a "rank", with 1 being the rank of the

shortest interspike interval.[47] Using the appropriate algorithm, the distance between "spikes" is used to group binding events into "burst" and "nonburst" time frames, where "bursts" are periods of time where the RNA is more accessible for probe binding, and "nonbursts" are periods of time where the RNA is less accessible for probe binding.[46] These groupings can then be analyzed through the fitting of separate cumulative distributions, with the representative dwell times for each conformation giving information on the structures being sampled.

In initial studies of probe binding behavior, spectroscopic techniques such as thermal denaturation monitored by absorbance and global conformational changes revealed by circular dichroism proved useful to tie together observations made through SiM-KARTS. While these techniques are not likely to reveal much about the structure of larger RNAs, they can be useful for experimental optimization and design. Thermal denaturation experiments report on the binding stability of an oligonucleotide to its target, such as a SiM-KARTS probe binding to the RNA of interest. This is particularly useful for SiM-KARTS as probe binding must be sufficiently strong to facilitate high quality data collection, but weak enough to be transient and minimize structural perturbations. Due to the hyperchromicity of nucleic acids, an increase in absorbance can be observed as a duplex dissociates into single strands.[48] The melting temperature of the complex, T_M, is the temperature where the duplex and single-stranded forms are equimolar.

Due to the complex nature of lncRNAs, methods for preliminary structural determination are useful to guide experimental design of SiM-KARTS. The larger the RNA, the more potential probing sites exist. However, many of these are likely double-stranded and fairly static in nature, so probing these regions with SiM-KARTS would not reveal any information beyond those determinations that could be made more easily by more traditional biochemical methods.

Chemical probing methods such as 2' selective hydroxyl acylation analyzed by primer extension (SHAPE) can be used to gain an understanding of the average secondary structure of a lncRNA. SHAPE reports on the flexibility of each nucleotide position of the RNA, which can be used to guide secondary structure prediction. Specific electrophilic reagents will react with a flexible 2' hydroxyl of RNA to form a 2'-*O*-adduct (Figure 2.3).[49] The reactivity of this functional group relies heavily on the conformation of the nucleotide where the reaction is occurring. Specific conformations enhance the reactivity of SHAPE reagents to the 2' hydroxyl, which are sampled most frequently by flexible nucleotides.[49] The adducts formed in the acylation reaction cause termination of reverse transcription, leading to fragments of complementary DNA with lengths corresponding to each flexible nucleotide. The locations of flexible nucleotides can then be used to guide further secondary structure predictions to obtain a more informed view of the average structure of an RNA.

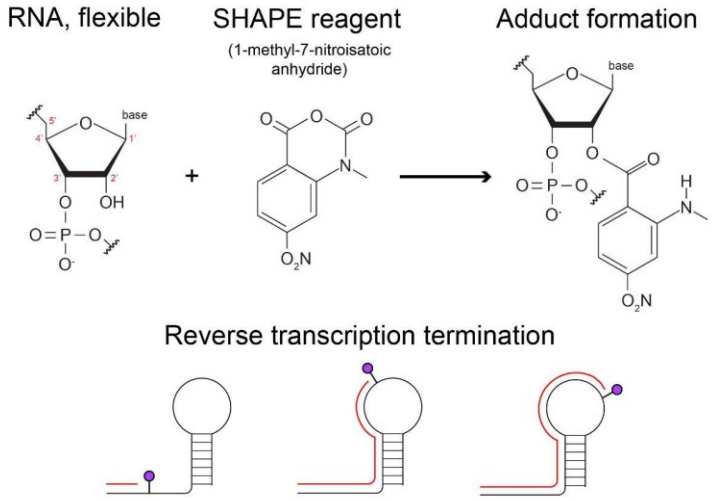


Figure 2.3. SHAPE adduct formation and reverse transcription termination. A SHAPE reagent reacts with RNA at a flexible nucleotide to form an adduct (top). Adducts along the RNA lead to reverse transcription termination and the production of DNA fragments corresponding to the flexible nucleotides (bottom).

5. Bridge to Chapter III

Studying non-coding RNA structure using single-molecule fluorescence microscopy can give a great depth of information about structure and dynamics, particularly when complementary methods such as steady state fluorescence spectroscopy or biochemical assays are considered in tandem. In Chapter III, I present an investigation of SiM-KARTS which will inform future studies of lncRNA structure and dynamics.

CHAPTER III: ALTERNATIVE PROBE CHEMISTRIES FOR SINGLE-MOLECULE ANALYSIS OF LONG NON-CODING RNA

1. Overview

This chapter contains previously published material, authored by Kalika R. Pai, Aimee Martin, Madison Kadrmas, and Julia R. Widom and published in *Journal of Biological Chemistry*. It describes studies done with the aim of adapting SiM-KARTS to study lncRNAs by studying the impact of alternative probe chemistries. I designed and performed all experiments, with contributions to single-molecule data collection and analysis by Aimee Martin and Madison Kadrmas. Data interpretation and writing was done in collaboration with Julia R. Widom.

2. Introduction

Long non-coding RNAs (lncRNAs) are RNAs that do not encode protein and are greater than 200 nucleotides in length. lncRNAs perform a variety of biological functions through interactions with DNA, RNA, or proteins. Specifically, many lncRNAs have regulatory functions, either epigenetic,[50] transcriptional,[51] or post-transcriptional.[52] Structure is an important factor in mediating the interactions that allow for these regulatory functions.[53] Additionally, many lncRNAs have been found to play roles in human disease, making the molecular details of lncRNAs interesting for the potential development of therapeutics.[17,54,55] However, there have not been in depth studies of the structure or dynamics of most lncRNAs,[56] which are difficult to study due to their large size and resulting structural complexity.

Single-molecule methods have strong potential for investigating the structure and dynamics of lncRNAs. Specifically, single-molecule microscopy avoids ensemble averaging and

allows for the investigation of the dynamics of heterogeneous systems on time scales relevant to large structural rearrangements.[57] However, common techniques, such as single-molecule Förster resonance energy transfer (smFRET),[8] are difficult to adapt to lncRNAs due to their length and structural complexity. For example, performing site-specific labeling of the RNA of interest can be costly, require complex experimental workflows, and impact the structures being studied.

An alternative to smFRET is single-molecule analysis of RNA transient structure (SiM-KARTS), which utilizes a short, fluorescently-labeled oligonucleotide probe that is complementary to a region of interest on the RNA.[44,46,58] By tracking binding and dissociation of this probe, the accessibility of the region of interest can be revealed. For example, frequent observation of probe binding indicates high accessibility, implying a single-stranded local structure. SiM-KARTS can also be used to investigate conformational changes of RNA by evaluating variations in probe binding frequency. Using spike train analysis,[47] "bursts" of probe binding can be distinguished from random events, uncovering transitions between conformations in which the target sequence is more or less accessible for probe binding. So far, SiM-KARTS has primarily been utilized to study riboswitches. For example, SiM-KARTS was performed on mRNA containing the $preQ_1$ riboswitch using a DNA probe complementary to the Shine-Dalgarno (SD) sequence (termed an "anti-SD" probe) to monitor ligand-induced pseudoknot docking.[44] It was determined that there are dynamic fluctuations in the accessibility of the target sequence, even at saturating ligand concentrations. In the case of the $preQ_1$ riboswitch, techniques such as small angle x-ray scattering (SAXS), smFRET, and molecular dynamics simulations aided in experimental design.[59,60]

SiM-KARTS shows promise as a method for probing lncRNA as it removes the requirement for site-specific labeling. The RNA can instead be transcribed and then end-labeled

using established chemistry performed under native conditions,[61,62] or a labeled capture oligonucleotide can be utilized to localize immobilized molecules without any modification to the target RNA. All previous reports of SiM-KARTS utilized an anti-SD probe, exploiting the presence of this motif in multiple riboswitches.[63–65] For lncRNA, no such common motif exists, and multiple structural domains may be of interest. For this reason, to adapt SiM-KARTS for applications to lncRNA, designing an appropriate probe is of high importance. Challenges in probe design include ensuring binding is strong enough to yield clear optical signals, yet transient enough to pose minimal disruption to the target RNA's dynamics. Lengthening the probe can increase binding stability, but also limits the precision with which specific regions of interest can be probed and increases the potential for the probe to disrupt the structure of the RNA. Mono- and divalent cations are well known to increase melting temperatures of DNA and RNA duplexes and hybrids[66] making adjusting cation concentrations a possible avenue for optimizing probe binding stability. However, as ionic conditions also have a large impact on stability of RNA secondary and tertiary structure,[67] controlling probe binding stability in this manner could distort the structure of the RNA being studied.

One solution to fine-tune probe binding strength without changing the length, location, or structure of a target site is to consider alternative probe chemistries. Specifically, the incorporation of locked nucleic acid (LNA) residues into DNA is known to increase the melting temperatures of its hybrids with RNA and DNA.[68] Incorporating one or two LNA residues into a short oligonucleotide can raise its predicted melting temperature by multiple degrees celsius, depending on placement within the strand.[69] The binding strength of short sequences can also be increased by the use of morpholino oligonucleotides, which are typically used as gene knockdown reagents due to their high binding specificity and stability.[70] In this study, we

investigated the impact of backbone chemistry on the binding behavior of SiM-KARTS probes in different structural contexts and under different ionic conditions using a model RNA based on the lncRNA *Braveheart* (Bvht; Fig. 3.1A).[16,18,19] Bvht is a heart-specific lncRNA that was identified to be a regulator of cardiovascular commitment, functioning in the same regulatory pathway as the transcription factor *MesP1* during embryonic stem cell differentiation in mice.[16] The structure of Bvht has been studied using chemical probing and SAXS, which revealed important features such as a 5' asymmetric G-rich internal loop that interacts with a zinc-finger transcription factor called cellular nucleic acid binding protein.[18] We show that probes with LNA residues incorporated demonstrate improved binding stability and structural sensitivity over DNA alone, while morpholino probes showed increased binding stability, but little structural sensitivity. We report a clustering approach in which the binding and unbinding rates observed within an individual trace are used as a kinetic fingerprint and show that these fingerprints effectively distinguish traces arising from different target RNA structures. Finally, we demonstrate that changes in the ionic environment can produce non-monotonic changes in probe binding kinetics, likely due to competing effects on the structure of the target RNA and the stability of the probe:RNA hybrid. These results will guide probe design for future studies utilizing SiM-KARTS, including investigations of lncRNA structure and dynamics.

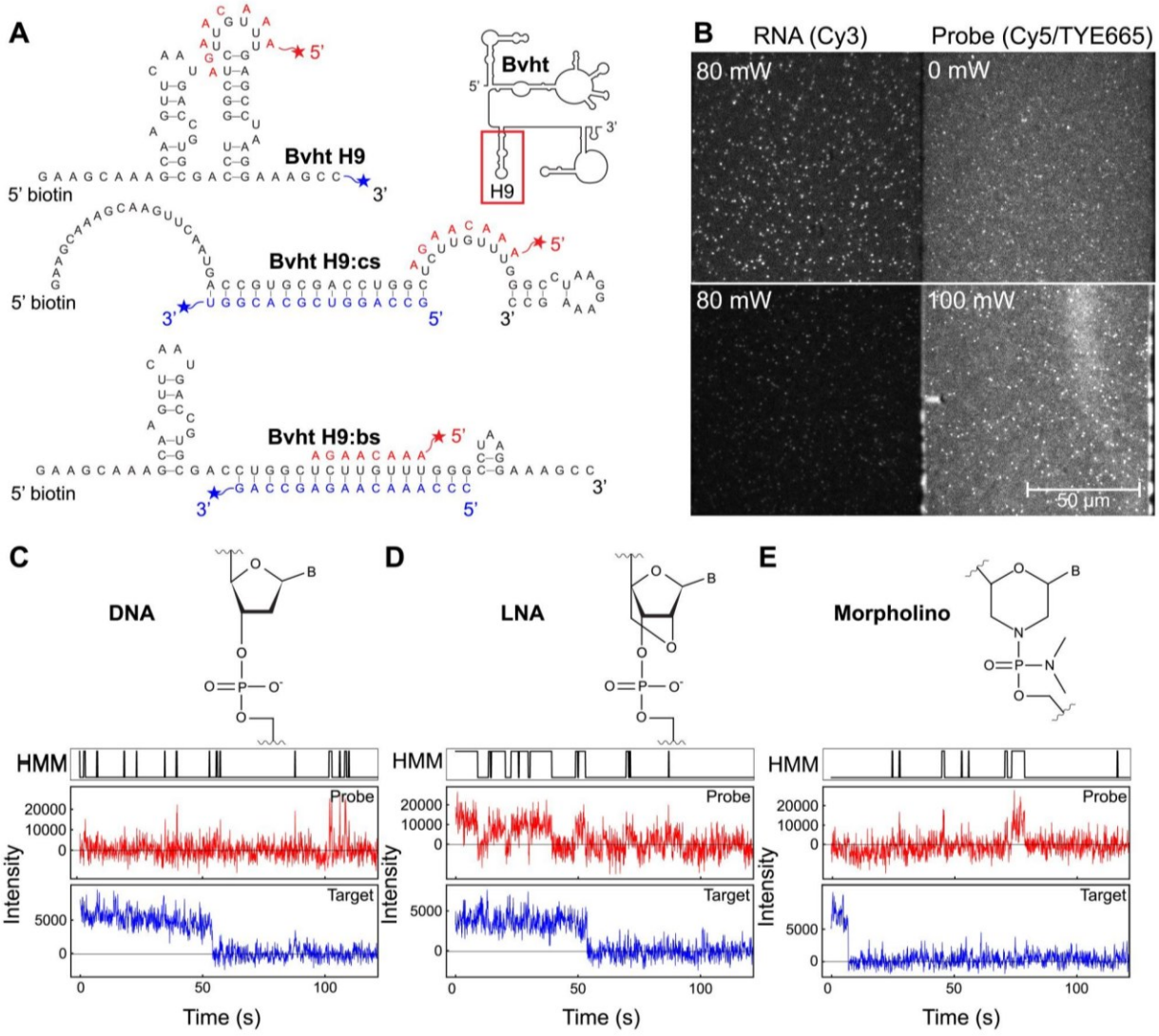
3. Results

A. Experimental Design

A 59-nucleotide sequence containing helix 9 of Bvht (hereafter called "Bvht H9") was selected from the full-length RNA as a model that allowed different structures to be obtained by hybridizing complementary RNA oligonucleotides (Fig. 3.1A; sequences in Table 3.1). Bvht was

chosen as an RNA of interest because, in addition to its biological significance, its structure has been well described by chemical probing and SAXS, providing the basis for design of a model RNA for our studies.[18,19] The native structure of Bvht, as previously determined by chemical probing, contains a flexible, single-stranded region to the 5' side of H9.[18] Using RNAstructure, [71] we found that outside the context of the full-length RNA, Bvht H9 is predicted to fold into a structure containing 2 hairpin loops. An 8-nucleotide segment of one loop was chosen as the probe binding site (Fig. 3.1A). The sequence of this loop is unique within the entirety of Bvht, and therefore is a feasible probing site for studies using the full lncRNA. The complementary strand "cs" was designed to generate a 9-nucleotide single-stranded region around the probing site upon hybridization to Bvht H9, and the blocking strand "bs" was designed to sequester the probing region in double-stranded form (Fig. 3.1A).

Figure 3.1. (next page) *A*, Model target RNA structures for SiM-KARTS with probe sequence in red adjacent to its binding site. Cy3-labeled complementary RNA strands (cs and bs) stably bound to the primary RNA (Bvht H9) are indicated in blue. Full structure of Bvht and location of H9 region are displayed in the upper right. *B*, Example image from SiM-KARTS data collection. RNA is detected in the left (green) channel via Cy3 fluorescence from either the RNA or a complementary strand (here, 3' Cy3 cs), while the probe is detected in the right (red) channel via Cy5 or TYE665 fluorescence (here, 5' TYE665 LNA2). Top: Portion of a field of view under only green laser illumination (source laser power of 80 mW), demonstrating fluorescence from immobilized RNA in the green channel. Spots are visible in the red channel due to spectral bleed-through from Cy3, and diffuse background in the red channel arises from a small amount of direct excitation of freely diffusing probe molecules by the green laser. Bottom: Remaining portion of the same field of view under simultaneous red (source laser power 100 mW) and green (source laser power 80 mW) illumination. Higher diffuse background in the red channel is due to more efficient excitation of freely diffusing probe molecules by the red laser. *C*, Backbone chemistry and example single-molecule trace for a DNA probe binding to Bvht H9:cs. Top: HMM idealization of red (probe) channel trace. Middle: Red (probe) channel trace; Bottom: Green (RNA) channel trace. *D*, Same as *C* for a DNA probe containing one LNA residue binding to Bvht H9:cs. *E*, Same as *C* for a morpholino probe binding to Bvht H9:cs.



Molecules were detected using wide-field prism-based total internal reflection fluorescence (TIRF) microscopy, with Bvht H9 detected in the green emission channel via a Cy3 label on its 3' end or the 3' end of cs or bs (ensuring that only hybridized molecules were detected in those cases), and 5' Cy5- or TYE665-labeled probe detected in the red channel (Fig. 3.1B). Probes were complementary to the 8-nucleotide target sequence and included one with a fully DNA backbone ("DNA"), one with an LNA residue at position 2 ("LNA2") in an otherwise DNA backbone, one with LNA residues at positions 2 and 3 in an otherwise DNA backbone

("LNA23") and one with a morpholino backbone ("MO") (Fig. 3.1C-E). The 8-nucleotide length and the number and location(s) of LNA residues were selected to achieve melting temperatures of the probe:RNA hybrid around or below room temperature. This is desirable for SiM-KARTS, as transient, reversible probe binding provides a kinetic signature of the structure of the target RNA without stably occluding the binding site and thereby impacting the structure of the target. Fluorescence intensity traces from the red (probe) channel were idealized using hidden Markov modeling (HMM) to define probe-bound and probe-unbound states (Fig. 3.1C-E). Ideal probe concentration was determined to be 60 nM by analyzing SiM-KARTS data collected with 20-100 nM LNA2 and comparing changes in bound and unbound dwell times (Fig. A1).

Table 3.1. Sequences of oligonucleotides. Probe binding sites are underlined in Bvht H9 and Bvht H9 34-45 sequences. "+X" indicates that nucleotide X is an LNA residue in an otherwise DNA oligonucleotide.

Oligo name	Oligo type	Sequence
Bvht H9	RNA	5'Bi GAAGCAAAGCAAGUUCAAUGACCGUGCGACC UG GCUCUUGUUUGGGCCUAAGGAAAGCC
cs	RNA	GCCAGGUCGCACGGU 3'Cy3
bs	RNA	CCCAAACAAGAGCCAG 3'Cy3
Bvht H9 34-45	RNA	GCUCUUGUUUGG
Bvht H7H8	RNA	5'Bi CAACAUGGGUUCUGGGAACUGAACCCAGGUC CUC UGGAAGGACAUCGAGUGCUC 3'Cy3
DNA	DNA	5'Cy5 AAACAAGA
LNA2	DNA w/ LNA	5'TYE665 A+AACAAGA
LNA23	DNA w/ LNA	5'TYE665 A+A+ACAAGA
MO	Morpholino	sulfo-Cy5 AAACAAGA

B. SiM-KARTS Probe Chemistry Impacts RNA Binding Properties

We first investigated the impact of probe chemistry on binding behavior to Bvht H9 (containing a partially occluded probe binding site) and Bvht H9:cs (containing a single-stranded probe binding site). The ideal probe will bind in a manner sensitive to structure, and have bound dwell times that are long enough to be readily distinguished from background noise, yet not so long that the binding impacts the structure of the target RNA. The rates of probe binding and dissociation were first quantified using cumulative distribution curves (Fig. 3.2; additional details in Fig. A2), which show the probability that a bound or unbound segment of a trace has a duration less than or equal to the corresponding time indicated on the x-axis. A slower rise from zero to 1 for unbound dwells indicates that longer intervals between binding events are typical. Similarly, for bound dwells, a slower rise from zero to 1 indicates that longer binding events are typical. Characteristic bound and unbound dwell times were quantified by fitting these curves with single- or double-exponential functions (Fig. A2G-J). Unbound dwell time distributions were well fit (R^2>0.97) with single-exponential functions in all cases examined, while bound dwell time distributions required double-exponential fits in many cases (Fig. A2I; Tables A1, A2), consistent with previous studies.[44] To facilitate direct comparison, bound dwells for all conditions were therefore modeled with double-exponential functions.

The *unbound* dwell times differ for each combination of target RNA structure (Bvht H9 or Bvht H9:cs) and probe chemistry (DNA, LNA2, LNA23 or MO) (Fig. 3.2A,C; Table 3.2). Unbound dwell times are longer in the presence of 1 mM Mg^{2+} for all probe types binding to Bvht H9:cs, while Mg^{2+} has little impact on binding to Bvht H9. The DNA probe generally has the longest unbound dwell times across all conditions, though the unbound dwell times for DNA and LNA2 probing Bvht H9 are very similar. LNA2 consistently binds Bvht H9:cs more quickly

than Bvht H9, showing strong discrimination between the two structures in both the absence and presence of 1 mM Mg^{2+}. LNA23 likewise exhibits faster binding to Bvht H9:cs than Bvht H9 in 0 mM Mg^{2+}, but has opposite behavior at 1 mM Mg^{2+}.

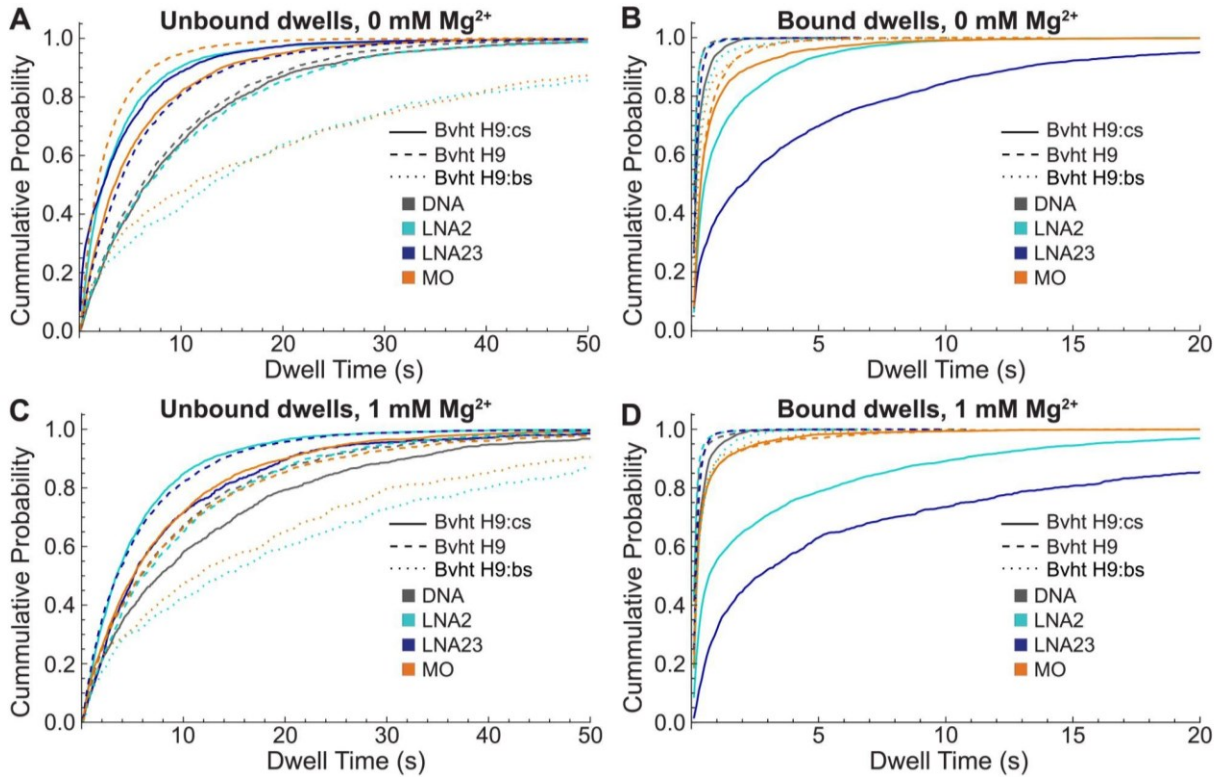


Figure 3.2. Impact of probe chemistry and Mg^{2+} on bound and unbound dwell times determined by HMM. *A*, Cumulative distributions of unbound dwell times obtained at 0 mM $MgCl_2$ for different probe types binding to Bvht H9 (dashed), Bvht H9:cs (solid) or Bvht H9:bs (dotted). Here and throughout, data with DNA probe are plotted in gray, LNA2 in variations of green and cyan, LNA23 in dark blue and MO in variations of orange and brown. *B*, Bound dwell times obtained at 0 mM $MgCl_2$ for different probe types binding to Bvht H9, Bvht H9:cs or Bvht H9:bs. *C*, Same as *A* but at 1 mM $MgCl_2$. *D*, Same as *B* but at 1 mM $MgCl_2$.

It is apparent that the increasing inclusion of LNA residues leads to longer *bound* dwell times, particularly with the target Bvht H9:cs (Fig. 3.2B,D; Table 3.2). There is also distinct separation between the bound dwell times of LNA probes for different RNA structures, with longer bound times observed for Bvht H9:cs, which has a more accessible binding site, than Bvht H9. Comparatively, the bound dwell times for the DNA and MO probes do not discriminate

between structures as effectively. Interestingly, the cumulative distributions of bound dwells for Bvht H9:cs are best described with a double-exponential fit, while bound dwells for Bvht H9 can be adequately described using a single exponential for all probes except MO. In these cases, the dominant component of a double-exponential fit (as reported for all bound dwells herein) has a weight of nearly 100% (Table A1). This distinction between target RNA structures is most evident with the LNA probes, where the heavier component is weighted at 99-100% for Bvht H9, and 51-62% for Bvht H9:cs (Table 3.2). The DNA probe exhibits the same trend but has a less drastic difference between structures, where the dominant component is weighted at 100% for Bvht H9 and the faster component is weighted at 81-88% for Bvht H9:cs.

SiM-KARTS is a high-background measurement due to the presence of freely diffusing fluorescent probe in the sample chamber (Fig. 3.1B). As a result, we performed extensive negative controls to ensure that the events identified by HMM were dominated by true probe binding rather than noise. Experiments with Bvht H9 hybridized to a blocking strand that completely occludes the binding site (Bvht H9:bs) showed that LNA2 and MO display significantly fewer binding events per trace (~6) and longer unbound dwell times (~20 s) than they do with Bvht H9 or Bvht H9:cs as the target (Fig. 3.2; Table 3.3), confirming that probe binding is strongly inhibited by competition for occupation of its complementary sequence. Further studies with LNA2 showed that in the absence of any target RNA, or in the presence of a target RNA lacking a complementary binding site (Bvht H7H8), the incidence of apparent probe binding events was lower still at only 1 or 2 bound dwells per trace, respectively, identified by HMM (example traces and quantification in Fig. A3). Similar numbers were observed for Bvht H9:cs with no probe in the sample chamber. Thus, a significant majority of bound dwells identified by HMM in cases where a probe and an accessible complementary sequence are both

available (Bvht H9, Bvht H9:cs) represent true probe binding events, as do many of the observed binding events to the occluded Bvht H9:bs binding site. Spurious fluctuations identified as binding events by HMM would have the effect of slightly decreasing the apparent bound and unbound dwell times relative to true probe binding behavior.

Table 3.2. Characteristic dwell times of probes binding to Bvht H9 or Bvht H9:cs. $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times in Fig. 3.2. τ_{bound} is presented as a weighted average determined from a double-exponential fit to the cumulative distribution of bound dwell times. Individual component parameters and R^2 values are reported in Tables S1-S2. The standard deviation σ, obtained from independently fitting three randomly selected subsets of the traces (bootstrapped sets), is noted in parentheses. "N" indicates the number of single-molecule traces included in each condition's analysis, "D/N" indicates the average number of bound dwells observed per trace, and the total number of bound dwells analyzed is the product of the two.

		Bvht H9				Bvht H9:cs			
Probe	$[Mg^{2+}]$ (mM)	$\tau_{unbound}$ (s) (σ)	$\tau_{bound,avg}$ (s) (σ)	N	D/N	$\tau_{unbound}$ (s) (σ)	$\tau_{bound,avg}$ (s) (σ)	N	D/N
DNA	0	9.0 (0.1)	0.146 (0.002)	118	18	9.5 (0.4)	0.24 (0.04)	93	16
DNA	1	8.9 (0.5)	0.170 (0.004)	97	16	12 (1)	0.28 (0.02)	101	12
LNA2	0	9.9 (0.2)	0.14 (0.01)	101	16	3.7 (0.4)	0.63 (0.09)	105	29
LNA2	1	9.4 (0.7)	0.15 (0.01)	109	17	5.2 (0.1)	0.70 (0.05)	166	18
LNA23	0	5.8 (0.2)	0.213 (0.004)	123	26	3.9 (0.1)	1.3 (0.4)	157	17
LNA23	1	5.4 (0.3)	0.222 (0.004)	102	28	8.0 (0.4)	2.4 (0.5)	108	8
MO	0	7.6 (0.6)	0.33 (0.02)	115	20	5.4 (0.5)	0.51 (0.03)	112	18
MO	1	9.0 (0.3)	0.30 (0.02)	117	15	7.6 (0.3)	0.31 (0.01)	126	20

Table 3.3. Characteristic dwell times for LNA2 and MO binding to Bvht H9:bs. $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times in Fig. 3.2. τ_{bound} is presented as a weighted average determined from a double-exponential fit to the cumulative distribution of bound dwell times. Individual component parameters and R^2 values are reported in Table A3. The standard deviation σ, obtained from independently fitting three bootstrapped sets, is noted in parentheses. "N" indicates the number of single-molecule traces included in each condition's analysis and "D/N" indicates the average number of bound dwells observed per trace.

		Bvht H9:bs			
Probe	$[Mg^{2+}]$ (mM)	$\tau_{unbound}$ (s) (σ)	$\tau_{bound,avg}$ (s) (σ)	N	D/N
LNA2	0	20 (2)	0.29 (0.02)	103	6
LNA2	1	22 (3)	0.33 (0.01)	102	5
MO	0	19 (2)	0.34 (0.01)	103	6
MO	1	18 (2)	0.29 (0.02)	100	7

We also considered whether Förster resonance energy transfer (FRET) could complicate the ability of HMM to identify binding events, given that the Cy3 label used to localize target RNA molecules is separated from the Cy5 or TYE665 label on a bound probe by less than 10 nm. We found that obvious evidence of FRET could be observed by visual inspection only in a small fraction of traces (~1%) as a decrease in Cy3 fluorescence during periods when the probe was bound. However, Cy3 was active (not photobleached) for only 10% of frames across Bvht H9 traces and 17% of frames across Bvht H9:cs traces, and there was no significant difference in the number of frames when a probe was determined to be bound while Cy3 was active (2.6% of frames and 39% of frames for LNA2 binding to Bvht H9 and Bvht H9:cs, respectively) compared to after photobleaching (2% and 39% of frames for Bvht H9 and Bvht H9:cs, respectively). This indicates that FRET does not impact our analysis of bound and unbound

dwell times. Potential impacts of FRET would be mitigated with longer target RNAs where the fluorophore used for localization can be placed farther from the probe binding site.

The precision of our SiM-KARTS measurements was further quantified by comparing dwell times extracted from movies collected on the same day, movies collected across different days (Fig. A4A-B; Table A4), and batches of traces (termed bootstrapped sets) selected randomly from within the combined data across all days (Fig. A4C-D; Table A5) (data on LNA2 binding to Bvht H9:cs). A small amount of variation in unbound and bound dwells is observed in data collected on the same day (Fig. A4A-B), with both quantities differing by 6-7% (standard deviation divided by mean) across movies on the day with higher variability (Table A4). Comparable variation of ~5% in unbound dwells was observed across bootstrapped sets (Table A5). In contrast, systematic variation of unbound dwell times is seen in the case of data collected on separate days, with a difference of 32% (difference divided by mean) between data pooled from movies on each day. This is likely due to inconsistencies in the concentration of binding-competent probe across freeze-thaw cycles, which could result from factors such as degradation (though this is unlikely for oligonucleotides with the backbones investigated here), pipetting uncertainty, adherence of the probe to walls of the tube, and gradual precipitation. In the results reported herein, all data for a given probe was collected on the same day to ensure accurate comparisons between target RNA structures and ionic conditions, except for MO, for which differences in unbound dwells should not be necessarily considered significant.

C. Melting Temperatures Confirm Increased Binding Stability of LNA and MO Probes

As an orthogonal metric of binding stability, thermal denaturation experiments were performed for each probe type bound to Bvht H9 34-45, a short RNA containing the probing sequence with 2-nucleotide overhangs on either end (Fig. 3.3). Melting curves were recorded in the absence or presence of 1 mM Mg^{2+} by monitoring absorbance at 260 nm, and were fitted (Fig. A5) to extract the melting temperature (T_M) and the width of the melting transition (ΔT) . As expected, the DNA probe has the lowest T_M, and the T_M values of LNA probes increase with the number of LNA residues incorporated (Table 3.4). The morpholino probe is similar in melting behavior to LNA2, but shows lower sensitivity to Mg^{2+}. The T_M obtained for the DNA probe represents an upper limit, as the slope of the melting curve indicates that the probe does not fully hybridize to the RNA even at 0 °C. The T_M values of probe hybrids with Bvht H9 34-45 follow the same trends as the bound dwells of probes binding to Bvht H9:cs observed by SiM-KARTS, with higher T_M (Table 3.4) values associated with longer bound dwell times (Table 3.2). Interestingly, the measured T_M values for LNA probes are significantly higher than their estimated values from IDT OligoAnalyzer.[72–75] However, the model used by that calculator is experimentally parameterized only for LNA:DNA hybrids, and approximates LNA nucleotides hybridizing with RNA as RNA nucleotides.[76] Notably, our results more closely approximate the predictions for LNA:DNA hybrids (Table 3.4). Additionally, no such melting temperature calculator exists for morpholinos, with most studies of morpholino-nucleic acid hybrids being limited to interactions with DNA only.[77,78] However, the MO:RNA hybrids that have been studied were shown to have higher T_M values than corresponding DNA:RNA hybrids by nearly 10 °C,[79] which is consistent with the MO:RNA melting temperatures measured here.

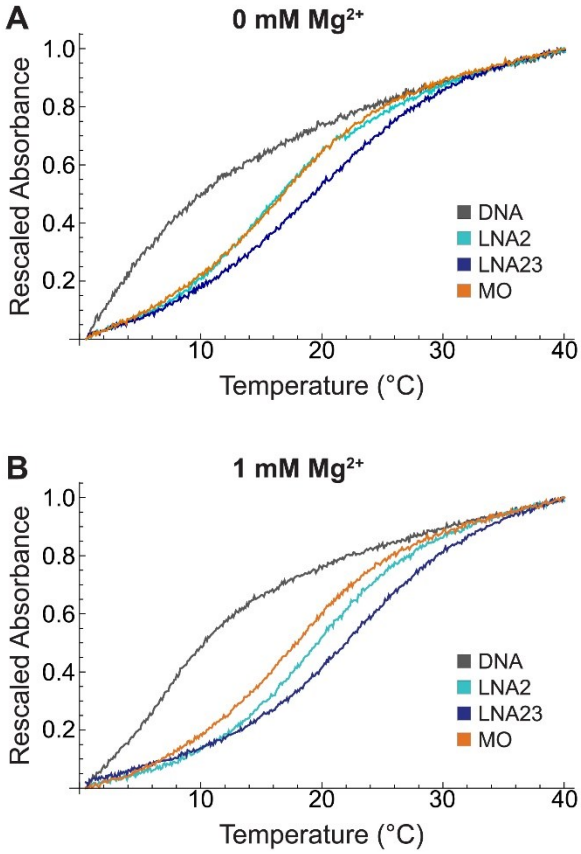


Figure 3.3. Melting curves for hybrids of each probe type and Bvht H9 34-45 measured by absorbance at 260 nm. For ease of comparison, curves are shifted to begin at an absorbance value of 0 and normalized to end at a value of 1. *A*, Melting curves for each probe type recorded at 0 mM $MgCl_2$. *B*, Melting curves for each probe type recorded at 1 mM $MgCl_2$.

Table 3.4. Melting temperature (T_M) and width of the melting transition (ΔT) values for each probe type hybridized to Bvht H9 34-45 in the absence or presence of 1 mM $MgCl_2$. Predicted T_M values were obtained using IDT OligoAnalyzer.

Probe	$[Mg^{2+}]$ (mM)	T_M from fit (°C)	ΔT from fit (°C)	T_M predicted (°C)
DNA	0	6.1[1]	2.5	4.1
DNA	1	7.2[1]	3.0	6.5
LNA2	0	16.9	3.5	5.4/18.8[2]
LNA2	1	19.4	3.5	7.8/21.5[2]
LNA23	0	20.2	4.1	9/22.4[2]
LNA23	1	20.9	3.9	11.5/21.5[2]
MO	0	17.6	3.4	-
MO	1	18.1	3.3	-

[1]Upper limit of T_M due to incomplete hybridization at 0 °C.
[2]RNA target calculation setting/DNA target calculation setting.

D. LNA Probes Distinguish Structures on a Trace-by-Trace Basis

The analyses of cumulative distribution functions described above demonstrated that all probe types exhibit distinguishable dwell time distributions when binding to pure BvhtH9 or pure BvhtH9:cs. However, it is particularly desirable for a probe to enable sorting of individual traces into structural categories within a heterogeneous mixture. To assess the ability of different probe types to distinguish traces arising from one target RNA structure from another, we developed a clustering approach in which each trace was assigned a kinetic signature characterized by the average unbound dwell time $\langle t_{unbound} \rangle$ and the average bound dwell time $\langle t_{bound} \rangle$ observed within it. We visualized the results using scatter plots with each trace represented by a point at its $(\langle t_{bound} \rangle, \langle t_{unbound} \rangle)$ coordinates (Fig. 3.4), then compared the $(\langle t_{bound} \rangle, \langle t_{unbound} \rangle)$ distributions observed with Bvht H9 or Bvht H9:cs as the target RNA for each

probe type in the absence and presence of 1 mM $MgCl_2$. All probe types yielded distributions for Bvht H9 and Bvht H9:cs whose medians were distinguishable to a high degree of statistical significance (Table 3.5), with p-values ranging from $7*10^{-6}$ for MO in 1 mM Mg^{2+} to $2*10^{-42}$ for LNA23 in 0 mM Mg^{2+} (multivariate Mann-Whitney test). However, the distributions for Bvht H9 and Bvht H9:cs overlap significantly in the cases of DNA and MO (Fig. 3.4A-B and G-H), suggesting that these probes would not enable traces obtained from a heterogeneous mixture to be assigned to specific structures.

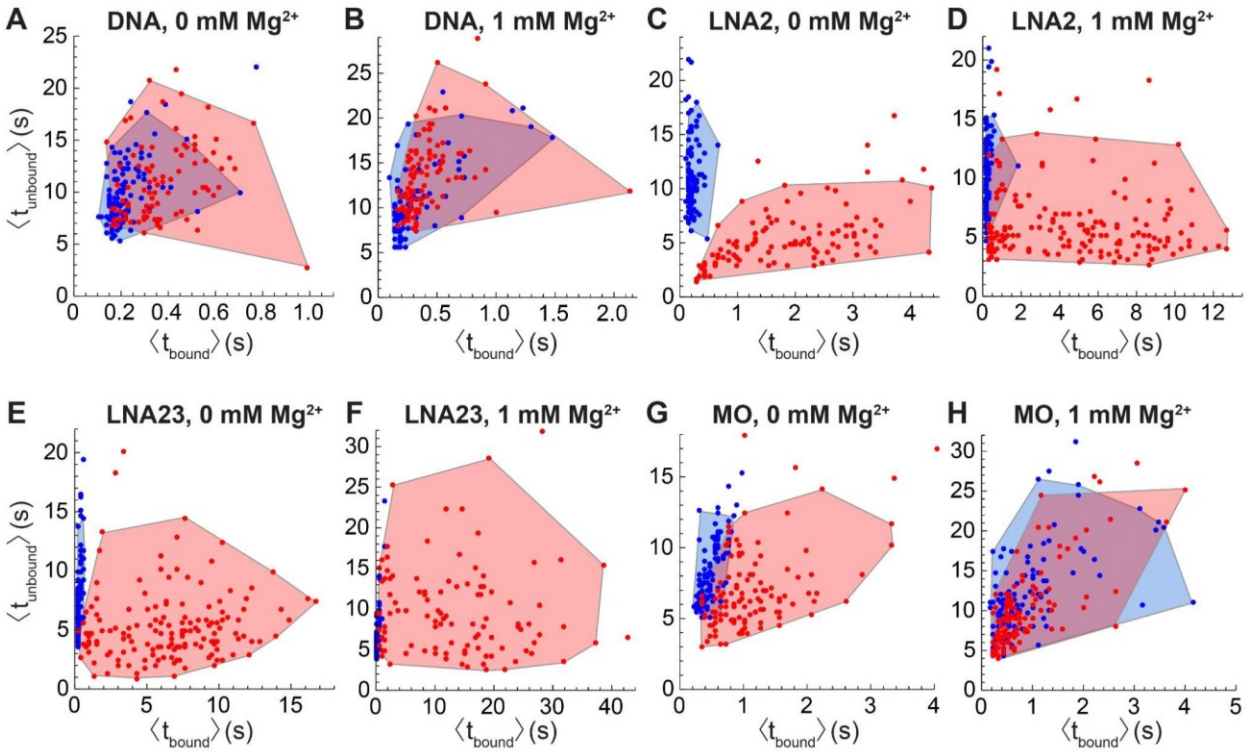


Figure 3.4. Clustering of traces according to their average bound and unbound dwell times. Each trace is represented by a point located at x-coordinate equal to the average duration of probe-bound segments within it, $\langle t_{bound} \rangle$, and y-coordinate equal to the average duration of unbound segments within it, $\langle t_{unbound} \rangle$. Shaded regions are convex hulls surrounding the 95% of points closest to the centroid, which is located at the median values of $\langle t_{bound} \rangle$ and $\langle t_{unbound} \rangle$. *A*, Scatter plot of $\langle t_{unbound} \rangle$ vs. $\langle t_{bound} \rangle$ and convex hulls for DNA probe binding to Bvht H9:cs (red) or Bvht H9 (blue) at 0 mM $MgCl_2$. *B*, Same as *A* but at 1 mM $MgCl_2$. *C*, LNA2 at 0 mM $MgCl_2$. *D*, LNA2 at 1 mM $MgCl_2$. *E*, LNA23 at 0 mM $MgCl_2$. *F*, LNA23 at 1 mM $MgCl_2$. *G*, MO at 0 mM $MgCl_2$. *H*, MO at 1 mM $MgCl_2$.

We quantified the degree of overlap by constructing a convex hull around a target RNA's ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) distribution, omitting the 5% of traces with the largest Euclidean distance from the centroid of the distribution (Fig. 3.4). The degree of overlap was defined as the fraction of one target RNA's traces whose coordinates fall within the boundary of the other target RNA's convex hull (by definition, 95% of a target RNA's traces fall within its own hull). For DNA and MO probes, we observed a high degree of overlap between the ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) distributions of Bvht H9 and Bvht H9:cs (Table 3.5). For example, for DNA in the absence of Mg^{2+}, 73% of Bvht H9 traces had ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) coordinates within the convex hull of Bvht H9:cs, and 73% of Bvht H9:cs traces fell within the hull of Bvht H9. In contrast, for LNA2 and LNA23 in the absence of Mg^{2+}, the percentage of traces falling within the convex hull of the other target RNA was 0% and 1%, respectively. Addition of 1 mM Mg^{2+} increased the overlap between Bvht H9 and Bvht H9:cs distributions for LNA2, LNA23 and MO, though LNA2 and LNA23 retained strong discrimination between the two (for example, only 10% of Bvht H9 traces fell within the convex hull of Bvht H9:cs for LNA2 at 1 mM Mg^{2+}). This suggests that probes containing LNA nucleotides could enable categorization of traces obtained from a heterogeneous mixture with a much higher degree of precision than DNA probes.

Table 3.5. Quantification of overlap between the ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) distributions for Bvht H9 and Bvht H9:cs.

Probe	[Mg^{2+}] (mM)	Bvht H9 centroid (s)[1]	Bvht H9:cs centroid (s)[1]	P-value[2]	% Bvht H9 traces within Bvht H9:cs hull[3]	% Bvht H9:cs traces within Bvht H9 hull[4]
DNA	0	(0.2, 9.2)	(0.4,10.4)	$3*10^{-15}$	73	73
DNA	1	(0.2, 9.9)	(0.4,13.5)	$2*10^{-8}$	87	60
LNA2	0	(0.2, 10.4)	(1.9,4.9)	$1*10^{-38}$	0	0
LNA2	1	(0.2, 10.3)	(4.5,5.2)	$4*10^{-37}$	10	6
LNA23	0	(0.3, 6.3)	(6.6,4.8)	$2*10^{-42}$	1	1
LNA23	1	(0.3, 5.7)	(13.0,8.7)	$1*10^{-32}$	2	15
MO	0	(0.5, 8.0)	(1.0,6.3)	$8*10^{-33}$	14	43
MO	1	(0.6, 10.9)	(0.6,9.0)	$7*10^{-6}$	87	81

[1](Median of $\langle t_{bound} \rangle$ across traces, median of $\langle t_{unbound} \rangle$ across traces).
[2]Probability that Bvht H9 and Bvht H9:cs distributions have the same centroid evaluated by the multivariate Mann-Whitney test.
[3]Percentage of Bvht H9 traces whose ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) coordinates fall within a convex hull surrounding the Bvht H9:cs ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) point cloud. The convex hull omits the 5% of Bvht H9:cs points furthest from the centroid.
[4]Same as [3] but with Bvht H9 and Bvht H9:cs reversed.

To test this idea, we divided the 95% of traces closest to the centroid into two pools (termed the "basis" pool and the "test" pool) and generated convex hulls using the traces in the basis pool. We then tested whether the ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) coordinates of each trace in the test pool localized within only the correct hull (e.g. Bvht H9 test traces within the Bvht H9 basis trace hull), only the incorrect hull (e.g. Bvht H9 test traces within the Bvht H9:cs basis trace hull), a region where the hulls overlapped, or outside of both hulls entirely. For LNA2 in the absence of Mg^{2+}, 85% of the test pool of Bvht H9 traces correctly localized only within the Bvht H9 basis

pool's hull, while none localized within the hull generated from the Bvht H9:cs basis pool (mean across 50 random allocations of traces into basis and test pools). The remaining 15% localized outside of both hulls, but this is expected due to the random allocation of outliers (which increase the footprint of the convex hull) into the basis and test pools, and could be mitigated by accumulating a larger basis pool. Conversely, 85% of Bvht H9:cs test pool traces correctly localized only within the Bvht H9:cs basis pool's hull and the remaining 15% localized outside of both. LNA23 exhibited almost identical behavior to LNA2 in its ability to discriminate between Bvht H9 and Bvht H9:cs traces. In contrast, for DNA in the absence of Mg^{2+}, only 29% of Bvht H9 test pool traces correctly localized only within the Bvht H9 hull, but 57% localized within both hulls due to the extensive overlap between them. Conversely, 27% of Bvht H9:cs traces correctly localized within only the Bvht H9:cs hull while 55% localized within both hulls. These results confirm that LNA probes offer the potential for trace-by-trace sorting into different structural categories, a unique capability among the probe types tested.

E. Divalent Cations Impact Probe Binding to Bvht H9:cs

The changes in T_M observed with and without 1 mM Mg^{2+} show a clear relationship between divalent cation concentration and probe:RNA hybrid stability. To investigate the impact of ionic conditions in SiM-KARTS, we chose to further study the binding of LNA2 and MO, as both showed promise as more stably-binding probes than DNA (Fig. 3.2; Table 3.2). For LNA2, the unbound dwell times for the more accessible Bvht H9:cs increase as Mg^{2+} concentration increases, with the less accessible Bvht H9 having consistent unbound dwell times regardless of Mg^{2+} concentration (Fig. 3.5A, Table 3.6). Unbound and bound dwell times display little dependence on Na^{+} concentration for either target RNA (Fig. 3.5C-D). Bound dwells for Bvht

H9 are short regardless of Mg^{2+} or Na^{+} concentration (Fig. 3.5B,D; Table 3.6). Bound dwells lengthen for Bvht H9:cs between 0 and 1 mM Mg^{2+}, but shorten substantially with the addition of 20 mM Mg^{2+}, which has previously been used for SiM-KARTS,[44] due to a significant increase in the weight of the faster component of the double-exponential fit (Table A7). This suggests that at high concentrations, Mg^{2+} is no longer just stabilizing the interaction between Bvht H9:cs and LNA2, but may be impacting the overall structure of Bvht H9:cs, affecting the accessibility of the probe binding site.

To further understand the impact of Mg^{2+} concentration on the nature of the binding site encountered by a probe, circular dichroism (CD) spectroscopy was performed on Bvht H9 34-45 and Bvht H9:cs at 0, 1, and 20 mM Mg^{2+}. Differences in the structure of Bvht H9 34-45 can be seen in the presence and absence of Mg^{2+}, but there are no appreciable differences between 1 mM and 20 mM Mg^{2+} (Fig. 3.6A). The lineshape of the CD spectrum of Bvht H9:cs remains consistent across Mg^{2+} concentrations, indicating that Mg^{2+} does not induce changes in its global architecture (Fig. 3.6B). However, the magnitude of the CD signal increases continuously with increasing Mg^{2+}, suggesting that this assembly adopts a more ordered structure in the presence of Mg^{2+}. This ordered structure may poise the target site in a configuration that is not amenable to probe binding, explaining the slower binding and more rapid dissociation observed at 20 mM Mg^{2+}.

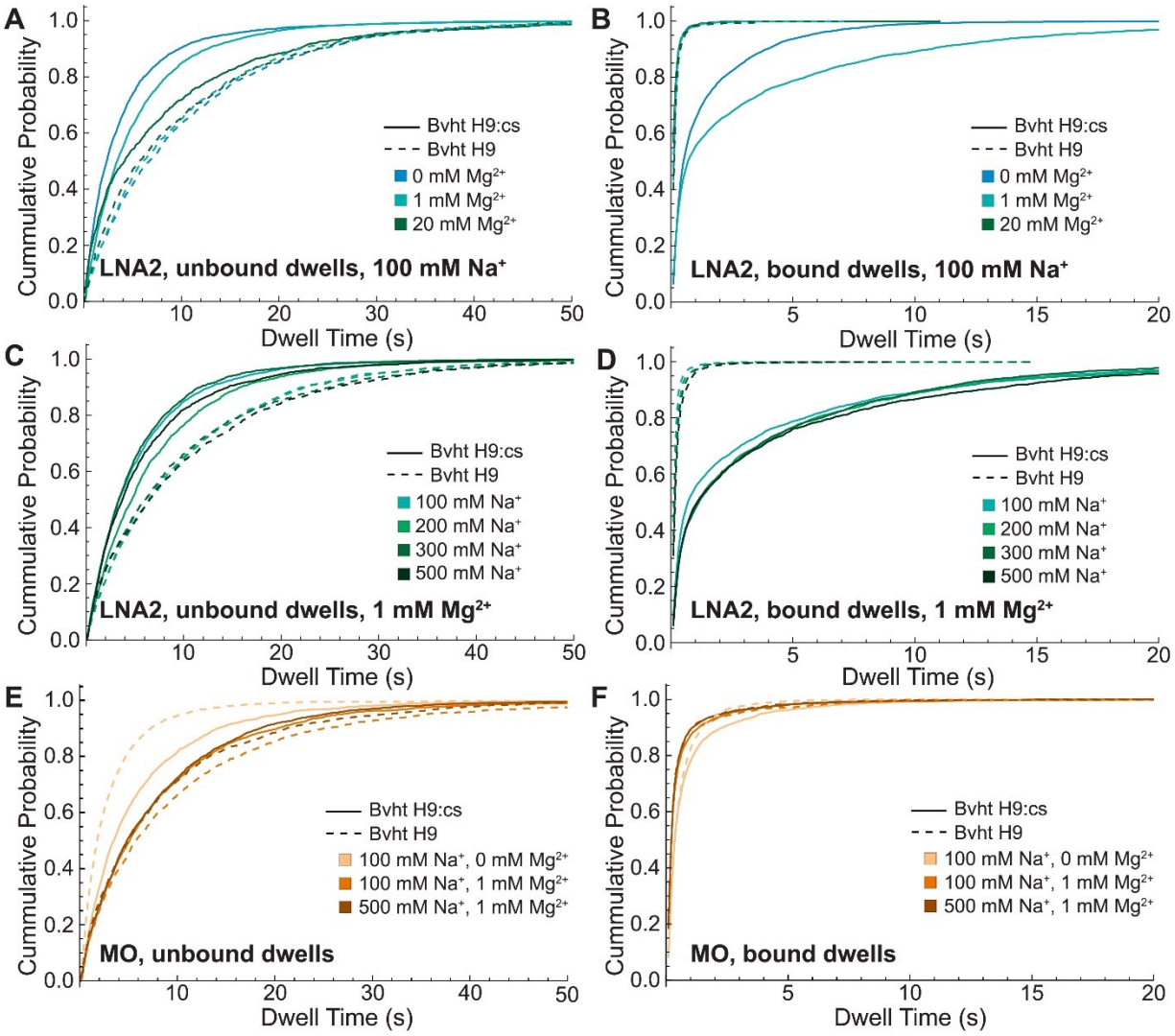


Figure 3.5. Impact of ionic concentration on dwell times of LNA2 or MO binding to Bvht H9 (dashed) or Bvht H9:cs (solid). *A*, Cumulative distributions of unbound dwell times for LNA2 at 100 mM NaCl and varying $MgCl_2$ concentration. *B*, Bound dwell times for LNA2 at 100 mM NaCl and varying $MgCl_2$ concentration. *C*, Unbound dwell times for LNA2 at 1 mM $MgCl_2$ and varying NaCl concentration. *D*, Bound dwell times for LNA2 at 1 mM $MgCl_2$ and varying NaCl concentration. *E*, Unbound dwell times for MO at indicated combinations of NaCl and $MgCl_2$ concentration. *F*, Bound dwell times for MO at indicated combinations of NaCl and $MgCl_2$ concentration.

Table 3.6. Characteristic dwell times of LNA2 or MO binding to Bvht H9 or Bvht H9:cs in indicated salt concentrations. $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times in Fig. 3.5. τ_{bound} is presented as a weighted average determined from a double-exponential fit to the cumulative distribution of bound dwell times. Individual component parameters and R^2 values are reported in Table A6-A9. The standard deviation σ, obtained from independently fitting three bootstrapped sets, is noted in parentheses. "N" indicates the number of single-molecule traces included in each condition's analysis and "D/N" indicates the average number of bound dwells observed per trace.

			Bvht H9				Bvht H9:cs			
Probe	$[Mg^{2+}]$ (mM)	$[Na^{+}]$ (mM)	$\tau_{unbound}$ (s) (σ)	$\tau_{bound,avg}$ (s) (σ)	N	D/N	$\tau_{unbound}$ (s) (σ)	$\tau_{bound,avg}$ (s) (σ)	N	D/N
LNA2	0[1]	100	9.9 (0.2)	0.14 (0.01)	101	16	3.7 (0.4)	0.63 (0.09)	105	29
	1[1]	100	9.4 (0.7)	0.15 (0.01)	109	17	5.2 (0.1)	0.70 (0.05)	166	18
	1	200	9.6 (0.2)	0.19 (0.01)	93	16	6.8 (0.5)	0.9 (0.1)	157	15
	1	300	9.3 (0.3)	0.21 (0.01)	105	16	5.0 (0.2)	0.89 (0.01)	146	18
	1	500	10.0 (0.7)	0.22 (0.03)	103	15	5.7 (0.6)	1.0 (0.4)	135	15
	20	100	9.1 (0.6)	0.169 (0.004)	89	16	7.4 (0.3)	0.11 (0.02)	62	19
MO	0[1]	100	7.6 (0.6)	0.33 (0.02)	115	20	5.4 (0.5)	0.51 (0.03)	112	24
	1[1]	100	9.0 (0.3)	0.30 (0.02)	117	15	7.6 (0.3)	0.31 (0.01)	126	18
	1	500	7.7 (0.4)	0.30 (0.02)	113	18	7.5 (0.7)	0.301 (0.004)	135	20

[1]Data in these rows reproduced from Table 3.2 for ease of comparison.

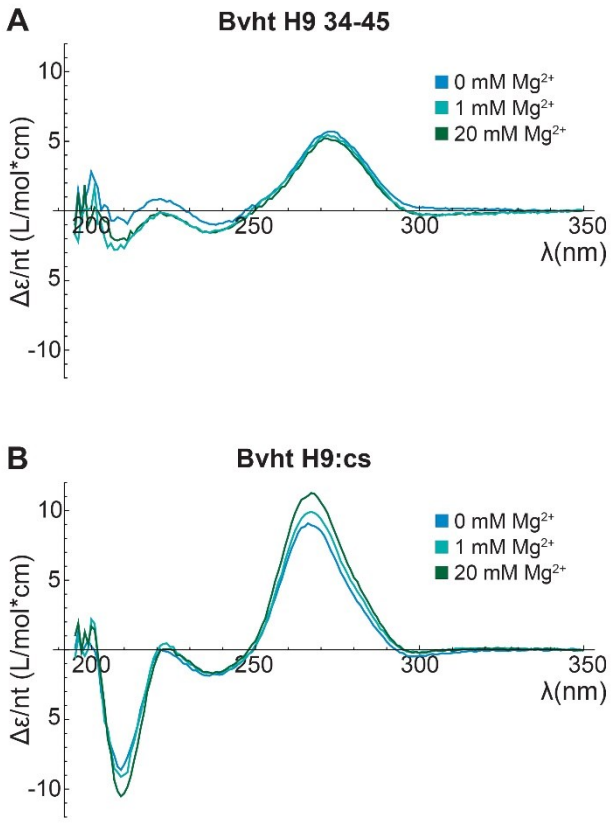


Figure 3.6. CD spectroscopy of RNA at varying $MgCl_2$ concentration. *A*, CD (expressed as Δε per nucleotide) of Bvht H9 34-45 at indicated $MgCl_2$ concentrations. *B*, CD of Bvht H9:cs at indicated $MgCl_2$ concentrations.

The binding of the morpholino probe was also of particular interest, as its neutral backbone could limit the effect that ions have on probe binding. Here, some variation is seen in the length of unbound dwells, without a clear trend between dwell time and structure probed or salt concentration (Fig. 3.5E-F; Table 3.6). The bound dwells are of consistent length and the fit components are of consistent weight for MO binding to Bvht H9 and Bvht H9:cs under all ionic conditions tested. This suggests that the binding stability of the morpholino probe is indeed less sensitive to ionic conditions, but, as noted earlier, is also less sensitive to the structure of the target RNA.

4. Discussion

RNA studies using SiM-KARTS have been limited to date, in part due to the optimization necessary to yield clear, interpretable signals from probe binding events. In this work, we found that probes with backbone chemistries beyond DNA can successfully report on RNA structure via SiM-KARTS. Specifically, we found that the incorporation of LNA residues in a DNA oligonucleotide leads to clearer discrimination between target RNA structures in which the probe binding sites exhibit different levels of accessibility. For LNA probes, bound dwell times were notably longer when probing a region that is fully single-stranded and accessible (as in Bvht H9:cs), than they were when probing a partially occluded region (as in Bvht H9). Although the bound dwell times of LNA probes on partially occluded regions were comparable to those observed for LNA2 on the fully occluded Bvht H9:bs complex, the latter was distinguished by much longer unbound dwell times. SiM-KARTS measurements utilizing LNA probes can therefore effectively discriminate among fully accessible, partially occluded and fully occluded binding sites on our model RNA. The optimal number and location of LNA residues is likely to depend on probe length and sequence. We show that with currently available models, melting temperatures predicted with a DNA target setting provide a starting point that can be used to guide design of LNA probes.

Previously, much of the interpretation of SiM-KARTS data was based on *unbound* dwell times being longer for structures with less accessible probing regions, compared to more accessible structures.[44,46,58] Interestingly, we found that this is not necessarily the case under all circumstances, with the dependence of unbound dwell times on the structure of the target RNA being inconsistent across different probe types. Here, the DNA probe demonstrated longer unbound dwells for the more accessible target RNA structure than the less accessible structure,

which is contrary to typical interpretation of data. However, LNA and MO probes generally exhibited longer unbound dwells with the less accessible target structure than the more accessible structure, as may be expected.

We have found that *bound* dwell times can also be used to discern information about the structure of the target RNA. There is a clear target structure-dependence of the bound dwell times when using LNA probes, with longer bound dwell times for the structure with a more accessible probe binding site in both the absence or presence of 1 mM Mg^{2+}. In the case of DNA, bound dwell times are 50-100% higher for the more accessible structure, but both are very short and are therefore difficult to distinguish from background noise. The morpholino probe, however, exhibits no appreciable difference between structures. We also found that additional information about target RNA structure can be gathered from the component weights of double-exponential fits to bound dwell time distributions. A single component is vastly dominant for DNA and LNA probes binding to more structured target RNA sites, such as Bvht H9 under all conditions and, to a lesser extent, Bvht H9:cs in 20 mM Mg^{2+}. In contrast, both components show significant weight for less structured target RNA such as Bvht H9:cs. While the exact mechanism underlying these differences is not known, it is clear that different target structures exhibit different fitting characteristics.

In addition to our analysis of the overall distribution of unbound and bound dwell times observed across many traces, we have found that it is possible, specifically when using LNA probes, to categorize the structures that give rise to individual traces. With DNA and MO probes, significant overlap exists in the behavior observed within individual traces with Bvht H9 and Bvht H9:cs as targets. LNA probes were found to show a higher degree of separation between distributions, allowing individual traces to be assigned to structural categories with higher

accuracy. This accuracy is achieved by constructing a kinetic fingerprint from a trace's bound *and* unbound dwell times, further highlighting the value of considering both quantities holistically.

Previous studies with SiM-KARTS have utilized Mg^{2+} concentrations above those generally considered physiological,[80] ranging from 5 to 20 mM,[44,58,64] to aid in stabilization of probe binding. We have found that for LNA probes, adjusting divalent cation concentration impacts both bound and unbound dwell times; however the impact on bound dwell times is not monotonic as Mg^{2+} concentration increases. Addition of Mg^{2+} increases the stability of the probe:RNA hybrid, as demonstrated by measured melting temperatures, but can also lead to structural changes in the target RNA, as observed by CD. While the exact nature of these conformational rearrangements is not known for Bvht H9, they likely play a role in the non-monotonic dependence of bound dwell times on Mg^{2+} concentration. Changes in monovalent salt concentration do not lead to significant changes in bound or unbound dwells.

Based on these results, we propose that to best optimize probe binding in SiM-KARTS experiments, it is advantageous to change the probe's chemistry rather than tune binding using environmental variables, such as salt conditions, or by increasing the probe's length, which may lead to reduced precision for targeting specific regions of interest. Additionally, it is worthwhile to consider both bound and unbound dwell times as well as their exponential fitting characteristics, to gain maximal information about the structure of the binding site and discriminate between similar structures. The probe design and data analysis approaches demonstrated here using a small, controllable model system will facilitate the application of SiM-KARTS to larger RNA targets of unknown structure.

5. Experimental Procedures

A. Oligonucleotides

To aid with oligonucleotide design, secondary structures of Bvht H9 alone and with various complementary strand bound were predicted using RNAstructure.[71] Bvht H9, Bvht H7H8, and the complementary strands cs and bs (Table 3.1) were purchased from Dharmacon, with 3' Cy3 labeling for cs and bs. Bvht H9 was purchased with 5' biotin, and was 3' end-labeled with Cy3 hydrazide (Lumiprobe) following established protocols[61] for experiments performed on this RNA without any complementary strand. Bvht H7H8, which has no more than three consecutive bases complementary to the probes used in this study throughout its sequence, was purchased with 5' biotin and 3' Cy3 labeling. Probes were designed to be complementary to nucleotides 36-43 of Bvht H9, and the 8-nucleotide length was chosen to achieve a low melting temperature, ensuring transient binding. The DNA probe was purchased from IDT with 5' Cy5 labeling and DNA/LNA chimera probes were purchased from Qiagen with 5' TYE665 labeling. TYE665 is a proprietary Cy5 substitute available from Qiagen, which does not offer Cy5 as an LNA modification. Per manufacturer recommendations for chemically modified oligonucleotides of these lengths, all of the above oligonucleotides were HPLC-purified, desalted and, for RNA, deprotected by the manufacturer. The morpholino probe was purchased from GeneTools with 5' sulfo-Cy5 labeling and was purified by filtration and precipitation by the manufacturer, per their recommendation for oligonucleotides of this type. For melting curves and CD, a truncated version of Bvht H9 was purchased from IDT, containing only the probing region with 2-nucleotide overhangs (Bvht H9 34-45).

B. RNA Preparation for SiM-KARTS

For single-molecule experiments, 1.5 μM Bvht H9 or Bvht H7H8 was heated to 90 °C in a buffer consisting of 50 mM Tris-HCl and 100 mM NaCl at pH 7.5 for 2 minutes. The RNA was then slowly cooled in air for 10 minutes before being placed on ice. For Bvht H9:cs or Bvht H9:bs, Bvht H9 was prepared at 1 μM in the presence of equimolar cs or bs then annealed as above. A 10 nM dilution of the RNA was prepared shortly before use in experiments.

C. SiM-KARTS Data Collection

Microscope slides were cleaned and passivated using the established dichlorodimethylsilane (DDS)-Tween20 method.[81] Briefly, quartz slides were cleaned and coated with DDS (Sigma Aldrich), then assembled into flow cells, which were stored at 4 °C for up to 3 weeks. Just prior to imaging, passivation was completed by incubating 0.2 mg/mL biotinylated bovine serum albumin (ThermoFisher) for 5 minutes, 0.2% v/v Tween-20 (Sigma Aldrich) for 10 minutes, and 0.2 mg/mL streptavidin (ThermoFisher) for 5 minutes, each in a buffer consisting of 10 mM Tris-HCl and 50 mM NaCl at pH 8. RNA was flowed on at a concentration of 10 pM in a buffer consisting of 50 mM Tris-HCl and 100 mM NaCl at pH 7.5 and allowed to incubate for 5-10 minutes until an adequate density of molecules, approximately 0.04 spots per μm^2, was achieved. Probes were flowed on in a buffer consisting of 50 mM Tris-HCl pH 7.5 and variable salt concentrations as specified in figures (100-500 mM NaCl and 0-20 mM $MgCl_2$), supplemented with 5 nM 3,4-dihydroxybenzoic acid (Sigma Aldrich), 5 mM Trolox (Sigma Aldrich), and 50 nM Protocatechuate 3,4-dioxygenase (MP Biomedicals) as an oxygen scavenging system. Probes were present at 60 nM unless otherwise noted. This mixture was incubated in the sample chamber for 10 minutes prior to data acquisition. Data were

collected at room temperature on a home-built prism-type total internal reflection fluorescence microscope (Leica DMi8 inverted microscope, 63x oil immersion objective). Fluorophores were excited continuously during data acquisition using a 532 nm laser (Coherent OBIS 532 LS) at 80 mW and a 637 nm laser (Coherent OBIS 637 LX) at 100 mW. Emission was split into red and green channels (Chroma T635lpxr longpass dichroic beamsplitter) and cleanup filters were placed in both channels (Chroma ET655lp in the red channel, Chroma ET585/65m in the green channel, and ZET635NF and ZET532NF in both channels). The channels were projected side-by-side onto an Andor iXon 888 EMCCD camera. Movies were 2,000 frames in duration and were recorded with an exposure time of 100 ms and an electron multiplication gain setting of 200.

D. SiM-KARTS Data Analysis

Coordinates of RNA molecules in the green channel were identified for each video file using the FIJI "find maxima" feature after rolling ball background subtraction. The tolerance for peak finding was defined for each movie by eye, taking into account the difference between signal and background. Custom MATLAB code was then used to extract fluorophore intensity traces for these locations and the corresponding locations in the red channel, which were identified by performing a calibration using fluorescent microspheres (FluoSpheres, 0.2 μm, red, carboxylate-modified microspheres, ThermoFisher) visible in both channels. Traces were corrected for spectral bleed-through of Cy3 emission into the red channel and were manually selected for further processing using custom MATLAB code if they exhibited single-step photobleaching of Cy3 and a flat noise floor in the red channel. For experiments with no immobilized RNA, traces were extracted from periodically distributed locations throughout the

field of view and ones with a flat noise floor in the red channel were selected for further analysis. Data from 2 to 3 movies taken on the same day were combined to yield the number of traces indicated by "N" in each data table. Selected traces were placed through a Wiener filter with a range of 1 to reduce background noise and then normalized to each have a noise floor of zero and a peak value of 2. Filtered traces underwent hidden Markov modeling in QuB Online software[43,82–85] using a 2-state model (probe bound and unbound) to obtain idealized traces. Starting parameter guesses for HMM were typically amplitudes of 0.2 and 1 with standard deviations of 0.2 and 0.6 for the unbound and bound states, respectively, and rates of k_{on}=0.3 s^{-1} and k_{off}=1 s^{-1}. The algorithm was typically allowed to re-estimate all of these parameters except for the standard deviation of the signal in the unbound state. Starting guesses and parameters allowed to float were altered in cases of obvious under- or over-fitting.

Transition rates were analyzed by obtaining the durations of dwells in the bound and unbound states, omitting segments that were truncated by the beginning or end of the trace. The cumulative distributions of each were fit with single- or double-exponential functions in Mathematica (Wolfram Alpha) (Equation 3.1 or Equation 3.2, respectively, where w_i is the weight of component i and τ_i is its time constant).

$$Fit(t) = w(1 - e^{-t/\tau}) \qquad \text{Equation 3.1}$$

$$Fit(t) = w_1(1 - e^{-t/\tau_1}) + w_2(1 - e^{-t/\tau_2}) \qquad \text{Equation 3.2}$$

To obtain uncertainties for dwell times, traces were randomly split into three pools (termed "bootstrapped sets"), and the bound and unbound dwell times observed within each set were independently assembled into cumulative distribution curves and fitted. Tables 3.2, 3.3, and 3.7 present the mean and standard deviation of the reported fitting parameters across the three bootstrapped sets.

Scatter plots were generated and analyzed in Mathematica. The durations of all bound segments of a trace were averaged together and the durations of all unbound segments were averaged together, omitting segments truncated by the beginning or end of the trace, to generate that trace's ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) coordinates. Scatter plots were constructed with each trace represented by a point at its ($\langle t_{bound} \rangle$,($\langle t_{unbound} \rangle$) coordinates. The coordinates of the centroid were identified as (median of $\langle t_{bound} \rangle$ across traces, median of $\langle t_{unbound} \rangle$ across traces). The medians of Bvht H9 and Bvht H9:cs ($\langle t_{bound} \rangle$,$\langle t_{unbound} \rangle$) distributions were compared using the multivariate Mann-Whitney test, which is appropriate for data with multiple dependent variables (here, $\langle t_{bound} \rangle$ and $\langle t_{unbound} \rangle$) that are not normally distributed. A convex hull was constructed around the points using the Mathematica built-in function "ConvexHullMesh", omitting the 5% of points with the largest Euclidian distance from the centroid. Each trace's coordinates were tested for whether they fell inside a dataset's convex hull using the Mathematica built-in function "RegionMember". To quantify the accuracy with which traces could be categorized, the 95% of traces closest to the centroid were randomly assigned to the basis pool or test pool (half in each pool), a convex hull was constructed around the coordinates of the basis pool, and the coordinates for traces in the test pool were queried using "RegionMember" to determine which hulls they localized within. The reported results are the average over 50 independent randomizations.

E. Melting Curves

Melting curves were recorded on a Jasco J-1500 CD spectrometer in a low headspace, 1 cm pathlength cuvette (Starna 26.100/LHS). 3 μM each of probe and Bvht H9 34-45 were combined in 20 mM sodium phosphate (pH 7.4) supplemented with NaCl to attain a total Na^+

concentration of 100 mM and $MgCl_2$ at 0 or 1 mM, as specified in figures. While Tris buffer was used for SiM-KARTS experiments to match earlier studies,[44,46] phosphate buffer was preferred for melting curves and CD spectroscopy due to the milder temperature-dependence of its pH and its superior ultraviolet transparency. Samples were incubated at 0 °C for 10 minutes before being heated at 0.2 °C/min from 0 °C to 40 °C, with absorbance monitored at 260 nm. A buffer scan was subtracted, and the resulting curves were shifted to coincide with zero at low temperature and normalized to end at 1 for display. T_M values were obtained by fitting the curves in Mathematica based on a previously described procedure.[86] First, the high-temperature segment of the curve with constant slope was fitted with a line, which was extrapolated to lower temperature. The fraction of strands that were hybridized was calculated using Equation 3.3:

$$\theta(T) = \frac{L(T) - A(T)}{L(T) - A(T_{min})} \qquad \text{Equation 3.3}$$

where L(T) is the value of the linear fit at temperature T, A(T) is the absorbance at temperature T, and T_{min} is the minimum temperature measured. θ(T) was then fitted using Equation 3.4:

$$\theta(T) = \frac{1}{1 + e^{(T - T_M)/\Delta T}} \qquad \text{Equation 3.4}$$

where T_M is the melting temperature and ΔT is the width of the melting transition. T_M values were predicted using the Integrated DNA Technologies Oligo Analyzer (https://www.idtdna.com/pages/tools/oligoanalyzer)[72] using an oligo concentration setting of 1.5 μM (following recommendations for equimolar strand concentrations) and ion concentrations to match desired conditions.

F. Circular Dichroism Spectroscopy

CD spectra were recorded on a Jasco J-1500 CD spectrometer in a 1 cm pathlength cuvette (Starna 26.100F-Q-10/Z15). Bvht H9:cs or Bvht H9 34-45 was prepared at an absorbance of ~0.8 AU in a buffer consisting of 20 mM sodium phosphate (pH 7.4) supplemented with NaCl to attain a total Na^+ concentration of 100 mM and $MgCl_2$ at 0 or 1 mM, as specified in figures. Bvht H9:cs was then annealed as described above by heating to 90 °C for 2 minutes and then slowly cooling in air for 10 minutes. Spectra were collected from 350-195 nm at 20 °C, at a 10 nm/min scanning speed. Spectra were processed by subtracting a buffer scan, zeroing the long-wavelength baseline, and converting to Δε per nucleotide using Equation 3.5, where CD is the CD signal in millidegrees, L is the pathlength, C is the RNA concentration determined by A260, and n is the number of nucleotides in the RNA.

$$\Delta\epsilon/nt = \frac{CD}{32980 \cdot L \cdot C \cdot n} \qquad \text{Equation 3.5}$$

6. Bridge to Chapter IV

With the right probe design, a greater depth of information can be acquired through SiM-KARTS than has previously been reported. Specifically, the use of locked nucleic acids within DNA probes stabilizes the transient binding events observed to a degree that allows trace-by-trace discrimination of different structures. This chapter has laid the foundation for the work that will be presented in Chapter IV, where the probe optimization considerations presented here are applied to the design of a SiM-KARTS probe for *Mhrt*.

CHAPTER IV: STRUCTURAL INVESTIGATIONS OF THE LNCRNA MHRT

1. Overview

In Chapter III, I reported on the incorporation of LNA residues in SiM-KARTS probes, and the depth of information that they can report. This chapter describes an experimental workflow to apply the conclusions of Chapter III to a heart-specific lncRNA, *Mhrt*. The use of chemical probing, specifically SHAPE, to aid in the determination of appropriate sequences within the RNA to target is presented, along with SiM-KARTS data obtained from following this approach. This chapter contains unpublished co-authored material. Experiments were conceived and planned in collaboration with Julia Widom. Contributions to chemical probing experiments were made by Natalie Sande, and tether binding experiments were performed with contributions from Megan Adamec.

2. Introduction

Mhrt RNAs are a group of heart-specific lncRNAs, which were called "antisense βMHC RNA" or β RNA upon initial discovery.[87] Both the initial and more recent naming conventions refer to the genetic origin of *Mhrt*, which originates from the locus encoding myosin heavy chains 6 and 7 (*Myh6* and *Myh7*).[17] Studies have shown altered levels of *Mhrt* in humans with conditions that induce cardiac stress, and it has been theorized that *Mhrt* has a cardioprotective function.[17,88] The exact mechanism behind this function remains unknown. Interactions between *Mhrt* and other RNAs and proteins have been reported, including the binding of methylated *Mhrt* to the protein MeCP2, and the binding of *Mhrt* to miRNA-3185.[89,90] One of the most extensively studied interactions is between *Mhrt* and the chromatin remodeling factor Brg1. It has been

shown that *Mhrt* can bind to Brg1, which is overexpressed under conditions of cardiac stress, and this interaction has been shown to affect the ratio of *Myh6* and *Myh7* transcripts.[17] In mice, it was shown that after induced cardiac stress, *Mhrt* expression and the ratio of *Myh6*/*Myh7* both drop. Taken all together, it is thought that Brg1 is able to bind both *Mhrt* and chromatinized DNA, including at the *Myh6* promoter. Binding of the DNA at this location represses the transcription of *Myh6*, and the antisense transcription of *Mhrt*. When levels are properly balanced, Brg1 is sequestered by the present *Mhrt*, allowing for normal transcription of both *Myh6* and *Mhrt* (Figure 4.1). Upon restoration of *Mhrt* levels, it has been shown that the heart is protected from hypertrophy and failure.[17] Despite its potential significance in cardiac health, molecular details of this interaction, and of the lncRNA *Mhrt* itself remain unknown.

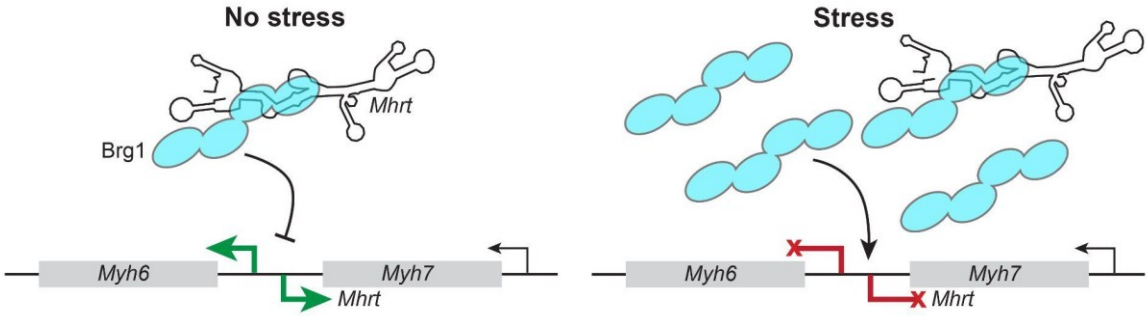


Figure 4.1. Proposed mechanism for the cardioprotective function of *Mhrt*. In conditions of no stress (left), the protein Brg1 is sequestered by *Mhrt* and normal transcription of *Myh6* and *Mhrt* occurs. In conditions of stress (right), Brg1 is overexpressed and not fully sequestered by *Mhrt*, allowing it to inhibit transcription of *Myh6* and *Mhrt*.

SiM-KARTS is well suited to explore the structure and dynamics of complex systems such as *Mhrt*, as it maintains the typical advantages of single-molecule measurements but does not require site specific labeling, and could also be used to study the interactions between *Mhrt* and Brg1. Utilizing different probing schemes, information about the structure and dynamics of different regions of the RNA, as well as protein binding can be studied.

The challenge of performing SiM-KARTS on lncRNAs is experimental design. Large RNAs present many potential probing sites, some of which may be more useful than others in yielding insights into biological function. Predicted secondary structures are a useful tool to begin selecting probing locations, but must be properly scrutinized through empirical investigations. RNA is known to fold co-transcriptionally, which is not accounted for in most calculations. Minimizing the free energy of a sequence may not always yield accurate results, particularly as the system becomes more complex with length. As the length of the RNA increases, it is reasonable to assume that it will encounter more kinetic traps during transcription, preventing the RNA from ever adopting its minimum free energy structure. Preliminary investigations of lncRNA structure utilizing chemical probing techniques can therefore greatly aid in the design of SiM-KARTS experiments.

Chemical probing techniques have been used to study RNA structure since 1980.[91] Today, a commonly used technique is 2' selective hydroxyl acylation analyzed by primer extension (SHAPE). SHAPE works by modifying the RNA through a reaction between a small-molecule electrophile and the 2' hydroxyl group of flexible nucleotides.[92] This reaction forms adducts which cause termination of reverse transcription (RT). When the reaction is properly optimized, each RNA molecule will have an average of one modification. Primer extension by reverse transcriptase generates complementary DNA fragments with lengths corresponding to the flexible, modified sites on the RNA. Analysis of the prevalence of fragments of different lengths is then used to generate SHAPE reactivities for each nucleotide of the RNA. Greater reactivity indicates that the nucleotide is more flexible (typically interpreted to be single-stranded), while low reactivities correspond to less flexible nucleotides (double-stranded).[93] I hypothesized that

SHAPE reactivities paired with the predicted secondary structure can guide the selection of regions of interest for further investigation by SiM-KARTS to reveal the dynamics of a lncRNA.

3. Structural Determination of *Mhrt* by Predicted Secondary Structure and SHAPE

RNA secondary structure can be easily predicted using readily available software or webservers.[71,94] The sequence of *Mhrt 779*, the most prevalent isoform of *Mhrt*,[17] was predicted using ViennaRNA,[94] and free energy minimization resulted in the structure shown in the lower panel of Figure 4.2. The construct *Mhrt 779 t1* was designed by adding a 22-nucleotide flexible sequence to the 3' end that can act as a tether for single-molecule experiments. It was checked that the addition of the tether sequence was not predicted to greatly impact the secondary structure. DNA was purchased including a T7 RNA polymerase promoter, and *Mhrt 779 t1* was obtained by *in vitro* transcription and native purification following previously published protocols.[62] The secondary structure of *Mhrt 779 t1* was then investigated further using SHAPE.

SHAPE has been utilized for lncRNA in the past.[62,95] Regardless, a significant amount of troubleshooting and optimization is needed to adapt a generalized protocol for SHAPE to a new RNA. Experimental design began with determining appropriate primers for the primer extension step. A properly binding primer is important to the SHAPE experiment because all results rely on RT working properly, which can only occur if primer is bound. Potential primers were purchased and annealed to *Mhrt 779 t1*, then run on a native agarose gel to test binding efficiency (Figure 4.2). While most primer remains unbound, a moderate amount of primer binding occurs for primers 1 and 2, while virtually no primer is bound for primers 3 and 4. Due to the highly structured nature of the predicted secondary structure of *Mhrt*, it was not unexpected that appropriate primer sequences are difficult to select.

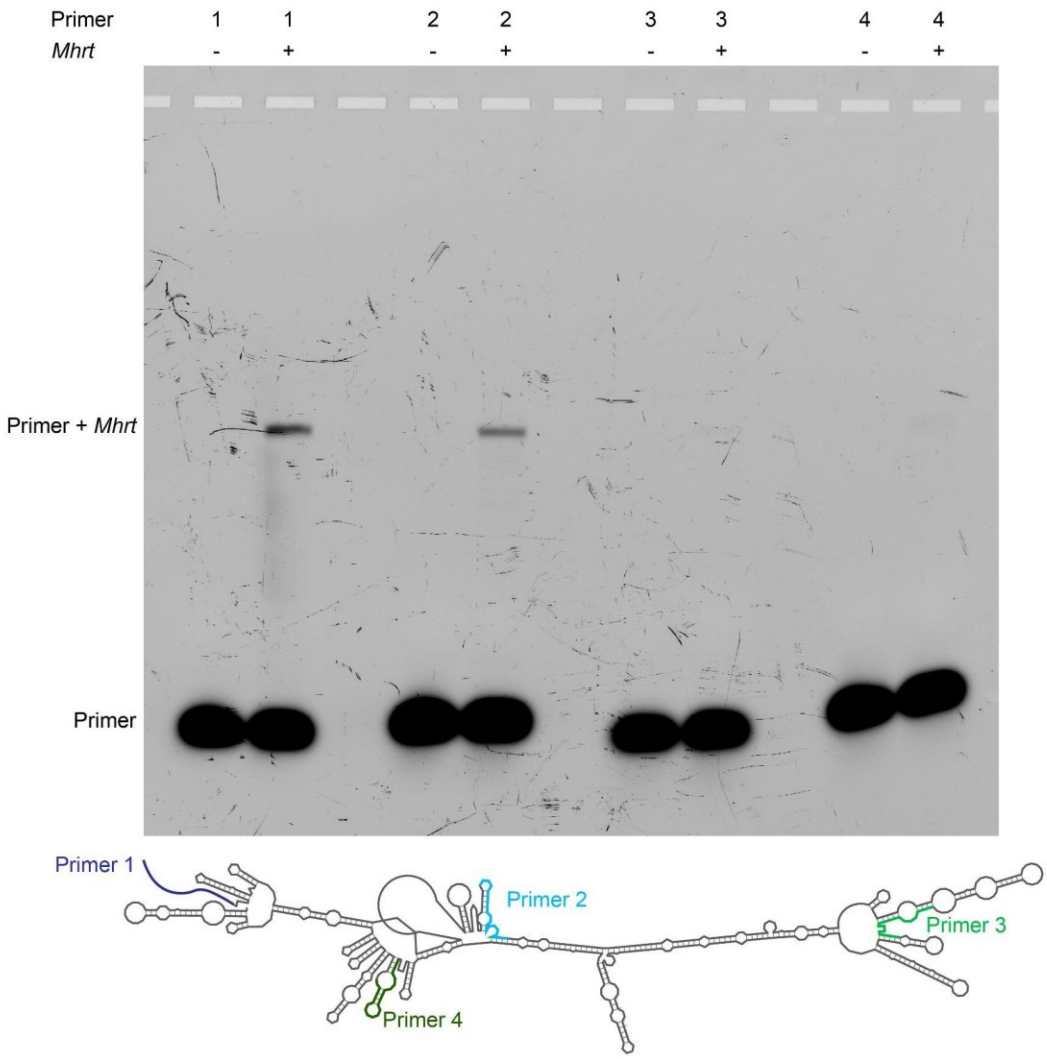


Figure 4.2. Native agarose gel (top) showing binding of each potential SHAPE primer to *Mhrt 779 t1*. The primer-*Mhrt 779 t1* complex is visualized in the upper band, while the lower band is unbound primer. Location of each primer binding site (bottom).

It is generally recommended that primers be designed approximately every 200 nucleotides to achieve single nucleotide resolution of SHAPE reactivities across the full length of the RNA.[62] In an attempt to improve primer binding, different annealing conditions were explored for primers 1 and 4. Higher temperatures aided in the binding of primer 1, but no binding of primer 4 occurred, regardless of the conditions of annealing (Figure 4.3). This is likely due to the long stretches within *Mhrt 779 t1* that are predicted to be double-stranded.

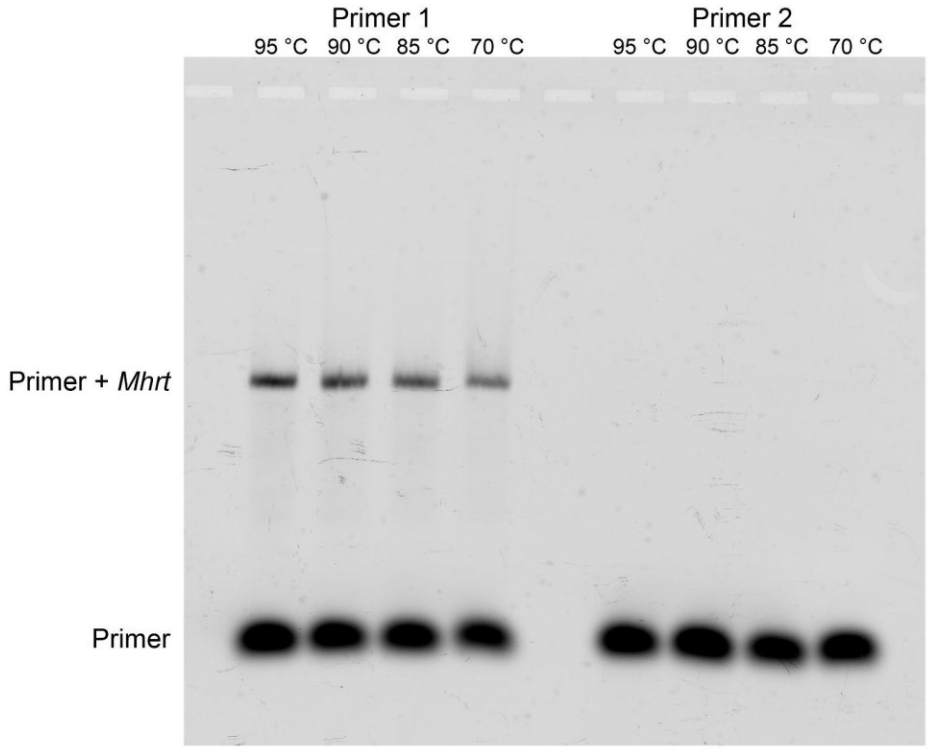


Figure 4.3. Native agarose gel exploring different annealing conditions of primer 1 or primer 4 to *Mhrt 779 t1*. All conditions resulted in binding of primer 1, with strong binding occurring at higher annealing temperatures. For primer 4, no binding was observed regardless of temperature.

To determine the identity of the nucleotides where adducts have formed and terminate the reverse transcriptase in SHAPE, unmodified RNA is used in RT reactions with dideoxynucleotide triphosphates (ddNTPs). During RT, if a ddNTP is incorporated, the primer extension is terminated, much like when a SHAPE adduct is encountered. This leads to fragments of complementary DNA of lengths corresponding to each instance of a certain nucleotide identity. For example, with the addition of ddTTP, fragments will be generated that terminate at the location of each A in the RNA. While a lack of sequence coverage is a concern for using SHAPE on *Mhrt 779 t1*, optimization of the sequencing reference lanes was able to reveal the extent of sequence coverage by RT. The ratio of ddNTP:dNTPs (reported as the ratio of ddNTP concentration relative to the concentration of each dNTP in an equimolar mixture)

incorporated in the RT reaction must be optimized to ideally allow for an even reporting of sequence along the full length of the RNA. In initial optimization experiments, a 1:1 ratio between the selected ddNTP and dNTPs was found to be ideal (Figure 4.4). These experiments also showed that with proper separation of bands, it seems likely that each primer will be able to report on more than a 200 nucleotide stretch of the RNA.

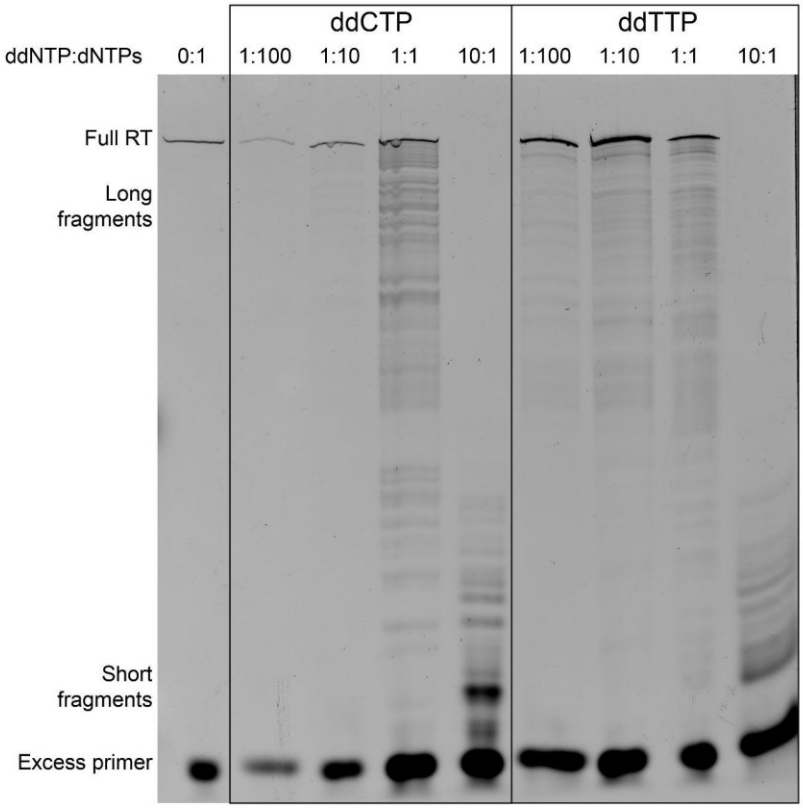


Figure 4.4. Denaturing polyacrylamide gel optimizing the ratio of ddNTP:dNTP for SHAPE sequencing. All samples were treated identically with the exception of ddNTP identity and concentration.

With a proper ddNTP:dNTPs ratio, the bands observed from modification with a SHAPE reagent can be appropriately assigned to the nucleotide they arise from. SHAPE reagents continue to be developed, as each has their benefits and drawbacks.[96] One popular reagent is 1-methyl-7-nitroisatoic anhydride (1M7). This reagent was synthesized specifically for SHAPE, and reduced the length of reaction needed for SHAPE, which was previously performed using N-

methylisatoic anhydride (NMIA). This was measured as the half-life of each reagent's hydrolysis reaction, which is only 14 s for 1M7, which allows for the modification step of the SHAPE reaction to take only 70 s as opposed to 20 minutes with NMIA.[92,97] Optimization of the concentration of 1M7 is necessary to achieve the modification of approximately 1 nucleotide on each RNA molecule. Too low a concentration of 1M7 will not exhibit an appreciable difference relative to a DMSO-only negative control, while too high of a concentration will lead to an over-representation of the nucleotides closest to the primer binding site. This delicate balance can only be found by testing various concentrations of 1M7 and determining the best concentration for the ideal distribution of adducts. For 1M7, a final concentration of 5 mM shows reasonable modification along the length of the RNA, while still having some RNA with no modification (as shown by the presence of full length complementary DNA) (Figure 4.5). A major issue using 1M7 arises from its short half-life. 1M7 undergoes a hydrolysis reaction with the 2' hydroxyl of RNA, but also with water.[97] To minimize hydrolysis with water, 1M7 is dissolved in anhydrous DMSO. However, we found that 1M7 was difficult to work with, likely due to hydrolysis before it comes into contact with the target RNA. Due to this, the effective concentration of 1M7 in the reaction depends on multiple different factors, especially the speed of reaction preparation and the person performing the reaction. For example, 5 mM 1M7 shows comparable band intensities to 7.5 and 10 mM for person 1, while 7.5 and 10 mM show primarily full length fragments for person 2 (Figure 4.5). The inconsistencies seen in 1M7 reactivity make it unreliable for the determination of lncRNA secondary structure.

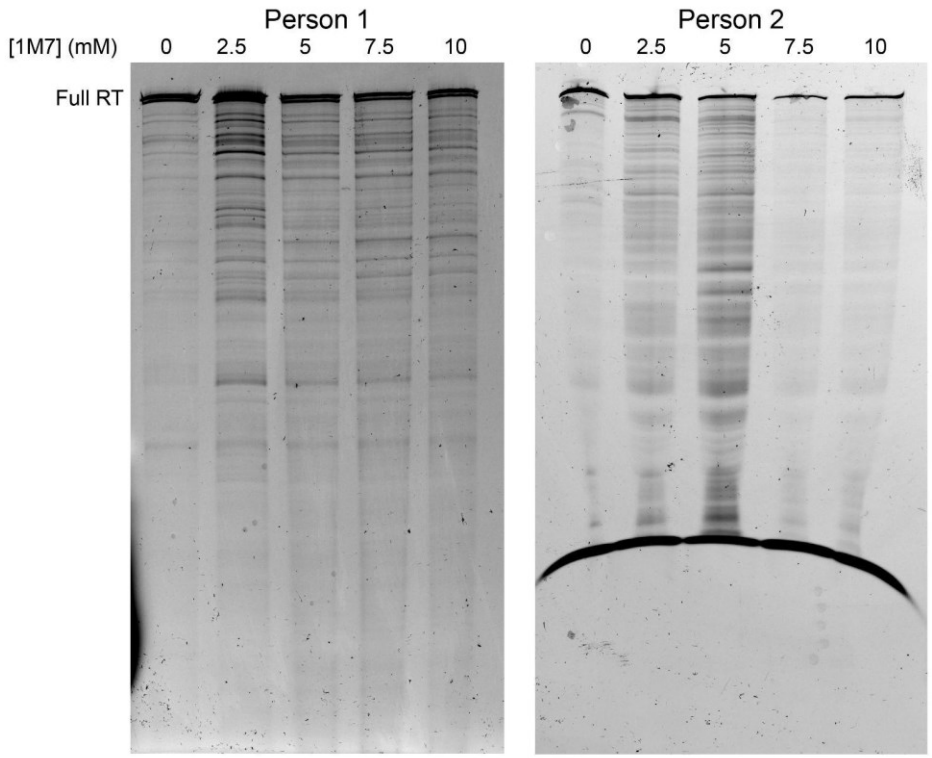


Figure 4.5. Optimization of 1M7 concentration visualized on denaturing polyacrylamide gels. Obvious differences exist between the same conditions when performed by different people and on different days, and are particularly obvious for the 5 mM 1M7 lanes.

Another SHAPE reagent that has shown promise is 2-methylnicotinic acid imidazolide (NAI).[96] NAI was initially designed as a SHAPE reagent for *in vivo* SHAPE, which requires a longer hydrolysis half-life due to the highly aqueous environment of cells. This reduction in reaction speed will reduce the effect of minor differences in timing between individuals or days of SHAPE experiments. As with 1M7, the concentration of NAI was optimized by modifying RNA with various concentrations of NAI. We found that 50 mM is an appropriate concentration for NAI reacting with *Mhrt 779 t1* (Figure 4.6).

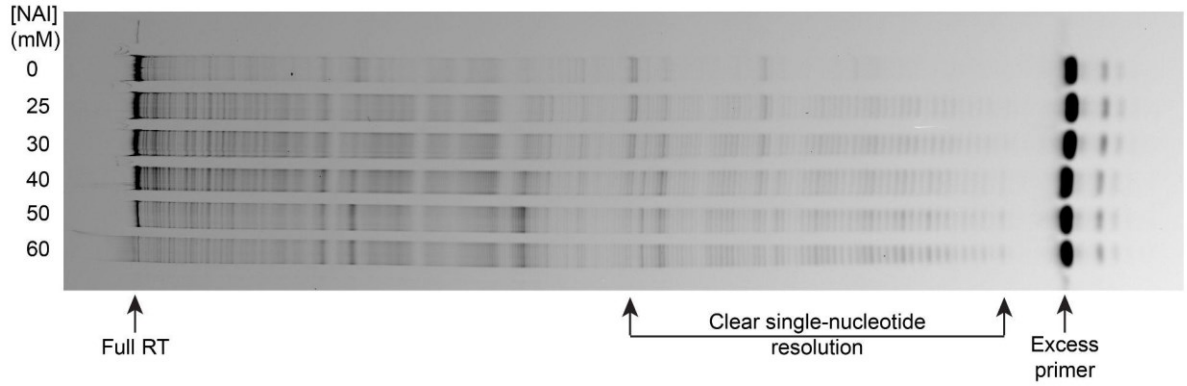


Figure 4.6. Optimization of NAI concentration visualized on a denaturing acrylamide gel.

We also investigated unexpected banding patterns that were sometimes observed in otherwise optimized reactions. An optimal SHAPE experiment will show minimal banding in the negative control, which was not always the case in our experiments. One preliminary gel for SHAPE of *Mhrt 779 t1* showed expected results from 50 mM NAI and the sequencing lanes, but the DMSO control reaction produced unexplained banding (Figure 4.7). Further troubleshooting revealed that this banding stemmed from the initial precipitation to remove the SHAPE reagent before RT, but the exact cause is not apparent. Fresh preparation of all components involved in this step were made, and the resulting negative control lane appeared as expected, with fewer short fragments present. However, there are also fewer short fragments present in the NAI lane, suggesting the precipitation step was impacting the RT of this sample as well. Due to the impact of the precipitation components on RT, the concentration of NAI used for this system should be reoptimized, as 50 mM NAI does not give single-nucleotide resolution of the shorter DNA fragments.

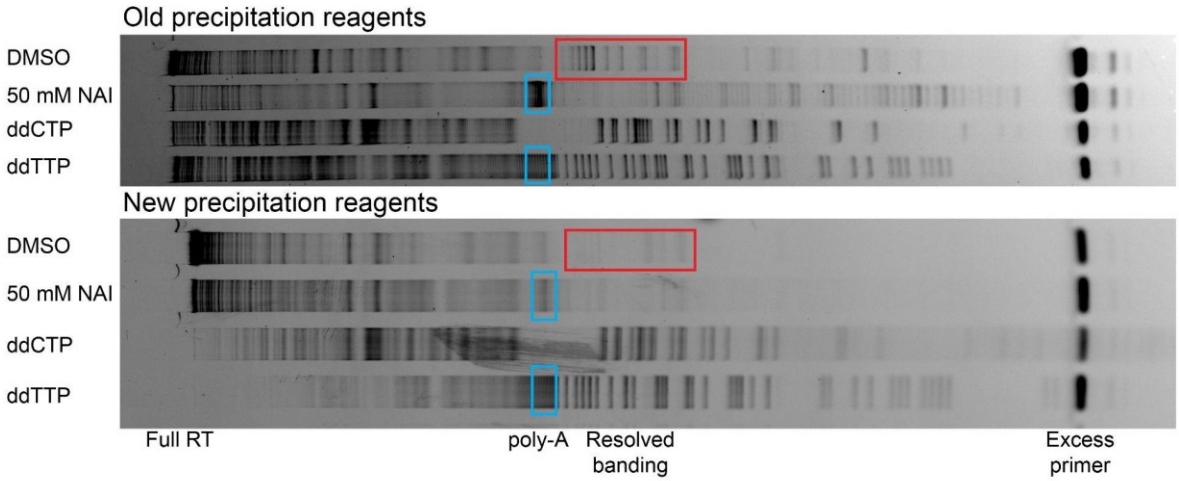


Figure 4.7. Preliminary results of *Mhrt 779 t1* SHAPE, before (top) and after (bottom) precipitation reagent troubleshooting. Before troubleshooting, the DMSO lane contained abnormal banding (red box), which resolved when fresh reagents were used. The poly-A region of *Mhrt 779 t1* shows higher SHAPE reactivity (cyan box), suggesting a flexible structure.

The higher SHAPE reactivity of certain nucleotides is obvious, even without a complete understanding of the secondary structure. Specifically, the poly-A region located from position 673 to 688 is shown to be more flexible than much of the rest of the RNA (Figure 4.7). This region is of interest as one of the few single-stranded regions in the predicted structure of *Mhrt 779*, and is also predicted to be flexible in most isoforms (Figure 4.8). A SiM-KARTS probe for this region was designed to be complementary to the 3' end of the poly-A region to incorporate some nucleotides other than A, ensuring a unique binding site. A single locked nucleic acid (LNA) residue was incorporated to bring the predicted T_M to around 20°C, with similar characteristics to LNA2 from Chapter 3.

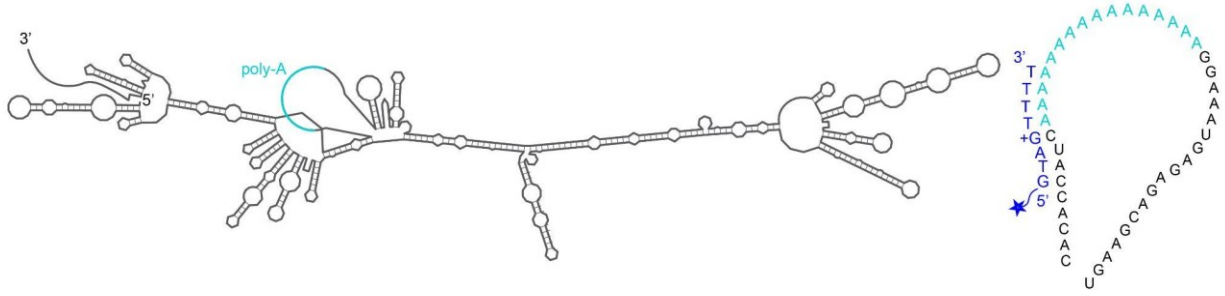


Figure 4.8. SiM-KARTS probe for poly-A region of *Mhrt 779 t1*. The probe sequence overlaps with the 3' end of the poly-A segment.

4. Preliminary Results of *Mhrt* Structure and Dynamics via SiM-KARTS

In addition to a single LNA residue, this probe was designed with a 5' Cy3 label, which is visible in the green channel, in contrast to the 5' TYE665 used for probes in Chapter 3, which were visible in the red channel. Due to optical factors, the green channel of our instrument has less inherent background. By using a fluorophore visible in the less noisy channel for the probe, this should lead to a more even background between channels. Additionally, instead of directly labeling *Mhrt 779 t1*, a tether sequence, "t1", containing a 5' biotin and 3' Cy5 was designed. This allows for the immobilization and visualization of *Mhrt 779 t1* without any post-processing of the transcribed RNA. The binding efficiency of *Mhrt 779 t1*:t1 was determined via native agarose gel electrophoresis. A variant of t1 without biotin was mixed in various ratios with *Mhrt 779 t1*, and binding was observed by Cy3 fluorescence in the *Mhrt 779 t1* band when t1 was bound. Two different incubation conditions were also considered. Ultimately, it was found that incubating *Mhrt 779 t1* in a 1.6 times excess of t1 at 37°C for 10 minutes was sufficient, resulting in ~95% of t1 being bound to *Mhrt 779 t1* (Figure 4.9). For preliminary SiM-KARTS results, this incubation scheme was utilized for sample preparation.

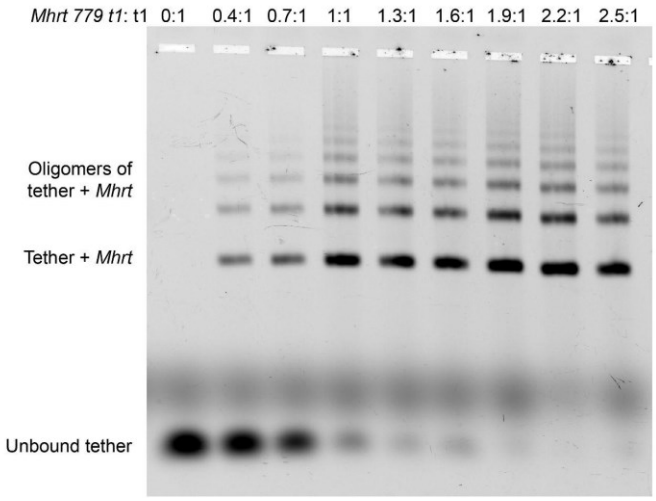


Figure 4.9. Native agarose gel showing the binding of *Mhrt 779 t1* to 50 nM Cy3-labeled t1, varying the amount of *Mhrt 779 t1* present.

Upon initial observations, this probe seems to be well designed for the *Mhrt 779 t1* system, supporting the probe design recommendations from Chapter 3. Traces extracted from movies show binding events that are clear, yet not excessively long (Figure 4.10). The lowered background of the green channel in comparison to the red leads to much cleaner tracking of probe binding. Despite visibly high-quality data, analysis of traces across 3 movies lead to the selection of only 61 molecules. While selection of a low fraction (~10%) of molecules is typical, this is significantly lower (~2%) than typical, and worth investigating further.

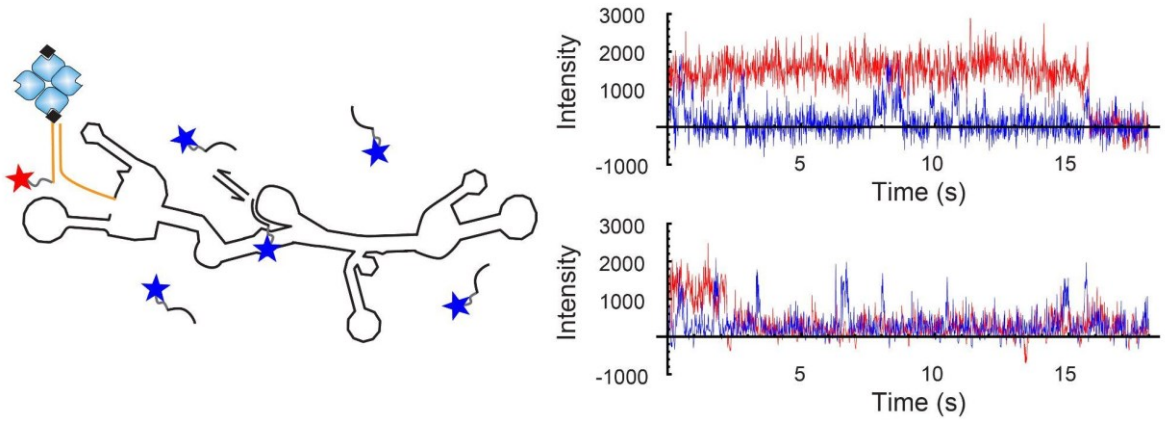


Figure 4.10. SiM-KARTS design (left) and example traces (right) of poly-A probe. *Mhrt 779 t1* is immobilized using a Cy5-labeled tether (red), and investigated with a Cy3-labeled probe (blue).

Analysis of dwell times after applying HMM to traces reveals the expected cumulative distribution curve shapes for an accessible, flexible structure when considering both the bound and unbound dwells, with reasonably short bound dwells and longer unbound dwells (Figure 4.11). Fitting these curves results in a characteristic unbound dwell time of ~11.5 seconds, which is long compared to the data in Chapter 3 for LNA2, where the most comparable case (Bvht H9:cs) has a characteristic unbound dwell time of 5.2 seconds. This suggests that the optimal concentration for this probe is higher than the probes used in Chapter 3, and that this variable should be optimized individually for every system and probing site. These long unbound dwells are likely due primarily to the rate of probe diffusion rather than the structure of the RNA itself, which complicates the use of the unbound dwell times as signatures of structure. This highlights the significance of our discovery reported in Chapter 3 that bound dwell times can provide rich information about the structure of the target site.

The characteristic bound dwell time is ~400 ms, which allows for clear distinction between the bound and unbound states, while not persisting so long as to significantly perturb the structure of the target. Additionally, the weights of each component were around 50%, which is consistent with the more flexible structures probed in Chapter 3. Interestingly, $\tau_{bound,avg}$ for this probe is shorter than for the most similar probe binding a flexible structure in Chapter 3, even though both have similar T_M, probe lengths, and GC content. Further investigations on the impact of RNA structure on length of bound dwells would be useful to determine if this difference in characteristic bound dwell time is due to the exact sequence of the probe binding site, or if it can be directly interpreted as a less accessible structure.

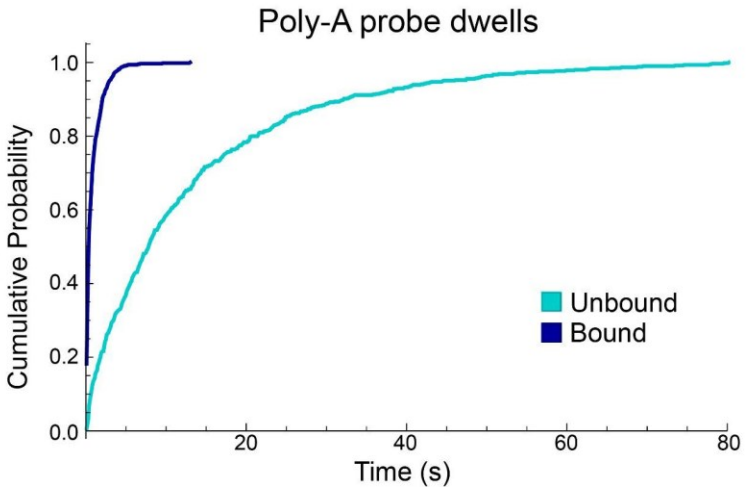


Figure 4.11. Cumulative distributions of poly-A probe unbound and bound dwell times.

Analysis of the unbound dwells using the Fano factor[46] suggests that dynamics may be occurring in this system on a timescale of around 20 seconds, as indicated by deviation of the Fano factor from 1 on this time frame (Figure 4.12). While diffusion of the probe is expected to lead to a random distribution of binding events in the absence of dynamics, heterogeneity of the probe binding kinetics is possible due to differences in local environment such as defects in the slide passivation or surface. While these results are not necessarily indicative of the RNA sampling two conformations, it is possible that structural rearrangements are in fact occurring.

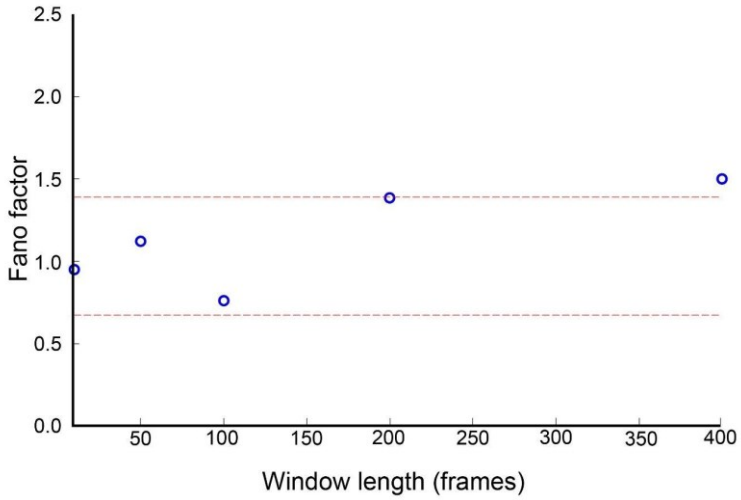


Figure 4.12. Fano factor of unbound dwell time of poly-A probe. Significant deviation away from 1 occurs when considering time scales above ~200 frames.

To further investigate potential structural rearrangements, the data was subjected to a burst analysis using previously published Matlab code,[46] which reported on the length of unbound dwells in and out of the bursts, as well as the length of each burst found. As seen in cumulative count curves, most unbound dwells were defined to be out-of-burst, with a smaller percentage assigned as in-burst (Figure 4.13). Cumulative distribution curves were fit to determine the characteristic interspike interval (ISI) for each group. The in-burst grouping ISI was found to be 300 ms, while the out-of-burst ISI was 17 s. Of 663 dwells analyzed, 208 were in-burst, while 455 were out-of-burst. This suggests a tendency towards a structure that is somewhat less accessible for binding, corresponding to the times the dwells are found to be out-of-burst, with less time spent in some more accessible structure that is sampled when the dwells are found to be in-burst.

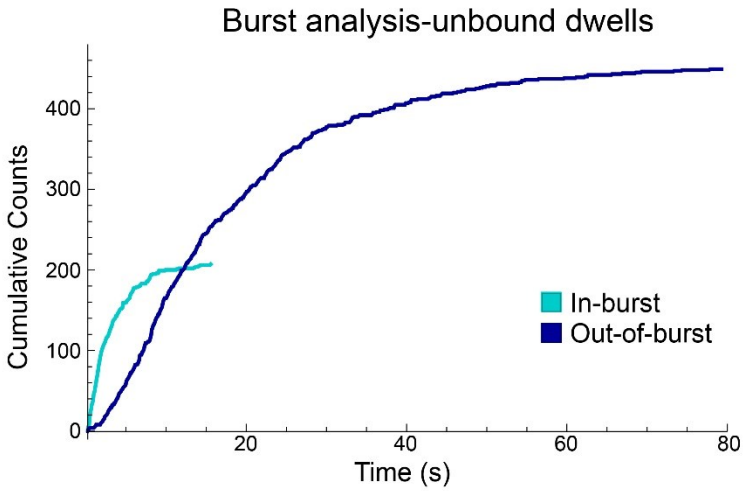


Figure 4.13. Cumulative counts of unbound dwells in- and out-of-burst.

5. Conclusions

This preliminary investigation of the structure of *Mhrt 779 t1* suggests that pairing SHAPE and SiM-KARTS is effective for determination of site-specific local structure of lncRNA. Specifically, the poly-A region identified as likely single-stranded in secondary

structure predictions is supported to be flexible by SHAPE and SiM-KARTS. While both techniques require further optimizations for a more holistic view of this lncRNA, the preliminary results reported here are promising. The use of LNA probes designed to have a predicted T_M of approximately 20°C resulted in analyzable data without further optimizations. The resulting bound dwell times are reasonable and detection of events is robust against noise.

6. Future Directions

Taken together, the work presented here and in Chapter 3 provide a base of knowledge for generalizing SiM-KARTS for use with other RNAs, and lncRNAs specifically. However, many optimizations and questions towards this goal remain. Specific to *Mhrt 779 t1*, further optimization of SHAPE is needed to obtain a full view of secondary structure. Reoptimization of NAI concentration, as well as the design of more primers for RT are needed. Ultimately, the use of capillary electrophoresis is the standard for resolving DNA fragments after primer extension,[98] and quantification of SHAPE reactivities would be aided by its use.

Further optimizations of SiM-KARTS for the specific *Mhrt 779 t1* system are necessary. Optimization of tether binding in situ would result in a higher percentage of usable molecules in single-molecule experiments. These optimizations could simply take the form of investigating *Mhrt 779 t1*:t1 annealing and immobilization. One step towards this optimization would be ensuring t1 alone is immobilizing well, and making any optimizations necessary for t1 immobilization before investigating *Mhrt 779 t1*:t1 immobilization further. Simple optimizations of instrumental variables are needed, as well as an optimization of probe concentration, as was done for the Bvht probes in Chapter 3. The signal to noise ratio (SNR) of SiM-KARTS data for *Mhrt 779 t1* so far is suboptimal, and should be optimized for future experiments by adjusting laser power and camera parameters. Specifically, raising the power of the green laser should

increase the SNR due to a physical increase in emitted photons. An increase of the EM gain level can also increase the SNR, as this defines the amount of signal amplification that occurs. However, a too high EM gain can lead to reduced dynamic range, ultimately reducing SNR, as well as accelerated aging of the camera. Similarly, a proper balance of probe concentration must be found, to maximize the frequency of binding events while maintaining a reasonable SNR.

Beyond these next steps, different probing schemes could be used to probe multiple aspects of the RNA at once. For example, using a Cy3-labeled probe in conjunction with a probe containing both Cy3 and Cy5, or a "FRET probe", would allow for the probing of two regions simultaneously. This probing scheme would include briefly illuminating the sample with a red laser to locate Cy5 labeled RNA, then performing the rest of the experiment with green laser illumination only. The Cy3 probe would show in the green channel, and the "FRET probe" would appear in the red channel, due to energy transfer between the two fluorophores held at a close distance. This scheme has the potential to reveal correlated dynamics, reporting on the impacts of specific local structures on other regions within the RNA. Similarly, protein binding could be correlated with probe binding at sites of interest by utilizing fluorescently labeled protein in addition to a probe. For *Mhrt 779 t1* specifically, the protein Brg1 is of interest, and attempts to express and purify this protein were made with limited success.

For SiM-KARTS, there are multiple expansions of analysis that could be beneficial. In its current form, the burst analysis method which has been reported is not easily generalizable. Here, the total number of bursts identified was 62, while only 61 molecules were input for analysis. The majority of these bursts were 100 ms in length, corresponding to a single frame. These very short "bursts" are responsible for the short in-burst ISI, skewing the analysis. Additionally, as is, the code for this analysis does not allow for direct tracking of bursts,[46]

meaning the events from which they arise are not reported. Due to this, it is difficult to determine issues in this analysis workflow. It seems likely that certain parameters need to be optimized to adjust the burst analysis algorithm to more generalized uses, and that the results obtained here are not physically meaningful. In contrast, the simple analysis of dwell times support the flexible structure predicted by free energy minimization and further suggested by SHAPE reactivity.

A method which directly reports on where bursts are identified and what events lie within each burst would be useful for data interpretation. Additionally, as is, the method does not report the bound dwell times in- and out-of-bursts. In Chapter 3 we showed that bound dwell time, particularly for probes containing LNA residues, is impacted by structure. It follows that when the RNA is "in-burst" with a highly accessible structure, binding events will likely have longer dwell times than when the RNA is "out-of-burst" with a less accessible structure. If this workflow is established, an additional analysis could be done utilizing the average bound and unbound dwells for each molecule in and out of burst, with a clustering analysis applied like in Chapter 3. This would show how distinct the states attributed to the in- or out-of-burst dwells are, and would provide additional detail for a well rounded view of RNA structure and dynamics.

7. Bridge to Chapter V

SiM-KARTS is a promising technique for the investigation of lncRNA, particularly when paired with chemical probing techniques such as SHAPE. For smaller non-coding RNA, smFRET is a powerful, well-established tool to study structure, dynamics, and interactions with other molecules. Chapter 5 will discuss investigations into two variants of riboswitch which interact with a biologically important small molecule using smFRET and bulk FRET measurements.

CHAPTER V: STUDIES OF LIGAND BINDING TO VARIANTS OF THE NAD^+-II RIBOSWITCH

1. Overview

Previous chapters have focused on the single-molecule microscopy technique SiM-KARTS, and studies of lncRNA. This chapter discussed studies of riboswitch variants using smFRET. This chapter contains unpublished co-authored material. Experiments were conceived and planned in collaboration with Julia Widom. Initial data collection and analysis of the NAD^+-II riboswitch were performed by Nico Conoan Nieves. I performed additional data collection and analysis. Single-molecule data of the mini-NAD^+-II riboswitch was collected in collaboration with Madison Kadrmas.

2. Introduction

Nicotinamide adenine dinucleotide (NAD^+) is an important coenzyme across species.[24] The conversion to its reduced form, NADH, acts as a key mechanism in numerous biological pathways, such as glycolysis. Numerous roles of NAD^+ in eukaryotes have been reported, with links to aging, DNA repair and metabolic pathways.[99] In prokaryotes, NAD^+ also plays a significant role in essential cellular pathways, such as metabolism, redox, and signal transduction.[100] The ratio of NAD^+ and NADH is critical for bacterial survival, with altered levels of each indicating a response to various external threats.[100]

Due to the importance of NAD^+ in bacteria, the biochemical pathways that regulate this molecule are of interest as potential targets for antibiotics. Riboswitches regulate bacterial gene expression in response to ligand binding, and the targeting of riboswitches could therefore lead

to the dysregulation of mRNA transcription or translation, ultimately impacting cell health or even survival.[23] Two classes of NAD^+ sensing riboswitches have been reported, each with different binding interactions to NAD^+. NAD^+-I riboswitches specifically bind the adenosine moiety of NAD^+,[25] while NAD^+-II riboswitches selectively bind the nicotinamide moiety.[26] Very recently, a sub-class of NAD^+-II riboswitches termed "mini-NAD^+-II" has been reported, which are found to be more abundant genetically.[31] Due to its recent discovery, relatively little is known about its structure and dynamics when compared to the full-length NAD^+-II riboswitch. The importance of NAD^+ motivates the development of tools to accurately sense the presence, as well as concentration of NAD^+ and its related molecules. The importance of NAD^+ justifies the development of tools to accurately sense the presence, as well as concentration of NAD^+ and its related molecules.[28] The development of such tools or therapies heavily relies on understanding the underlying structure, dynamics, and binding mechanisms of the riboswitch in question. The work reported here aims to further understand the behaviors of NAD^+-II variants (Figure 5.1).

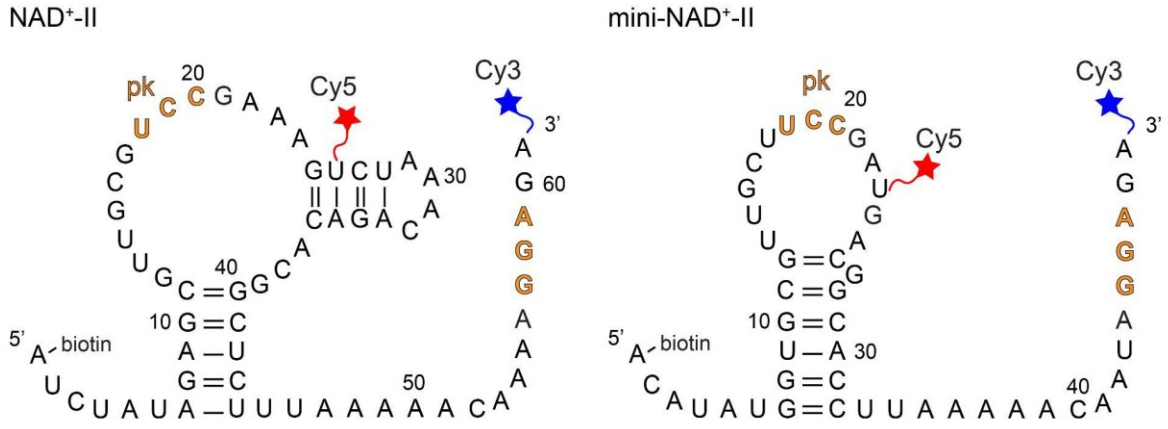


Figure 5.1. Secondary structures and labeling sites of riboswitch constructs for FRET. NAD^+-II (left) and mini-NAD^+-II (right) both have 5' biotin modifications, with 3' Cy3 labels. Cy5 labels are located at the bases shown (red star).

Initial studies of the NAD^+-II riboswitch reported the secondary structure of the RNA with and without ligand present, as well as its affinity for different ligands related to NAD^+, namely β-nicotinamide mononucleotide (NMN) and nicotinamide riboside (NR).[26] It was found that NMN binds the most tightly of the three, followed by NAD^+ then NR, suggesting that NAD^+-II recognizes the nicotinamide moiety of its ligand for binding. Further studies have determined crystal structures of NAD^+-II, which showed that the adenosine moiety of NAD^+ undergoes only nonspecific interactions with the riboswitch, and binding is driven by stacking and hydrogen bonding interactions with the nicotinamide moiety.[28] While these studies yielded a general understanding of this RNA's structure and interactions with ligands, details on conformational dynamics of the RNA were not investigated.

To address this gap, our lab previously investigated NAD^+-II using smFRET, biolayer interferometry, and steady-state fluorescence in the form of bulk FRET measurements establishing methodology and optimized oligonucleotide design for further studies.[101] This work revealed that the riboswitch undergoes fast transitions between multiple conformational states. However, the analysis of the smFRET data was limited, specifically due to the use of a theoretical correction for γ-factor, which only considered instrumental optics in the calculation. In reality, additional factors impact the observed γ-factor, such as the environment of the fluorophores. This correction has the ability to significantly change the appearance of smFRET histogram peak locations and widths. Published FRET data investigated interactions of NAD^+-II with NMN, leaving open the question of how NAD^+-II interacts with NAD^+, which was proposed to be the biologically relevant ligand due to its higher prevalence in the cell. Additionally, interactions between NAD^+-II and etheno-NAD^+ (ε-NAD^+), an emissive analogue

of NAD^+, are of interest for future studies that exploit the fluorescence of ε-NAD^+ to investigate how the ligand changes structure upon binding and how it interacts with NAD^+-II.

In addition to a more complete analysis of the interactions of NAD^+-II with various ligands, investigations into the newly reported mini-NAD^+-II riboswitch are reported here. The conformational dynamics and ligand binding behaviors of a mini-NAD^+-II riboswitch will be compared to the better understood but less biologically prevalent NAD^+-II riboswitch.

3. NAD^+-II Riboswitch

The published analysis of NAD^+-II smFRET data suggested the presence of at least three conformational states, with different distributions in different conditions.[101] With RNA in the absence of Mg^{2+} and NMN, a mid-FRET state of approximately E_{FRET} = 0.3-0.4 was most prevalent. The prevalence of this state lowers in the presence of 20 mM Mg^{2+}, with an increase in the prevalence of the high-FRET state between E_{FRET} = 0.8-1.0. Additionally, it is worth noting that the peak of each state shifts to slightly higher E_{FRET} as Mg^{2+}, then ligand, are added. I followed up this work by refining the processing and analysis of existing smFRET traces, including unpublished data collected with NAD^+ and ε-NAD^+,[102] and by performing bulk FRET measurements to analyze the affinity of NAD^+ to our fluorescently-labeled riboswitch.

I first refined the processing of existing traces by applying an empirically determined value of 2.1 as the γ-factor and applying a bilateral filter to reduce noise in traces. A larger γ-factor correction has the direct consequence of shifting all peaks to lower E_{FRET} values (Figure 5.2). Application of the bilateral filter also leads to an apparent narrowing of peaks, with generally sharper features present across the histograms. In the case of NAD^+-II without Mg^{2+} or ligand, the mid-FRET peak shifts from ~0.35 in the uncorrected histogram to ~0.2 after

corrections, with an increased height and decreased width. This corrected histogram suggests the existence of four distinct FRET states, with peaks centered at approximately 0.05, 0.2, 0.4, and 0.7. These E_{FRET} values are supported by visual inspection of traces, where sampling of these states can be observed. Fitting the histogram with a sum of four Gaussians centered at these E_{FRET} values allows for the determination of the area under the curve for each state, which expresses the prevalence of each state.

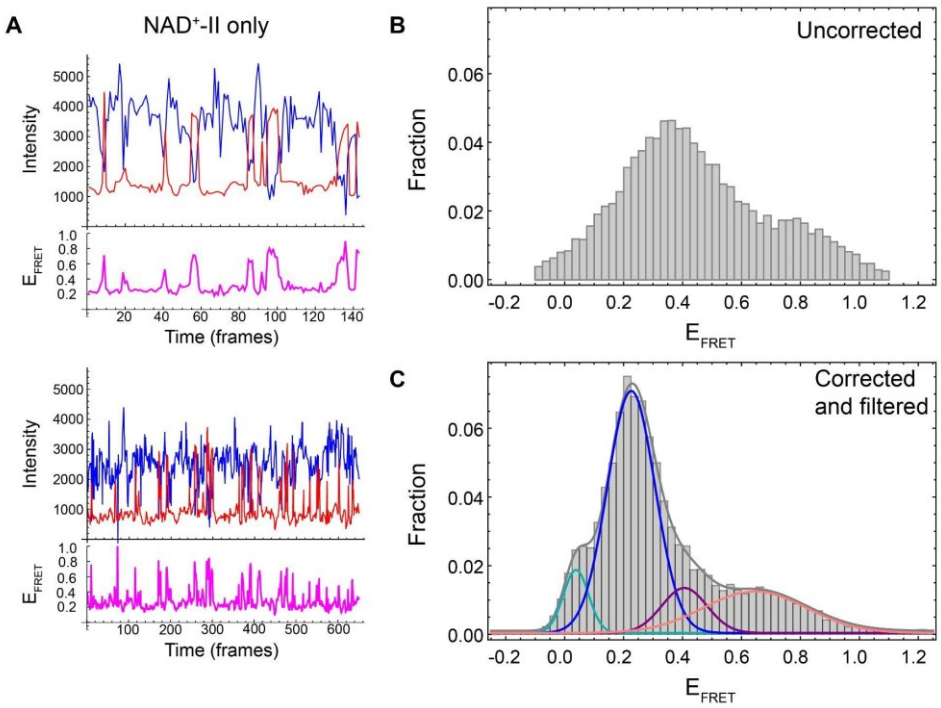


Figure 5.2. smFRET data from NAD^+-II data collected in the absence of ligand and Mg^{2+}. Representative traces (A) show frequent and fast transitions between multiple FRET states. Uncorrected data result in a histogram with broad distributions at ~0.4 and ~0.8 (B), while data with the corrections and filtering discussed result in more distinct states (C).

Fitting the histogram of all cases with four states leads to reasonable identification of peaks at similar FRET efficiencies, suggesting four primary conformational states for NAD^+-II, which are suppressed or enhanced under various conditions (Figure 5.3). In the presence of 20 mM Mg^{2+}, the prevalence of the high-FRET state increases from ~22% without Mg^{2+} to ~50%,

signifying a more folded structure is induced by the addition of Mg^{2+}. When NAD^+-II interacts with NMN, NAD^+, or ε-NAD^+, the high-FRET state becomes clearly favored. This is the most drastic for NMN, which is consistent with NAD^+-II binding NMN more strongly than NAD^+.[26] The high-FRET state has a prevalence of ~77% for NAD^+-II in the presence of NMN, while the prevalence is ~60% in the presence of NAD^+ or ε-NAD^+. It is important to note that despite these shifts towards the high-FRET state, NAD^+-II remains a dynamic system that samples each of the four states identified in all cases. Further analysis is in progress to uncover the conformational dynamics that result from the binding of each ligand.

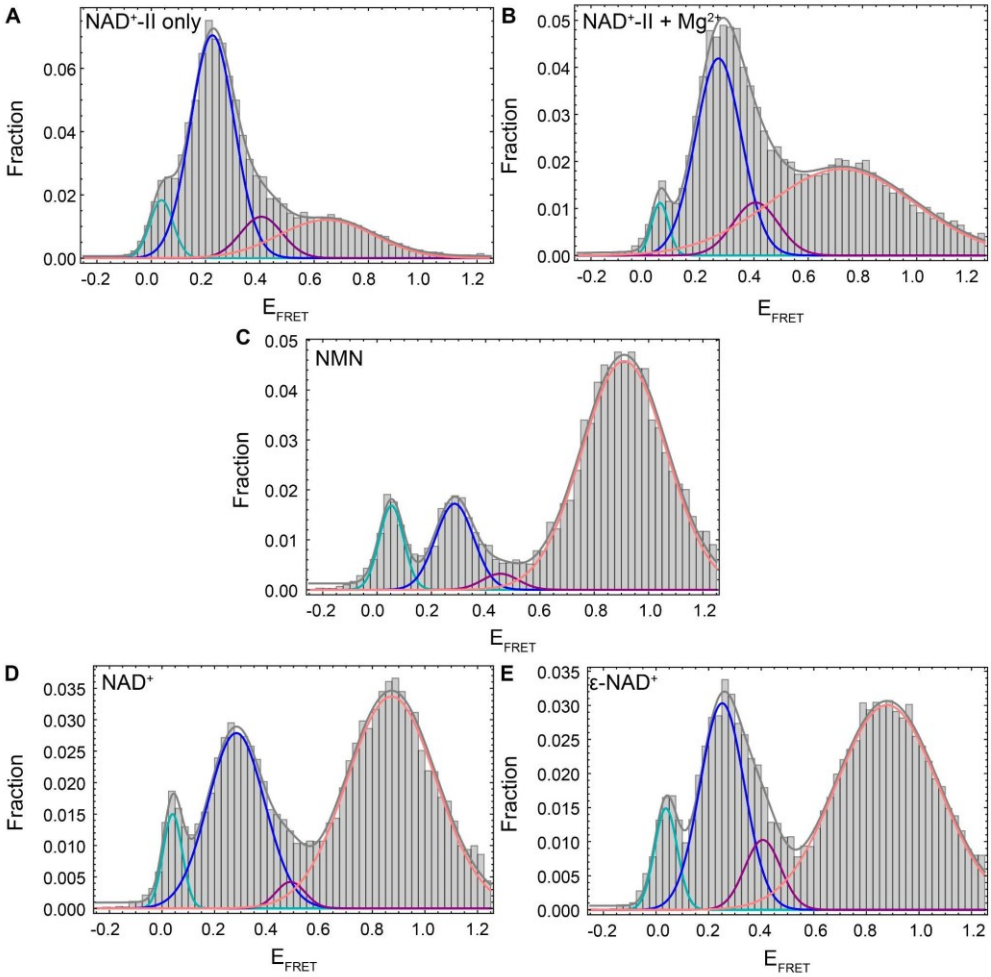


Figure 5.3. Histograms for NAD^+-II only (A), NAD^+-II + 20 mM Mg^{2+} (B), NAD^+-II + 20 mM Mg^{2+} + NMN (C), NAD^+-II + 20 mM Mg^{2+} + NAD^+ (E), and NAD^+-II + 20 mM Mg^{2+} + ε-NAD^+ (E), all fit with a sum of four Gaussians.

While the K_D of both NMN and NAD^+ have been reported,[26] it is important to consider potential impacts of oligonucleotide design on the binding affinity of the RNA. Specifically, the fluorophores added to facilitate smFRET could impact the competency of the RNA for ligand binding. For a direct comparison to published K_D values, additional bulk FRET titrations of NAD^+ were performed, resulting in a K_D of 84 μM (Figure 5.4). By the same method, the K_D of NAD^+-II binding to NMN was previously reported as 32 μM.[101] These results of these titrations support stronger binding of NMN than NAD^+ to the riboswitch, as suggested by the smFRET histograms. These binding affinities are consistent with previous reports of NAD^+-II riboswitches, with K_D values of approximately 100 μM for NAD^+ and 40 μM for NMN.[31]

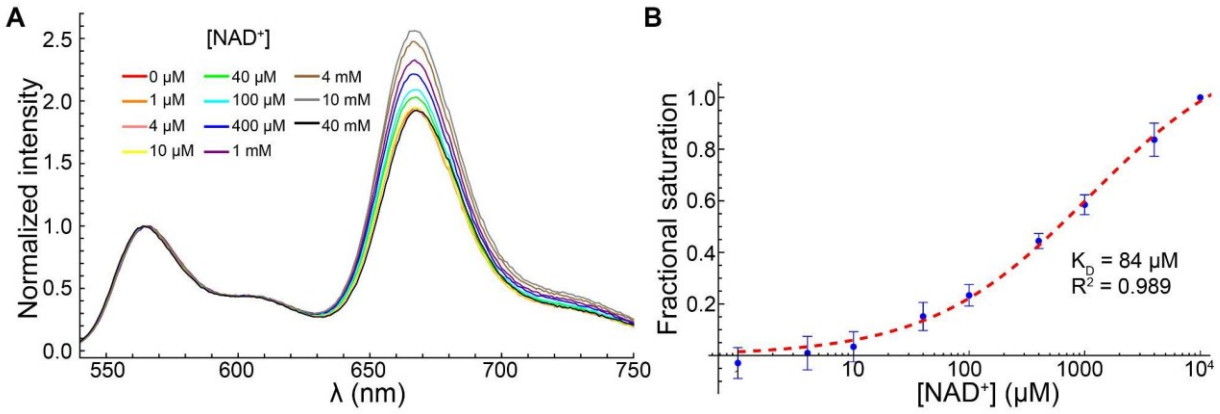


Figure 5.4. Bulk FRET titration of NAD^+. Emission scans of NAD^+-II at increasing concentrations of NAD^+, normalized to the intensity of Cy3 (A). The black spectrum, taken at 40 mM NAD^+, indicates quenching of Cy5 by NAD^+, and was not used for the determination of K_D. Fractional saturation of NAD^+-II at each NAD^+ concentration from three replicates, fit with the Hill equation (B).

4. Mini-NAD^+-II Riboswitch

The mini-NAD^+-II riboswitch for FRET studies (mini-NAD^+-II) was designed to match the form of NAD^+-II reported above (Figure 5.1). To confirm that the chosen fluorophore modification sites are not impacting the ability of miniNAD$^+$-II to bind ligand, and that the

fluorophore locations report sensitively on the conformation of the RNA, initial steady-state fluorescence measurements were performed. With increasing concentrations of Mg^{2+}, the observed FRET efficiency increases, as marked by an increase in fluorescence around 670 nm relative to the intensity at 565 nm. This signifies a response of the FRET efficiency to the addition of ligand, validating the design of the oligonucleotide.

Bulk FRET titrations of NMN to mini-NAD^+-II were performed (Figure 5.5). Data taken at 20 mM Mg^{2+} revealed a K_D of 2.14 μM, which is comparable to the published 4.6 μM K_D from in-line probing.[31] Published measurements of ligand binding affinity for the NAD^+-II and mini-NAD^+-II riboswitches are limited to 20 mM Mg^{2+}, which is a high relative to the typical physiological concentration of less than 1 mM free Mg^{2+},[103] and biolayer interferometry data of NAD^+-II showed no ligand binding in the absence of Mg^{2+} or in 1 mM Mg^{2+}.[101] In contrast, titrations of *mini*-NAD^+-II with NMN recorded at 1 mM Mg^{2+} show response to ligand, with a K_D of 19.8 μM.

Preliminary smFRET data indicates that mini-NAD^+-II is fairly static in nature, with less frequent sampling of different states when compared to NAD^+-II (Figure 5.6). Traces obtained from smFRET also suggest high FRET efficiencies on average, which is supported by steady-state fluorescence data. This work is ongoing, with further data collection and analysis necessary to better understand the conformational dynamics of mini-NAD^+-II.

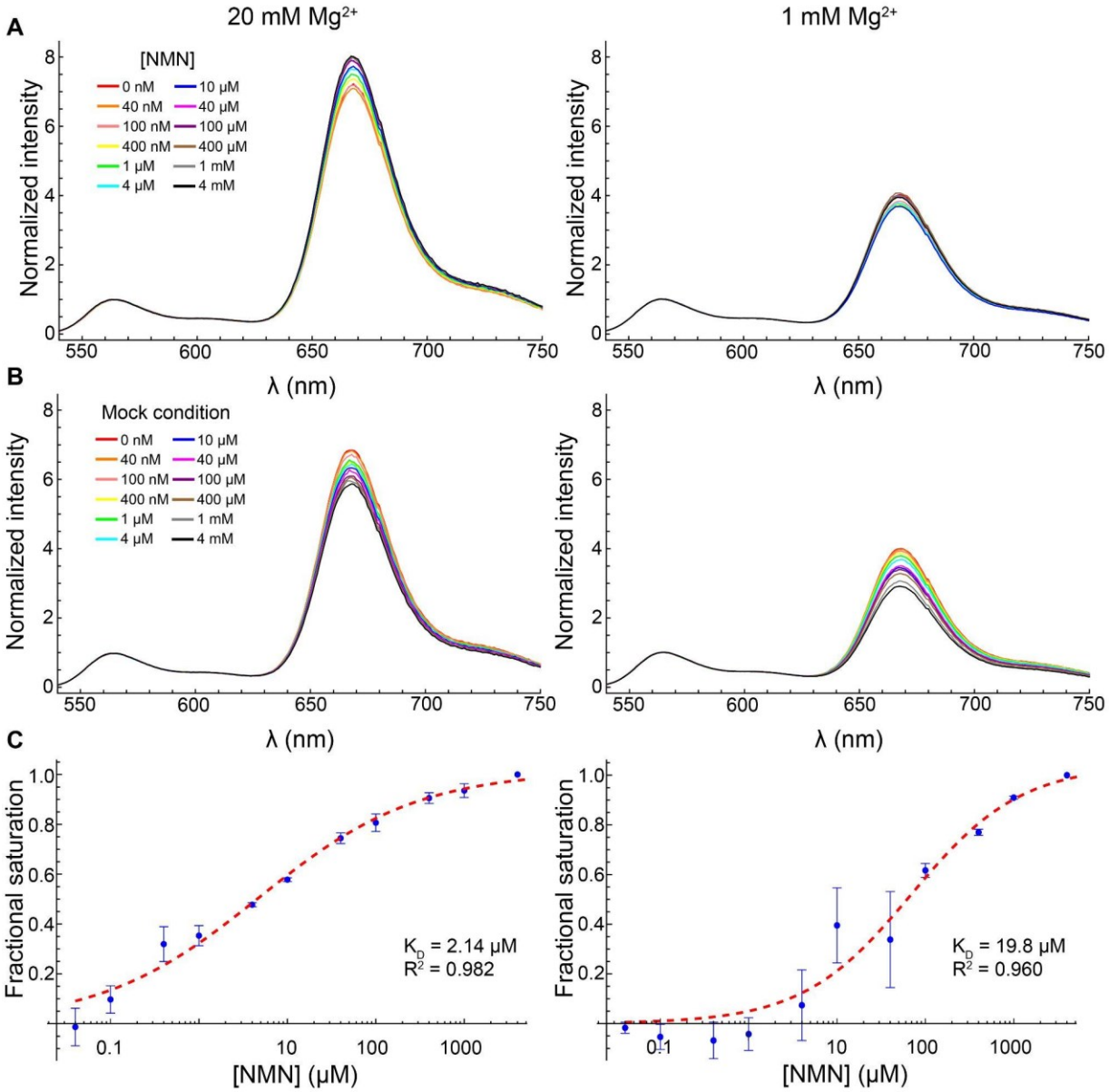


Figure 5.5. Mini-NAD$^+$-II bulk FRET titrations of NMN at 20 mM (left) and 1 mM Mg^{2+} (right). Emission scans of mini-NAD$^+$-II at increasing concentrations of NMN, normalized to the intensity of Cy3 (A). Emission scans from mock titrations of mini-NAD$^+$-II to correct for Cy3 and Cy5 photobleaching (B). Fractional saturation of mini-NAD$^+$-II at each NMN concentration from three replicates, fit with the Hill equation (C).

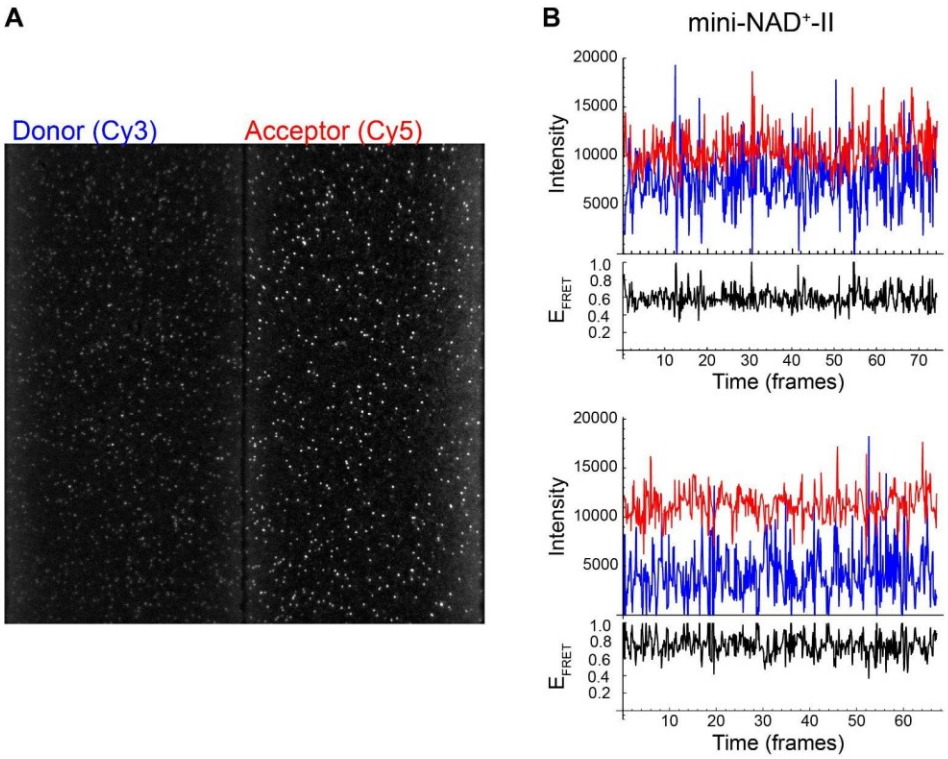


Figure 5.6. Example smFRET data of mini-NAD$^+$-II. Example single-molecule image (A), with donor (blue) signal in the left channel and acceptor (red) signal in the right channel. Example traces (B) with donor signal in blue, acceptor signal in red, and calculated E_{FRET} in black.

5. Conclusions

The NAD$^+$-II riboswitch has complex, fast dynamics, as observed by smFRET. In comparison, the mini-NAD$^+$-II riboswitch, which is more abundant across species, appears in preliminary smFRET data to be more static in nature, with less overall sampling of different states. Bulk FRET titrations support reports that mini-NAD$^+$-II binds NMN more strongly than the NAD$^+$-II riboswitch. Additionally, the ability of mini-NAD$^+$-II to bind ligand at physiological Mg^{2+} concentrations may explain its greater prevalence across species.

6. Future Directions

Many questions remain unanswered, in regard to both the NAD^+-II riboswitch and the mini-NAD^+-II riboswitch. For the NAD^+-II riboswitch, further analysis of smFRET data is required to determine the kinetics of conformational transitions, including refinement of the corrections applied and reported above. These corrections have the effect of adjusting the observed FRET efficiencies of states, and appear to uncover a fourth state that was not explicitly modeled in previous work.[101] The models used for HMM will be re-worked to account for these adjustments, and additional analyses will be done to investigate the underlying conformational dynamics of NAD^+-II.

The mini-NAD^+-II riboswitch was first reported very recently, and there are many aspects of its structure and function to investigate further. Bulk FRET titrations with NAD^+ will be performed and compared to NAD^+-II bulk FRET data and to mini-NAD^+-II smFRET. Additional smFRET studies are planned for mini-NAD^+-II, to uncover any conformational dynamics which might exist, and how they change in the presence of NMN, NAD^+, and ε-NAD^+. The same corrections mentioned for NAD^+-II above will be determined and applied here, and the data will be subjected to HMM to determine the kinetics of any transitions. In addition to conditions matching those explored for the NAD^+-II riboswitch, data taken at lower Mg^{2+} concentrations will highlight a distinction between the two of potential biological relevance.

For both the NAD^+-II and mini-NAD^+-II riboswitches, further investigations utilizing bulk fluorescence spectroscopy with ε-NAD^+, an emissive analogue of NAD^+ are of interest. Due to its emissive nature, ε-NAD^+ can be used to report on changes to ligand structure upon binding, as well as interactions with the RNA that alter its photophysical properties. Initial investigations showed that ε-NAD^+ binding exerts the same impacts on the structure and dynamics of NAD^+-II

as NAD^+, suggesting it acts as a faithful ligand analogue.[102] These experiments will be performed on mini-NAD^+-II as well to assess how analogous the two ligands are, adding context to future results. This insight could inform the design of small molecules used to target NAD^+ riboswitches as antibiotics, or tools to accurately sense NAD^+ and related molecules.

CHAPTER VI: CONCLUDING REMARKS

Non-coding RNAs have diverse biological functions, which are directly related to their structures. Due to this diversity, there is not a singular best technique for studying RNA structure. The work I presented in this dissertation investigated multiple non-coding RNAs, using different single-molecule techniques, as well as spectroscopy and chemical probing. Here I summarize these investigations.

In Chapter 3, I presented a study of SiM-KARTS, a single-molecule microscopy technique well suited for studying complex systems such as lncRNA. Using Bvht, a model RNA based on the lncRNA *Braveheart*, we explored the design of SiM-KARTS experiments to adapt the technique to lncRNA. Specifically, we found that the incorporation of LNA residues in SiM-KARTS probes is beneficial, and these probes are able to report additional structural information through their bound dwell times. Further analysis of SiM-KARTS data collected using LNA probes revealed that such probes allow for the trace-by-trace determination of structure. Spectroscopic techniques including absorbance based thermal denaturation and circular dichroism were utilized to support the observations made in the SiM-KARTS experiments.

In Chapter 4, I applied the principles of SiM-KARTS probe design reported in Chapter 3 to study the structure and dynamics of *Mhrt*, a heart-specific lncRNA. Due to its complexity, preliminary investigations of secondary structure utilized the chemical probing technique SHAPE. While additional SHAPE experiments could reveal more about the secondary structure, the data reported here showed a flexible poly-A region within *Mhrt*. This region was chosen as a SiM-KARTS probe target site, and an LNA containing probe was designed. I showed that this probe is well designed for the system, supporting the conclusions of Chapter 3.

In Chapter 5, I presented an investigation of a different type of non-coding RNA, the riboswitch, using smFRET and bulk fluorescence spectroscopy. Previously reported smFRET data for NAD^+-II proposed at least three distinct FRET states which undergo rapid transitions. By applying an additional correction factor and filter to the data, I proposed that a model containing four states is more representative of the system. I presented additional bulk FRET measurements which indicate the NAD^+-II construct used retains its ability to bind NAD^+. I reported preliminary investigations of the newly reported mini-NAD^+-II riboswitch. Bulk FRET measurements showed the labeled construct retains its ability to bind ligands with similar affinities to literature values. Importantly, I showed that, unlike NAD^+-II, mini-NAD^+-II can bind ligands in the presence of low concentrations of magnesium ions. Mini-NAD^+-II smFRET studies are ongoing, but I presented initial data which suggest the RNA is primarily static in nature.

APPENDIX: SUPPORTING INFORMATION FOR CHAPTER III- ALTERNATIVE PROBE CHEMISTRIES FOR SINGLE-MOLECULE ANALYSIS OF LONG NON-CODING RNA

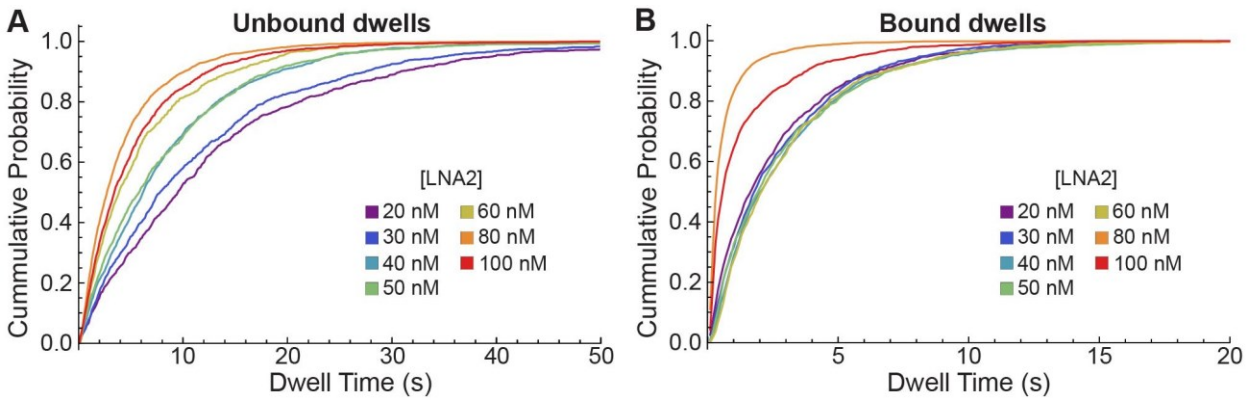
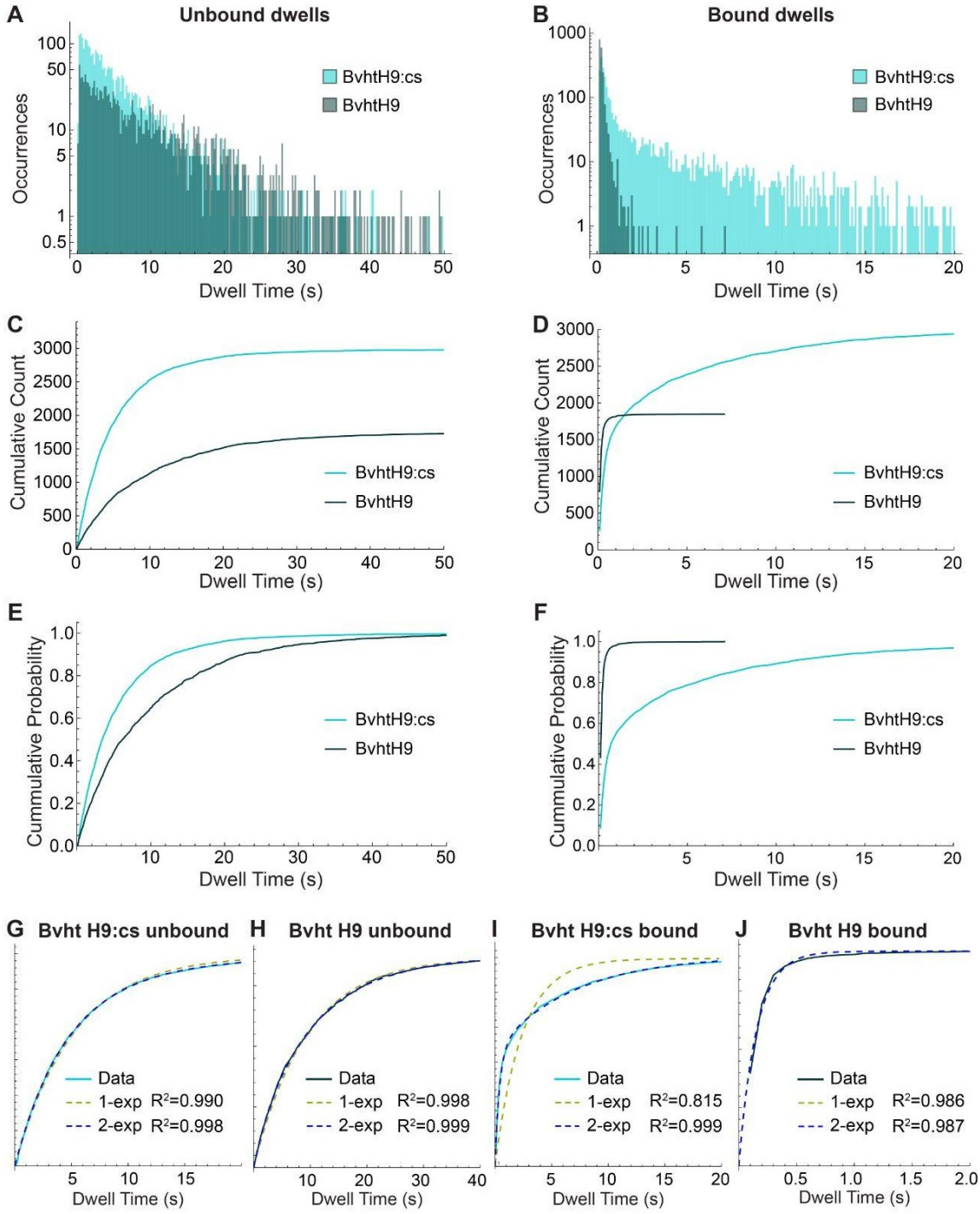


Figure A1. Impact of probe concentration on dwell times of LNA2 binding to Bvht H9:cs. *A*, Cumulative distributions of unbound dwell times at indicated LNA2 concentrations. *B*, Bound dwell times at indicated LNA2 concentrations. As expected, unbound dwell time decreases with increasing probe concentration while bound dwell time remains unchanged up to 60 nM probe. The drastic increase in short bound dwells at 80 and 100 nM LNA2 indicates that fluctuations in background noise due to the fluorescence of freely diffusing probe are being identified as binding events. Hence, 60 nM was chosen as a concentration that yielded a substantial number of binding events per trace (quantified in main text tables) without introducing excessive noise.

Figure A2. (next page) Overview of dwell time distribution analysis. Data on LNA2 in the presence of 1 mM Mg^{2+} is used as an example. *A*, Histogram displaying the number of times an unbound segment of a trace lasted for the corresponding dwell time indicated on the x-axis. For clearer visualization, dwell times are binned into 2-frame (0.2 s) increments. Includes all "bracketed" unbound segments of all traces recorded under this experimental condition (segments that are not truncated by the beginning or end of the trace). The y-axis is on a logarithmic scale. *B*, Same as *A* for bound segments of traces. *C*, Cumulative distribution plot displaying the *number* of unbound trace segments with duration less than or equal to the corresponding dwell time indicated on the x-axis. Curves begin at the x-value corresponding to the frame rate of the movie (0.1 s), indicating a dwell time of 1 frame. This plot preserves all of the information present in an unbinned dwell time histogram. *D*, Same as *C* but for bound segments of traces. Curves end at the x-value corresponding to the longest observed bound dwell time. *E*, Cumulative distribution plot displaying the *probability* that an unbound segment of a trace has a duration less than or equal to the corresponding time indicated on the x-axis. This plot preserves all of the information in the histogram except for the total number of dwells that were analyzed. *F*, Same as *E* but for bound segments of traces. *G-J*, Exponential fitting of cumulative distribution functions, showing cumulative count curves (solid) and single- (dashed yellow) and double- (dashed blue) exponential fits. The R^2 value for each fit is provided. *G*, Bvht H9:cs unbound dwells. *H*, Bvht H9 unbound dwells. *I*, Bvht H9:cs bound dwells. *J*, Bvht H9 bound dwells.


A
Unbound dwells
Occurrences
Dwell Time (s)
BvhtH9:cs
BvhtH9
B
Bound dwells
C
Cumulative Count
D
E
Cummulative Probability
F
G Bvht H9:cs unbound
Data
1-exp R²=0.990
2-exp R²=0.998
H Bvht H9 unbound
1-exp R²=0.998
2-exp R²=0.999
I Bvht H9:cs bound
1-exp R²=0.815
2-exp R²=0.999
J Bvht H9 bound
1-exp R²=0.986
2-exp R²=0.987

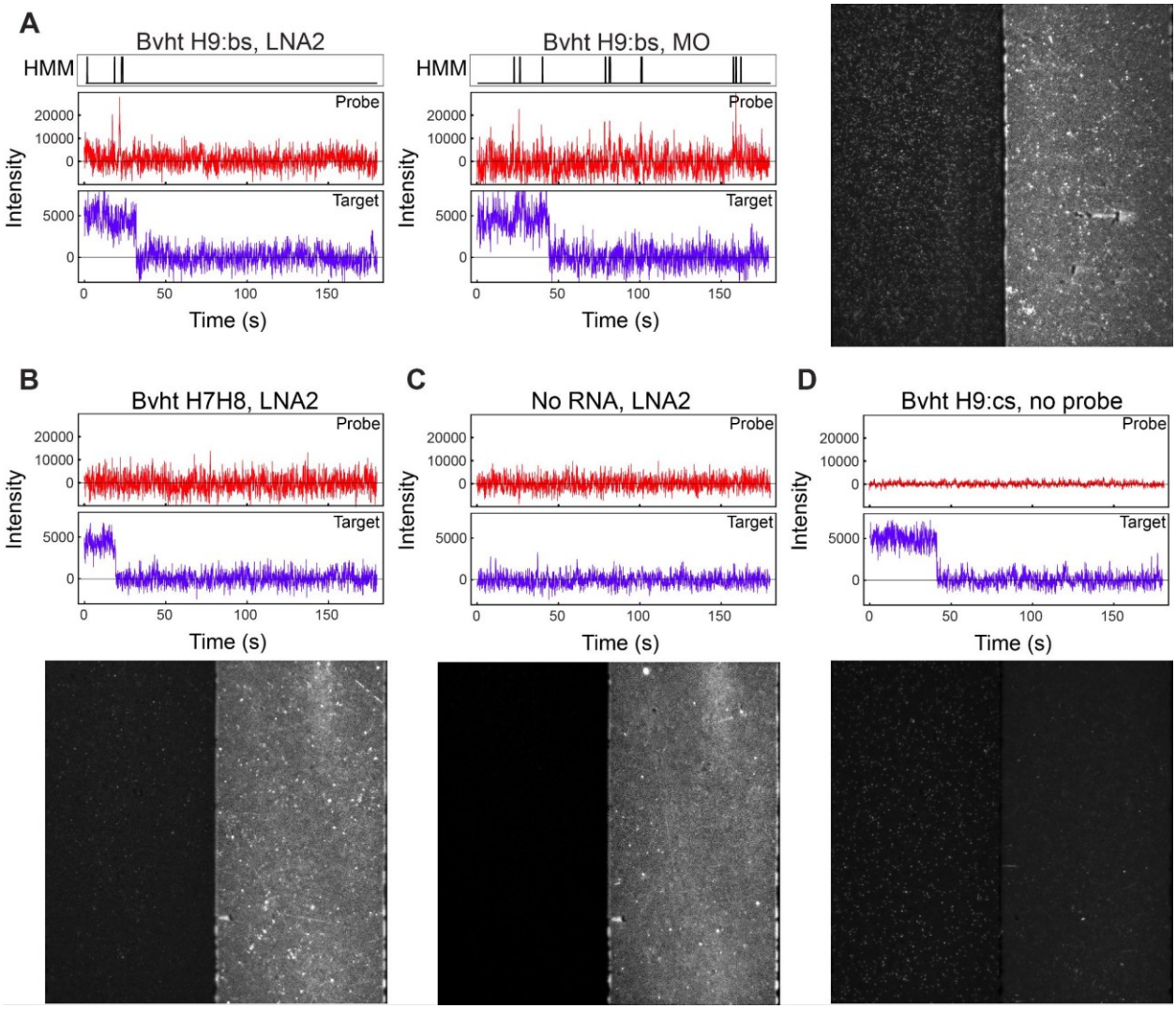


Figure A3. Representative single-molecule traces of control conditions. *A*, Representative single-molecule traces for LNA2 (left) or MO (middle) probing Bvht H9:bs. Top: HMM idealization of red (probe) channel trace. Middle: Red (probe) channel trace; Bottom: Green (target) channel trace; Right: example image from SiM-KARTS data collection of LNA2 probing Bvht H9:bs.*B,* Same as *A* but for LNA2 probing Bvht H7H8, with example image below. (N=63 traces analyzed; D/N=2 average bound dwells per trace). *C,* Same as *B* but for LNA2 with no target RNA present (N=167; D/N=1). *D,* Same as *B* but for Bvht H9:cs in the absence of probe (N=124; D/N=1).

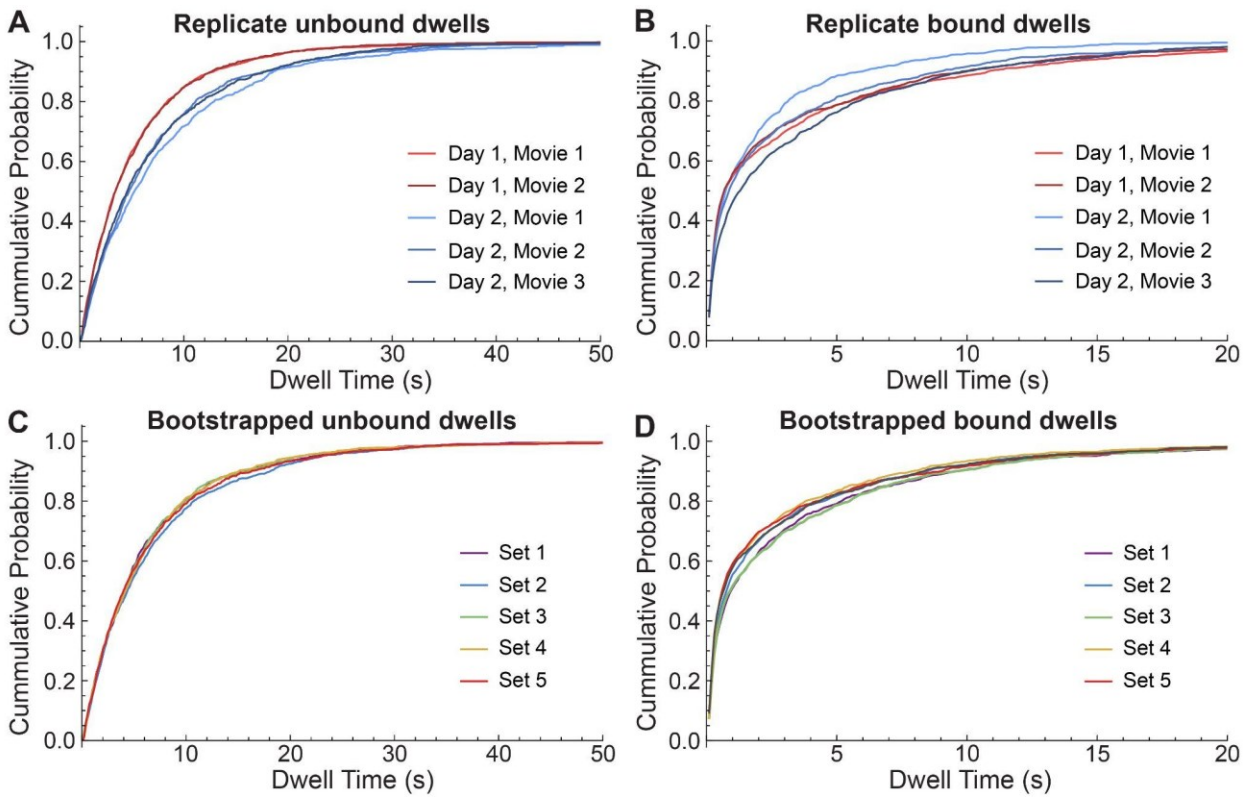


Figure A4. SiM-KARTS data reproducibility. *A*, Cumulative distributions of unbound dwell times for LNA2 binding to Bvht H9:cs at 1 mM Mg^{2+} across replicate movies (variations in shade) recorded on two days (red vs blue). *B*, Bound dwell times for LNA2 binding to Bvht H9:cs at 1 mM Mg^{2+} across replicate movies recorded on two days. *C,* Data from *A* after bootstrapping analysis performed by randomly dividing all traces included in panels *A*-*B* into five sets and analyzing each independently. *D*, Data from *B* after bootstrapping analysis.

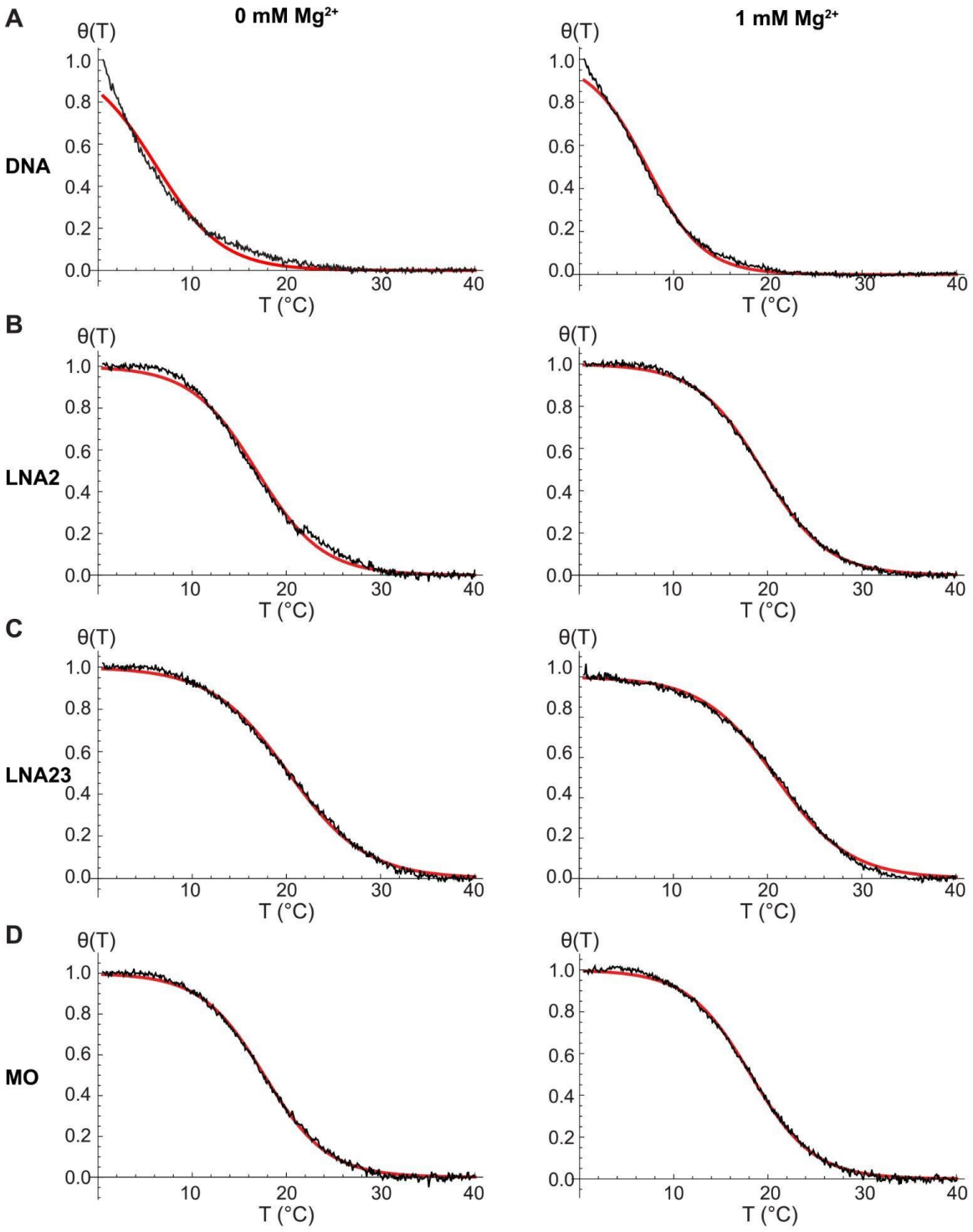


Figure A5. Processed melting curves represented as fraction bound θ vs. temperature T (black) and fits (red) recorded at 0 mM $MgCl_2$ (left column) or 1 mM $MgCl_2$ (right column). *A*, DNA probe. *B*, LNA2. *C*, LNA23. *D*, MO.

Table A1. Detailed fitting results for the Bvht H9 data reported in Table 3.2. $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are the time constants and associated weights obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit is presented as R^2.

Probe	$[Mg^{2+}]$ (mM)	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}
DNA	0	9.0	0.999	0.146	1.00	-	-	0.988
DNA	1	8.9	0.994	0.170	1.00	-	-	0.978
LNA2	0	9.9	0.998	0.14	1.00	-	-	0.989
LNA2	1	9.4	0.998	0.15	1.00	-	-	0.987
LNA23	0	5.8	0.997	0.213	1.00	-	-	0.978
LNA23	1	5.4	0.996	0.222	1.00	-	-	0.977
MO	0	7.6	0.997	1.22	0.13	0.30	0.87	0.997
MO	1	9.1	0.992	2.92	0.14	0.26	0.86	0.997

Table A2. Detailed fitting results for the Bvht H9:cs data reported in Table 3.2. $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are the time constants and associated weights obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit is presented as R^2.

Probe	$[Mg^{2+}]$ (mM)	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}
DNA	0	9.5	0.999	0.69	0.19	0.21	0.81	0.996
DNA	1	12	0.993	0.96	0.12	0.25	0.88	0.996
LNA2	0	3.7	0.992	2.60	0.45	0.39	0.55	0.998
LNA2	1	5.2	0.998	6.93	0.45	0.41	0.55	0.999
LNA23	0	3.9	0.972	7.2	0.62	0.6	0.38	0.998
LNA23	1	8.0	0.996	16.1	0.49	1.3	0.51	0.999
MO	0	5.4	0.995	2.96	0.20	0.42	0.80	0.995
MO	1	7.6	0.997	2.39	0.15	0.26	0.85	0.992

Table A3. Detailed fitting results for the Bvht H9:bs data reported in Table 3.3. $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are the time constants and associated weights obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit is presented as R^2.

Probe	$[Mg^{2+}]$ (mM)	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}
LNA2	0	19.6	0.980	2.11	0.11	0.26	0.89	0.992
LNA2	1	21.5	0.982	1.21	0.18	0.28	0.82	0.994
MO	0	18.5	0.963	1.66	0.18	0.29	0.82	0.993
MO	1	17.8	0.983	2.77	0.13	0.26	0.87	0.992

Table A4. Detailed fitting results for experimental replicates of LNA2 binding to Bvht H9:cs at 1 mM Mg^{2+}. $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are the time constants and associated weights obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit is presented as R^2. "N" indicates the number of single-molecule traces included in each condition. "D/N" indicates the average number of bound dwells observed per trace.

Day	Movie	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}	N	D/N
1	1	5.19	0.998	7.02	0.46	0.41	0.54	0.998	95	18
1	2	5.16	0.999	6.68	0.44	0.40	0.56	0.998	71	19
1[1]	1,2	5.20	0.998	6.93	0.45	0.41	0.55	0.999	166	18
2	1	6.97	0.998	5.51	0.48	0.44	0.52	0.999	71	15
2	2	7.71	0.998	3.64	0.50	0.44	0.50	0.997	63	15
2	3	6.91	0.998	5.71	0.58	0.43	0.42	0.999	73	15
2	1,2,3	7.23	0.998	5.13	0.50	0.46	0.50	0.999	207	15

[1]Data in this row reproduced from Table A2.

Table A5. Detailed fitting results for bootstrapped data sets of LNA2 binding to Bvht H9:cs at 1 mM Mg^{2+}. $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are the time constants and associated weights obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit is presented as R^2. "N" indicates the number of single-molecule traces included in each condition. "D/N" indicates the average number of bound dwells observed per trace.

Set	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}	N	D/N
1	6.05	0.997	6.73	0.47	0.40	0.53	0.998	74	13
2	5.99	0.996	6.29	0.48	0.40	0.52	0.999	74	14
3	6.03	0.997	5.58	0.45	0.37	0.55	0.999	75	14
4	5.94	0.999	5.40	0.50	0.41	0.50	0.997	75	14
5	5.55	0.996	6.47	0.49	0.34	0.51	0.998	75	13

Table A6. Detailed fitting results for the Bvht H9 data reported in Table 3.6 (LNA2 salt series). $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit is presented as R^2.

$[Mg^{2+}]$ (mM)	$[Na^+]$ (mM)	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}
0[1]	100	9.9	0.998	0.14	1.00	-	-	0.989
1[1]	100	9.4	0.998	0.15	1.00	-	-	0.987
1	200	9.6	0.999	0.19	1.00	-	-	0.990
1	300	9.3	0.993	0.21	1.00	-	-	0.989
1	500	10.0	0.993	0.21	0.94	1.23	0.06	0.995
20	100	9.1	0.993	0.17	1.00	-	-	0.988

[1]Data in these rows reproduced from Table A1.

Table A7. Detailed fitting results for the Bvht H9:cs data reported in Table 3.6 (LNA2 salt series). $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit is presented as R^2.

$[Mg^{2+}]$ (mM)	$[Na^+]$ (mM)	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}
0[1]	100	3.7	0.992	2.60	0.45	0.39	0.55	0.998
1[1]	100	5.2	0.998	6.93	0.45	0.41	0.55	0.999
1	200	6.8	0.999	6.6	0.51	0.5	0.49	0.999
1	300	5.0	0.999	6.12	0.54	0.44	0.46	0.999
1	500	5.7	0.996	7.7	0.48	0.5	0.52	0.998
20	100	7.4	0.971	0.47	0.11	0.10	0.89	1.000

[1]Data in these rows reproduced from Table A2.

Table A8. Detailed fitting results for the Bvht H9 data reported in Table 3.6 (MO salt series). $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit is presented as R^2.

$[Mg^{2+}]$ (mM)	$[Na^+]$ (mM)	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}
0[1]	100	7.6	0.997	1.22	0.13	0.30	0.87	0.997
1[1]	100	9.0	0.992	2.92	0.14	0.26	0.86	0.997
1	500	7.7	0.991	2.65	0.12	0.33	0.88	0.990

[1]Data in these rows reproduced from Table A1.

Table A9. Detailed fitting results for the Bvht H9:cs data reported in Table 3.6 (MO salt series). $\tau_{unbound}$ is obtained from a single-exponential fit to the cumulative distribution of unbound dwell times. $\tau_{bound,1}$, $\tau_{bound,2}$, w_1, and w_2 are obtained from a double-exponential fit to the cumulative distribution of bound dwell times. Goodness of fit to the respective cumulative distribution is presented as R^2.

$[Mg^{2+}]$ (mM)	$[Na^+]$ (mM)	$\tau_{unbound}$ (s)	$R^2_{unbound}$	$\tau_{bound,1}$ (s)	$w_{bound,1}$	$\tau_{bound,2}$ (s)	$w_{bound,2}$	R^2_{bound}
0[1]	100	5.4	0.995	2.96	0.20	0.42	0.80	0.995
1[1]	100	7.6	0.997	2.39	0.15	0.26	0.85	0.992
1	500	7.5	0.998	2.82	0.11	0.27	0.89	0.991

[1]Data in these rows reproduced from Table A2.

REFERENCES CITED

1. Long Non Coding RNA Biology | SpringerLink. https://link.springer.com/book/10.1007/978-981-10-5203-3 (accessed 2025-12-15).

2. Zaug, A. J.; Cech, T. R. In Vitro Splicing of the Ribosomal RNA Precursor in Nuclei of Tetrahymena. Cell 1980, 19 (2), 331–338. https://doi.org/10.1016/0092-8674(80)90507-3.

3. Poller, W.; Sahoo, S.; Hajjar, R.; Landmesser, U.; Krichevsky, A. M. Exploration of the Noncoding Genome for Human-Specific Therapeutic Targets—Recent Insights at Molecular and Cellular Level. Cells 2023, 12 (22), 2660. https://doi.org/10.3390/cells12222660.

4. Ray, S.; Widom, J. R.; Walter, N. G. Life under the Microscope: Single-Molecule Fluorescence Highlights the RNA World. Chem. Rev. 2018, 118 (8), 4120–4155. https://doi.org/10.1021/acs.chemrev.7b00519.

5. Johnsson, P.; Lipovich, L.; Grandér, D.; Morris, K. V. Evolutionary Conservation of Long Non-Coding RNAs; Sequence, Structure, Function. Biochim. Biophys. Acta BBA - Gen. Subj. 2014, 1840 (3), 1063–1071. https://doi.org/10.1016/j.bbagen.2013.10.035.

6. Yang, Y.; Zhan, L.; Zhang, W.; Sun, F.; Wang, W.; Tian, N.; Bi, J.; Wang, H.; Shi, D.; Jiang, Y.; Zhang, Y.; Jin, Y. RNA Secondary Structure in Mutually Exclusive Splicing. Nat. Struct. Mol. Biol. 2011, 18 (2), 159–168. https://doi.org/10.1038/nsmb.1959.

7. Jhunjhunwala, A.; Ali, Z.; Bhattacharya, S.; Halder, A.; Mitra, A.; Sharma, P. On the Nature of Nucleobase Stacking in RNA: A Comprehensive Survey of Its Structural Variability and a Systematic Classification of Associated Interactions. J. Chem. Inf. Model. 2021, 61 (3), 1470–1480. https://doi.org/10.1021/acs.jcim.0c01225.

8. Lai, W.-J. C.; Kayedkhordeh, M.; Cornell, E. V.; Farah, E.; Bellaousov, S.; Rietmeijer, R.; Salsi, E.; Mathews, D. H.; Ermolenko, D. N. mRNAs and lncRNAs Intrinsically Form Secondary Structures with Short End-to-End Distances. Nat. Commun. 2018, 9 (1), 4328. https://doi.org/10.1038/s41467-018-06792-z.

9. Vicens, Q.; Kieft, J. S. Thoughts on How to Think (and Talk) about RNA Structure. Proc. Natl. Acad. Sci. 2022, 119 (17), e2112677119. https://doi.org/10.1073/pnas.2112677119.

10. Lorenz, C.; Lünse, C. E.; Mörl, M. tRNA Modifications: Impact on Structure and Thermal Adaptation. Biomolecules 2017, 7 (2), 35. https://doi.org/10.3390/biom7020035.

11. Green, R.; Noller, H. F. Ribosomes and Translation. Annu. Rev. Biochem. 1997, 66, 679–716. https://doi.org/10.1146/annurev.biochem.66.1.679.

12. Zhang, Y.; Dong, X.; Guo, X.; Li, C.; Fan, Y.; Liu, P.; Yuan, D.; Ma, X.; Wang, J.; Zheng, J.; Li, H.; Gao, P. LncRNA-BC069792 Suppresses Tumor Progression by Targeting KCNQ4 in Breast Cancer. Mol. Cancer 2023, 22 (1), 41. https://doi.org/10.1186/s12943-023-01747-5.

13. Zhang, X.; Hong, R.; Chen, W.; Xu, M.; Wang, L. The Role of Long Noncoding RNA in Major Human Disease. Bioorganic Chem. 2019, 92, 103214. https://doi.org/10.1016/j.bioorg.2019.103214.

14. Zhou, H.; Wang, B.; Yang, Y.; Jia, Q.; Zhang, A.; Qi, Z.; Zhang, J. Long Noncoding RNAs in Pathological Cardiac Remodeling: A Review of the Update Literature. BioMed Res. Int. 2019, 2019, e7159592. https://doi.org/10.1155/2019/7159592.

15. Effect of lncRNA MALAT1 on Rats with Myocardial Infarction through Regulating ERK/MAPK Signaling Pathway.

16. Klattenhoff, C. A.; Scheuermann, J. C.; Surface, L. E.; Bradley, R. K.; Fields, P. A.; Steinhauser, M. L.; Ding, H.; Butty, V. L.; Torrey, L.; Haas, S.; Abo, R.; Tabebordbar, M.; Lee, R. T.; Burge, C. B.; Boyer, L. A. Braveheart, a Long Noncoding RNA Required for Cardiovascular Lineage Commitment. Cell 2013, 152 (3), 570–583. https://doi.org/10.1016/j.cell.2013.01.003.

17. Han, P.; Li, W.; Lin, C.-H.; Yang, J.; Shang, C.; Nurnberg, S. T.; Jin, K. K.; Xu, W.; Lin, C.-Y.; Lin, C.-J.; Xiong, Y.; Chien, H.-C.; Zhou, B.; Ashley, E.; Bernstein, D.; Chen, P.-S.; Chen, H.-S. V.; Quertermous, T.; Chang, C.-P. A Long Noncoding RNA Protects the Heart from Pathological Hypertrophy. Nature 2014, 514 (7520), 102–106. https://doi.org/10.1038/nature13596.

18. Xue, Z.; Hennelly, S.; Doyle, B.; Gulati, A. A.; Novikova, I. V.; Sanbonmatsu, K. Y.; Boyer, L. A. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. Mol. Cell 2016, 64 (1), 37–50. https://doi.org/10.1016/j.molcel.2016.08.010.

19. Kim, D. N.; Thiel, B. C.; Mrozowich, T.; Hennelly, S. P.; Hofacker, I. L.; Patel, T. R.; Sanbonmatsu, K. Y. Zinc-Finger Protein CNBP Alters the 3-D Structure of lncRNA Braveheart in Solution. Nat. Commun. 2020, 11 (1), 148. https://doi.org/10.1038/s41467-019-13942-4.

20. Kontaraki, J. E.; Marketou, M. E.; Kochiadakis, G. E.; Maragkoudakis, S.; Konstantinou, J.; Vardas, P. E.; Parthenakis, F. I. The Long Non-Coding RNAs MHRT, FENDRR and CARMEN, Their Expression Levels in Peripheral Blood Mononuclear Cells in Patients with Essential Hypertension and Their Relation to Heart Hypertrophy. Clin. Exp. Pharmacol. Physiol. 2018, 45 (11), 1213–1217. https://doi.org/10.1111/1440-1681.12997.

21. Zhang, L.; Wu, Y.-J.; Zhang, S.-L. Circulating lncRNA MHRT Predicts Survival of Patients with Chronic Heart Failure. J. Geriatr. Cardiol. 2019, 16 (11), 818–821. https://doi.org/10.11909/j.issn.1671-5411.2019.11.006.

22. Xuan, L.; Sun, L.; Zhang, Y.; Huang, Y.; Hou, Y.; Li, Q.; Guo, Y.; Feng, B.; Cui, L.; Wang, X.; Wang, Z.; Tian, Y.; Yu, B.; Wang, S.; Xu, C.; Zhang, M.; Du, Z.; Lu, Y.; Yang, B. F. Circulating Long Non-Coding RNAs NRON and MHRT as Novel Predictive Biomarkers of Heart Failure. J. Cell. Mol. Med. 2017, 21 (9), 1803–1814. https://doi.org/10.1111/jcmm.13101.

23. Panchal, V.; Brenk, R. Riboswitches as Drug Targets for Antibiotics. Antibiotics 2021, 10 (1). https://doi.org/10.3390/antibiotics10010045.

24. Chiarugi, A.; Dölle, C.; Felici, R.; Ziegler, M. The NAD Metabolome — a Key Determinant of Cancer Cell Biology. Nat. Rev. Cancer 2012, 12 (11), 741–752. https://doi.org/10.1038/nrc3340.

25. Malkowski, S. N.; Spencer, T. C. J.; Breaker, R. R. Evidence That the nadA Motif Is a Bacterial Riboswitch for the Ubiquitous Enzyme Cofactor NAD+. RNA 2019, 25 (12), 1616–1627. https://doi.org/10.1261/rna.072538.119.

26. Panchapakesan, S. S. S.; Corey, L.; Malkowski, S. N.; Higgs, G.; Breaker, R. R. A Second Riboswitch Class for the Enzyme Cofactor NAD+. RNA 2021, 27 (1), 99–105. https://doi.org/10.1261/rna.077891.120.

27. Chen, H.; Egger, M.; Xu, X.; Flemmich, L.; Krasheninina, O.; Sun, A.; Micura, R.; Ren, A. Structural Distinctions between NAD+ Riboswitch Domains 1 and 2 Determine Differential Folding and Ligand Binding. Nucleic Acids Res. 2020, 48 (21), 12394–12406. https://doi.org/10.1093/nar/gkaa1029.

28. Xu, X.; Egger, M.; Li, C.; Chen, H.; Micura, R.; Ren, A. Structure-Based Investigations of the NAD+-II Riboswitch. Nucleic Acids Res. 2023, 51 (1), 54–67. https://doi.org/10.1093/nar/gkac1227.

29. Eichhorn, C. D.; Kang, M.; Feigon, J. Structure and Function of preQ1 Riboswitches. BBA - Gene Regul. Mech. 2014, 1839 (10), 939–950. https://doi.org/10.1016/j.bbagrm.2014.04.019.

30. Hertz, L. M.; White, E. N.; Kuznedelov, K.; Cheng, L.; Yu, A. M.; Kakkaramadam, R.; Severinov, K.; Chen, A.; Lucks, J. B. The Effect of Pseudoknot Base Pairing on Cotranscriptional Structural Switching of the Fluoride Riboswitch. Nucleic Acids Res. 2024, 52 (8), 4466–4482. https://doi.org/10.1093/nar/gkae231.

31. King, C. G.; Cheng, K. P.; Breaker, R. R. Miniature NAD+-II Riboswitches Control Bacterial Genes for Nicotinamide Salvage and de Novo NAD+ Biosynthesis. RNA 2026, 32 (2), 162–170. https://doi.org/10.1261/rna.080744.125.

32. Cao, X.; Zhang, Y.; Ding, Y.; Wan, Y. Identification of RNA Structures and Their Roles in RNA Functions. Nat. Rev. Mol. Cell Biol. 2024, 25 (10), 784–801. https://doi.org/10.1038/s41580-024-00748-6.

33. Hingerty, B.; Brown, R. S.; Jack, A. Further Refinement of the Structure of Yeast tRNAPhe. J. Mol. Biol. 1978, 124 (3), 523–534. https://doi.org/10.1016/0022-2836(78)90185-7.

34. Chen, Y.; Pollack, L. SAXS Studies of RNA: Structures, Dynamics, and Interactions with Partners. WIREs RNA 2016, 7 (4), 512–526. https://doi.org/10.1002/wrna.1349.

35. Gabashvili, I. S.; Agrawal, R. K.; Spahn, C. M. T.; Grassucci, R. A.; Svergun, D. I.; Frank, J.; Penczek, P. Solution Structure of the E. Coli 70S Ribosome at 11.5 Å Resolution. Cell 2000, 100 (5), 537–549. https://doi.org/10.1016/S0092-8674(00)80690-X.

36. Barnwal, R. P.; Yang, F.; Varani, G. Applications of NMR to Structure Determination of RNAs Large and Small. Arch. Biochem. Biophys. 2017, 628, 42–56. https://doi.org/10.1016/j.abb.2017.06.003.

37. Low, J. T.; Weeks, K. M. SHAPE-Directed RNA Secondary Structure Prediction. Methods 2010, 52 (2), 150–158. https://doi.org/10.1016/j.ymeth.2010.06.007.

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

39. McCann, J. J.; Choi, U. B.; Zheng, L.; Weninger, K.; Bowen, M. E. Optimizing Methods to Recover Absolute FRET Efficiency from Immobilized Single Molecules. Biophys. J. 2010, 99 (3), 961–970. https://doi.org/10.1016/j.bpj.2010.04.063.

40. McKinney, S. A.; Joo, C.; Ha, T. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. Biophys. J. 2006, 91 (5), 1941–1951. https://doi.org/10.1529/biophysj.106.082487.

41. Baum, L. E.; Petrie, T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. Ann. Math. Stat. 1966, 37 (6), 1554–1563.

42. Yoon, B.-J. Hidden Markov Models and Their Applications in Biological Sequence Analysis. Curr. Genomics 2009, 10 (6), 402–415. https://doi.org/10.2174/138920209789177575.

43. Milescu, L. S.; Nicolai, C.; Bannen, J. QuB Software, 2000. qub.mandelics.com/online/.

44. Rinaldi, A. J.; Lund, P. E.; Blanco, M. R.; Walter, N. G. The Shine-Dalgarno Sequence of Riboswitch-Regulated Single mRNAs Shows Ligand-Dependent Accessibility Bursts. Nat. Commun. 2016, 7 (1), 8976. https://doi.org/10.1038/ncomms9976.

45. Eden, U. T.; Kramer, M. A. Drawing Inferences from Fano Factor Calculations. J. Neurosci. Methods 2010, 190 (1), 149–152. https://doi.org/https://doi.org/10.1016/j.jneumeth.2010.04.012.

46. Welty, R.; Schmidt, A.; Walter, N. G. Probing Transient Riboswitch Structures via Single Molecule Accessibility Analysis. In RNA Structure and Dynamics; Ding, J., Stagno, J. R., Wang, Y.-X., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2023; Vol. 2568, pp 37–51. https://doi.org/10.1007/978-1-0716-2687-0_4.

47. Gourévitch, B.; Eggermont, J. J. A Nonparametric Approach for Detection of Bursts in Spike Trains. J. Neurosci. Methods 2007, 160 (2), 349–358. https://doi.org/10.1016/j.jneumeth.2006.09.024.

48. Schroeder, S. J.; Turner, D. H. Optical Melting Measurements of Nucleic Acid Thermodynamics. Methods Enzymol. 2009, 468, 371–387. https://doi.org/10.1016/S0076-6879(09)68017-4.

49. McGinnis, J. L.; Dunkle, J. A.; Cate, J. H. D.; Weeks, K. M. The Mechanisms of RNA SHAPE Chemistry. J. Am. Chem. Soc. 2012, 134 (15), 6617–6624. https://doi.org/10.1021/ja2104075.

50. Much, C.; Lasda, E. L.; Pereira, I. T.; Vallery, T. K.; Ramirez, D.; Lewandowski, J. P.; Dowell, R. D.; Smallegan, M. J.; Rinn, J. L. The Temporal Dynamics of lncRNA Firre-Mediated Epigenetic and Transcriptional Regulation. Nat. Commun. 2024, 15 (1), 6821. https://doi.org/10.1038/s41467-024-50402-0.

51. Dykes, I. M.; Emanueli, C. Transcriptional and Post-Transcriptional Gene Regulation by Long Non-Coding RNA. Genomics Proteomics Bioinformatics 2017, 15 (3), 177–186. https://doi.org/10.1016/j.gpb.2016.12.005.

52. He, R.-Z.; Luo, D.-X.; Mo, Y.-Y. Emerging Roles of lncRNAs in the Post-Transcriptional Regulation in Cancer. Genes Dis. 2019, 6 (1), 6–15. https://doi.org/10.1016/j.gendis.2019.01.003.

53. Mattick, J. S.; Amaral, P. P.; Carninci, P.; Carpenter, S.; Chang, H. Y.; Chen, L.-L.; Chen, R.; Dean, C.; Dinger, M. E.; Fitzgerald, K. A.; Gingeras, T. R.; Guttman, M.; Hirose, T.; Huarte, M.; Johnson, R.; Kanduri, C.; Kapranov, P.; Lawrence, J. B.; Lee, J. T.; Mendell, J. T.; Mercer, T. R.; Moore, K. J.; Nakagawa, S.; Rinn, J. L.; Spector, D. L.; Ulitsky, I.; Wan, Y.; Wilusz, J. E.; Wu, M. Long Non-Coding RNAs: Definitions, Functions, Challenges and Recommendations. Nat. Rev. Mol. Cell Biol. 2023, 24 (6), 430–447. https://doi.org/10.1038/s41580-022-00566-8.

54. Jiang, Z.; Yu, Q.; Luo, X. Identification of Long Non-Coding RNA MVIH as a Prognostic Marker and Therapeutic Target in Acute Myeloid Leukemia. J. Clin. Lab. Anal. 2020, 34 (4), e23113. https://doi.org/10.1002/jcla.23113.

55. Lu, G.; Zhang, Y. Long Non-Coding RNA ATB Promotes Human Non-Small Cell Lung Cancer Proliferation and Metastasis by Suppressing miR-141-3p. PLOS ONE 2020, 15 (2), e0229118. https://doi.org/10.1371/journal.pone.0229118.

56. McDonel, P.; Guttman, M. Approaches for Understanding the Mechanisms of Long Noncoding RNA Regulation of Gene Expression. Cold Spring Harb. Perspect. Biol. 2019, 11 (12), a032151. https://doi.org/10.1101/cshperspect.a032151.

57. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S. Toward Dynamic Structural Biology: Two Decades of Single-Molecule Förster Resonance Energy Transfer. Science 2018. https://doi.org/10.1126/science.aan1133.

58. Chauvier, A.; Cabello-Villegas, J.; Walter, N. G. Probing RNA Structure and Interaction Dynamics at the Single Molecule Level. Methods 2019, 162–163, 3–11. https://doi.org/10.1016/j.ymeth.2019.04.002.

59. Jenkins, J. L.; Krucinska, J.; McCarty, R. M.; Bandarian, V.; Wedekind, J. E. Comparison of a PreQ1 Riboswitch Aptamer in Metabolite-Bound and Free States with Implications for Gene Regulation*. J. Biol. Chem. 2011, 286 (28), 24626–24637. https://doi.org/10.1074/jbc.M111.230375.

60. Suddala, K. C.; Rinaldi, A. J.; Feng, J.; Mustoe, A. M.; Eichhorn, C. D.; Liberman, J. A.; Wedekind, J. E.; Al-Hashimi, H. M.; Brooks, C. L., III; Walter, N. G. Single Transcriptional and Translational preQ1 Riboswitches Adopt Similar Pre-Folded Ensembles That Follow Distinct Folding Pathways into the Same Ligand-Bound Structure. Nucleic Acids Res. 2013, 41 (22), 10462–10475. https://doi.org/10.1093/nar/gkt798.

61. Awwad, D. A.; Rahmouni, A. R.; Aboul-ela, F. Protocol for Efficient Fluorescence 3′ End-Labeling of Native Noncoding RNA Domains. MethodsX 2020, 7, 101148. https://doi.org/10.1016/j.mex.2020.101148.

62. Chillón, I.; Marcia, M.; Legiewicz, M.; Liu, F.; Somarowthu, S.; Pyle, A. M. Native Purification and Analysis of Long RNAs. Methods Enzymol. 2015, 558, 3–37. https://doi.org/10.1016/bs.mie.2015.01.008.

63. Chauvier, A.; Dandpat, S. S.; Romero, R.; Walter, N. G. A Nascent Riboswitch Helix Orchestrates Robust Transcriptional Regulation through Signal Integration. Nat. Commun. 2024, 15 (1), 3955. https://doi.org/10.1038/s41467-024-48409-8.

64. Blechar, J.; de Jesus, V.; Fürtig, B.; Hengesbach, M.; Schwalbe, H. Shine–Dalgarno Accessibility Governs Ribosome Binding to the Adenine Riboswitch. ACS Chem. Biol. 2024, 19 (3), 607–618. https://doi.org/10.1021/acschembio.3c00435.

65. Suddala, K. C.; Price, I. R.; Dandpat, S. S.; Janeček, M.; Kührová, P.; Šponer, J.; Banáš, P.; Ke, A.; Walter, N. G. Local-to-Global Signal Transduction at the Core of a Mn2+ Sensing Riboswitch. Nat. Commun. 2019, 10 (1), 4304. https://doi.org/10.1038/s41467-019-12230-5.

66. Nakano, S.; Fujimoto, M.; Hara, H.; Sugimoto, N. Nucleic Acid Duplex Stability: Influence of Base Composition on Cation Effects. Nucleic Acids Res. 1999, 27 (14), 2957–2965. https://doi.org/10.1093/nar/27.14.2957.

67. Draper, D. E. A Guide to Ions and RNA Structure. RNA 2004, 10 (3), 335–343. https://doi.org/10.1261/rna.5205404.

68. Singh, S. K.; Koshkin, A. A.; Wengel, J.; Nielsen, P. LNA (Locked Nucleic Acids): Synthesis and High-Affinity Nucleic Acid Recognition. Chem. Commun. 1998, No. 4, 455–456. https://doi.org/10.1039/a708608c.

69. Owczarzy, R.; You, Y.; Groth, C. L.; Tataurov, A. V. Stability and Mismatch Discrimination of Locked Nucleic Acid–DNA Duplexes. Biochemistry 2011, 50 (43), 9352–9367. https://doi.org/10.1021/bi200904e.

70. Karkare, S.; Bhatnagar, D. Promising Nucleic Acid Analogs and Mimics: Characteristic Features and Applications of PNA, LNA, and Morpholino. Appl. Microbiol. Biotechnol. 2006, 71 (5), 575–586. https://doi.org/10.1007/s00253-006-0434-2.

71. Ali, S. E.; Mittal, A.; Mathews, D. H. RNA Secondary Structure Analysis Using RNAstructure. Curr. Protoc. 2023, 3 (7), e846. https://doi.org/10.1002/cpz1.846.

72. Owczarzy, R.; Tataurov, A. V.; Wu, Y.; Manthey, J. A.; McQuisten, K. A.; Almabrazi, H. G.; Pedersen, K. F.; Lin, Y.; Garretson, J.; McEntaggart, N. O.; Sailor, C. A.; Dawson, R. B.; Peek, A. S. IDT SciTools: A Suite for Analysis and Design of Nucleic Acid Oligomers. Nucleic Acids Res. 2008, 36 (suppl_2), W163–W169. https://doi.org/10.1093/nar/gkn198.

73. Sugimoto, N.; Nakano, S.; Katoh, M.; Matsumura, A.; Nakamuta, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. Thermodynamic Parameters to Predict Stability of RNA/DNA Hybrid Duplexes. Biochemistry 1995, 34 (35), 11211–11216. https://doi.org/10.1021/bi00035a029.

74. Owczarzy, R.; You, Y.; Moreira, B. G.; Manthey, J. A.; Huang, L.; Behlke, M. A.; Walder, J. A. Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. Biochemistry 2004, 43 (12), 3537–3554. https://doi.org/10.1021/bi034621r.

75. Owczarzy, R.; Moreira, B. G.; You, Y.; Behlke, M. A.; Walder, J. A. Predicting Stability of DNA Duplexes in Solutions Containing Magnesium and Monovalent Cations. Biochemistry 2008, 47 (19), 5336–5353. https://doi.org/10.1021/bi702363u.

76. Primer Analysis Tool and Tm Calculator | IDT. Integrated DNA Technologies. https://www.idtdna.com/pages/tools/oligoanalyzer (accessed 2026-02-12).

77. Gong, P.; Wang, K.; Liu, Y.; Shepard, K.; Levicky, R. Molecular Mechanisms in Morpholino-DNA Surface Hybridization. J. Am. Chem. Soc. 2010, 132 (28), 9663–9671. https://doi.org/10.1021/ja100881a.

78. Rihon, J.; Mattelaer, C.-A.; Montalvão, R. W.; Froeyen, M.; Pinheiro, V. B.; Lescrinier, E. Structural Insights into the Morpholino Nucleic Acid/RNA Duplex Using the New XNA Builder Ducque in a Molecular Modeling Pipeline. Nucleic Acids Res. 2024, 52 (6), 2836–2847. https://doi.org/10.1093/nar/gkae135.

79. Summerton, J. E. Morpholinos and PNAs Compared. Lett. Pept. Sci. 2003, 10 (3), 215–236. https://doi.org/10.1007/s10989-004-4913-y.

80. Leamy, K. A.; Assmann, S. M.; Mathews, D. H.; Bevilacqua, P. C. Bridging the Gap Between In Vitro and In Vivo RNA Folding. Q. Rev. Biophys. 2016, 49, e10. https://doi.org/10.1017/S003358351600007X.

81. Hua, B.; Han, K. Y.; Zhou, R.; Kim, H.; Shi, X.; Abeysirigunawardena, S. C.; Jain, A.; Singh, D.; Aggarwal, V.; Woodson, S. A.; Ha, T. An Improved Surface Passivation Method for Single-Molecule Studies. Nat. Methods 2014, 11 (12), 1233–1236. https://doi.org/10.1038/nmeth.3143.

82. Nicolai, C.; Sachs, F. SOLVING ION CHANNEL KINETICS WITH THE QuB SOFTWARE. Biophys. Rev. Lett. 2013, 08 (03n04), 191–211. https://doi.org/10.1142/S1793048013300053.

83. Qin, F.; Auerbach, A.; Sachs, F. Estimating Single-Channel Kinetic Parameters from Idealized Patch-Clamp Data Containing Missed Events. Biophys. J. 1996, 70 (1), 264–280. https://doi.org/10.1016/S0006-3495(96)79568-1.

84. Hawkes, A. G.; Jalali, A.; Colquhoun, D. The Distributions of the Apparent Open Times and Shut Times in a Single Channel Record When Brief Events Cannot Be Detected. Philos. Trans. R. Soc. Lond. Ser. Phys. Eng. Sci. 1997, 332 (1627), 511–538. https://doi.org/10.1098/rsta.1990.0129.

85. Qin, F.; Auerbach, A.; Sachs, F. Maximum Likelihood Estimation of Aggregated Markov Processes. Proc. R. Soc. Lond. B Biol. Sci. 1997, 264 (1380), 375–383. https://doi.org/10.1098/rspb.1997.0054.

86. Mergny, J.-L.; Lacroix, L. Analysis of Thermal Melting Curves. Oligonucleotides 2003, 13 (6), 515–537. https://doi.org/10.1089/154545703322860825.

87. Haddad, F.; Bodell, P. W.; Qin, A. X.; Giger, J. M.; Baldwin, K. M. Role of Antisense RNA in Coordinating Cardiac Myosin Heavy Chain Gene Switching *. J. Biol. Chem. 2003, 278 (39), 37132–37138. https://doi.org/10.1074/jbc.M305911200.

88. Zhang, J.; Gao, C.; Meng, M.; Tang, H. Long Noncoding RNA MHRT Protects Cardiomyocytes against H2O2-Induced Apoptosis. Biomol. Ther. 2016, 24 (1), 19–24. https://doi.org/10.4062/biomolther.2015.066.

89. Harikrishnan, K. N.; Okabe, J.; Mathiyalagan, P.; Khan, A. W.; Jadaan, S. A.; Sarila, G.; Ziemann, M.; Khurana, I.; Maxwell, S. S.; Du, X.-J.; El-Osta, A. Sex-Based Mhrt Methylation Chromatinizes MeCP2 in the Heart. Iscience 2019, 17, 288-+. https://doi.org/10.1016/j.isci.2019.06.031.

90. Lang, M.; Ou, D.; Liu, Z.; Li, Y.; Zhang, X.; Zhang, F. LncRNA MHRT Promotes Cardiac Fibrosis via miR-3185 Pathway Following Myocardial Infarction. Int. Heart. J. 2021, 62 (4), 891–899. https://doi.org/10.1536/ihj.20-298.

91. Peattie, D. A.; Gilbert, W. Chemical Probes for Higher-Order Structure in RNA. Proc. Natl. Acad. Sci. 1980, 77 (8), 4679–4682. https://doi.org/10.1073/pnas.77.8.4679.

92. Chamberlin, S. I.; Weeks, K. M. Mapping Local Nucleotide Flexibility by Selective Acylation of 2‘-Amine Substituted RNA. J. Am. Chem. Soc. 2000, 122 (2), 216–224. https://doi.org/10.1021/ja9914137.

93. Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. Selective 2’-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE): Quantitative RNA Structure Analysis at Single Nucleotide Resolution. Nat. Protoc. 2006, 1 (3), 1610–1617. https://doi.org/10.1038/nprot.2006.249.

94. Lorenz, R.; Bernhart, S. H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.; Hofacker, I. L. ViennaRNA Package 2.0. Algorithms Mol. Biol. AMB 2011, 6, 26. https://doi.org/10.1186/1748-7188-6-26.

95. Novikova, I. V.; Hennelly, S. P.; Sanbonmatsu, K. Y. Structural Architecture of the Human Long Non-Coding RNA, Steroid Receptor RNA Activator. Nucleic Acids Res. 2012, 40 (11), 5034–5051. https://doi.org/10.1093/nar/gks071.

96. Busan, S.; Weidmann, C. A.; Sengupta, A.; Weeks, K. M. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. Biochemistry 2019, 58 (23), 2655–2664. https://doi.org/10.1021/acs.biochem.8b01218.

97. Mortimer, S. A.; Weeks, K. M. A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. J. Am. Chem. Soc. 2007, 129 (14), 4144–4145. https://doi.org/10.1021/ja0704028.

98. Lusvarghi, S.; Sztuba-Solinska, J.; Purzycka, K. J.; Rausch, J. W.; Grice, S. F. J. L. RNA Secondary Structure Prediction Using High-Throughput SHAPE. JoVE J. Vis. Exp. 2013, No. 75, e50243. https://doi.org/10.3791/50243.

99. Covarrubias, A. J.; Perrone, R.; Grozio, A.; Verdin, E. NAD+ Metabolism and Its Roles in Cellular Processes during Ageing. Nat. Rev. Mol. Cell Biol. 2021, 22 (2), 119–141. https://doi.org/10.1038/s41580-020-00313-x.

100. Chen, Y.; Ying, Y.; Lalsiamthara, J.; Zhao, Y.; Imani, S.; Li, X.; Liu, S.; Wang, Q. From Bacteria to Biomedicine: Developing Therapies Exploiting NAD+ Metabolism. Bioorganic Chem. 2024, 142, 106974. https://doi.org/10.1016/j.bioorg.2023.106974.

101. Nieves, N. E. C.; R. Widom, J. Ligand Binding Characteristics of an NAD + Riboswitch Revealed by FRET and Biolayer Interferometry. 2024. https://doi.org/10.1039/D4CC04261A.

102. Conoan Nieves, N. E. Utilizing Single-Molecule Fluorescence Microscopy to Elucidate Conformational Dynamics of RNA. Ph.D., University of Oregon, United States -- Oregon, 2025.

103. Hartwig, A. Role of Magnesium in Genomic Stability. Mutat. Res. - Fundam. Mol. Mech. Mutagen. 2001, 475 (1), 113–121. https://doi.org/10.1016/S0027-5107(01)00074-4.