Establishing Ligand Mediated RNA Folding of Translational Riboswitches as Genetic Regulators using Single Molecule Microscopy

by

Arlie Jihan Rinaldi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2013

Doctoral Committee:

Professor Nils G. Walter, Chair
Professor Hashim M. Al-Hashimi
Professor Carol A. Fierke
Professor George A. Garcia
Professor Janet L. Smith

Dedication

To my parents, who always continued to pick up the phone.

Acknowledgements

Throughout my six years in graduate school, I have had the absolute pleasure of meeting and working with an enormous number of people who have shaped and changed my life forever. I would first like to thank my graduate advisor, Dr. Nils Walter. Nils took a chance on me when I was in a tough spot, and I will forever be overwhelmingly thankful for his trust, patience and guidance throughout my time in the lab. I am, without any trace of a doubt, a better scientist and person due to my experience under his mentorship. I would also like to thank my equally wonderful committee: Dr. Carol Fierke, Dr. Hashim Al-Hashimi, Dr. Janet Smith and Dr. George Garcia. During our (many!) meetings, I have learned a great deal from such exceptional people, and I will always be thankful for your encouragement, helpful comments and continued faith in me throughout this journey.

I would also like to thank my fellow lab members, past and present. I have formed a unique relationship with each and every one of you, and you have made this experience fun and stimulating. I truly believe that this group is one of the most well-rounded and helpful groups I could have ever had the pleasure of working with. I would particularly like to thank Krishna for showing me the ropes during my first few months in the lab, and who continues to be a great colleague. To Paul – thank you for answering every silly question I've ever come to you with without judgment, and for helping me take this project in a direction that would not have been possible without your valuable insight. To the Walter lab ladies, thank you for the lunch trips and endless encouragement. You've given me a level of sanity I needed to maneuver my way through this process. To Heidi – oh my goodness, what would I have done without you? Thank

you for every lunch, every coffee trip, and every glass of wine. You've been a great friend! To Matt Marek – thank you for being my stand-in big brother when my real big brother can't be around to look out for me. And finally, to Mario – you've become one of my best friends. Thank you so much, not only for helping me with the microscope a million times, and with Matlab scripts, and with a million other things, but also for all the fun times we've had. I'm so happy to have gotten to know you as well as I have, and I can't wait to see where the future takes us!

I came to Ann Arbor in the summer of 2007 without knowing anyone, and I was so lucky to get to know the incoming chemistry class that year. To my fellow first years, I had so much fun getting to know you all. To Crystal – you were the first person I met when I walked through the door of the hotel during our orientation at 1:00 in the morning after getting flown all over the northeast and I'm so happy to have spent the first three years of my graduate career with you right by my side. I truly don't know what I would have done without you throughout all these years. And I can't wait for this summer! To Devon – some of my best memories of Ann Arbor were spent getting to know you. We may not speak very much, but you will always be one of my best and most treasured friends. To the rest of the first years – Noah Gardner, Allison Knauff, Jenna Majchrzak, Eric Majchrzak, Ted Boron, Kevin Hartman and Tom Slaney, thank you for a great time, and for some really great tailgates.

To all my friends from back home, you mean so much to me and thank you for not getting mad at me when we go months in between phone calls. To Angela – you are my oldest friend and I don't know what I would do if I couldn't give you a call whenever I need. I'm so happy for everything that you've done in the past few years, and I can't wait to come visit you in San Francisco! To my other, much smaller half, Lesa – I can't even begin to describe how much I miss spending every day with you and your incessant sarcasm; no one understands me quite

about we never go another four years without talking? Don't make me throw another CRC in your direction! I'm so happy we caught up over the last few months.

Last and certainly not least, I'd like to thank my wonderful family. No one has shown more compassion and more patience during what has undoubtedly been the most mentally, emotionally and physically trying experience of my life. I would never have been able to accomplish what I have without your love and never-ending support. Thank you to my big brother, Rory for being such a great role-model to me. Outside of our father, you are the best dad in the world, and I've learned a great deal about selflessness from you. Also, thanks for all the great music suggestions. They help me get through long days in the microscope room like you don't even know! Thank you to my sister-in-law, Natalie – you've always treated me as a sister, and I certainly love you as my sister. Thank you to my nephews, Xander, Torin, and my newest nephew, Roklan, for being a constant source of cuteness and pure joy in my life. I can't wait to watch the three of you grow up! Thank you to my grandparents for their continued support, I love you all! And finally, thank you from the bottom of my heart, to my parents Nancy and Gary. I am at a loss for words when I try to put into writing the amount of love and admiration I have for the both of you. There is no telling where I would be if I had not had the confidence that you instilled in me that I can do whatever I want to do for as long as I can remember. Everything I do, including this dissertation, is and always has been with you in mind. Thank you.

Table of Contents

List of Figures

List of Tables

Abstract

Riboswitches are non-coding RNA regulatory elements located primarily in the 5' UTR of bacterial messenger RNAs. Most commonly, they regulate gene expression of the downstream gene through transcriptional attenuation or through translational initiation inhibition in response to changing concentrations of a cellular signal. Riboswitches are composed of a ligand binding aptamer domain and an expression platform which confers the genetic decision. One of the most challenging questions remaining in the riboswitch field is how a ligand binding event can confer a large-scale conformational change, and how this change can effect a genetic decision.

In this study, we demonstrate that two structurally similar transcriptional (*Bsu*) and translational (*Tte*) preQ$_1$ riboswitches adopt similar pre-folded ensembles in the absence of ligand using a combination of single molecule techniques, molecular dynamics simulations and NMR. This result is in contrast to previous studies which suggested a largely unfolded and a loose pseudoknot for the transcriptional and translational riboswitch, respectively. Gō-model simulations of the two aptamers suggest that the ligand binds late (*Bsu*) and early (*Tte*) relative to pseudoknot folding, suggesting the two riboswitches tend to fold via conformational selection and induced fit, respectively. Finally, we show through the rational design of a single nucleotide swap distal from the ligand binding pocket that we find to predictably control the aptamers' pre-folded states and their ligand binding affinities.

Additionally, we have developed a single molecule footprinting assay to probe the accessibility of the ribosomal binding site of two translational riboswitches as a function of ligand concentration. The feasibility of the assay is demonstrated using the *V. vulnificus* adenine

riboswitch, in which we use a fluorophore labeled nucleic acid oligo to mimic the 3’ end of the 16S rRNA. Finally, we probe the accessibility of the Shine-Dalgarno sequence of the *Tte* $preQ_1$ riboswitch in the context of its full-length mRNA, which contains two overlapping open reading frames. Our results provide a novel single molecule assay which can be extended to study other biologically relevant molecules, and provide, for the first time, a molecular basis for the occlusion of the Shine-Dalgarno sequence by translational riboswitches to effect gene regulation.

CHAPTER 1

Introduction

1.1 Importance of the RNA structure-function relationship

1.1.1 Biological significance of non-coding RNA

As long as 60 years ago, it was speculated by James Watson that RNA played an important role in the translation of the genetic material encoded in our DNA into protein[1]. Since the discovery of messenger RNA (mRNA) as the liaison of information between nuclear DNA and cytoplasmic ribosomes[2,3], our knowledge of the vast variety of biological functions of non-protein-coding RNAs (ncRNAs) in the cell has exploded. One of the first breakthroughs to challenge the relatively mundane role of RNA as a simple genetic messenger was the discovery of RNA-based enzymes, or ribozymes, in particular RNase P and the *Tetrahymena* Group I intron[4,5]. Similarly, the discovery that the catalytic portion of the ribosome is composed of RNA[6,7] ushered in a new train of thought in which an ancient "RNA world" may have existed, wherein only RNA molecules existed and catalyzed all necessary chemical reactions required for basic cellular function.

Following the discovery of tRNA "adaptors" required for translating the information in a DNA codon into the encoded amino acid during protein synthesis[1,8], the initial consensus for the role of RNA was as an accessory to protein function. This notion was challenged when it was

discovered that less than ~2% of the human genome encodes protein[9,10]. As a rebuttal, it was suggested that the remaining "junk" DNA was remnants of previous viral infections or non-viable mutations. However, when it was recently suggested that up to ~90% of the genome is actively transcribed[11], it became clear that these ncRNAs must serve some important biological function. Since then, a wide variety of ncRNA classes has been elucidated, and shown to perform a multitude of roles within the cell, including the regulation of gene expression[12,13], translation[14], splicing[15] and maintenance of chromosome ends[16].

The wide variety of biological roles played by ncRNA seems counterintuitive considering the small diversity of functional groups utilized in RNA as compared to protein (4 nucleobases in RNA versus 20 amino acids in proteins). However, what RNA lacks in sequence diversity, it makes up for in structural complexity. An inherently flexible, single-stranded molecule, RNA does not generally fold into a single native conformation, but rather folds into many different conformations along a rugged folding free energy landscape[17]. RNA folding is a hierarchal process in which large complex folds begin by local secondary structure formation, which then proceeds by the nucleation of tertiary contacts that allows it to perform its biological function in the cell[18]. This structure-function relationship of RNA has been at the center of RNA biology research since the discovery of the double helix and continues to be a significant focus in the field.

1.1.2 Riboswitches as genetic regulators

Beginning in the early 1990s, researchers began utilizing *in vitro* selection, or Systematic Evolution of Ligands by Exponential Enrichment (SELEX), techniques to uncover RNA molecules that bind specifically to a cognate ligand with high affinity[19,20]. A couple of years later, the first naturally occurring RNA "aptamer" was discovered in the 5' UTR of the *tyrS* gene

of *B. subtilis*. It was found that uncharged $tRNA^{Tyr}$ was able to bind the so-called T-box region of the 5' UTR, resulting in the formation of an antiterminator helix, allowing read-through of the downstream gene by RNA polymerase. Alternatively, charged $tRNA^{Tyr}$ was not able to interact completely with the RNA, resulting in the formation of a terminator helix, aborting transcription[21]. Following this discovery, the term "riboswitch" was first coined by Breaker and coworkers to describe a region in the 5' UTR of an *E. coli* mRNA that binds thiamine derivatives, which regulates gene expression through occlusion of the Shine-Dalgarno sequence in the presence of thiamine pyrophosphate (TPP)[22]. Today, the term riboswitch has expanded to include mRNAs that respond to a variety of cellular signals, including, but not limited to nucleobases[23,24], amino acids[25,26], cofactors[27,28], pH[29], temperature[30], metal ions[31,32], and fluoride ions[33].

Riboswitches are gene regulatory elements most commonly found within the 5' UTR of bacterial mRNAs, but can also be found more scarcely within archaea[27], plants[34], fungi[35] and eukaryotes[35,36]. Riboswitches are composed of two usually mutually exclusive structural elements: a highly conserved, ligand sensing aptamer domain and a less conserved expression platform that either enables or attenuates expression of the downstream gene without the need for proteins. This decision is achieved most commonly through transcription termination or inhibition of translation initiation. In the former, ligand binding to the aptamer domain induces a downstream conformational change in the expression platform, resulting in the formation of a stable terminator hairpin (**Figure 1.1A**). This hairpin results in premature dissociation of RNA polymerase and subsequent mRNA degradation by the cell. Alternatively, ligand binding can cause a conformational change resulting in the occlusion of the Shine-Dalgarno site by Watson-Crick base pairing (**Figure 1.1B**). This conformational change blocks the 30S ribosomal binding

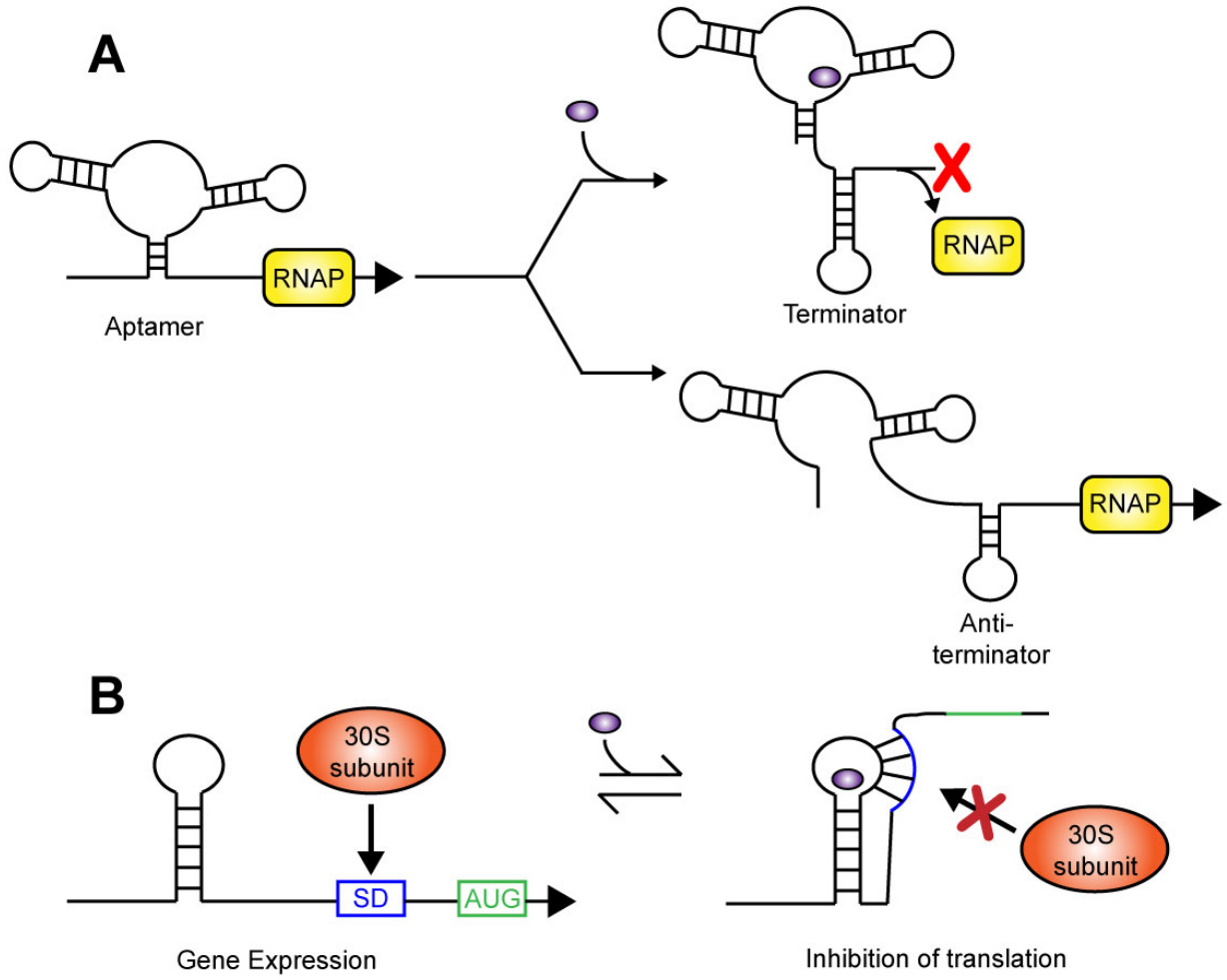

Figure 1.1 Modes of genetic regulation by riboswitches.
(A) In transcriptional riboswitches, RNA polymerase first transcribes the aptamer region of the RNA. Once cleared from the aptamer, the riboswitch makes a genetic decision based on the availability of its cognate ligand. If sufficient ligand is not present, or does not bind in a timely fashion, a stable terminator helix is formed, and RNA polymerase dissociates, causing an aborted transcript. However, if ligand does bind the aptamer domain, this causes alternative base-pairing in which an anti-terminator is formed, and RNA polymerase transcribes the downstream gene. (B) In translational riboswitches, the RNA is transcribed and a genetic decision can be made in a reversible manner. If ligand is present and is bound by the aptamer domain, the ribosomal binding site, or Shine-Dalgarno site is occluded, resulting in inhibition of translation.

site, thereby inhibiting initiation of translation of the nascent transcript[13]. Alternative modes of riboswitch-mediated gene regulation have recently been reported including mRNA degradation[37-39] and alternative self-splicing[40]. Myriad studies have been conducted focusing on the conformational disparity between the ligand-bound and ligand-free states[41-46], and are beginning to yield a detailed picture of the mechanisms by which riboswitches can regulate gene expression. However, the question of how a ligand binding event can elucidate such a large-scale conformational change remains unanswered. To better understand the complex function of RNA secondary and tertiary structure motifs in gene regulation by riboswitches, it is of great importance to study the conformational changes and folding dynamics incurred upon ligand binding.

As mentioned above, translationally operating riboswitches achieve gene regulation through the sequestration of the ribosome binding site, or Shine-Dalgarno (SD) sequence through alternative base pairing interactions as a function of ligand binding. A great deal of effort has been concentrated on observing this occlusion through structural studies on the ligand-free and ligand-bound states, although the former has proven to be more difficult to obtain, probably due to the dynamic nature of the ligand-free conformation. In particular, crystal structures of translational riboswitches have demonstrated ligand-mediated occlusion of the SD sequence in the SAM-II[47], SAM-III[48] and $preQ_1$[42] riboswitches. Furthermore, the central guanosine of the SD sequence in the SAM-III riboswitch directly binds the SAM ligand, indicating that the SD sequence is directly involved in ligand recognition[48]. Translational riboswitches have been observed to adopt RNA folds commonly found in nature, such as H-type pseudoknots and 3-way junctions. In contrast to transcriptional riboswitches, which utilize two mutually exclusive domains, it is common for translational riboswitches to have partially overlapping ligand-sensing

aptamer domains and decision making expression platforms, thereby achieving genetic regulation from a compact structural motif. This may be due in part to the longer time window over which translational riboswitches can bind ligand to effect gene regulation, as compared to transcriptional riboswitches that need to bind ligand in a narrow time window before the RNA polymerase completes transcription of the expression platform. These structural studies have been imperative in the understanding of translational regulation, but the question of how the ligand-mediated conformational change occurs persists due to the relative lack of structural information regarding the ligand-free conformations of these riboswitches.

While the structure of only one ligand-free translational riboswitch has been solved[42], a great deal of work has been performed to understand the conformational disparity between the apo and holo conformations of translational riboswitches using techniques such as 2-aminopurine fluorescence assays[41,49,50], chemical probing techniques[51-53], NMR[41,53-55], SAXS[42,53,54], single molecule optical trapping[56,57] and single molecule FRET[58]. These studies have revealed a number of features common to riboswitches, despite their varying structural characteristics. Metal ions, in particular, Mg^{2+} cations are important for stabilizing secondary and tertiary structures within RNA. The positively charged metal cations serve to stabilize the highly negatively charged phosphate backbone of the polymeric RNA. Perhaps not surprisingly, metal ions have shown to play an important role in proper riboswitch folding, which may be important for efficient function. For example, 2-aminopurine fluorescence assays showed the importance of Mg^{2+} cations in the structural preorganization of secondary and tertiary RNA elements, particularly in the binding pocket of the aptamer domains. In many cases, the cognate ligand of riboswitches is almost completely encapsulated by RNA, making this preorganization imperative for solvent accessibility and efficient ligand recognition. For instance, a Mg^{2+}-induced nucleation

of secondary and tertiary structural elements was observed in the TPP riboswitch using single molecule FRET, particularly in the core of the three-way junction surrounding the ligand binding site[58]. Similarly, structural reorganizations were mediated by the addition of Mg^{2+} to the *V. vulnificus* adenine riboswitch as observed by fluorescence changes to 2-aminopurine containing riboswitch variants. These structural changes occurred locally in the ligand binding pocket, as well as globally through the formation of kissing loop interactions between loops L2 and L3 of the adenine riboswitch's "tuning fork" architecture[50].

Similar experiments have been performed to show further RNA compaction and folding due to ligand addition after Mg^{2+}-mediated folding. SAXS analysis on the SAM-II riboswitch showed a ligand dependent compaction of the RNA following SAM addition from 31.7 Å to 20.7 Å. However, in the presence of 10 mM Mg^{2+}, the RNA only compacted slightly from 21.5 Å to 19.5 Å following ligand addition, indicating that Mg^{2+} alone can have similar impacts on RNA folding and compaction as ligand[54]. Single molecule force spectroscopy was used to study how ligand alters the relative stability of helices within the *V. vulnificus* adenine aptamer. It was shown that upon ligand addition, the overall characteristics of unfolding remained relatively unchanged with one striking difference: the stability of the P1 helix increased drastically upon adenine addition[57]. Relative P1 stability as a function of ligand concentration is a common feature for translational riboswitches that contain a more complex global fold with mutually exclusive aptamer domains and expression platforms. Finally, work on the translationally operating SAM-III riboswitch indicated that this riboswitch can act as a reversible switch that can respond to varying physiological changes including availability of ligand. Structural changes within the RNA were monitored by 2-AP fluorescence in the presence of high ligand concentrations and a competitor RNA, that sequestered the pool of ligand[59]. These results

highlight the global and reversible conformational changes that occur upon ligand binding to translational riboswitches and stress the importance of physiological conditions upon ligand binding. However, the literature still lacks a molecular basis for the ligand-dependent communication between aptamer domain and expression platform to regulate gene expression. To this end, a major challenge lies ahead to understand the dynamic processes utilized by riboswitches to effect genetic regulation as a function of ligand binding.

1.1.3 Observation of translational regulation by riboswitches

A significant portion of work done to date on riboswitches has focused primarily on the global conformational changes induced by a single ligand binding event through three-dimensional structural information obtained via X-ray crystallography and NMR. Several structures of ligand bound riboswitches demonstrate an occlusion of the Shine-Dalgarno sequence through Watson-Crick and non-canonical base pairing[42,47,48,60] and sometimes through a direct interaction with the ligand itself[48]. In addition to these structural studies, a series of chemical probing experiments in the presence and absence of ligand have shown a differential pattern of protection surrounding the SD sequence and the AUG start codon[47,49,52,60,61]. These studies have provided indirect evidence of a genetic mode of regulation in which a ligand binding event occludes the SD sequence to prevent the association of 30S subunits, inhibiting translation of the downstream gene. The only direct evidence of a sequestered 30S ribosomal binding site was demonstrated in the SAM-III riboswitch by Henkin and coworkers[62]. Through the use of nitrocellulose filter binding assays, it was shown that incubation of the riboswitch RNA with its cognate ligand reduced ribosome retention by ~4-fold. Additionally, ribosomal toeprinting assays in the presence of SAM show a 2-fold amelioration of a primer extension stop corresponding to 16

nucleotides downstream of the AUG start codon. This distance is consistent with the start codon positioned within the P-site of the ribosome[63]. These results indicate that addition of ligand to a SAM riboswitch results in an occlusion of the SD sequence, which in turn sequesters the ribosomal binding site. However, this is the only evidence for 30S ribosomal occlusion, and further studies need to be performed to directly demonstrate this mode of genetic regulation as a common theme for translational riboswitches.

1.2 Single molecule microscopy

1.2.1 Why single molecule microscopy instead of classical ensemble techniques?

Within the last twenty years, the development of single molecule microscopy has begun to revolutionize the biosciences[64,65]. Classically, questions regarding complex biological systems have been investigated using bulk ensemble assays that average out rare but often significant idiosyncrasies within a given population of molecules. By contrast, single molecule experiments have the unique capability of revealing the diversity that exists from molecule to molecule, and make it an attractive means to study RNA, an implicitly dynamic molecule that often does not adopt a single conformation, but typically folds along a frustrated folding landscape into multiple heterogeneous conformations[17]. Observation of one molecule for an extended period of time and/or many molecules for limited periods of time will yield information related to that of the entire population as observed in a bulk ensemble experiment. However, single molecule observation will additionally access the interconversion kinetics of different species at equilibrium as well as reveal transient species, both features difficult to detect otherwise. Moreover, single molecule experiments are advantageous due to their use of very small quantities of sample, eliminating the multimerization effects that bulk experiments may suffer[64].

To this end, a variety of techniques have been developed utilizing the single molecule regime, including single molecule fluorescence microscopy[66,67], super-resolution imaging and super-accuracy co-localization techniques[68-71], and force microscopy[72-74]. The broad array of applications available today equips scientists with the ability to obtain kinetic and mechanistic information regarding biological systems that was not available only a few years ago.

1.2.2 Single molecule fluorescence resonance energy transfer: Observing conformational changes one molecule at a time

Single molecule fluorescence resonance energy transfer (smFRET) has been applied to study the folding of many biologically-relevant molecules[75,76], including RNA[77-89]. smFRET constitutes a non-radiative energy transfer process in which a single dye molecule is excited by a laser. The distance between this "donor" fluorophore and a second "acceptor" fluorophore is reported in real-time through transfer of excitation energy to the acceptor based on an overlap of the donor emission with the acceptor excitation spectrum. As such, the smaller the donor-acceptor distance, the lower (more quenched) the emission of the donor dye and the greater the emission of the acceptor dye. FRET is generally quantified as a FRET efficiency (E), which is given by the Förster equation, $E = \left(1 + \left(\frac{R}{R_0}\right)^6\right)^{-1}$, where R_0 is the Förster radius (or distance) for the specified donor-acceptor dye pair at which their FRET efficiency is 50%, and R is the donor-acceptor distance. Since every donor and acceptor dye pair has its own specific Förster radius, based on the given overlap of their spectra and the assumption of isotropic relative spatial orientation of their transition dipole moments, they must be chosen carefully and in a manner that will report on the specific distance to be studied. Conversely, the molecule of interest must be labeled in such a way that the distance between the fluorophores is within ~2-fold of their

Förster radius, differences in smFRET efficiency are expected between conformational states of interest, and the biological function of the RNA is retained. The process of choosing appropriate labeling sites in an RNA is ameliorated when a crystal or NMR structure is available and hypotheses exist on conformational changes of interest. If no such information is available, chemical probing methods under varying conditions may reveal residues that are accessible to solvent as suitable for labeling, as well as changes in base-pairing and tertiary interactions that lead to hypotheses on biologically relevant shape changes amenable to smFRET probing.

smFRET in general requires doubly fluorophore labeled RNA molecules. In addition, for TIRF-based smFRET, RNA molecules must also be tethered to the surface of a quartz slide, usually through a streptavidin-biotin interaction, leading to the overall requirement of three RNA modifications. TIRF microscopy utilizes an evanescent field created by the total internal reflection of a laser beam at the quartz-water interface of the slide surface[75]. This results in illumination of only molecules that are directly tethered to the slide surface, as the field penetrates only ~100 nm into the solution, thus reducing background signal. This method of reducing the illuminated volume is advantageous to others, such as diffusion smFRET in which the molecules are freely diffusing through a focused laser beam[85], as TIRF-based smFRET allows for the detection of molecules for extended periods of time and is not limited by the time a molecule spends in the confocal volume. Instead, TIRF-based smFRET observation is limited by the longevity of the fluorophores before they photobleach, which can be extended by the addition of an oxygen scavenging system, such as the protocatechuate-3,4-dioxygenase system[90] supplemented with Trolox to decrease photoblinking of the acceptor dye[91], even in complex reaction mixtures[92]. Taken together, this TIRF-based smFRET approach allows for the detection of hundreds of single molecules in one field of view for several tens of seconds at a time. From

these data, FRET efficiency histograms are constructed to reveal conformational changes, as well as the dwell times before interconversion rates to gain information on individual rate constants[93], a difficult task to accomplish using ensemble techniques.

The goal of this thesis is utilize a combination of single molecule techniques in combination with computational molecular dynamics simulations and NMR spectroscopy to study the ligand-mediated folding dynamics of two structurally similar transcriptional and translational $preQ_1$ riboswitches. We have found that the folding dynamics of these two riboswitches that employ two different modes of regulation only differ in subtle but distinct ways. Additionally, the ligand binding properties of the riboswitches can be controlled via a mutation distal from the ligand binding site. Finally, a single molecule footprinting assay is developed to probe the accessibility of the ribosomal binding site of two translational riboswitches, one in the context of the full mRNA. This assay provides, for the first time, direct evidence of the genetic mode of regulation by translational riboswitches in single molecules. Altogether, this dissertation highlights the importance of studying RNA using methods other than classical ensemble techniques due to its inherent heterogeneous qualities and the many idiosyncrasies within an RNA population. Also, this dissertation sets the framework for a single molecule footprinting assay, which could be extended to various biologically-relevant molecules.

CHAPTER 2

Single Molecule Characterization of the Ligand Induced Folding Behavior of Transcriptional and Translational preQ$_1$ Riboswitches[1]

2.1 Introduction

Riboswitches are highly structured, non-coding RNA motifs that are found in up to 4% of all 5'-untranslated regions of bacterial messenger RNAs and respond to cellular metabolites or other signals to control gene expression[39,94-97]. Riboswitches are composed of a highly conserved aptamer domain, which binds a ligand, and a downstream variable expression platform that modulates the genetic on/off switch. Regulation of gene expression is achieved through one of multiple possible modes, most commonly transcription attenuation and inhibition of translation initiation. While the general principles of genetic regulation by riboswitches are understood, the molecular basis of their action remains largely elusive.

Riboswitches respond to a variety of ligands including nucleobases[23,24], amino acids[25,26], cofactors of metabolic enzymes[27,28] and metal ions[31,32], and are found in a multitude of bacterial species[94]. 7-Aminomethyl-7-deazaguanine, or preQ$_1$, is one such ligand (**Figure 2.1A**). It is derived from guanine and is an intermediate in the queuosine biosynthetic pathway in

[1] Arlie Rinaldi performed all smFRET experiments and data analysis on the *Tte* riboswitch. Krishna C. Suddala performed all smFRET experiments and data analysis on the *Bsu* riboswitch. Jun Feng performed Gō model simulations in **Figures 2.14** and **2.15**, Anthony M. Mustoe performed coarse-grained RNA simulations in **Figure 2.6**, and Catherine D. Eichhorn performed NMR experiments in **Figure 2.9**. This work has been submitted to *Nucleic Acids Research*.

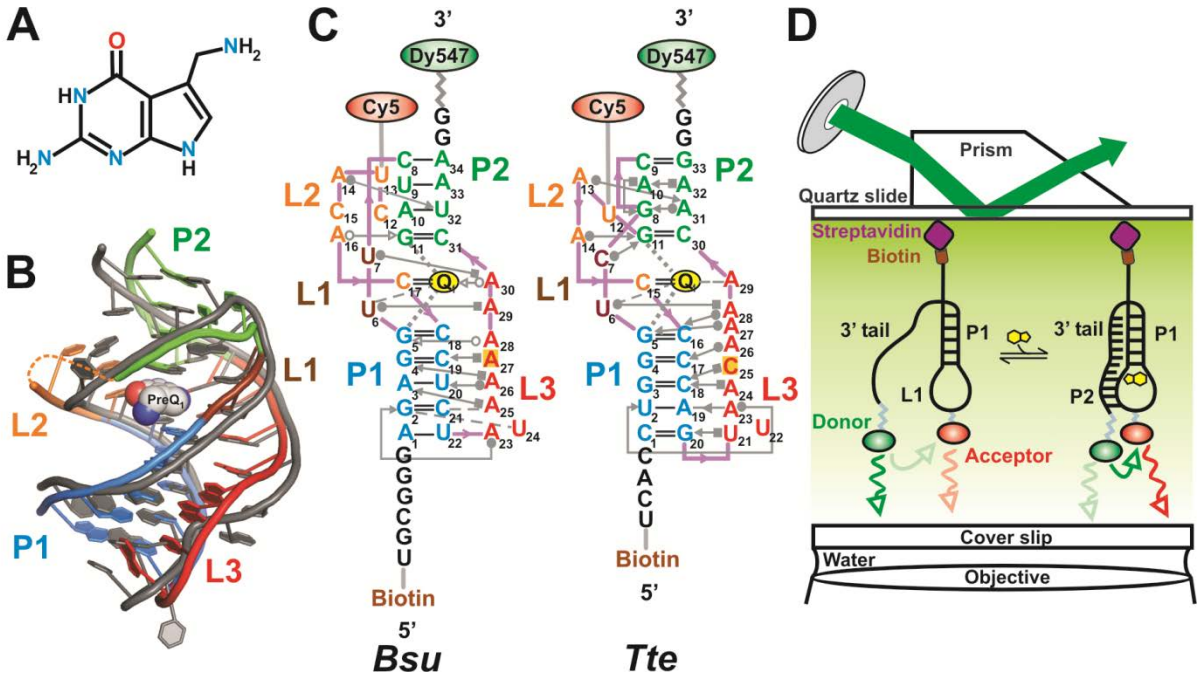

Figure 2.1 Structural comparison of the Bsu and Tte preQ$_1$ riboswitches.
(A) Structure of preQ$_1$ (7-aminomethyl-7-deazaguanine). (B) Structural overlay of the Bsu (colored, PDB ID 3FU2, chain A) and Tte (gray, PDB ID 3Q50) riboswitch crystal structures. The sugar-phosphate backbone is shown as a single ribbon. preQ$_1$ is space-filled and colored as in A. Secondary structure elements are color-coded as indicated. (C) Secondary structure maps of the Bsu and Tte riboswitches with interactions shown in Leontis-Westhof nomenclature[98]. Individual secondary structures are color-coded as in B and the locations of fluorophores and biotin modifications are indicated. (D) Prism-based TIRFM setup for smFRET.

bacteria[99,100]. Queuosine is found in bacteria and eukaryotes at the wobble position of tRNAs for His, Tyr, Asn, and Asp[101], where it is thought to be essential for translational fidelity[102-104] as well as bacterial virulence[105]. The family of $preQ_1$ riboswitches encompasses some of the smallest ligand-responsive RNAs found in nature, making them an ideal target for mechanistic studies. The *Bacillus subtilis* (*Bsu*) and related $preQ_1$ riboswitch aptamers in particular have been studied extensively[43,106-114] and are known to regulate the *queCDEF* operon through transcription termination[115]. In contrast, the $preQ_1$ riboswitch aptamer from *Thermoanaerobacter tengcongensis* (*Tte*) controls the expression of a putative $preQ_1$ transporter and has been implicated as translationally operating wherein ligand binding sequesters the first two nucleotides of the Shine-Dalgarno sequence through Watson-Crick base pairing[42,116].

In the *Bsu* riboswitch, the aptamer has to bind the ligand and stably fold into the pseudoknot structure to switch off gene expression before the competing anti-terminator hairpin is formed in the expression platform, whereas in the *Tte* riboswitch the expression platform partially overlaps with the ligand-binding aptamer domain. In either case, ligand binding and folding of the aptamer domain is a key event in the regulation of gene expression that is not well understood. Crystal structures of the $preQ_1$ bound *Bsu*[110] and *Tte*[42] aptamer domains (referred to henceforth simply as "riboswitches") overlay closely with an all-C3' unit-vector RMSD (URMSD) of only 1.8 Å[117] (**Figure 2.1B**), comparable to a URMSD of 1.6 Å between the lowest-energy NMR[109] and crystal[110] structures of the *Bsu* riboswitch. Both riboswitches adopt classic H-type pseudoknot structures containing a 5-base pair (bp) stem P1, a 2-nucleotide (nt) loop L1, followed by a 4-bp stem P2 and loops L2 and L3 (**Figure 2.1C**). In both riboswitches, the last nucleotide of L2 is a cytidine that recognizes $preQ_1$ through Watson-Crick base pairing (**Figure 2.1C**). There are only a few relatively subtle differences between the riboswitches

(**Figure 2.1C**). For example, the P2 stem of the *Bsu* riboswitch bears three Watson-Crick and a single non-canonical C8-A34 bp, whereas the *Tte* riboswitch has two Watson-Crick and two non-canonical bp, G8-A31 and A10-A32. In addition, the L2 loops in the *Bsu* and *Tte* riboswitches are 6- and 4-nt in length, respectively, and the 6-adenosine containing L3 loop in the *Bsu* riboswitch is interrupted by a C insertion in the *Tte* riboswitch.

Despite their structural similarities in the ligand bound state, discrepancies between the conformational behaviors of these two riboswitches have been reported for the ligand free form. In particular, while NMR studies of the *Bsu* riboswitch[109,118] suggested a largely unfolded conformation in which its 3' tail (encompassing L3 and the 3' segment of P2, **Figure 2.1C**) does not form tertiary interactions, X-ray crystallography and SAXS studies on the *Tte* riboswitch[42] suggest that it does form a pre-folded tertiary conformation poised for recognition in the absence of ligand. These results are surprising considering that the transcriptional and not necessarily the translational riboswitch must bind ligand efficiently during a narrow commitment time window, yet apparently only the translational riboswitch adopts a pre-folded conformation poised for ligand recognition. Consequently, even though structural data are available for both riboswitches, how the RNAs transition from the ligand-free to the ligand-bound form to affect gene expression is still unclear.

Here, we combine single molecule fluorescence resonance energy transfer (smFRET)[75,89,92] with computational techniques to compare the folding behavior of the *Bsu* and *Tte* riboswitches. We show that in the absence of ligand, both the *Bsu* and *Tte* riboswitches exist in a similar ensemble of conformations, contrary to previous studies. Both riboswitches exhibit a major population of a 'pre-folded' state ensemble wherein their 3' tail adopts transient interactions with the P1-L1 stem-loop, reminiscent of the fully folded state, and a minor

population of a folded-like state. The pre-folded state is poised to bind ligand and, for the *Tte* riboswitch, we find evidence that it can sense $preQ_1$. Subtle differences exist in the folding behavior of the transcriptional and translational $preQ_1$ riboswitches that can in part be attributed to differential local flexibility of the single-stranded, A-rich 3' tail[118], which leads to differences in long-range transient interactions. Together with results from Gō-model folding simulations, we show that these differences lead to distinct folding pathways for the transcriptional and translational riboswitches that can be classified as biased towards late ligand binding to an already preformed RNA pocket (conformational selection) and early ligand binding followed by the binding pocket folding around it (induced fit), respectively. Our study unveils design principles for pseudoknot folding that are dependent on the dynamic properties of the single-stranded 3' tail, a notion we put to the test by rationally reengineering the riboswitches' pre-folded states and ligand binding affinities through point mutations in this tail. Our results highlight the benefits of comparative studies, establish a framework for delineating conformational selection and induced fit pathways, and contribute to an emerging view that environmental conditions as well as distal sequence variations fine-tune the ligand binding properties of riboswitches and RNA in general.

2.2 Materials and Methods

2.2.1 Preparation of RNAs for smFRET

All RNA constructs were synthesized by Dharmacon Inc. (Fayette, CO) with a 5' biotin modification, 3' DY547 label and 5-aminoallyl-uridine (5NU) label at position U12 (*Tte*) and U13 (*Bsu*) for later functionalization with Cy5 (**Figure 2.1C**). Oligonucleotides were deprotected following the manufacturer's instructions. For labeling each construct, one dye pack of the Cy5-

NHS ester (GE Healthcare) was dissolved in 30 µL DMSO and used to label ~3.4 nmol RNA in a total reaction volume of 50 µL containing 0.1 M sodium bicarbonate buffer, pH 8.7. The reactions were incubated and tumbled at room temperature in the dark for 4 h. Reaction volumes were adjusted to 500 µL with deionized water and loaded onto a Nap-5 gel filtration column (GE Healthcare) for desalting and removal of excess free dye. Fractions containing the RNA were collected, ethanol precipitated and pellets were resuspended in 50 µL deionized water.

2.2.2 Single molecule FRET

We assembled a microfluidic channel on a quartz slide with an inlet and outlet port and coated it with biotinylated-BSA, followed by streptavidin, as previously described[75,86]. We folded the RNA by heating at 70 ^{o}C for 2 min and allowing it to cool to RT for at least 20 min in near-physiological smFRET buffer (50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM $MgCl_2$) in the presence or absence of $preQ_1$. 100 µL of 10-50 pM of the heat annealed RNA was flowed onto the slide and incubated for 10 min for binding. Excess RNA was removed by flowing 100-200 µL smFRET buffer with or without $preQ_1$ through the channel. An oxygen scavenging system was included in smFRET buffer (+/- $preQ_1$), consisting of 5 mM protocatechuic acid and 50 nM protocatechuate-3,4-dioxygenase to slow photobleaching and 2 mM Trolox to reduce photoblinking[91]. DY547 was directly excited using a 532 nm laser and emission from DY547 and Cy5 fluorophores was simultaneously recorded using an intensified CCD camera (I-Pentamax, Princeton Instruments) at 100 ms time resolution[79,80,83,84,86,92]. Experiments at 33 ms data acquisition were performed on a similar prism-based TIRF setup with an EMCCD camera (iXon, Andor Technology). smFRET time traces were extracted from the raw movie files using IDL (Research Systems) and analyzed using Matlab (The Math Works) scripts. Genuine

fluorescence traces were selected manually based on the following features: single-step photobleaching, a signal-to-noise ratio of >5:1, a total (donor + acceptor) fluorescence intensity of > 300 (arbitrary units), and a total fluorescence duration of >10 s. The FRET ratio was calculated as $I_A/(I_A + I_D)$, where I_A and I_D represent the background corrected fluorescence intensities of the acceptor (Cy5) and donor (DY547) fluorophores, respectively. FRET distribution histograms were plotted using OriginLab 8.1. HMM analysis was performed on smFRET traces using the segmental k-means algorithm in the QuB software suite as described[93]. We used a two-state model (with mid-FRET and high-FRET states) to idealize the data; a third zero-FRET state was included to account for a few blinking events. Transition Occupancy Density Plots (TODPs) were then generated from the idealized data using Matlab[92].

2.2.3 TOPRNA Simulations

TOPological modeling of RNA (TOPRNA) uses three primary pseudo atoms to represent the base (B), sugar (S), and phosphate (P) moieties of an RNA nucleotide. Base pairs are treated as permanent bonds between paired B atoms and regions of contiguous base pairs are parameterized through dihedral potentials to assume standard A-form helical structure. A small filler atom was also placed between each set of paired bases to more accurately reproduce the steric profile of a base pair. Non-base-paired residues were parameterized to maintain RNA-like bond lengths and angles between pseudo-atoms, but were otherwise treated as freely rotatable chains. Both base-paired and non-base-paired residue potential parameters were derived from fits of CHARMM[119] potential functions to structural-database derived statistical potentials. Electrostatic interactions were ignored and, with the exception of a small attractive force between base-paired B atoms meant to simulate intra-helix base stacking, all other non-bonded interactions were solely

repulsive in nature. The steric radii of these repulsive interactions were approximated from the minimum dimension of the chemical moiety each pseudo-atom represents. Initial coordinates for both the *Bsu* and *Tte* riboswitches were obtained by equilibrating the initially linear chains of the same sequences as the two RNAs used for smFRET (**Figure 2.1C**) with constraints that forced the formation of an A-form helical P1 stem. The base paired residues of P1 were then 'bonded' together and additional simulation-dependent dihedral and distance restraints were applied to the 3' tail and/or residues of the P2 stem, followed by further equilibration (see Supplementary Tables S1 and S2 for restraint details). Temperature replica exchange simulations were then performed in CHARMM[119] with the MMTSB replica exchange server[120] using four temperature windows from 300 K to 400 K for 50,000 exchange cycles so as to achieve exhaustive sampling at each condition. 1,000 timesteps of Langevian dynamics with 5 ps^{-1} friction coefficient and 0.02 ps integration time step were performed in between each exchange cycle. Neither the fluorophores nor linkers were included in these simulations. Estimates of the inter-fluorophore distances expected in solution were then obtained by measuring the distances sampled over the length of the simulations between the base pseudo-atoms of U13 and G36 and U12 and G35 for the *Bsu* and *Tte* molecules, the sites of fluorophore attachment, respectively.

Table 2.1 TOPRNA simulation parameters for the *Bsu* riboswitch system

Simulation Description	**P1 Stem**[1]	**P2 Stem**[2]	**Tail Stacking Dihedrals**[3]	**Tail-P1 Distance Constraints**[4]
Unstacked 3' tail (red)	Yes	No	---	---
Stacked 3' tail (green)	Yes	No	A26 to C31	---
Stacked 3' tail with lower tail-P1 interactions (blue)	Yes	No	A26 to C31	U20-A26 (B,S) A3-A26 (B) G4-A27 (B) C19-A27 (B,S) G5-A28 (B,S) C18-A28 (B)
Unstacked 3' tail with lower tail-P1 interactions (purple)	Yes	No	---	U20-A26 (B,S) A3-A26 (B) G4-A27 (B) C19-A27 (B,S) G5-A28 (B,S) C18-A28 (B)
Stacked 3' tail with upper tail-P1 interactions (cyan)	Yes	No	A26 to C31	U6-A29 (B,S) U7-A30 (B,S)
Stacked 3' tail with all tail-P1 interactions (orange)	Yes	No	A26 to C31	U20-A26 (B,S) A3-A26 (B) G4-A27 (B)

				C19-A27 (B,S) G5-A28 (B,S) C18-A28 (B) U6-A29 (B,S) U7-A30 (B,S)
Ligand-Bound (black)	Yes	Yes	A26 to C31	U20-A26 (B,S) A3-A26 (B) G4-A27 (B) C19-A27 (B,S) G5-A28 (B,S) C18-A28 (B) U6-A29 (B,S) U7-A30 (B,S)

[1]Base pairs A1-U22, G2-C22, A3-U20, G4-C19, and G5-C18 are physically bonded together

[2]Base pairs G11-C31, A10-U32, U9-A33 are physically bonded together, and base pair C8-A34 is enforced through B-atom to B-atom distance constraints.

[3]Backbone dihedral potentials parameterized to enforce A-form helical conformation were added to the residues within the range listed.

[4]Teriary contacts were enforced through the use of flat-well NOE distance constraints with $k_{max}=k_{min}=f_{max}=2.0$ kcal/mol. The constraints were centered on the interaction distances found in chain A of the 3FU2 crystal structure and the well width was set to 1 Å. Letters in parentheses

denote whether the constraint used was between two base atoms (B), or between a base and a sugar atom (S), or both (B,S).

Table 2.2 TOPRNA simulation parameters for the *Tte* riboswitch system

Simulation Name	**P1 Stem Paired**[1]	**P2 Stem Paired**[2]	**Tail Stacking Dihedrals**[3]	**Tail-P1 Distance Constraints**[4]
Unstacked 3' Tail (red)	Yes	No	---	---
Stacked 3' Tail (green)	Yes	No	A24 to C30	---
Stacked 3' tail with lower tail-P1 interactions (blue)	Yes	No	A24 to C30	U2-A23 (B,S) A19-A23 (B) G4-A26 (B) C17-A26 (B,S) G5-A27 (B,S) C16-A27 (B)
Unstacked 3' tail with lower tail-P1 interactions (purple)	Yes	No	---	U2-A23 (B,S) A19-A23 (B) G4-A26 (B) C17-A26 (B,S) G5-A27 (B,S) C16-A27 (B)
Stacked 3' tail with all tail-P1 interactions (orange)	Yes	No	A24 to C30	U2-A23 (B,S) A19-A23 (B) G4-A26 (B) C17-A26 (B,S) G5-A27 (B,S) C16-A27 (B)

				U6-A28 (B,S)
Ligand-Bound	Yes	Yes	A24 to C30	U2-A23 (B,S) A19-A23 (B) G4-A26 (B) C17-A26 (B,S) G5-A27 (B,S) C16-A27 (B) U6-A28 (B,S) A10-A32 (B)

[1]Base pairs C1-G20, U2-A19, G3-C18, G4-C17, and G5-C16 are physically bonded together.

[2]Base pairs C9-G33 and G11-C30 are physically bonded together, and non-canonical base pair A10-A32 is enforced through B-atom to B-atom distance constraints.

[3]Backbone dihedral potentials parameterized to enforce A-form helical conformation were added to the residues within the range listed.

[4]Flat-well NOE distance constraints were used as described in Table S2. Constraint centers were derived from the 3Q50 crystal structure.

Table 2.3 FRET values computed from TOPRNA simulated distance distributions

	Bsu				***Tte***			
	R_0=51 Å		R_0=57 Å		R_0=51 Å		R_0=57 Å	
Simulation	d_l=0 Å	d_l=10 Å	d_l=0 Å	d_l=10 Å	d_l=0 Å	d_l=10 Å	d_l=0 Å	d_l=10 Å
Ligand-Bound (black)	0.96	0.82	0.98	0.89	0.99	0.93	0.99	0.96
Stacked 3’ tail with all tail-P1 interactions (orange)	0.84	0.65	0.90	0.76	0.87	0.70	0.92	0.79
Stacked 3’ tail with lower tail-P1 interactions (blue)	0.81	0.62	0.88	0.73	0.87	0.70	0.92	0.80
Stacked 3’ tail with upper tail-P1 interactions (cyan)	0.82	0.62	0.89	0.74	--	--	--	--
Unstacked 3’ tail with lower tail-P1 interactions (purple)	0.73	0.52	0.82	0.64	0.77	0.57	0.85	0.68
Stacked 3’ tail (green)	0.38	0.24	0.48	0.33	0.39	0.25	0.48	0.34
Unstacked 3’ Tail (red)	0.38	0.23	0.49	0.32	0.39	0.24	0.50	0.34

The mean FRET value (E) corresponding to each distance distribution was obtained by assuming complete averaging over the TOPRNA generated ensembles using the following equation[121]:

Here, r is the end-to-end $E = \left\langle 1 \Big/ 1 + \left(\frac{r + d_l}{R_0} \right)^6 \right\rangle$ distance of a given conformer, R_0 is the approximate Förster radius indicated, and d_l is an additional distance added to estimate possible increases in fluorophore-fluorophore distances that may be expected from the unsimulated linkers.

2.2.4 Gō- model RNA simulations

The function of the Gō model follows the potential form:

$$V = \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{\substack{improper \\ dihedrals}} K_\chi (\chi - \chi_0)^2$$

$$+ \sum_{dihedrals} K_\phi [1 - \cos(\phi - \phi_0) + 0.5(1 - \cos(3 \times (\phi - \phi_0)))]$$

$$+ \sum_{i<j-3} \left\{ \varepsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \varepsilon_{i,j} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} \right\}$$

where the equilibrium distances (r_0) and angles (θ_0, χ_0 and ϕ_0) were determined by the native RNA structure. The equilibrium force constants (K_r, K_θ, and K_χ) were adopted from the CHARMM[119] force field based on atom types. Since multiplicity exists in the CHARMM force field for the force constants of dihedrals, we defined K_ϕ as the barrier height from the minimum to the maximum in the CHARMM dihedral potential. It should be noted that, although we are using the native dihedrals in the potential function, the RNA stability is largely dictated by the native contacts as follows. All heavy atoms within 4 Å distance were considered native contact pairs, excluding pairs that are connected within 3-bonds. The native contact potential takes the form of a Lennard-Jones 6-12 potential, where σ is the contact distance in the native structure and ε is the well depth defining the strength of the native interactions, while all non-native contacts are mutually repulsive with $\varepsilon\sigma^{12} = 1.3 \times 10^{-9} kcal\text{Å}^2 / mol$. We further partitioned the native contacts into van der Waals contacts and hydrogen bonding contacts. A hydrogen bond was identified when the distance between the acceptor (A) and donor (H-D) was less than 2.4 Å and the A-H-D angle was >120°. For contacts within the RNA, the strength of the interaction ε was 0.1 kcal/mol for van der Waals interactions, 2.15 kcal/mol for G-C hydrogen bonds, and 1.58 kcal/mol for all other hydrogen bonds. For contacts between ligand and RNA, ε of van der Waals and hydrogen bond interactions were 0.15 and 2.89 kcal/mol, respectively.

Each of the 51 folding simulations of the transcriptional and translational riboswitches were performed using the GROMACS simulation package[122] and each simulation was carried out with a different unfolded starting conformation. Stochastic dynamics were performed with a coupling time constant of 1.0 ps and a time step of 2 fs. All bonds were constrained in the simulations.

2.2.5 NMR Spectroscopy

All NMR experiments were performed at 298 K on an Avance Bruker 600 MHz spectrometer equipped with a triple-resonance cryogenic (5 mm) probe. NMR spectra were analyzed using NMR Draw[123] and Sparky 3. Uniformly labeled $^{13}C/^{15}N$ samples were prepared by *in vitro* transcription using T7 RNA polymerase as described previously[124]. RNA samples were repeatedly exchanged into NMR buffer (25 mM NaCl, 15 mM Na_iPO_4, pH 6.4, 0.1 mM EDTA) using an Ultra-4 Amicon (Millipore Corp.). Final RNA concentrations were 1-2 mM. $MgCl_2$ titrations of the *Bsu* aptamer were performed by the incremental addition of $MgCl_2$ to a 0.2 mM RNA sample. A 4:1 ratio of $preQ_1$ was added to a ~0.5 mM RNA sample to obtain the $preQ_1$-bound sample. Chemical shift differences were determined through 2D ^{1}H-^{13}C and ^{1}H-^{15}N HSQC experiments.

2.3 Results

2.3.1 smFRET of the ligand-free *Bsu* and *Tte* riboswitches

To exploit smFRET[75,89,92] for its ability to elucidate even subtle conformational differences, we chemically modified the crystallized riboswitch sequences by attaching Dy547 as a donor

fluorophore at the 3' terminus and Cy5 as an acceptor fluorophore on a uracil residue of L2 (U13 of *Bsu* and U12 of *Tte*, **Figure 2.1C** and Materials and Methods). Both of these uracils were chosen since they are weakly conserved, not involved in any intramolecular interactions, and extrude into solvent[109-111]. In addition, the donor-acceptor pair is positioned such that pseudoknot formation upon ligand binding is expected to result in close proximity (~20-30 Å) and thus high FRET, whereas extended or unfolded conformations should result in considerably longer distances and lower FRET (**Figure 2.1D**). Finally, we introduced a 5' biotin to immobilize the RNA on a quartz slide for observation of single molecules by prism-based total internal reflection fluorescence microscopy (TIRFM, **Figure 2.1D**), essentially as described[69,89,92], and used a buffer approximating physiological conditions (50 mM Tris-HCl, pH 7.5 100 mM K^+, 1 mM Mg^{2+}) at room temperature.

In the absence of $preQ_1$ ligand, FRET histograms of a few hundred molecules (as indicated in **Figure 2.2A**), which survey conformational sampling of the entire population[93], exhibit a major broad peak around a FRET value of 0.72 (mean) ± 0.13 (standard deviation, SD) and 0.70 ± 0.13 for the *Bsu* and *Tte* riboswitches, respectively (**Figure 2.2A**). Additionally, both ligand-free riboswitches contain a minor population of a shorter-distance conformation as indicated by a higher FRET value; the *Bsu* riboswitch shows ~9% with a FRET value of 0.89 ± 0.05, whereas the *Tte* riboswitch shows ~11% with a FRET value of 0.90 ± 0.05. Notably, the width (SD) of the 0.7 (or mid)-FRET state is larger than that of the 0.9 (or high)-FRET state, suggesting that especially the former represents a broader dynamic ensemble of structures (or possibly one that our FRET probes more sensitively report on) rather than a single defined conformation. Examination of individual traces shows that the high-FRET state arises in part from relatively short-lived excursions of single molecules up from the mid-FRET state (**Figure**

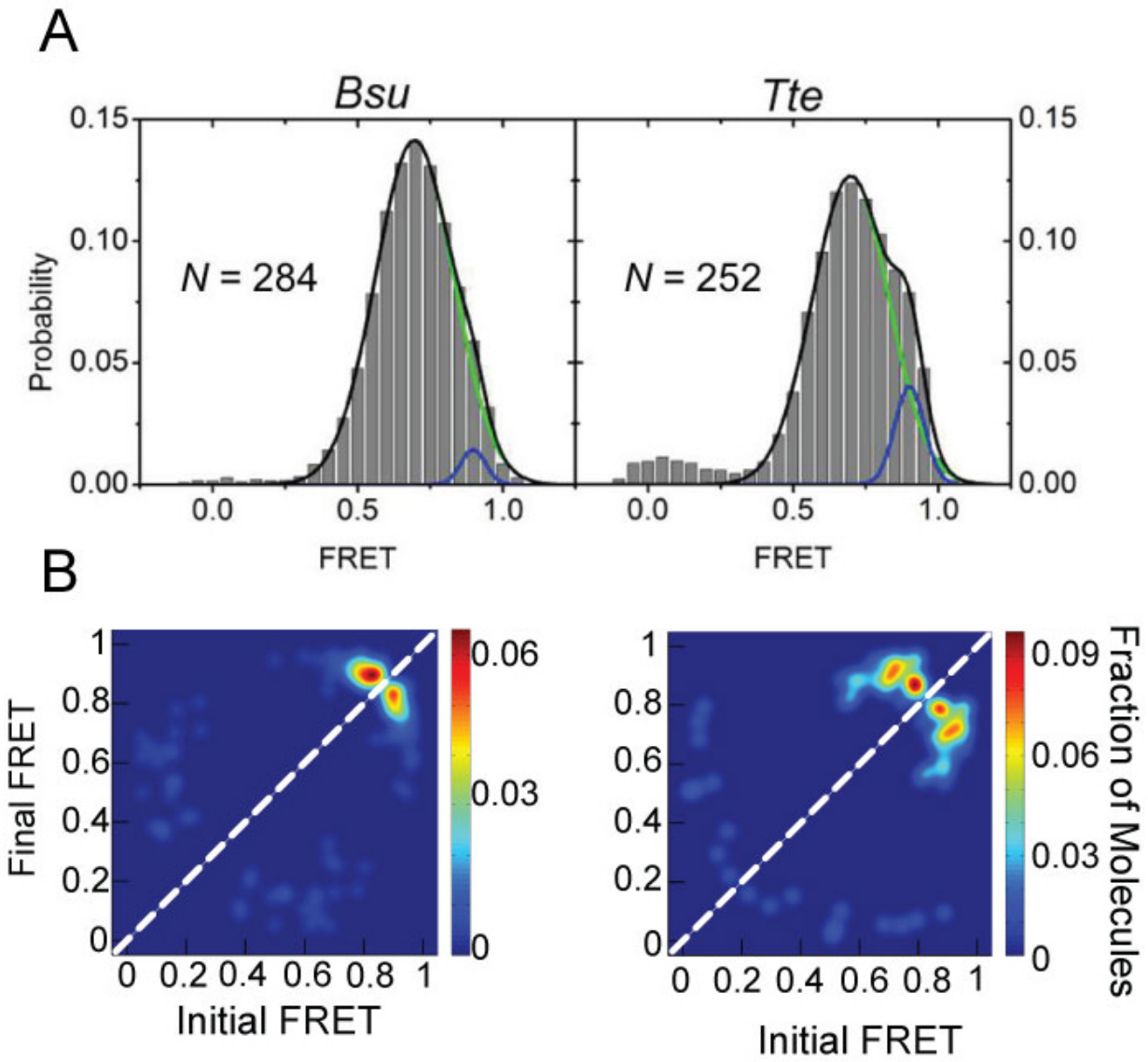

Figure 2.2 FRET behavior of *Bsu* and *Tte* aptamers in the absence of ligand.
(A) FRET population histograms for *Bsu* and *Tte* riboswitch aptamers in the absence of ligand in near-physiological buffer. *N* corresponds to the number of molecules sampled. Green and blue lines indicate the Gaussian fits of the mid- and high-FRET states, respectively. Black lines indicate the cumulative fit. (B) Transition occupancy density plots (TODPs) of *Bsu* and *Tte* riboswitch aptamers in the absence of ligand. TODPs are displayed as heat maps illustrating the fraction of all molecules exhibiting a transition from the mid-FRET to high-FRET state, displayed as off-diagonal contours.

2.2B and **2.3**). Additionally, some molecules appear to stably persist in the high-FRET state with dwell times of up to tens of seconds (**Fig. 2.3**, panel IV). Heat annealing at 90 °C in the absence of Mg^{2+} (rather than at 70 °C in the presence of Mg^{2+}) to more rigorously remove residual structure only caused slight changes in the FRET distributions of the two riboswitches (**Figure 2.4** and Materials and Methods).

A distinction exists when comparing the conformational dynamics of the two riboswitches in the absence of ligand. Almost half of all *Tte* riboswitch molecules undergo dynamic switching between the mid- and high-FRET states, as clearly distinguished at our time resolution (100 ms, **Figure 2.3B**). In contrast, ~13% of all *Bsu* riboswitch molecules undergo such observable conformational switching (**Figure 2.3A**). Closer inspection reveals anti-correlation between the donor and acceptor intensities when conformational switching does occur in either of the two riboswitches, implying true transitions between conformational states of distinct fluorophore distance rather than local quenching effects of just one of the fluorophores. For the *Bsu* riboswitch, the timescales of these transitions are faster than those of the *Tte* riboswitch and are close to our time resolution (**Figure 2.3A**), suggesting that we may miss a significant number of even faster transitions. We therefore performed cross-correlation analysis[125] on these data, which revealed that an additional 15% of all *Bsu* molecules show anti-correlation between the donor and acceptor signals without transitions revealed by Hidden Markov Modeling (HMM) (Materials and Methods, **Figure 2.5A**). To further evaluate the underlying dynamics, we measured the *Bsu* riboswitch in the absence of ligand at 33 ms time resolution. As expected, this faster time resolution increases the population of molecules with HMM-resolved transitions to 64%, with an additional 10% displaying anti-correlation without discernible transitions (**Figure 2.5B**). Even at this faster time resolution, however, 26% of all

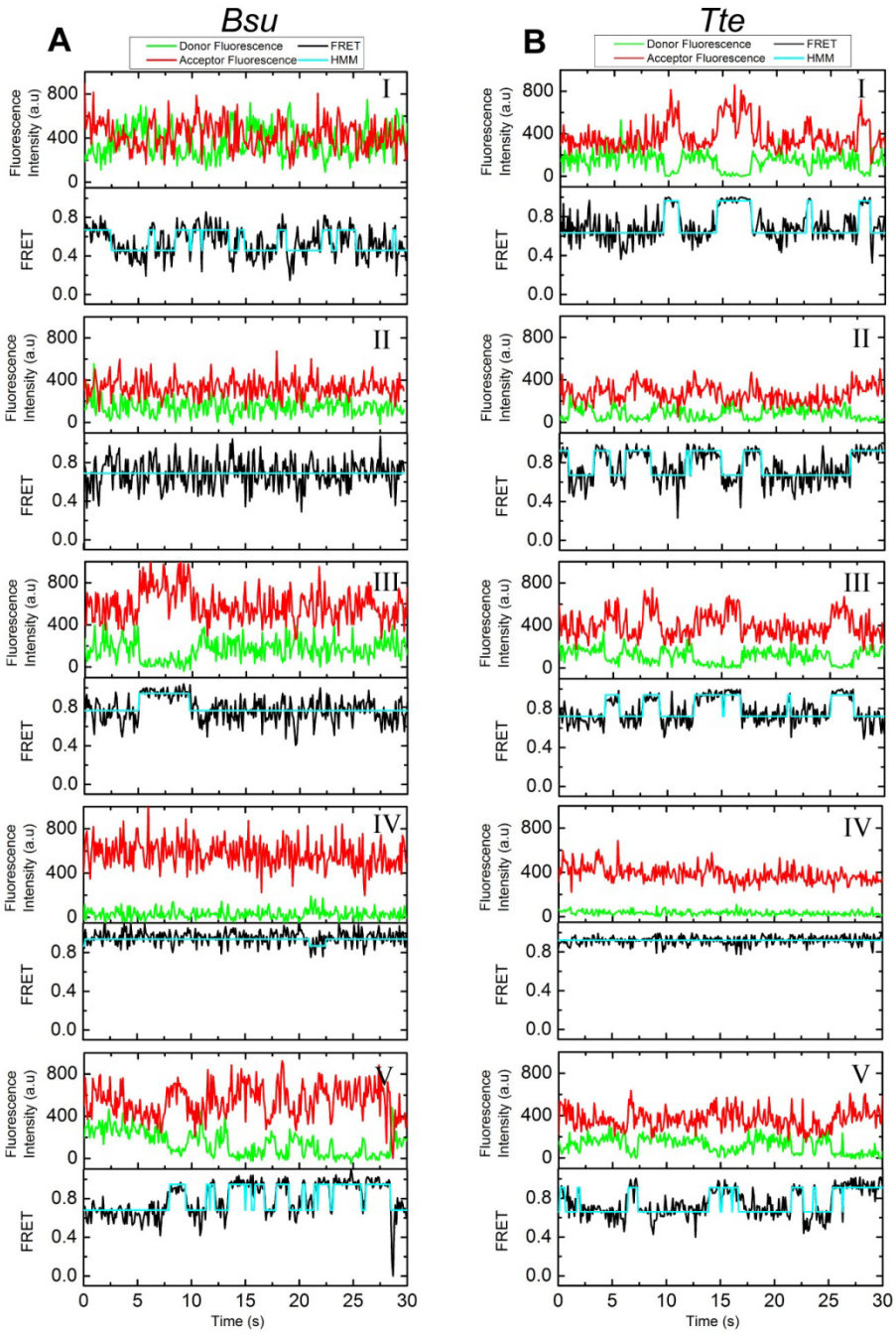

Figure 2.3 Representative raw smFRET traces of the *Bsu* and *Tte* riboswitch aptamers in the absence of ligand.

(A) Five representative time traces illustrating the donor (green) and acceptor (red) intensities with corresponding FRET (black) traces for the *Bsu* riboswitch. Computed HMM fits (cyan) are overlaid on the FRET trace. (B) Same as (A) but for *Tte*.

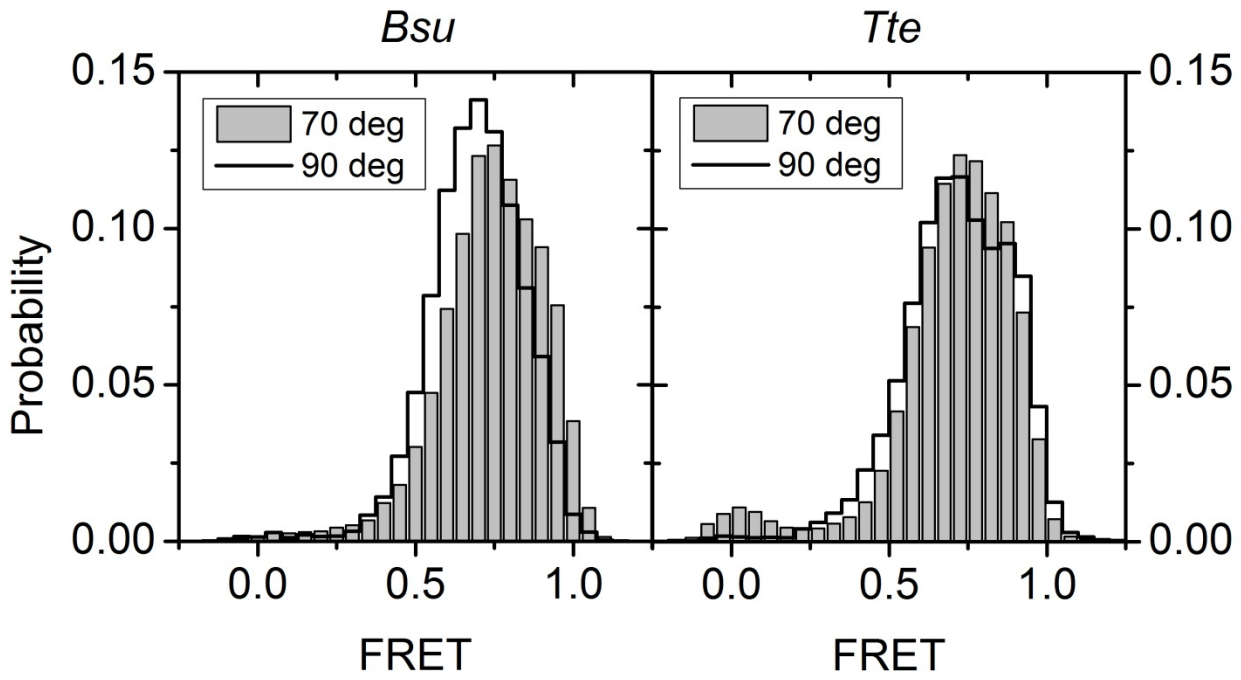

Figure 2.4 Effect of differing heat annealing conditions of the preQ$_1$ riboswitches.
smFRET histograms of the *Bsu* and *Tte* riboswitches in the absence of ligand under different heat annealing conditions. RNA was either heated to 70°C in smFRET buffer and slow cooled to room temperature (gray bars) or heat annealed at 90°C in the absence of Mg^{2+}, and slow cooled to room temperature followed by 1 mM Mg^{2+} addition (black lines).

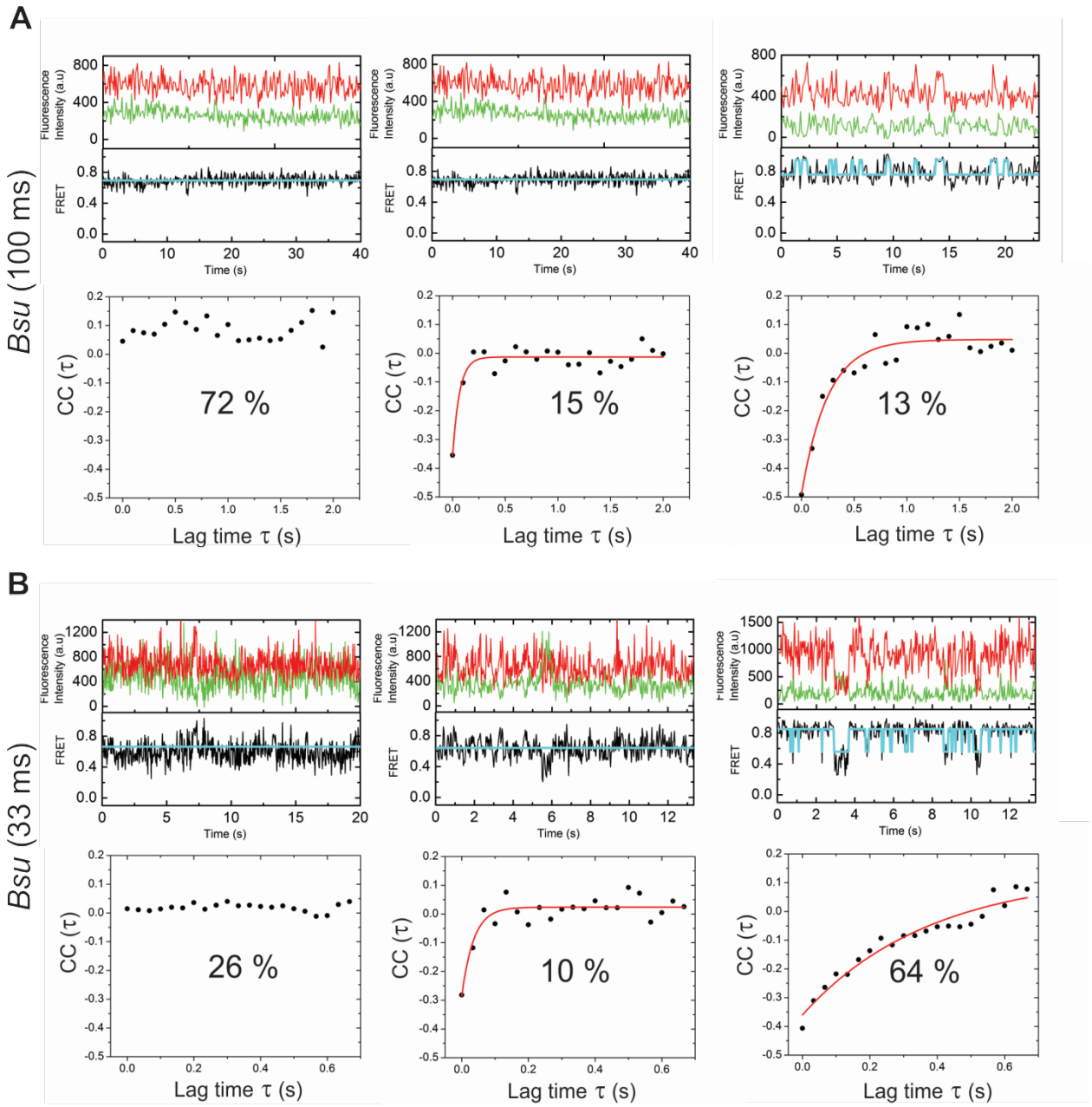

Figure 2.5 Donor-acceptor cross-correlation analysis of exemplary smFRET traces of the *Bsu* riboswitch aptamer in the absence of ligand.

(A) Cross-correlation analysis at 100 ms time resolution, showing smFRET traces with their HMM fits (cyan, top panel) and cross-correlation functions, fit with single-exponentials (red, bottom panel). Fractions are given for each of three observed behaviors: left, trace with no detectable dynamics; middle, trace with fast dynamics as shown by cross-correlation between the donor and acceptor signals; right, trace with slow dynamics as identified by the HMM. (B) Same as in A but at 33 ms time resolution. The time constants (τ) for the single exponential fits are: A, 0.065 s and 0.082 s; B, 0.035 s and 0.39 s.

Bsu molecules in the absence of $preQ_1$ reveal no detectable dynamics (**Figure 2.5B**), suggesting that at least this fraction of molecules undergoes transitions that are faster still, similar to observations on the *Fnu* $preQ_1$ riboswitch[43].

A first-order analysis of the mean FRET values associated with the two histogram peaks suggests that the high-FRET state of both riboswitches is consistent with the range of distances expected between fluorophores for the native ligand bound states, which is significantly below the Förster distance of the Dy547-Cy5 fluorophore pair (~50 Å)[126]. The distance of ≤30 Å associated with a FRET value of 0.9 agrees well with the crystal structures of the ligand-bound *Bsu*[110] and *Tte*[42] riboswitches, as well as with that of the ligand-free *Tte* riboswitch[42]. As discussed above, the broad mid-FRET state likely reflects a dynamic ensemble of partially unfolded states. However, its relatively high FRET value of 0.7, corresponding to a donor-acceptor distance of ~45 Å, appears much higher than that expected if the conformation of the A-tail were truly random.

2.3.2 Coarse-grained RNA simulations provide conformational insight into FRET states

To gain insight into the conformational ensembles underlying the mid- and high-FRET states, we performed simulations of both riboswitches with a coarse-grained RNA model we term TOPRNA. TOPRNA uses three pseudo-atoms per nucleotide of an RNA parameterized as a freely rotatable polymer with RNA-consistent bond lengths and angles and with only repulsive non-bonded van der Waals interactions (**Materials and Methods**). Known base-paired regions are parameterized to assume A-form helical structure, and tertiary structures are modeled using dihedral and distance restraints based on the available crystal structures. TOPRNA simulations thus allow us to build a comprehensive picture of the 3D conformational ensemble of the $preQ_1$

riboswitches subject to given sets of tertiary structure constraints and inherent space-filling and chain-connectivity properties. Consistent with our initial expectations, we found that ensembles generated without any enforced tertiary restraints possess inter-dye distance distributions that are too long to give a mid-FRET value of ~ 0.7 (**Figure 2.6** and **Table 2.3**). In contrast, the ensembles with either a partially P1- or P2-docked 3' tail lead to mean donor-acceptor distances in the range of 35-45 Å, highly consistent with that observed for the mid-FRET state (**Figure 2.6** and **Table S3**). We note that these partially folded conformations are still flexible enough to lead to a broad distance (and therefore FRET) distribution as observed for the mid-FRET state (**Figure 2.2A**). Finally, our simulations of the fully folded, ligand-bound state produced a narrower and shorter distance distribution with means around ~25 Å, as expected for the high-FRET state (**Figure 2.6** and **Table 2.3**). Altogether, we have strong evidence that the mid-FRET state in both riboswitches consists of a pre-folded conformational ensemble in which the A-tail transiently interacts with the minor groove of P1, P2 is partially formed, or a combination of both, whereas the high-FRET state represents a compact, less flexible ligand-bound state (or, in the absence of ligand, a folded-like state).

2.3.3 Buffer conditions impact the ligand-free *Bsu* and *Tte* riboswitches

Despite subtle differences in their dynamics, the *Bsu* and *Tte* riboswitches in the absence of ligand exhibit surprisingly similar conformational distributions. These observations contrast with previous suggestions from NMR spectroscopy that the ligand-free *Bsu* riboswitch largely resides in a partially folded, open conformation[109,118], whereas the crystallized ligand-free *Tte* riboswitch is found in a conformation highly similar to that when ligand is bound, which becomes only slightly less compact in solution[42]. To resolve this apparent discrepancy, we studied the buffer

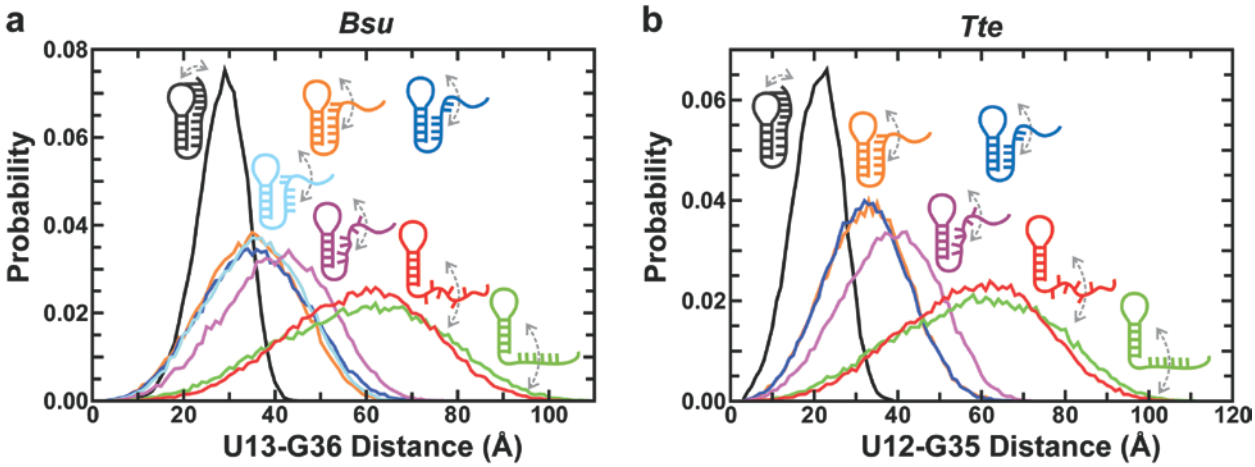

Figure 2.6 Coarse grained TOPRNA simulations.
TOPRNA simulations predict the distance distributions between the fluorophore labeled residues as a function of specific interactions in the *Bsu* (A) and *Tte* (B) riboswitch aptamers. Color code is as follows: green, stacked 3' tail; red, unstacked 3' tail; purple, blue, orange and cyan (*Bsu* only), partially docked into the P1 and/or P2 stem with varying degrees of intersegmental and stacking interactions as indicated; black, fully folded as found in the ligand-bound crystal structures.

dependence of the *Bsu* riboswitch using smFRET. We observed that by lowering the Mg^{2+} concentration and pH from those of our near-physiological smFRET buffer to those of a typical NMR buffer (15 mM Na_iPO_4, pH 6.4, 25 mM NaCl, 0.1 mM EDTA), the mid-FRET peak of the pre-folded state decreases significantly to a FRET value of 0.62 ± 0.20 with an increased SD (**Figure 2.7**). These observations suggest that the tertiary interactions between the 3' tail and P1-L1 stem-loop become less favorable, consistent with previous NMR studies that could not detect them[109,118]. Furthermore, when this NMR buffer is supplemented with 2 mM Mg^{2+}, the mid-FRET ensemble shifts back up to 0.69 ± 0.16, close to the smFRET buffer value (**Figure 2.7**). These observations suggest that the addition of Mg^{2+}, and to an extent the increase to a near-physiological pH, favors a more compact conformational ensemble with transient interactions of the 3' tail with the P1-L1 stem-loop.

Although Mg^{2+} is dispensable for $preQ_1$ binding and recent work found no specific binding sites for Mg^{2+} in the ligand-bound *Bsu* riboswitch[114], its role in the ligand-free state of both the riboswitches remains unclear. One Mg^{2+} ion was found adjacent to the ligand binding pocket of the ligand-free *Tte* structure[42], indicating Mg^{2+} may be important in stabilizing the ligand-free state. To further pinpoint the effect of Mg^{2+} on the ligand-free *Bsu* and *Tte* riboswitches, we performed Mg^{2+} titrations in our smFRET buffer and monitored the resulting conformational populations. In the absence of Mg^{2+}, the smFRET population histogram for the *Bsu* riboswitch exhibits a major broad peak around a FRET value of 0.61 ± 0.20 and a minor peak around 0.92 ± 0.06 (**Figure 2.8A**). Similarly, the *Tte* riboswitch shows a FRET peak around 0.71 ± 0.17, accompanied by a peak at 0.94 ± 0.06 (**Figure 2.8B**). The lower mean FRET value and larger width of the *Bsu* mid-FRET state is consistent with our observations in NMR buffer (**Figure 2.7**) and shows that the ligand-free *Tte* riboswitch is more compact than the *Bsu*

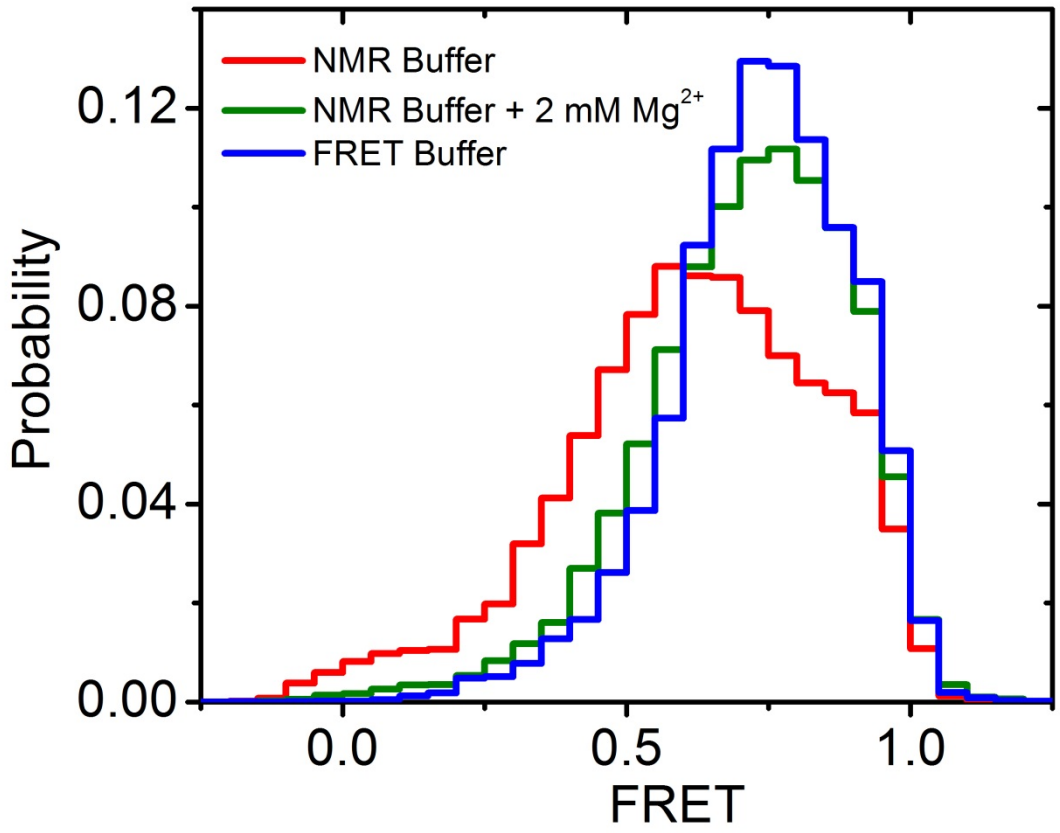

Figure 2.7 Buffer dependence of the *Bsu* riboswitch aptamer in the absence of ligand. Low ionic strength NMR buffer (15 mM Na_iPO4, pH 6.4, 25 mM NaCl, 0.1 mM EDTA; red) shifts the mean FRET value of the mid-FRET state down to 0.61. Supplementing NMR buffer with 2 mM Mg^{2+} (green) shifts the value back up to 0.69, close to the mean FRET value of the mid-FRET state in near-physiological smFRET buffer (50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM $MgCl_2$; blue).

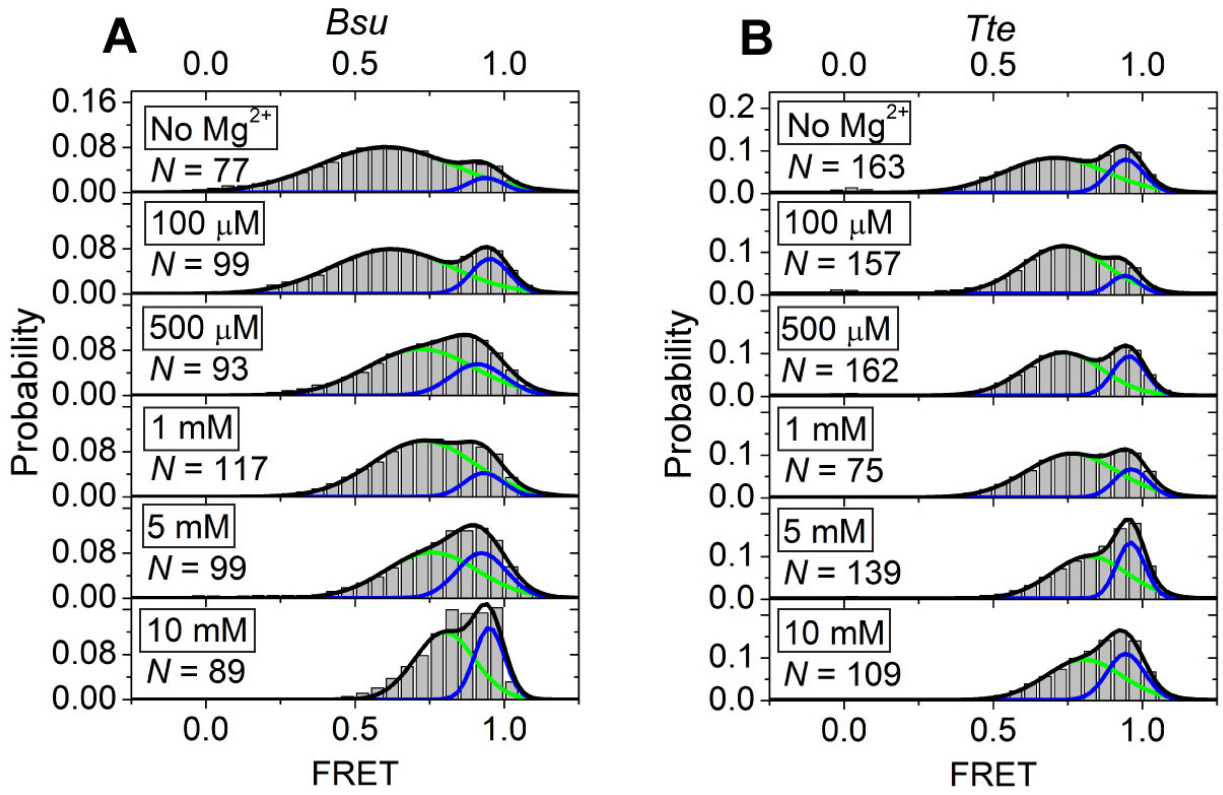

Figure 2.8 Mg^{2+} titration of the *Bsu* (A) and *Tte* (B) riboswitches.
The Mg^{2+} concentrations are indicated in the boxes. *N*, number of molecules sampled per condition. Green and blue lines indicate individual Gaussian fits of the mid-FRET and high-FRET states, respectively. Black lines indicate cumulative fits.

riboswitch in the absence of Mg^{2+}. Increasing the Mg^{2+} concentration at constant ionic strength results in a higher mean FRET value and smaller width of particularly the mid-FRET state for both riboswitches (**Figure 2.8**), suggesting that this ensemble becomes more compact and ordered. In addition, the relative fraction of the high-FRET state increases, approaching ~34% and ~36% at 10 mM Mg^{2+} for the *Bsu* and *Tte* riboswitches, respectively. These values are similar to the fractions of high-FRET state at high ligand concentrations. We conclude that high concentrations of Mg^{2+} alone, in the absence of ligand, can induce compact folded-like conformations, consistent with recent studies of the SAM-II riboswitch[41,54].

To further reconcile our smFRET data with those from the previous NMR study of the ligand-free *Bsu* riboswitch[109,118], we studied the buffer dependence of the *Bsu* riboswitch using solution-state NMR. Our previous NMR studies of the *Bsu* aptamer revealed a kissing-dimer interaction involving the palindromic L2 loop sequence 5'-U_9AGCUA_{14}-3' (**Figure 2.1C**)[118], as observed also for other $preQ_1$ riboswitches[112]. A double C12U/C15U mutant eliminates dimer formation at the high concentrations used for NMR[109,118]. We observed that, in contrast to the wild-type *Bsu* aptamer, Mg^{2+} addition now causes significant chemical shift perturbations in the absence of ligand; notably, residues around the ligand binding site move to unusual spectral positions that are typically associated with tertiary interactions (**Figure 2.9**), suggesting that the ligand binding pocket is in a folded-like conformation. Yet some of these new NMR resonances differ from those in the ligand-bound conformation (**Figure 2.9**), indicating a distinct Mg^{2+}-dependent conformation. In addition, resonances corresponding to the 3' tail interacting with the P1 helix are not observed, indicating that it does not stably dock with the helix even upon addition of Mg^{2+}. These findings agree closely with our smFRET-monitored Mg^{2+} titration (**Figure 2.8**) and together provide strong evidence that the capacity to form dimers at the high

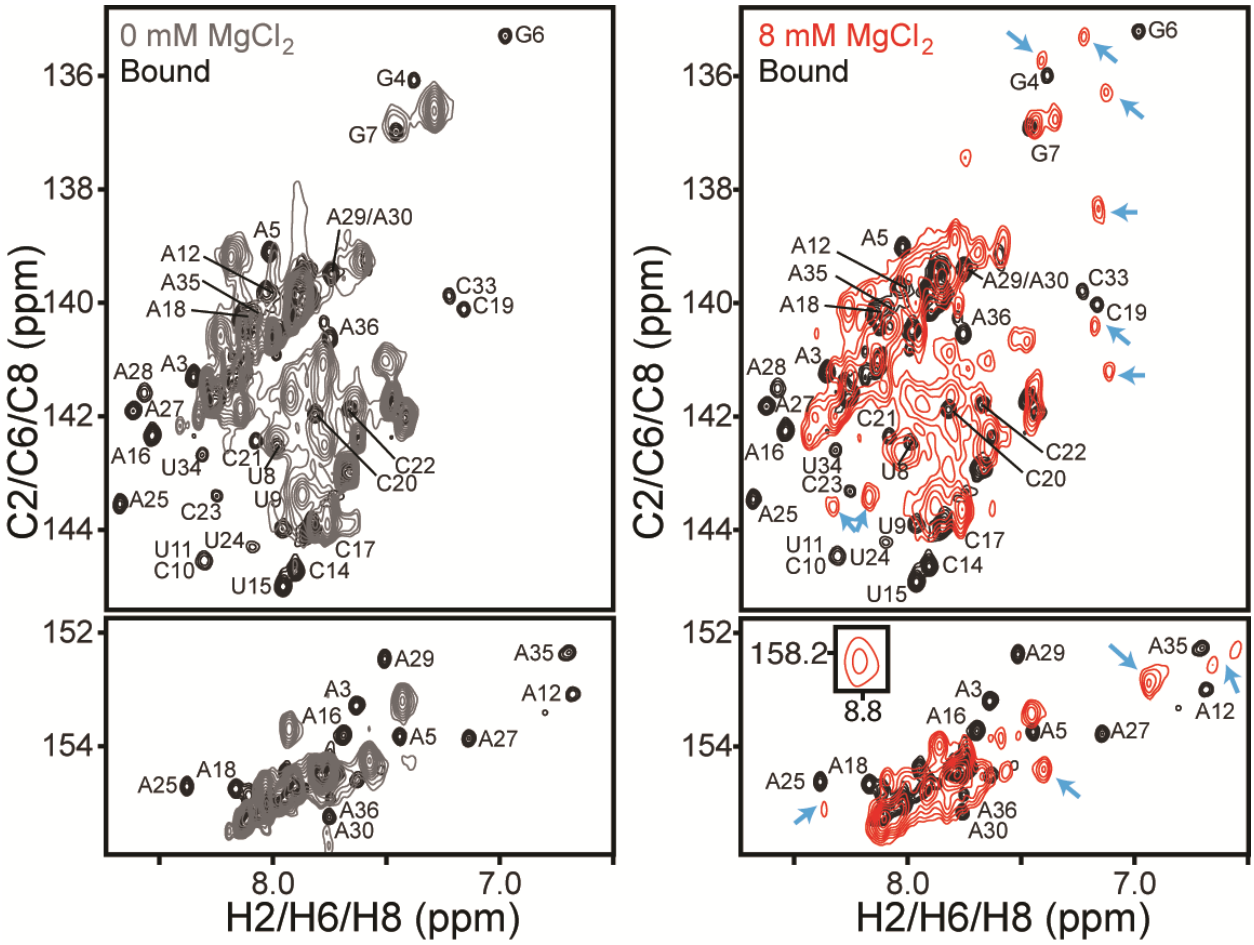

Figure 2.9 NMR characterization of the *Bsu* preQ$_1$ riboswitch – effect of Mg^{2+} on the ligand-free conformation.
From left to right: 2D ^{1}H-^{13}C HSQC comparison of free and preQ$_1$-bound *Bsu* aptamer; comparison of preQ$_1$-free *Bsu* aptamer in the absence of $MgCl_2$ (grey), with 8 mM $MgCl_2$ (red), and with preQ$_1$-bound *Bsu* aptamer in the absence of $MgCl_2$ (black). Addition of $MgCl_2$ gives rise to new peaks that are similar to preQ$_1$-bound chemical shifts in the ligand binding pocket, indicating that Mg^{2+} pre-organizes the ligand-free conformation. Arrows point to chemical shifts indicative of tertiary interactions that are different from those of the ligand bound conformation. Nucleotides are numbered following previous NMR studies on the *Bsu* riboswitch.

concentration used for NMR as well as buffer differences, mainly Mg^{2+} account for the discrepancies between smFRET and NMR studies of the ligand-free *Bsu* riboswitch. This is also clearly demonstrated in a recent NMR study on the *Fnu* $preQ_1$ riboswitch[43].

2.3.4 Effect of $preQ_1$ on the *Bsu* and *Tte* FRET profiles

To confirm our preliminary assignments of the mid- and high-FRET states, we studied the effect of ligand on the conformational sampling of the two riboswitches. As the $preQ_1$ concentration increases, so does the high-FRET state population, at the expense of that of the mid-FRET state (**Figrues 2.10A,B**). We fitted each FRET histogram with a sum of two Gaussian functions (**Figures 2.10A,B**) and plotted the fraction of the high-FRET state as a function of ligand concentration (**Figure 2.11**). For both riboswitches, the high-FRET population increases with ligand concentration and saturates with half-titration points of $K_{1/2}$ = 134 ± 45 nM and 69 ± 22 nM for the *Bsu* and *Tte* riboswitches, respectively (**Figure 2.11A**). To rule out the possibility that the varying number of molecules across our ligand titrations skews our results, we performed an analysis wherein 100 molecules were randomly chosen for each condition, then analyzed as in **Figures 2.10** and **2.11**. We found no difference in the Gaussian distributions and $K_{1/2}$ values outside the stated errors. We note that the apparent $preQ_1$ affinity of the two riboswitches differs somewhat from previously reported values ($K_{1/2}$ = 50 nM in 50 mM Tris-HCl, pH 8.3, 20 mM $MgCl_2$, 100 mM KCl[115] and 2 nM in 10 mM sodium cacodylate, pH 7.0, 3 mM $MgCl_2$[42] for the *Bsu* and *Tte* riboswitches, respectively). This may be due to differences in technique (in-line probing, surface immobilization) and/or buffer conditions, especially pH and Mg^{2+} concentration, which we have shown to have a significant effect on the compactness of both riboswitches (**Figures 2.7** and **2.8**).

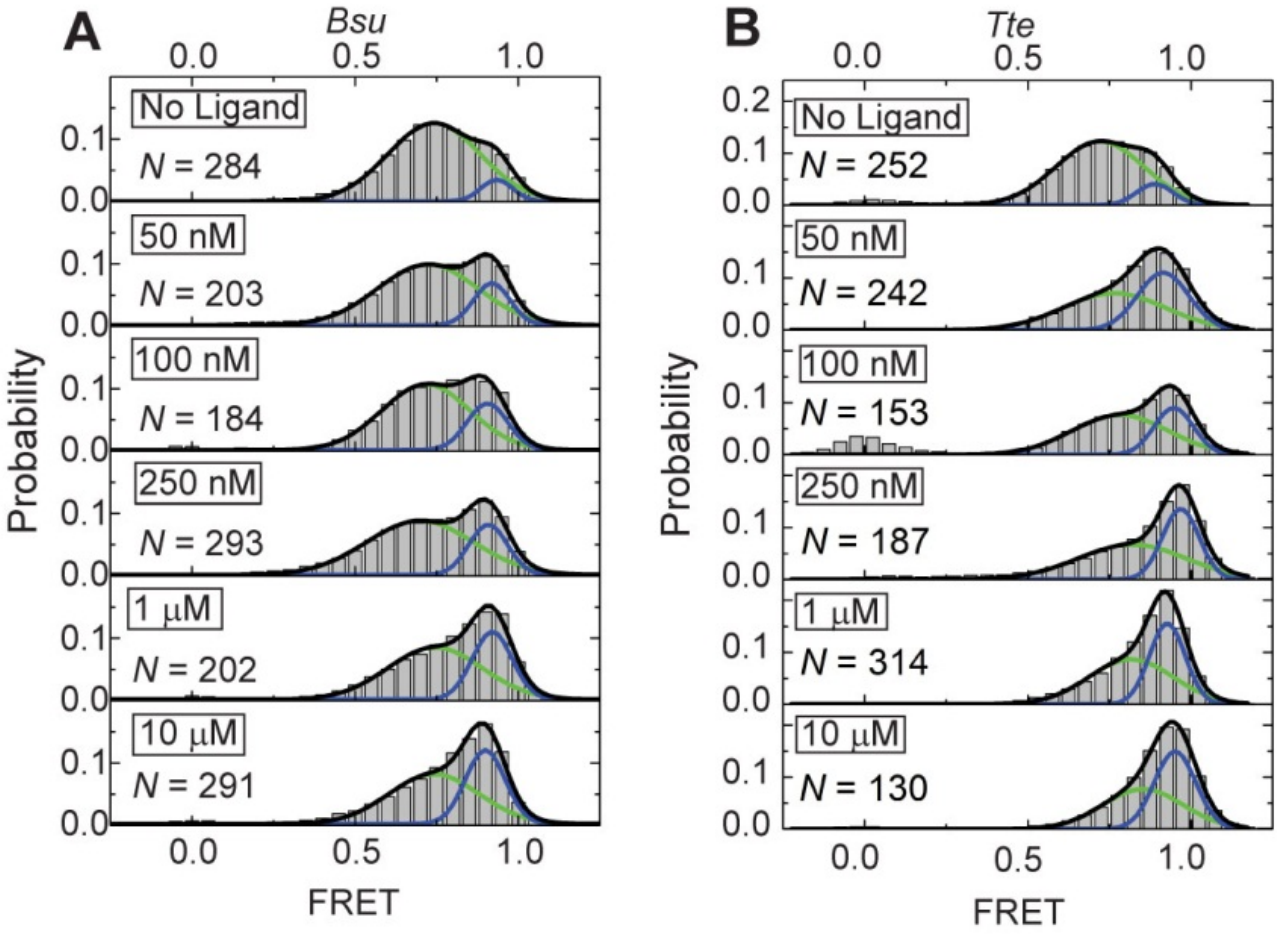

Figure 2.10 Effect of ligand on single preQ$_1$ riboswitch molecules.
(A) smFRET population histograms of the *Bsu* riboswitch aptamer with increasing ligand concentration as indicated. *N*, number of molecules sampled. Green and blue lines indicate Gaussian fits of the mid- and high-FRET states, respectively. Black lines indicate cumulative fits. (B) Same as in (A), but for the *Tte* riboswitch aptamer.

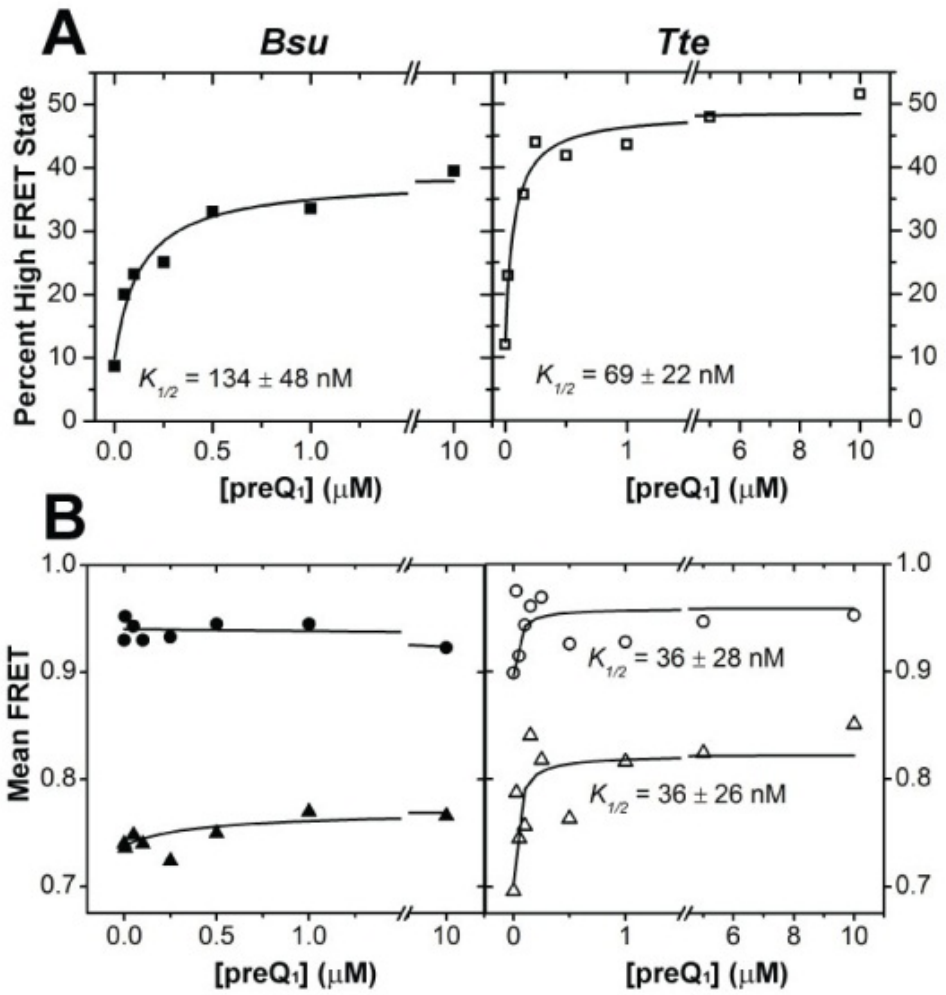

Figure 2.11 Effect of ligand on the distribution of the mid- and high-FRET states.
(A) The FRET histograms of Figure 2 were quantified and the percent high-FRET state was plotted as a function of ligand concentration. The data were fit with a non-cooperative binding isotherm and the respective apparent $K_{1/2}$ values are indicated for both *Bsu* (closed symbols) and *Tte* (open symbols). (B) The centers of the Gaussian fits for the mid-FRET (green) and high-FRET (blue) states from Figure 2 were plotted as a function of ligand concentration and fit with a non-cooperative binding isotherm, yielding the $K_{1/2}$ values indicated for the *Tte* riboswitch.

These observations strongly implicate the high-FRET state as the ligand-bound, fully folded state. We note that even at $preQ_1$ concentrations as high as 10 μM, however, the fraction of this folded state does not shift above ~50% for either of the two riboswitches, in part because both riboswitches remain dynamic and sample the pre-folded state, perhaps due to the existence of alternative RNA conformations incapable of binding ligand[89,109]. In both riboswitches, ligand addition also causes changes in the transition kinetics between the mid- and high-FRET states (**Figures 2.3, 2.12** and **2.13**). Whereas the *Tte* riboswitch becomes less dynamic with slower transitions, the *Bsu* riboswitch displays increased dynamics with clear two-state transitions. These observations further support the notion that the ligand-free *Bsu* riboswitch undergoes very fast transitions that are slowed down by the ligand through stabilization of the folded state, thus enabling detection at our 100 ms time resolution.

When fitting the FRET histograms from the $preQ_1$ titration in Figure 2 with Gaussian functions, we noticed another difference between the riboswitches. In the case of the *Bsu* riboswitch, the mean FRET values of both the pre-folded and folded states vary very little with increasing $preQ_1$ concentration (**Figure 2.11B**). In contrast, for the *Tte* riboswitch the mean FRET value of particularly the pre-folded state significantly increases upon ligand titration, with a $K_{1/2}$ of 36 ± 26 nM (**Figure 2.11B**), within error of that observed for the increase in fraction of high-FRET state (**Figure 2.11A**). This observation provides evidence that the *Tte* pre-folded state with a partially formed binding site (without P2 formed yet) "senses" rising $preQ_1$ concentrations in that its broad conformational ensemble (likely a time average of many conformers) is increasingly biased towards folded-like (more compact) conformations with increasing ligand concentration, thus narrowing the gap between the mean FRET values of the pre-folded and folded states. The folded state also shifts to higher FRET values, but less so

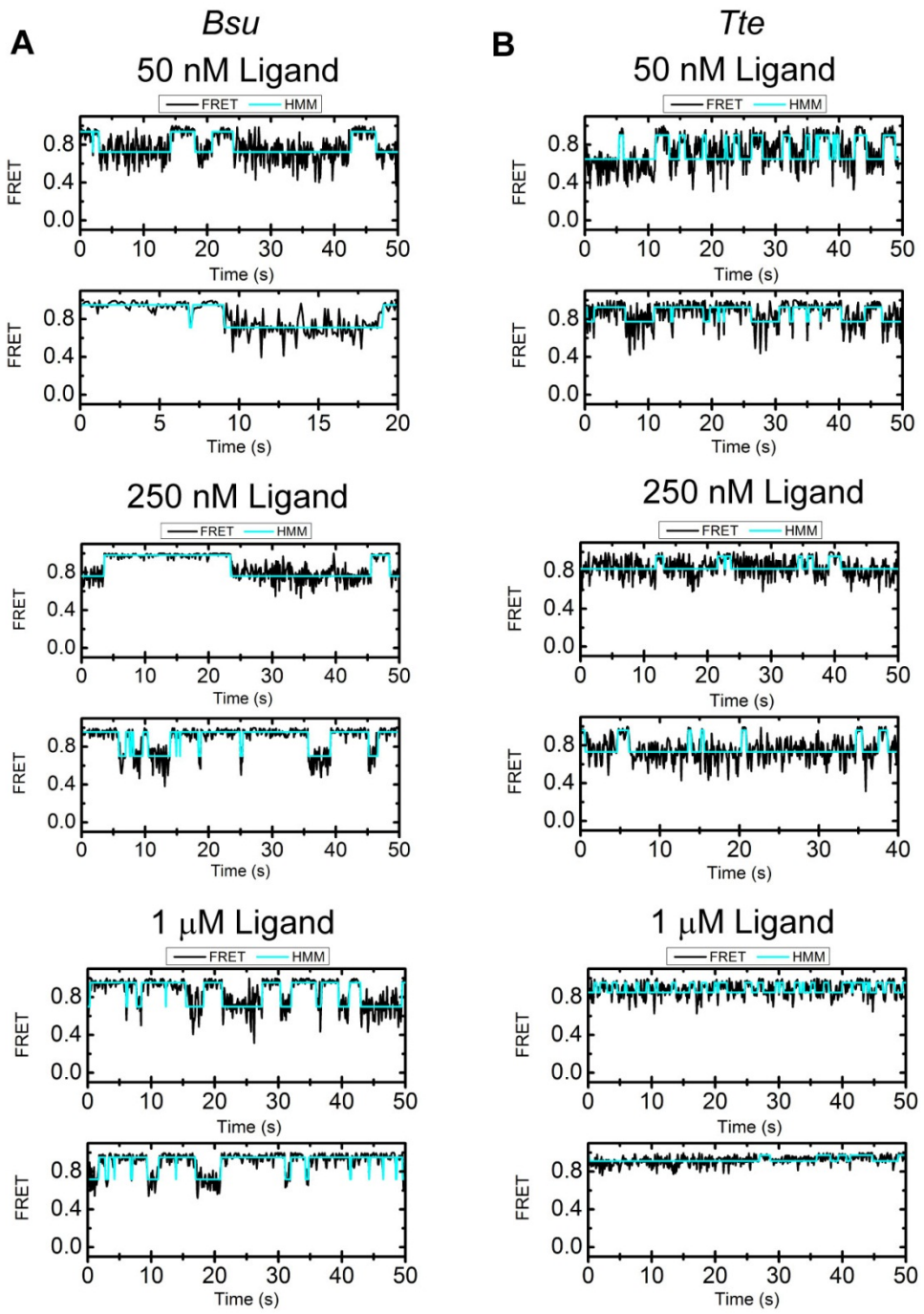

Figure 2.12 Exemplary smFRET traces of the *Bsu* and *Tte* riboswitches in the presence of preQ$_1$.
(A) smFRET traces (black) for the *Bsu* riboswitch with HMM fits (cyan) overlaid. (B) Same as A, but for the *Tte* riboswitch. The *Bsu* riboswitch shows fewer transitions at low (50 nM) ligand concentration than the *Tte* riboswitch. Conversely, at intermediate (250 nM) and high ligand concentrations (1 μM), the *Tte* riboswitch shows less dynamics than the *Bsu* riboswitch.

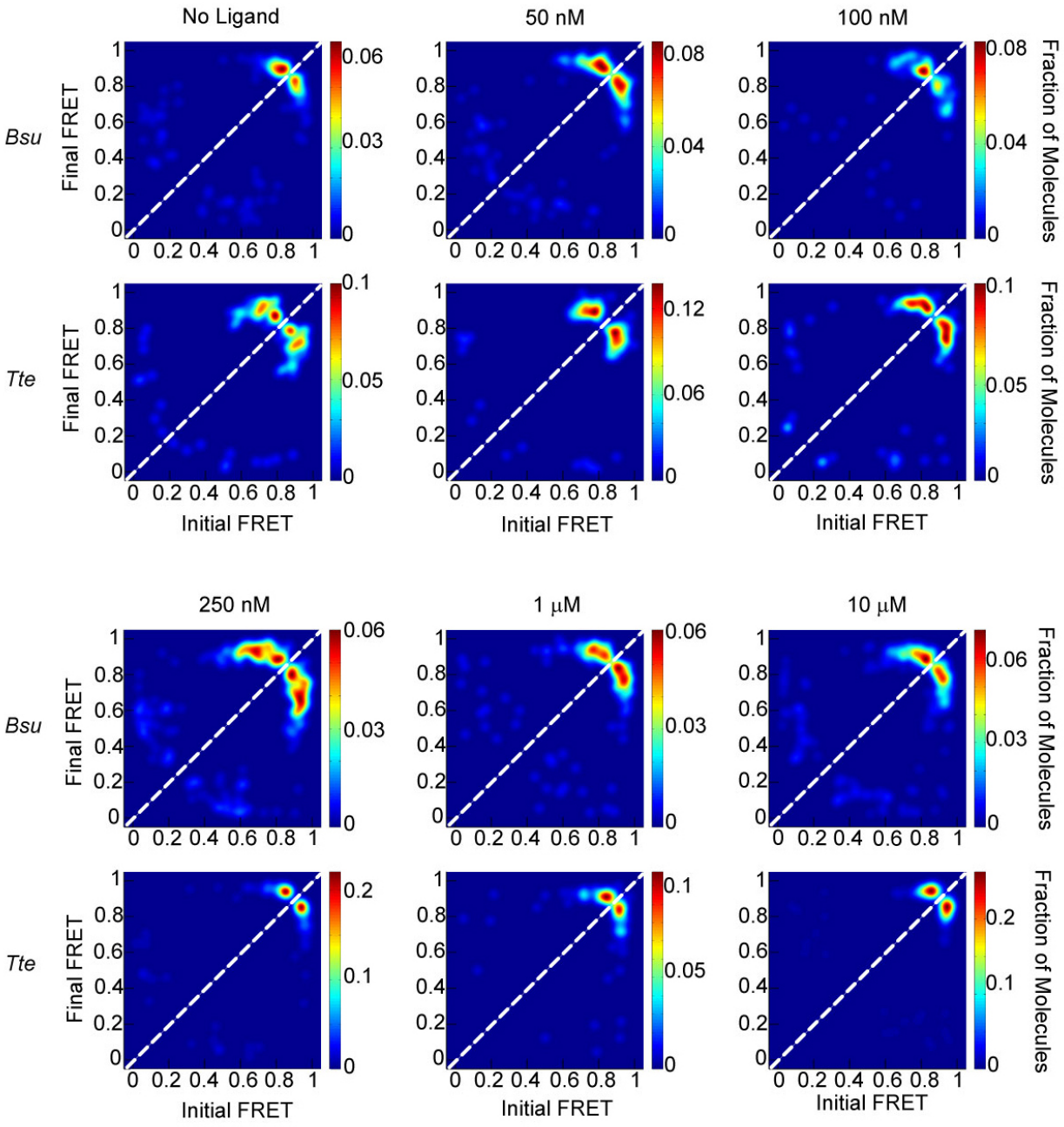

Figure 2.13 Transition occupancy density plots (TODPs) of the preQ$_1$ riboswitches at varying ligand concentrations.

TODPs[93] are displayed as heat maps illustrating the fraction of all molecules that exhibit a specific transition from an initial FRET state to a final FRET state for the *Bsu* and *Tte* riboswitches, as indicated. The plots highlight the differences in the transitions between the pre-folded (~0.7 FRET) and folded (~0.9 FRET) states as a function of ligand concentration; transitions between the pre-folded and folded states are seen as off-diagonal contours. In the ligand-free *Tte* riboswitch, these contours move closer to the diagonal (dashed line) upon increasing the ligand concentration, indicative of the pre-folded and folded states becoming structurally more similar. By contrast, the ligand-free *Bsu* riboswitch has fewer observed

transitions between the pre-folded and folded states and the TODP with no ligand displays contours close to the diagonal that move farther away with increasing ligand concentration (until at least ~250 nM ligand). This behavior is likely caused by stabilization of the folded high-FRET state by the ligand, resulting in slower transitions that become increasingly resolved at our 100 ms time resolution.

(**Figure 2.11B**). There are several ligand-binding mechanisms that could result in these observations. For one, the pre-folded state may rapidly (on the smFRET timescale) sample the native ligand bound state, and the ligand shifts this pre-existing equilibrium by increasing the population of bound conformation, either via conformational selection (ligand binding to an already pre-folded binding pocket) or induced fit (ligand binding to a relatively disordered pocket, thereby inducing folding of the binding pocket), thus resulting in the observed increased FRET value with increasing ligand concentration. Alternatively, although less likely, ligand binding may stabilize increasingly more native-like pre-folded states of the RNA corresponding to different levels of tail docking. Although our smFRET data themselves cannot resolve the ligand-binding mechanism, they reveal significant differences in the folding behavior of the two riboswitches on the path towards structurally similar ligand-bound end states.

2.3.5 Gō-model simulations suggest alternative ligand-mediated folding pathways

To better elucidate the differences in *Bsu* and *Tte* riboswitch folding, we utilized all-atom Gō-model simulations to compare the folding pathways of the *Bsu* and *Tte* riboswitches (**Figure 2.14**), essentially as described[107]. Gō models are biased toward the fully folded state, leading to smooth folding free energy landscapes and have been used to study the folding of biomolecules, especially in cases where assessing folding pathways experimentally is challenging[107,127,128]. By performing 51 single molecule simulations for each riboswitch and averaging the fraction of native contacts, Q, as a measure of folding progress, we explored the prevailing folding pathways (**Figures 2.14** and **2.15**). We found that both riboswitches follow a similar folding pathway initially as formation of the local P1 stem precedes that of the distal P2 stem. The order

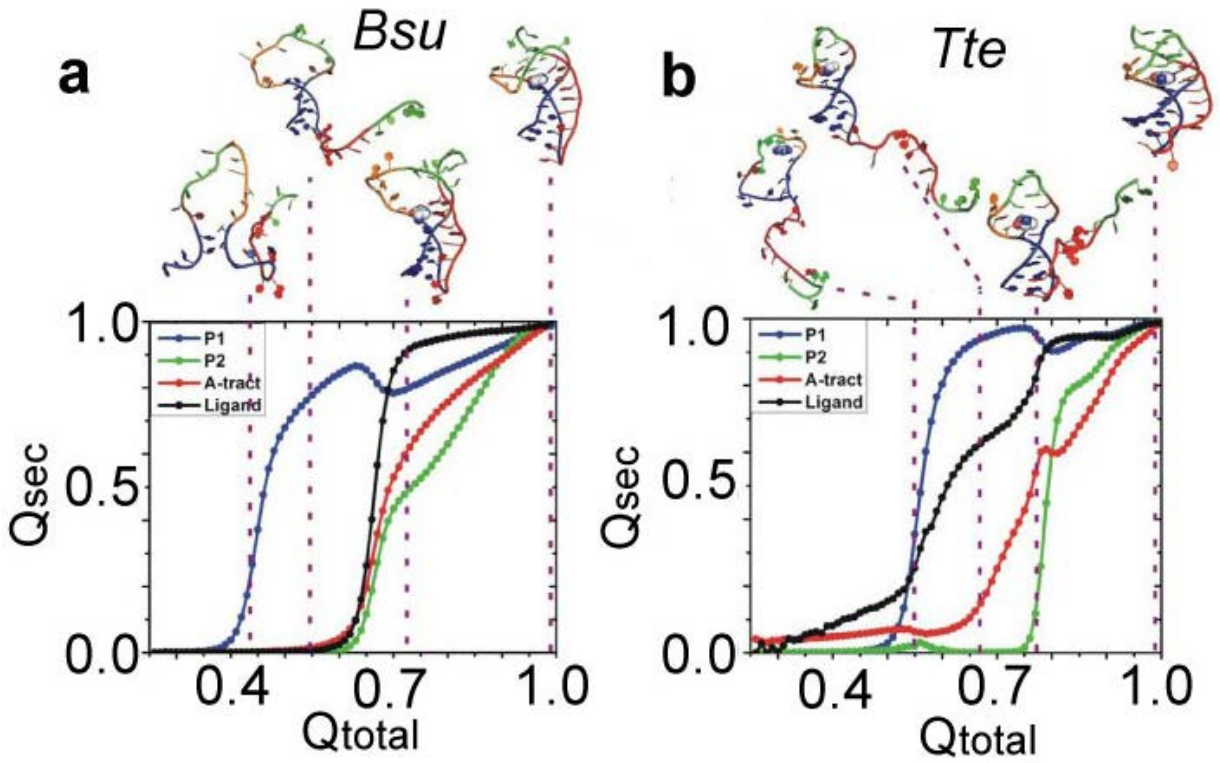

Figure 2.14 Gō-model simulations of single *Bsu* and *Tte* riboswitch aptamers.
(A) Fraction of native contacts for each structural component, Q_{sec} (P1, blue; P2, green; 3' tail, red; ligand, black), averaged over 51 simulations and plotted as a function of the total contacts observed in the native folded structure, Q_{total}. Above characteristic points along the folding pathway one representative conformation is shown. (B) Same as (A), but for *Tte*.

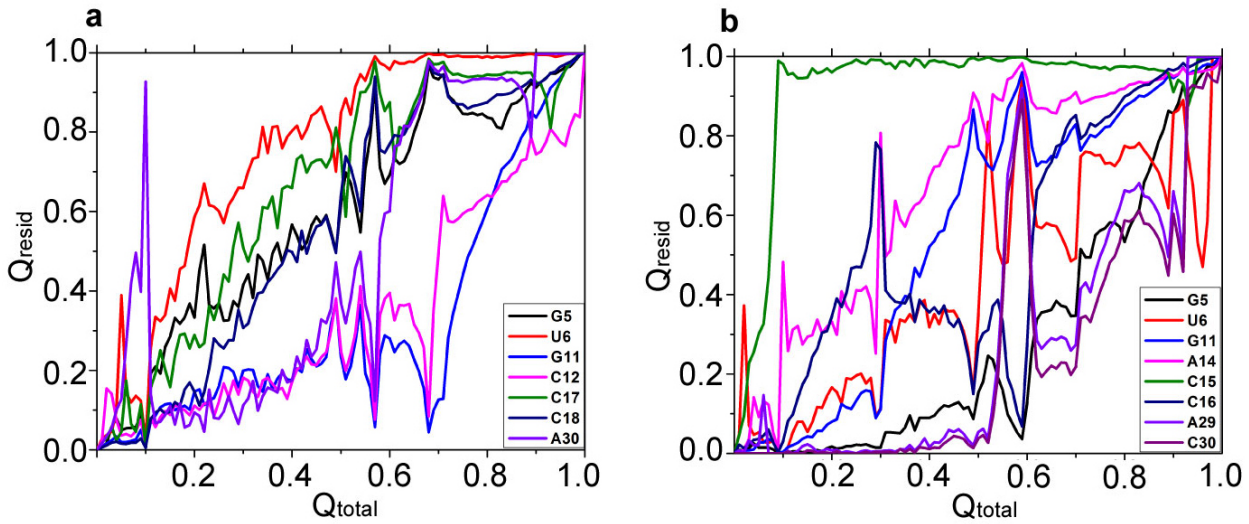

Figure 2.15 Gō model simulations of ligand binding to the *Bsu* (A) and *Tte* (B) riboswitches. The fraction of native contacts, Q_{resid}, formed with preQ$_1$ by each nucleotide, as indicated, is plotted as a function of the fraction of total ligand contacts, Q_{total}. Nucleotides are numbered as in **Figure 2.1C**.

of formation of the remaining contacts, however, significantly differs. In the transcriptional *Bsu* riboswitch, ligand binding, stable docking of the A-tail, and folding of the P2 stem all occur late and almost concomitantly (**Figure 2.14A**), whereas in the translational *Tte* riboswitch, ligand binding to the top of P1 occurs early, trailed by A-tail docking just prior to folding of the P2 stem (**Figure 2.14B**). Furthermore, C15 of the *Tte* riboswitch (**Figure 2.1C**) forms its Watson-Crick base pair with $preQ_1$ very early in the folding pathway, clearly preceding folding of the remaining binding pocket around the ligand (**Figure 2.15B**). By contrast, the corresponding C17 of the *Bsu* riboswitch forms its Watson-Crick base pair with $preQ_1$ concomitantly with folding of the other binding pocket residues (**Figure 2.15A**). These divergent tendencies in ligand-mediated folding of the *Bsu* and *Tte* riboswitches essentially recapitulate the classical mechanisms of conformational selection and induced fit, respectively.

2.3.6 A mutation in the 3'-tail controls the ligand-binding properties of the riboswitches

Our data collectively show that the 3' tail transiently interacts with the P1 stem in the absence of ligand. However, the observed shift in mean FRET of the *Tte* mid-FRET state and the lack of such a shift in the *Bsu* riboswitch, coupled with the differences in the Gō-model-derived folding pathways, suggest that the nature of this pre-folded state differs in subtle but potentially important ways. Our previous NMR studies of the isolated 3' tail of the *Bsu* riboswitch showed that disruption of its contiguous A-tract by an A-to-C mutation resulted in weakened stacking interactions[118]. We therefore propose that the uniform A-tract in the 3' tail of the *Bsu* riboswitch renders it more ordered than the *Tte* riboswitch tail with its C insertion, and thus allows for efficient ligand recognition and binding. This proposal presents a clear, testable hypothesis — that one can modulate the ligand binding properties of the $preQ_1$ riboswitch by promoting either

relative order or disorder of the 3' tail. To test this hypothesis we introduced opposite A27C and C25A mutations into the equivalent positions of the *Bsu* and *Tte* riboswitches, respectively. Indeed, disruption of the 3' tail stacking interactions through the A27C mutation in the *Bsu* riboswitch resulted in a marked decrease in ligand binding affinity by two orders of magnitude relative to wild-type ($K_{1/2}$ = 11 μM versus 134 nM, **Figures 2.16A, 2.11A** and **2.17A**), despite the distal nature of the mutation (**Figure 2.1C**). In addition, we observed a notable decrease in the mean FRET value of the pre-folded state in the absence of ligand (0.74 to 0.65, **Figure 2.16B**), which only slightly varies across a broad ligand concentration (**Figure 2.16C**). This decrease is most likely due to a (partial) loss of A-rich tail rigidity as demonstrated by NMR[118], which our TOPRNA simulations predict will lead to an increase in the inter-fluorophore distance monitored by smFRET (purple line in **Figure 2.6A**). By contrast, enhancement of the 3' tail stacking interactions through the C25A mutation in the *Tte* riboswitch results in no change in binding affinity relative to wild-type ($K_{1/2}$ = 64 nM versus 69 nM, **Figures 2.16A, 2.11A** and **2.17B**), accompanied by an increase in the mean FRET value of the pre-folded state (0.75 compared to 0.70 of the wild-type in the absence of ligand, **Figure 2.16B**), which only slightly varies across $preQ_1$ concentrations (**Figure 2.16C**). Consistent with our hypothesis, these results indicate that a distal point mutation in the A-rich 3' tail of the $preQ_1$ riboswitches can change the nature of pre-folded ligand-free conformational ensemble in subtle but powerful ways, affecting not only the ligand binding affinity, but also the overall compactness of the pre-folded state.

2.4 Discussion

Despite acting through completely different modes of gene regulation, the *Bsu* and *Tte* $preQ_1$ riboswitches are strikingly similar in their aptamer sequence and structure (**Figures 2.1B, C**). By

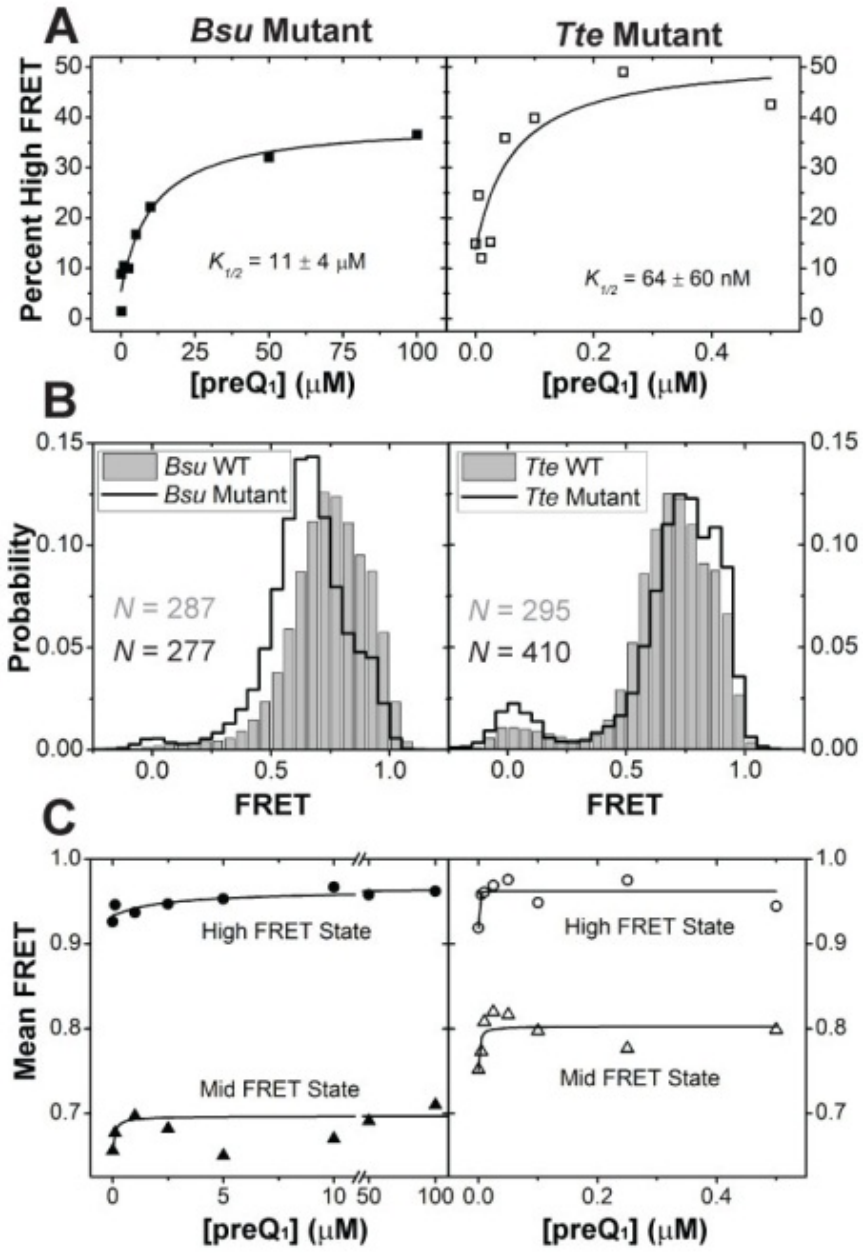

Figure 2.16 smFRET characterization of riboswitch mutants.
(A) The Gaussian distributions from **Figure 2.17** were quantified and the fraction high-FRET state was plotted as a function of $preQ_1$ concentration for the *Bsu* (closed symbols) and *Tte* (open symbols) riboswitches. (B) smFRET histograms of the *Bsu* and *Tte* riboswitches in wild-type (WT, gray bars) and mutant (black line) forms, in the absence of $preQ_1$. (C) The centers of the mid-FRET (triangles) and high-FRET (circles) states in **Figure 2.17** were plotted as a function of ligand concentration for both the *Bsu* (closed symbols) and *Tte* (open symbols) riboswitches

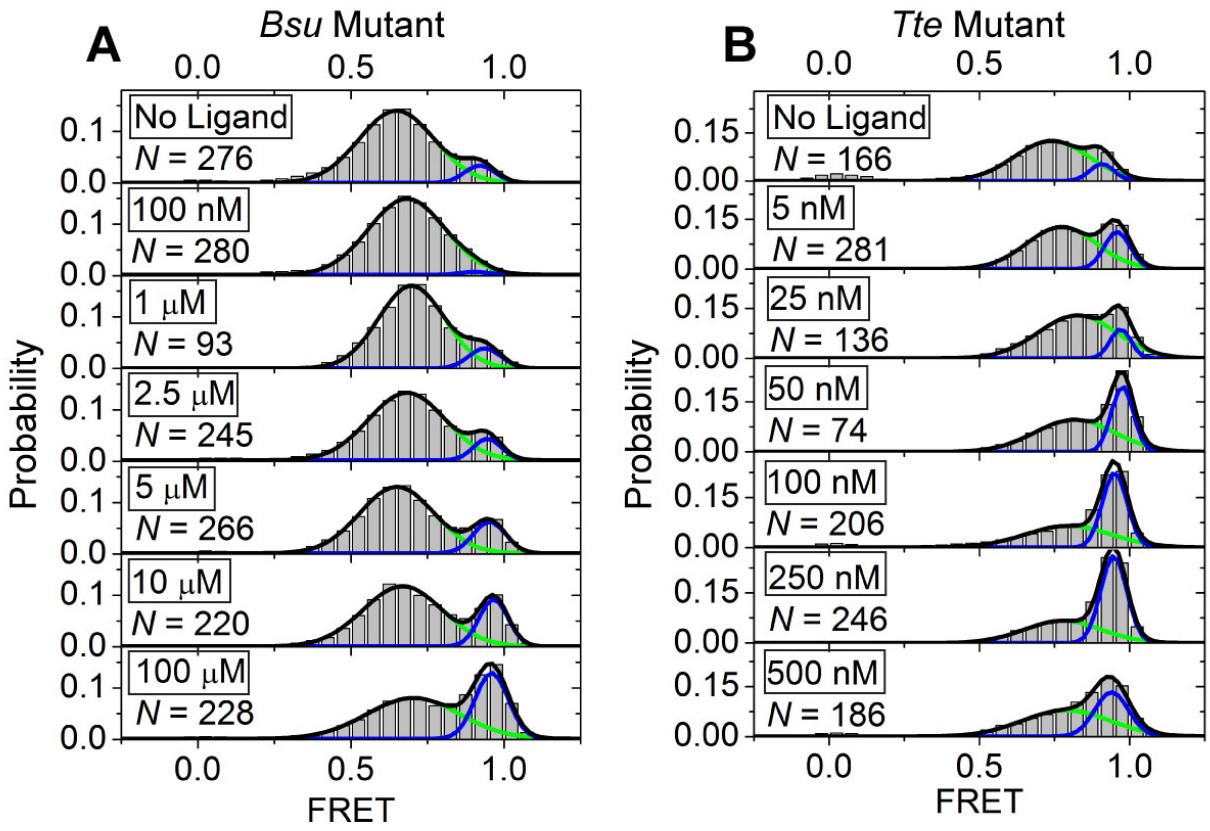

Figure 2.17 smFRET characterization of preQ$_1$ binding to the *Bsu* (A) and *Tte* (B) riboswitch mutants.

Ligand concentrations are indicated in the boxes. *N*, number of molecules sampled per condition. Green and blue lines indicate individual Gaussian fits of the mid-FRET and high-FRET states, respectively. Black lines indicate cumulative fits.

contrast, diverse ensemble-averaging techniques such as NMR spectroscopy, X-ray crystallography and SAXS initially led to distinct structural models for the ligand-free states of the riboswitch, depicting them as either a hairpin with a non-interacting and dynamic 3' tail (*Bsu*) or as a loose pseudoknot (*Tte*), and left the question of how ligand binding leads to the compact ligand-bound state largely unanswered[42,109]. Here, we have used smFRET and both coarse-grained and atomistic (Gō-model) simulations to carry out a detailed side-by-side comparison of the dynamics and ligand-mediated folding of the two riboswitches at the single molecule level. We show that under near-physiological conditions both the ligand-free riboswitches similarly adopt two distinct FRET states – a major, already pre-folded state that, in the case of the *Tte* riboswitch, directly senses ligand, and a minor folded-like state that becomes more populated with increasing ligand concentration. Transitions between the pre-folded and folded states are observed even in the absence of ligand, in the *Bsu* riboswitch particularly at 33 ms time resolution. Our coarse-grained simulations suggest that the pre-folded state is an ensemble of conformations with varying degrees of interaction between the (partially) stacked A-rich 3' tail and the P1-L1 stem-loop. Despite their similarities, smFRET and all-atom Gō-model simulations show that the two riboswitches follow, on average, distinct ligand-mediated folding mechanisms, wherein the *Bsu* riboswitch tends to fold more by conformational selection and the *Tte* riboswitch has a relatively greater tendency to fold by induced fit. We also note that both mechanisms appear to be utilized by both riboswitches, just to differing extents, and that it is difficult to speculate whether one is more advantageous than the other with respect to their overall modes of genetic regulation. Our results support the unifying model in **Figure 2.18**, where the differences between the structurally similar transcriptional and translational $preQ_1$ riboswitches reduce to subtle, yet significant, relative shifts in their conformational sampling

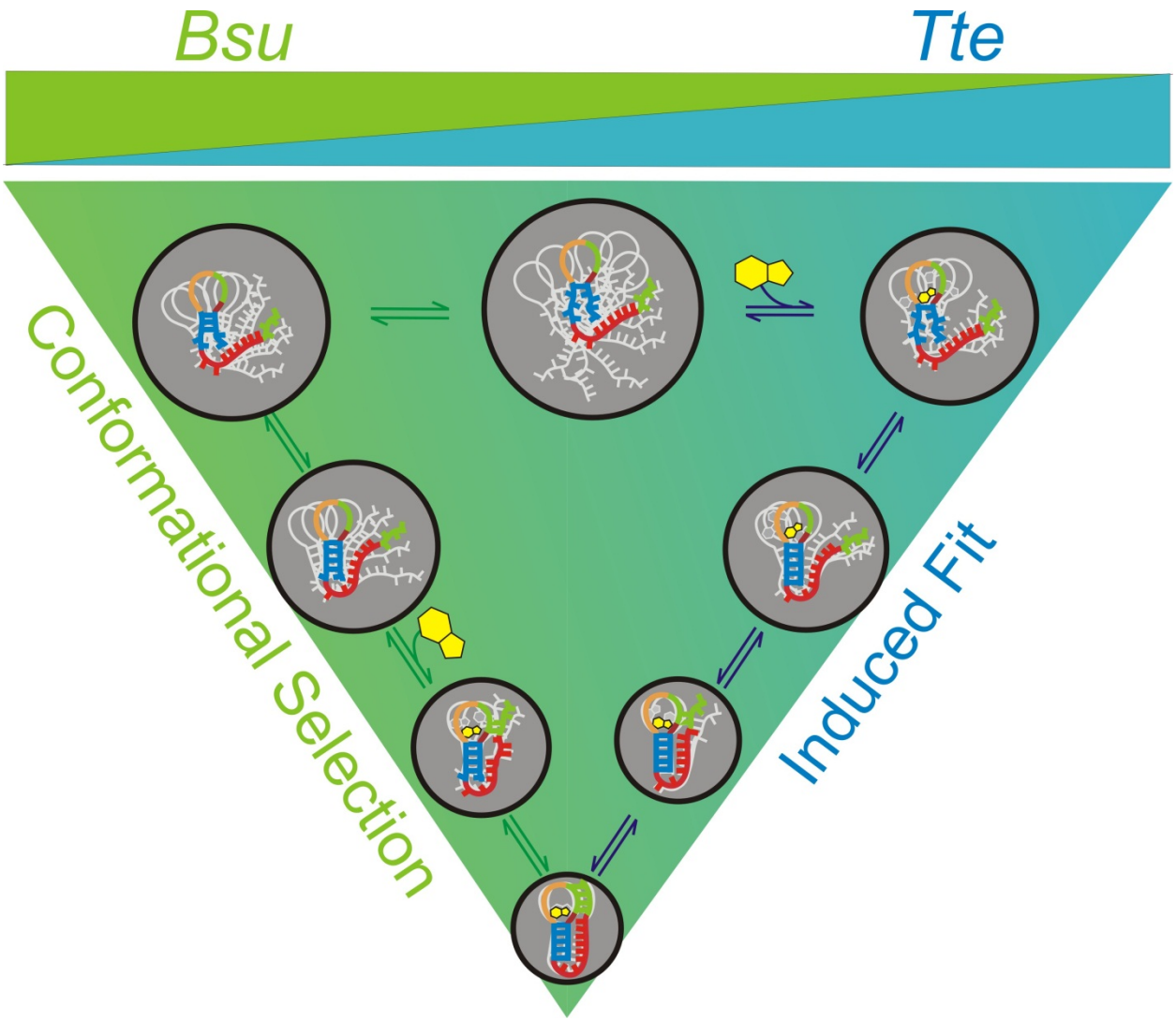

Figure 2.18 Parsimonious folding model of the *Bsu* and *Tte* preQ$_1$ riboswitch aptamers. A combination of smFRET and computational simulations support a model in which the preQ$_1$ ligand binds late and concomitantly with the docking of the 3' tail and formation of the P2 stem in the *Bsu* riboswitch, signifying conformational selection (represented in green). By contrast, early binding of the preQ$_1$ ligand to a partially unfolded conformation induces folding into the bound structure of the *Tte* riboswitch, consistent with an induced fit model (represented in blue). The size of the white circle and the number of gray outlines describe the extent of conformational dynamics of each state. Both mechanisms are not mutually exclusive and it is plausible that a combination of both induced fit and conformational selection mechanisms are at work in both riboswitches.

upon ligand binding. This model finds further support as we show that remote mutations in the 3' A-rich tail, which we have previously shown to diminish or enhance stacking of the A's[118], have significant effects on the ligand binding properties of the riboswitches by shifting their conformational sampling as predicted by the model.

Previous studies of the transcriptional *Bsu* preQ$_1$ riboswitch[109] and the closely related *Fnu* (*Fusobacterium nucleatum*) riboswitch[112,113] led to models wherein ligand binding and RNA folding largely occur concurrently, essentially due to a failure to observe interactions between the 3' tail and the P1-L1 stem-loop by NMR in the absence of ligand. This apparent discrepancy turns out to be due to competing RNA dimerization facilitated by kissing-loop interactions (between U$_9$AGCUA$_{14}$ in L1 loop of the *Bsu* riboswitch) at the high RNA concentrations used for NMR, as recent NMR studies attest[43,118], as well as differences in buffer conditions as we show by decreasing the Mg^{2+} concentration and pH (**Figures 2.7, 2.8** and **2.9**). The buffer dependence can be rationalized since 1 mM Mg^{2+} in smFRET experiments is in very large stoichiometric excess over the RNA used (10-50 pM during slide binding, which is further lowered as the excess of RNA not bound to the slide is washed away). By comparison, the close-to-millimolar concentration of a 36-nt RNA during standard NMR experiments can render even 10 mM Mg^{2+} sub-stoichiometric relative to the backbone phosphates that need to be charge neutralized to stabilize an RNA's tertiary structure. In addition, lowering the pH from 7.5 (our near-physiological smFRET buffer) to 6.4 (the standard NMR buffer) likely results in a small population of protonated nucleobases, which is expected to destabilize hydrogen bonding and stacking. Our results are supported by a recent NMR study on the *Fnu* riboswitch that showed that adding Mg^{2+} to the buffer and avoiding dimerization results in a pre-organized pseudoknot-like conformation in the absence of ligand[43]. Our observation of a ligand-free 'pre-folded' state

in the *Bsu* and *Tte* riboswitch aptamers are also supported by recent computational simulations that showed that the 3'- tail interactions with P1-L1 stem-loop are stable in both riboswitches at 300 K even in the absence of ligand[108,111,129].

Our observations are also generally consistent with prior X-ray crystallographic studies of the translational *Tte* riboswitch, which showed only minor differences between the ligand-free (pre-folded) and ligand-bound (folded) states. Complementary SAXS experiments further showed that the ligand-free state additionally opens up in solution compared to the crystal lattice[42]. This observation is in accordance with our data that indicate a shift in the mid-FRET state to a more compact structure when increasing the ligand concentration. Here, we have added a dynamic picture for this riboswitch by showing that ligand binding typically occurs early in the RNA folding pathway, albeit with transient interactions of the 3' tail with the P1-L1 stem-loop already in place, followed by closure of the binding pocket around the ligand through induced fit (**Figure 2.18**). By contrast, conformational selection (capture) was proposed for the translational SAM-II riboswitch that also folds into an H-type pseudoknot[41]. This conclusion was primarily derived from the slightly (<2-fold) faster ensemble-averaged relaxation kinetics upon ligand addition of 2-aminopurine stacking when incorporated into the P2 stem as compared to the P1 stem or 3' tail. Yet, since ligand binding was only indirectly monitored, it is difficult to establish the exact sequence of ligand binding and P2 stem formation as a way to unambiguously distinguish between conformational selection and induced fit. In fact, the authors correctly speculated that multiple pathways may coexist[41], a notion that we here have expanded upon by showing that the structurally related *Bsu* and *Tte* riboswitches are, on average, opposing representatives on a sliding scale of only relative tendencies to fold via ligand-mediated conformational selection and induced fit, respectively (**Figure 2.18**).

Our model in **Figure 2.18** is also consistent with the expectation that both transcriptional and translational riboswitches can be poised to bind ligand. Both the *Bsu* and *Tte* riboswitches adopt a prominent, dynamic pre-folded conformation in the absence of ligand, with the 3' tail already pre-positioned, through transient interactions with the P1-L1 stem-loop, close to its eventual placement in the ligand-bound folded state. This poised state is aided by stacking in the A-rich 3' tail, as evident when we mutate the central A27 to C (*Bsu*), which diminishes the nucleotide's stacking interactions while keeping its sugar edge intact for hydrogen bonding. Consequently, this mutation dramatically lowers the ligand binding affinity (by ~130-fold) (**Figure 2.16A**). Yet, we find that similar predisposition of the 3' tail for ligand binding still allows for either conformational selection (*Bsu*) or induced fit (*Tte*). The main distinction between these folding mechanisms lies in the relative longevity of the complex between ligand and the pre-folded conformation with a disordered binding pocket (**Figure 2.18**), making the two mechanisms notoriously difficult to distinguish[130]. We observe here that the *Tte* riboswitch is characterized by comparably slow, and thus more easily detectable, conformational exchange between the pre-folded and folded states as well as a shift of the pre-folded conformational ensemble towards more compact, higher-FRET conformers upon ligand encounter. It is tempting to speculate that these experimental distinctions from the *Bsu* riboswitch may provide a general signature for a riboswitch favoring ligand-induced fit. Other signatures are the early ligand binding compared to binding pocket folding, as observed in our Gō-model simulations, and the relative fractional flux through the pathway, which requires a full assessment of all pathway rates that has not yet been experimentally accomplished[130]. In general, a combination of both mechanisms is often utilized in complex bimolecular binding processes and the major mechanism followed depends on many factors including ligand and RNA concentrations[130]. For

instance, when assessing the kinetic parameters required to assign one mechanism over another, the relative tendency for a molecule to select one pathway over the other can be affected by ligand concentration. More specifically, at high ligand concentrations, the flux throughout the pathway would shift towards induced fit. While the cellular concentration of $preQ_1$ has not been presented in the literature, a reasonable upper limit can be provided by the K_m of the *E. coli* transglycosylase enzyme (390 nM)[131], which is responsible for incorporating $preQ_1$ into tRNA, which is slightly higher than the $K_{1/2}$ values measured for both riboswitches. Furthermore, even when a ligand primarily selects a specific conformation with a preformed binding pocket, as appears to be the case for the *Bsu* riboswitch, the binding pocket still has to close to entirely envelop the ligand. Caution is therefore warranted when assigning one or the other mechanism to a specific riboswitch.

Interestingly, the *Tte* riboswitch has a shorter L2 loop and lacks two nucleotides that are unresolved in the *Bsu* crystal structure[110], suggesting they are flexible. The pre-folded state ensemble of the *Tte* riboswitch thus may require less sampling and be able to fold more efficiently around a transiently bound ligand than the *Bsu* riboswitch to achieve induced fit. Alternatively, this feature may be related to either its function as a translational riboswitch or its origin from a thermophilic bacterium. Conversely, the comparably faster transitions of the *Bsu* riboswitch between the pre-folded and folded states may help it rapidly bind ligand by conformational selection within the short time window (<2 s) before the transcribing RNA polymerase clears the downstream expression platform. Single molecule force experiments on the structurally similar *pbuE* and *addA* adenine riboswitch aptamers showed that they sample similar conformational ensembles but differ subtly in their folding pathways and dynamics that relate to their distinct functions as transcriptional and translational riboswitches, respectively[57].

Given that our work also shows similarities in the conformational distributions of the *Bsu* and *Tte* riboswitches with subtle differences in their dynamics and folding, this appears to be a common feature of structurally similar but functionally different riboswitches. More such comparative studies are needed to understand how riboswitch structure and dynamics are fine-tuned by nature to function through different gene regulation mechanisms. Given the great impact that we find environmental (buffer) conditions to have, future studies should ideally test such hypotheses under the growth conditions of the bacterium from which a given riboswitch is derived.

In summary, we have utilized smFRET, NMR and computational simulations to characterize the folding behaviors of two structurally similar but functionally distinct transcriptional and translational riboswitch aptamers. Our work presents direct evidence for a pre-folded intermediate in the folding pathway of both riboswitches, poised to bind ligand. Furthermore, our data yield evidence that the pre-folded state of the translational riboswitch with only a partially formed binding site directly senses and binds ligand. Our work thus reveals that even small, structurally similar RNAs can adopt distinguishable folding mechanisms, consistent with recent observations for highly homologous proteins[132]. We also demonstrate the fine-tuned conformational sampling of these riboswitches, as mutation of a single nucleotide distal from the ligand binding pocket has dramatic effects on ligand binding and, therefore, gene regulation. This may be exploited in the future engineering of riboswitches for gene regulatory functions.

2.5 Acknowledgements

We thank Dr. George Garcia for a generous gift of $preQ_1$, Dr. David Rueda for the use of his prism-based TIRF microscopy setup for the 33 ms data acquisition and Kyle Vrtis for assistance

using it, Dmitri Ermolenko for advice on heat annealing of the $preQ_1$ riboswitches, and Mario Blanco for assistance in the HMM analysis of smFRET data.

CHAPTER 3

Establishment of a Single Molecule Footprinting Assay for the Observation of Translational Regulation by an Adenine Riboswitch[2]

3.1 Introduction

The purine family of riboswitches is the largest group of ligand sensing riboswitches currently known in nature. Effectors of this family of riboswitches include adenine, guanine, 2'-deoxyguanosine, cyclic-di-GMP and $preQ_1$. The first purine binding riboswitches were uncovered by Breaker and coworkers in a bioinformatics study comparing conserved sequences upstream of the *xpt-pbuX* operon that encodes for a xanthine phosphoribosyltransferase and a xanthine permase, respectively[133]. This study revealed 32 representative mRNA domains in a variety of Gram-positive and Gram-negative bacteria that conform to a consensus "G-box" domain. Biochemical analyses were conducted on the *Bacillus subtilis xpt* 5' UTR to probe the predicted secondary structure, which consists of three helices, P1, P2 and P3, as well as determining its binding affinity for various closely related purine bases. Of these, guanine displayed the highest binding affinity, leading the authors to correctly identify this aptamer as a guanine-binding riboswitch. Of these 32 representatives, it was noted that three sequences

[2] Arlie J. Rinaldi performed all experiments and data analysis, with the exception of *in vitro* translation of the ADD riboswitch, which was done by Martina Schönnenbeck.

carried a C-U mutation in the conserved P3-P1 junction. It was correctly hypothesized that these delegates could be adenine specific riboswitches.

In the years that followed, myriad studies were conducted on the guanine-sensing *xpt* riboswitch, as well as the adenine-sensing *pbuE* and *add* riboswitches. Unlike the S-adenosylmethionine family of riboswitches, the guanine and adenine riboswitches adopt a similar global architecture, as evidenced by their crystal structures[55]. In the ligand bound state, both riboswitches adopt a three-helix junction resembling a "tuning fork," in which the two arms of the fork, helices P2 and P3, form a coaxial stack on helix P1. The terminal loops, L2 and L3 form a kissing loop interaction mediated by two G-C Watson-Crick base pairs. Significant sequence conservation is present in the three-helix junction, at the ligand-binding site, with the exception of the cytosine or uridine residue, which confers ligand specificity. In fact, a single C74U mutation in the guanine riboswitch efficiently changes its ligand discrimination from guanine to adenine, owing to the overall structural similarity between the two[134].

One of the most extensively studies adenine-binding riboswitches is the translationally acting *add* riboswitch from *Vibrio vulnificus*. The *add* riboswitch is located directly upstream from the *add* gene, which encodes for an adenosine deaminase. It regulates gene expression as a translational activator, where adenine binding is thought to cause release of the Shine-Dalgarno sequence, allowing initiation of translation to occur (**Figure 3.1**). Alternatively, in the absence of adenine, the RNA adopts an alternative conformation in which the P1 stem is destabilized, and a downstream helix forms, occluding the Shine-Dalgarno sequence and the AUG start codon (Figure 3.1)[55]. Single molecule force extension experiments demonstrate that the *add* riboswitch folds in a hierarchal manner in which local secondary structural elements in the P2 and P3 stems fold initially, followed by tertiary interactions between the L2 and L3 loops[57].

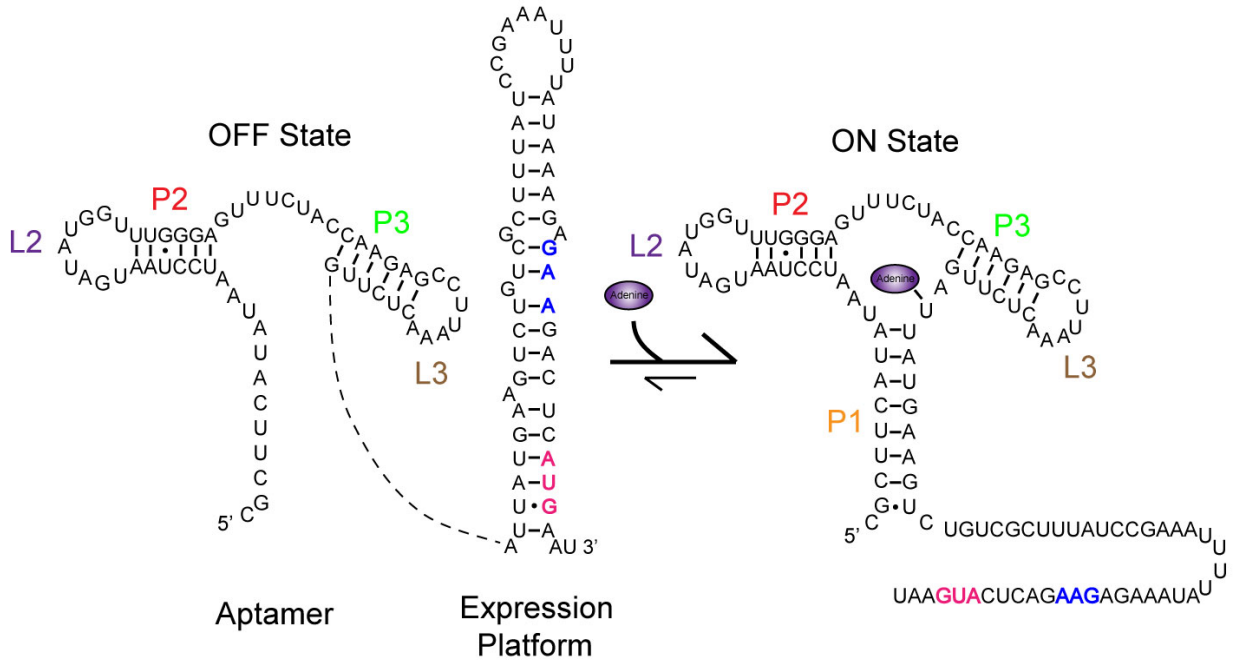

Figure 3.1 *add* riboswitch secondary structure.
The structure of the ligand-bound aptamer is known, however, the orientation of the expression platform and the ligand-free structure has only been predicted by chemical probing and bioinformatics searches. In the ligand-free state, stable helices P2 and P3 form, but the Shine-Dalgarno sequence (blue) and AUG start codon (pink) are engaged in Watson-Crick base pairing, and therefore, sequestered from ribosome binding. Upon ligand binding, however, helix P1 is stabilized, thus exposing the SD sequence and start codon, making them accessible to 30S binding.

Finally, the P1 stem is stabilized by the presence of adenine; however, the P1 helix is much more dynamic in the absence of ligand and the aptamer fluctuates rapidly between an ON and OFF state[57], providing further evidence that translational riboswitches may be able to reversibly bind ligand depending on changing physiological conditions[50,59]. Importantly, this study was extended to include the expression platform containing the ribosomal binding site and the AUG start codon (**Figure 3.1**). These results show that in the presence of adenine, the expression platform is mostly unstructured as the force extension curve population histogram mimics that of the aptamer alone. However, in the absence of adenine, a new population arises that unfolds at low force, matching that of the expression platform alone[57]. These results are significant because the authors were able to speculate towards the time required for equilibration of the ON and OFF states, providing an insight towards the kinetics and thermodynamics of 30S ribosome binding to the expression platform. While these results are compelling and provide a solid framework towards the molecular mechanism of translational regulation by riboswitches, direct 30S subunit binding to riboswitches as a function of ligand concentration has yet to be demonstrated.

While the results established heretofore provide valuable insight into the dynamic conformational changes occurring during the folding of the adenine riboswitch as a function of ligand concentration and how they may affect ribosomal binding, the methods employed do not always mimic physiological conditions and/or yield information regarding subtle heterogeneities within a system. For example, ensemble assays based on, for example, 2-aminopurine fluorescence involve labeling the RNA with a non-natural nitrogenous base, and the bulk nature of the assay averages out possibly important idiosyncrasies within the system, rendering them invisible. Also, force extension by optical trapping involves rigorously pulling on the 5’ and 3’ ends of an RNA molecule until all secondary and tertiary structural elements are completely

unfolded, a process that does not mimic RNA folding *in vivo*, since RNAs are folding cotranscriptionally in a 5' to 3' direction[135], often with the help of folding chaperones[136-139]. For these reasons, novel techniques are needed to probe the sensitive dynamics of RNA and riboswitch folding under physiological conditions without perturbing their inherent conformations.

To this end, we have developed a single molecule footprinting technique derived from the high-resolution imaging technique called points accumulation for imaging in nanoscale topography (PAINT) method[71]. PAINT utilizes small fluorescently-labeled nucleic acid probes that hybridize to specific sites within a target molecule, which is also fluorescently labeled for localization purposes. Under total internal reflection fluorescence microscopy (TIRFM) conditions[75], the target molecule is hybridized to the quartz slide surface, most commonly through a biotin-streptavidin interaction. Next, the fluorescently labeled probe is flowed over the surface of the slide and the evanescent field is illuminated with the excitation wavelengths of the target and probe fluorophore. The probe is designed with a low melting temperature such that its residence time on the target molecule is short. Therefore, a diffraction-limited colocalization signal appears only upon probe binding to the target molecule since molecules within the evanescent field are illuminated while those remaining in solution contribute to background signal. Due to the low thermodynamic stability of the interaction of the probe with its target, it will dissociate quickly, depending on the time resolution of the observation window, thus resulting in a loss of signal.

This technique was originally utilized as a high-resolution imaging technique that has been applied to image large unilamellar vesicles[71] and DNA origami[140,141]. We have adapted this technique to probe the dynamic changes in the accessibility of the Shine-Dalgarno sequence of

the adenine riboswitch as a function of ligand binding. To mimic 30S subunit binding, we observed the transient binding of a Cy5-labeled DNA probe with the exact sequence of the 3' end of the 16S rRNA of *Vibrio vulnificus*, mimicking the anti-Shine Dalgarno of the ribosome. 1-ethyl-3-(dimethylaminopropyl)carboiimide (EDC) mediated coupling was used to covalently link a 5' biotin to the *add* riboswitch, and periodate oxidation followed by Cy3-hydrazide labeling was used to label the 3' end. Using the single molecule footprinting assay, which combines the structural information of an ensemble chemical probing assay with the molecule to molecule sensitivity of a single molecule technique, we have shown that addition of adenine to the doubly labeled *add* riboswitch increases the binding rate of the probe. Our results demonstrate that adenine modulates the accessibility of the Shine-Dalgarno sequence of a translational riboswitch.

3.2 Materials and Methods

3.2.1 5' Biotin labeling

Lyophilized gel purified *add* riboswitch RNA was resuspended to 1 mg/mL concentration in sterile water. Since EDC coupling requires a monophosphate at the 5' end, 530 nmol (~ 20 µg) of RNA was dephosphorylated in 1X Antarctic phosphatase buffer (New England Biolabs) using 25 U of Antarctic phosphatase (New England Biolabs) in a total reaction volume of 100 µL. The reaction was incubated at 37 °C for 30 minutes. After incubation, the enzyme was heat inactivated at 65 °C for 5 minutes. Monophosphate addition was achieved using 20 U of T4 Polynucleotide Kinase (New England Biolabs) in 1X T4 DNA Ligase buffer (New England Biolabs) in a total reaction volume of 150 µL. The reaction was incubated at 37 °C for 30 minutes. Enzyme was removed from the mixture by extraction with a 25:24:1

phenol:chloroform:isoamyl alcohol mix. RNA was precipitated by ethanol precipitation and resuspended in 75 µL of 10 mM K-phosphate, pH 7.0, 150 mM NaCl, 10 mM EDTA. To this mix, 50 mg EDC, 10 µL of 44 mg/mL biotin hydrazide slurry in water and 315 µL of 0.1 M imidazole, pH 6.0 were added. The mixture was incubated at room temperature for 12-16 hours overnight. RNA was precipitated by ethanol precipitation.

3.2.2 3' end Cy3 labeling via periodate oxidation

Pelleted RNA from the previous step was resuspended in 100 mM sodium acetate, pH 5.2 and 2.5 mM sodium periodate in a total volume of 400 µL. The reaction mix was on ice in the dark for one hour. RNA was ethanol precipitated and resuspended in 100 mM sodium acetate, pH 5.2. To this mix, 40 µL of 10 mM Cy5 hydrazide (GE Life Sciences) in DMSO was added to a total volume of 400 µL. The reaction mix was tumbled in the dark at room temperature for four hours. Unreacted Cy5 was removed using a Nap5 gel filtration column (Illustra) equilibrated in deionized water. Fractions containing RNA were ethanol precipitated and resuspended in TE buffer, pH 7.0.

3.2.3 Single molecule footprinting

We assembled a microfluidic channel on a quartz slide with an inlet and outlet port and coated it with biotinylated-BSA, followed by streptavidin, as previously described[75,86]. 50 fmol of doubly labeled RNA was heat annealed at 70 °C for 2 minutes in 50 mM Tris-HCl, pH 7.5, 0.1 M KCl, 2 mM $MgCl_2$, +/- adenine (Sigma) and allowed to cool to room temperature for 20 minutes. 100 µL of heat annealed RNA was flowed over the slide and allowed to equilibrate for 5 minutes. Excess RNA was washed off the slide with buffer +/- adenine. An oxygen scavenging system consisting of 5 mM protocatechuic acid and 50 nM protocatechuate-3,4-dioxygenase, +/- adenine, to slow photobleaching, 2 mM Trolox to reduce photoblinking[91] and 50 nM Cy5-probe

was flowed over the slide and allowed to equilibrate for 5 minutes. Probe sequence is as follows: 5'-GGA TCA CCT CCT TT-Cy3 3'. Both Cy5 and Cy3 dyes were directly excited simultaneously using a 638 nm red diode laser and 532 nm green laser, respectively. Emission from both fluorophores was simultaneously recorded using an intensified CCD camera (I-Pentamax, Princeton Instruments) at 100 ms time resolution using Micro-Manager software. Fluorescence traces were extracted from the raw movie files using IDL (Research Systems) and analyzed using Matlab (The Math Works) scripts. Genuine traces exhibiting binding were manually selected using the following criteria: a single photobleaching step of the Cy5 signal, Cy5 fluorescence intensity of >200 intensity units and at least two binding events per trajectory with a signal to noise ratio of at least 3:1. A lower signal to noise ratio was utilized for these experiments than traditional smFRET experiments due to the higher than normal background noise in the Cy3 channel from the high concentration of probe in solution. Suitable traces were compiled and a nonlinear filter was applied to reduce the probability of false transitions in the data processing. Hidden Markov Modeling analysis was performed on the donor intensity using the segmental k-means algorithm in the QuB software suite as described[93]. A two-state model was used with an unbound and bound state to idealize the data. Transition density plots were constructed to extract the dwell times in the bound and unbound states, as described[93]. Dwell times were fit to a single exponential in OriginLab 8.1 from which on and off rates were calculated.

3.3 Results

3.3.1 Development of a doubly labeled *add* riboswitch construct

In order to probe the accessibility of the Shine-Dalgarno sequence of the *add* riboswitch as a function of adenine concentration using a fluorescently labeled anti-SD probe, we decided to

utilize total internal reflection fluorescence microscopy (TIRFM). TIRFM creates an evanescent fluorescent field that penetrates only about ~100 nm into the slide surface[75]. For this reason, only those molecules immobilized on the slide will fluoresce, greatly reducing background fluorescence, an important feature for these experiments that include high concentrations of fluorescently labeled nucleic acid probes. To render the *in vitro* transcribed *add* riboswitch viable for TIRFM measurements, we 3' end labeled the RNA with Cy5 using a periodate oxidation reaction (see **Materials and Methods**). This reaction oxidizes vicinal diols at the 3' end of RNA to aldehydes, which can then be conjugated to Cy5 hydrazide. Following the 3' end labeling step, the 5' end was labeled with biotin using EDC coupling in the presence of imidazole and biotin hydrazide. This doubly labeled riboswitch construct is then immobilized to the quartz slide via the 5' biotin modification, and localized via its Cy5 label on the 3' end.

To determine whether labeled *add* molecules were able to bind to the slide and be visualized by direct Cy5 excitation, 250 pM RNA was flowed over the slide and visualized using the Micro-Manager software package. Indeed, when riboswitch molecules were added to the slide, several hundred spots were observed in the Cy5 channel upon illumination, with very few spots observed in the Cy3 channel (**Figure 3.2**). The spots that were observed in the Cy3 channel can be attributed to impurities within the biotinylated BSA or streptavidin solutions that do not contribute to data analysis as Cy3 and Cy5 signals are colocalized during trace selection (see **Materials and Methods**). The presence of Cy5 signal in the presence of RNA is well above the background with buffer alone, indicating that the biotin and Cy5 labeling scheme was successful, and rendering the RNA suitable for single molecule analysis.

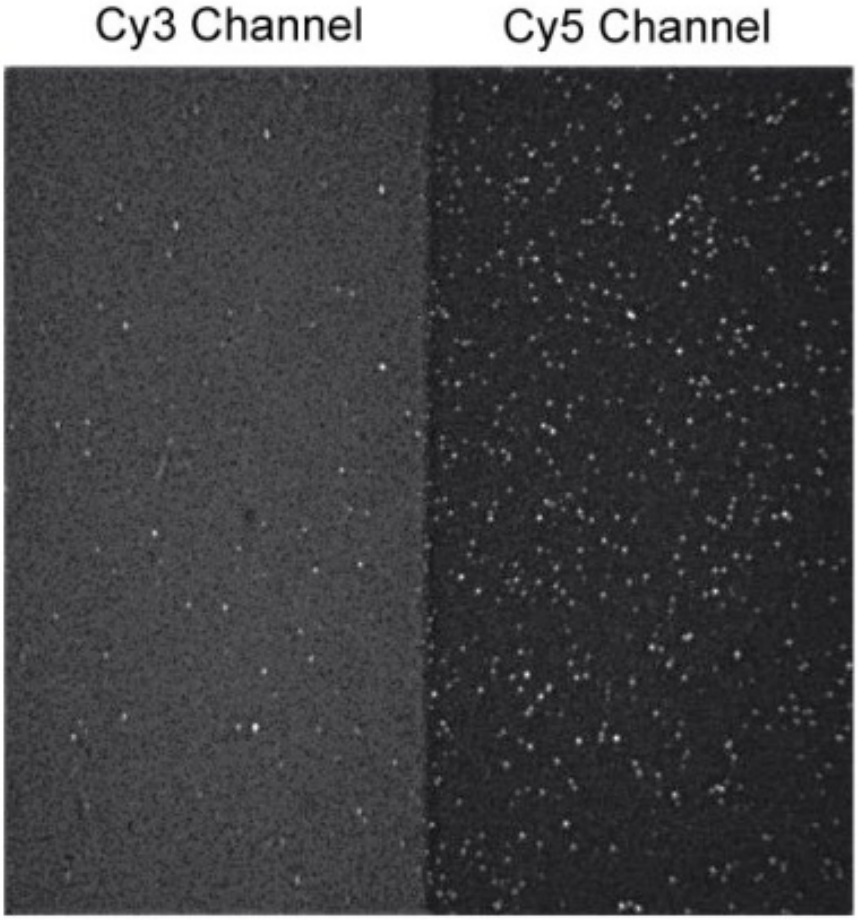

Figure 3.2 Representative field of view.
Add riboswitch RNA that has been modified on the 5' and 3' ends with biotin and Cy5, respectively are observed as single molecule spots in the Cy5 channel upon illumination by a 640 nm laser.

3.3.2 Single molecule footprinting of the *add* riboswitch

To exploit single molecule footprinting as a means to probe the accessibility of the SD sequence of the *add* riboswitch, we exploited the experimental scheme shown in **Figure 3.3**. To begin,we created a biotin streptavidin coating on a quartz slide, as previously described[75,86], to which heat annealed RNA in the presence or absence of adenine was immobilized. After a short equilibration period, Cy3 labeled probe was flowed over the molecules. Since the probe was designed to have a low melting temperature, the residence time of the probe on the RNA should ideally be <1 s. For this reason, transient binding events of the probe should not affect the inherent global fold of the RNA, and should not compete with more thermodynamically stable conformations. Additionally, since the binding rate of the probe with the RNA target depends linearly on the concentration of the probe, the binding rate can be controlled and can be interpreted as a measure of accessibility of the target sequence. Therefore, this method will provide a non-invasive method to observe real-time conformational changes of single RNA molecules without the averaging effects of traditional bulk footprinting assays.

To determine whether we could visualize binding of the anti-SD probe to the RNA, we began under conditions in which the SD sequence of the riboswitch should be completely exposed. The affinity of the complete *add* aptamer and expression platform is not known; however, the K_d of a 2-aminopurine derivative was determined, and found to be 2.3 μM[50]. Therefore, to achieve complete accessibility of the SD sequence in the wild-type riboswitch, the RNA was folded in the presence of 200 μM adenine. Indeed, hundreds of transient binding events of the Cy3-labeled probe on the RNA were collectively detected at our 100 ms time resolution (**Figure 3.4**). All observed binding events were collected and compiled, and the dwell

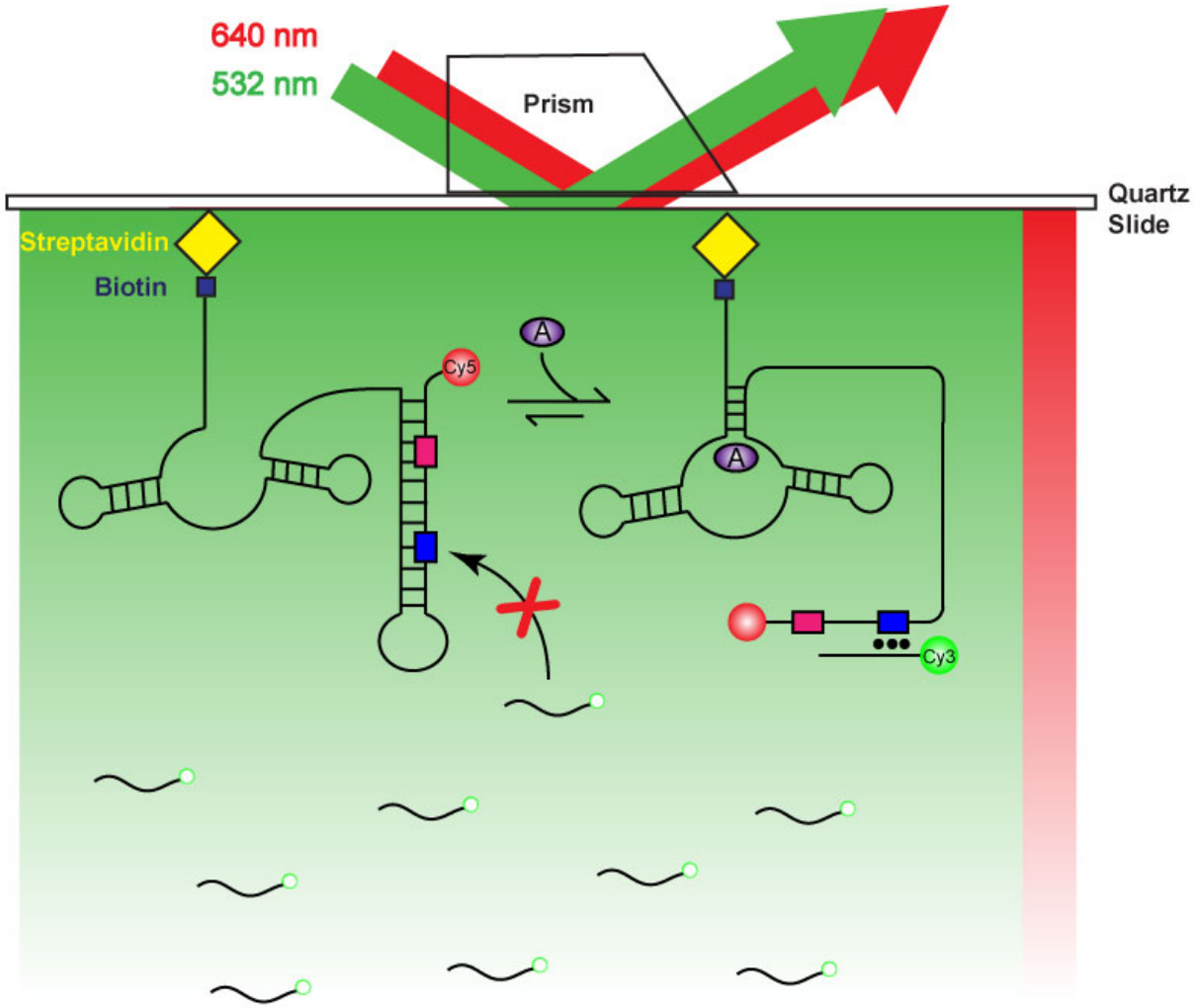

Figure 3.3 Single molecule footprinting experiment for the *add* riboswitch.
Cy5 labeled *add* RNA is immobilized on the slide via a biotin streptavidin interaction. In the absence of adenine, the SD sequence (blue box) and AUG start codon (pink box) are occluded by Watson-Crick base pairing. When adenine is present, however, the P1 stem is stabilized and the SD sequence and AUG start codon, are released, allowing for probe to bind. Binding events are determined by colocalizing the Cy3 and Cy5 signals.

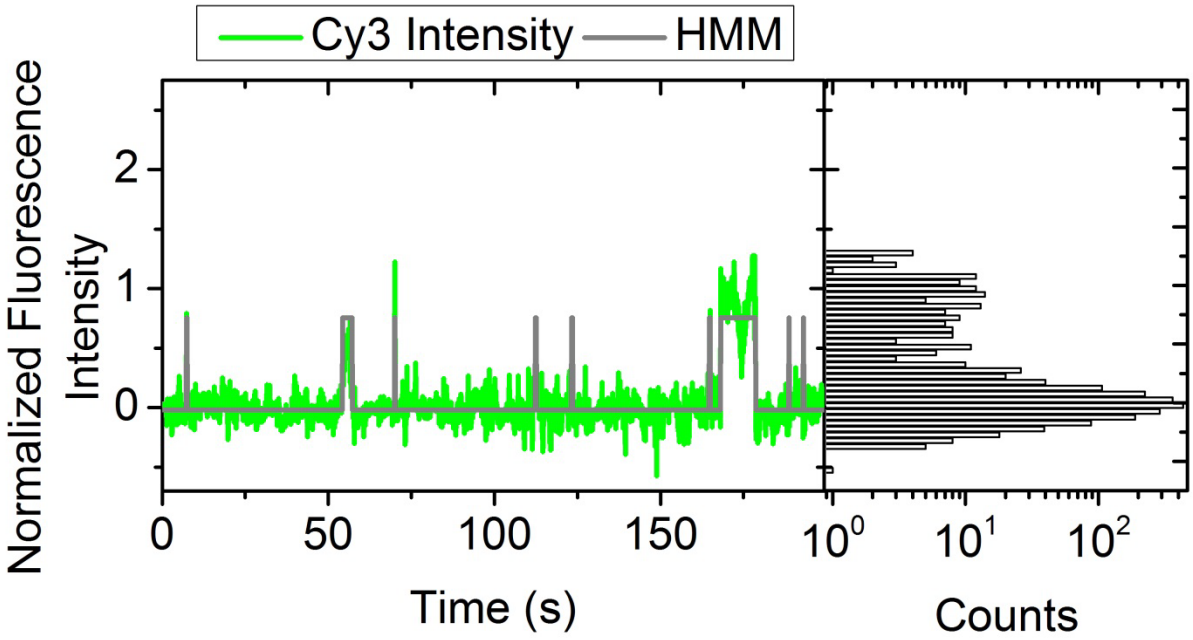

Figure 3.4 Representative binding trajectory.
Cy3 intensity is shown as normalized to the entire data set. Idealized Hidden Markov Model (HMM) is shown in gray. Cy3 intensity is plotted as a histogram, which displays a clear shoulder.

times of the probe in the bound and unbound state were fit with a single-exponential association curve (**Figure 3.5**). By taking the inverse of the time constant calculated by the equation, rate constants for the association and dissociation of the probe onto and from the RNA can be calculated. Under conditions in which the riboswitch is in a constitutive ON state (**Figure 3.1**), and therefore, the SD sequence is completely exposed, the association rate constant (k_{on}) is $1.16x10^{6}$ $M^{-1}s^{-1}$ and the rate of dissociation (k_{off}) is 1.66 s^{-1}.

3.3.3 Effect of adenine on SD sequence accessibility

After establishing the feasibility of the single molecule footprinting assay, we decided to test the effect of adenine on the rate constants for association and dissociation of the probe to the *add* riboswitch RNA. We performed the exact experiment outlined above with successively decreasing adenine concentrations and analyzed each data set. As the concentration of adenine decreases, one would expect the adenine riboswitch to progress towards the OFF state, where the SD sequence becomes occluded through Watson-Crick base-pairing in the expression platform (**Figure 3.1**). As the adenine concentration was sequentially decreased from saturating adenine (200 µM) to 0 µM, the k_{on} indeed decreased from $1.2x10^{6}$ $M^{-1}s^{-1}$ to $8.4x10^{5}$ $M^{-1}s^{-1}$ (**Figure 3.6**, right panel). The slower binding rate constant in the absence of adenine indicates more time between binding events, suggesting the SD sequence becomes less accessible to probe binding. Suprisingly, the rate constant for dissociation of the probe, k_{off}, shows a slight increase with adenine concentration (**Figure 3.6**, left panel). It is difficult to speculate as to the molecular reasoning behind this trend, and further investigation is needed to expand on these preliminary results.

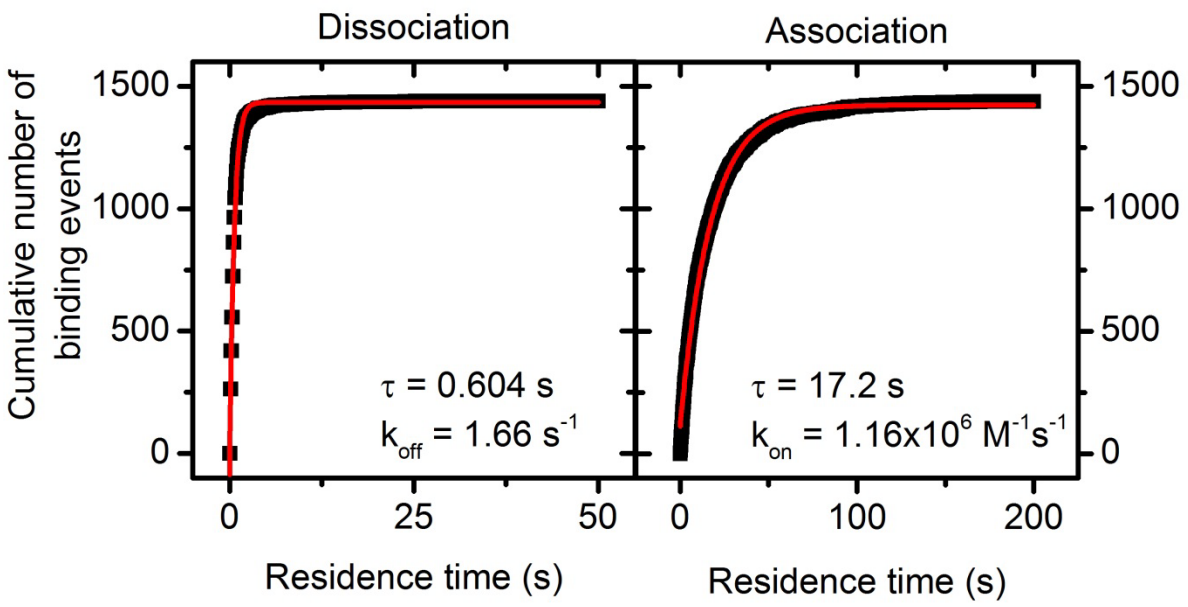

Figure 3.5 Single exponential fits of binding events to the *add* riboswitch in the ON state. Dwell times were extracted from the HMM fits of all molecules displaying binding events at 200 μM adenine concentration, and compiled. Red line indicates the single exponential fit. Dwell times and rate constants are as shown.

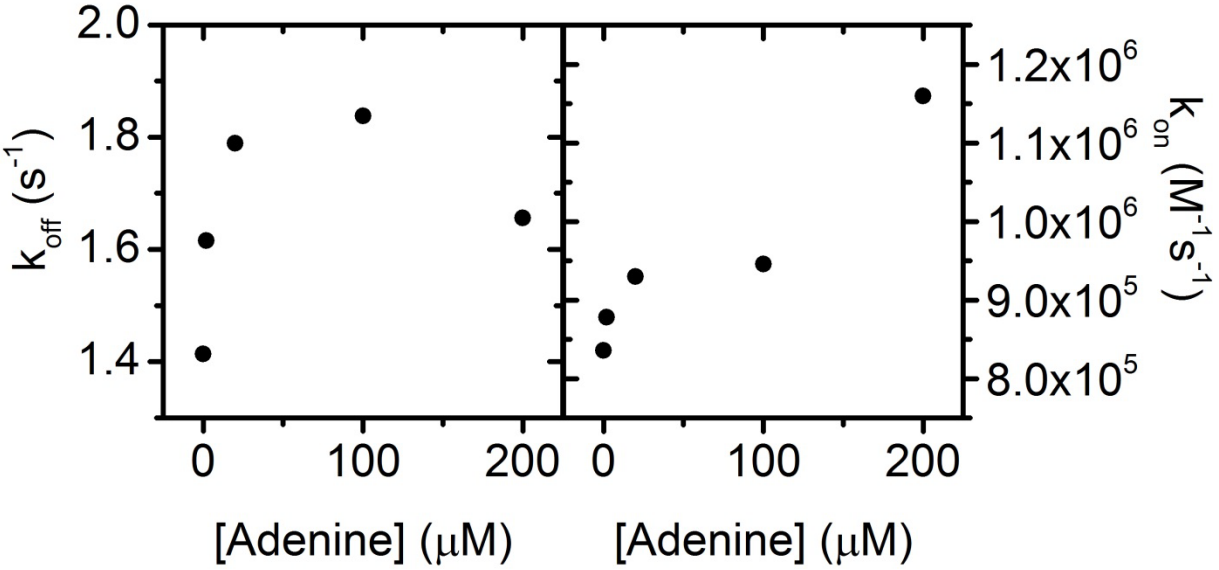

Figure 3.6 Effect of adenine on the rate constants for binding and dissociation of probe. The rate constants for binding (k_{on}) and for dissociation (k_{off}) are plotted as a function of adenine concentration.

3.4 Discussion

Since their initial discovery about ten years ago, researchers in the riboswitch field have largely resorted to classical bulk footprinting techniques to gain an understanding of the conformational ensembles adopted by these structured, dynamic RNA molecules. Additionally, insights into the ligand binding properties of riboswitches were gleaned from X-ray crystallography[23,31,36,42,44,47,48,54,55,60,110,116,142] and NMR[41,43,53,54,112-114,118] structural studies. While these studies provided the foundation for the global folds and ligand recognition motifs utilized by riboswitches, they did not address a large question still remaining in the field: What is the molecular basis for the conformational change conferred upon a single ligand binding event? Also, how is this change translated into a genetic decision?

Here we have described a novel single molecule assay that has the potential to answer these questions directly for the first time. We have taken the basic qualities of a classical chemical footprinting technique, and adapted them for single molecule studies to describe the local conformational changes that occur around the Shine-Dalgarno (SD) region of a translational riboswitch. Translational riboswitches operate through the ligand-dependent occlusion or release of the SD sequence, thereby inhibiting or allowing initiation of translation to occur, respectively. In the case of the *V. vulnificus add* riboswitch, adenine binding induces the formation of a P1 helix, releasing the SD sequence and AUG start codon from alternative base-pairing (**Figure 3.1**). To mimic binding of the 30S small subunit, we have designed a small fluorescently labeled DNA oligonucleotide containing the endogenous sequence of the *V. vulnificus* 16S rRNA[143]. By observing a colocalization signal between the Cy5 labeled riboswitch RNA and the Cy3 labeled rRNA oligonucleotide mimic, we have shown that this single molecule footprinting technique is feasible. Finally, we show a quantifiable difference in

binding rates in the presence and absence of adenine, suggesting that this technique can be applied to other dynamic, structured RNAs.

One slight drawback of this technique over classical footprinting techniques is the need for a doubly labeled RNA construct; the RNA must contain a biotin or other capturing modification in order to immobilize it to the slide surface for TIRFM[75], and a fluorophore label to localize its position on the slide. We have utilized a labeling procedure that produces a doubly labeled RNA construct with sufficient yield for single molecule studies. RNA contains two intrinsic functional groups, the 5' phosphate and 3' vicinal diols, which can be exploited to perform chemical modifications. First, labeling of the 5' end of RNA can be accomplished by reacting the phosphate group with the carbodiimide cross-linker EDC and imidazole[144] to generate a reactive phosphorylimidazolide group that can be coupled to any primary amine containing molecule, such as biotin hydrazide. The 5' modified RNA can then be reacted with sodium periodate, which oxidizes the ribose 2',3'-diol in RNA to aldehyde groups that are reacted with a hydrazide modified fluorophore, such as Cy5 hydrazide[144]. This tandem strategy provided us with a TIRF-viable construct without internal modifications that may interfere with secondary or tertiary structure formation. While these modifications do take some time and effort to generate, they are accomplished using readily available reagents, and do not require internal labels, which can be quite challenging to incorporate. Additionally, the amount of material required for single molecule studies is significantly lower (~100 pM) than required other ensemble techniques, such as NMR (close to millimolar).

Our single molecule footprinting data in the presence of adenine demonstrate multiple probe binding events per time trajectory. The time between each event gives important kinetic information regarding the accessibility of the SD sequence. A faster k_{on} implies the SD sequence

is exposed, thus allowing for rapid association of the 16S rRNA mimic. Indeed, we see a faster k_{on} in the presence of adenine. When the adenine concentration is decreased, the k_{on} also decreases, indicating a sequestration of the target SD sequence as a direct consequence of ligand concentration. Further concentrations of adenine will be tested to further prove the feasibility of this assay. While these data are very preliminary, we believe that this single molecule footprinting assay provides a direct readout on the occlusion and release of structural elements within biologically relevant RNAs, and can be extended to other dynamically structured RNAs.

3.5 Acknowledgements

We thank Dr. Harald Schwalbe and Martina Schonnenbeck for their generous gift of the *add* riboswitch RNA.

CHAPTER 4

Direct Observation of Ligand-Mediated Shine-Dalgarno Sequestration of Single preQ$_1$ Riboswitches in the Full-length mRNA[3]

4.1 Introduction

Almost forty years ago, Shine and Dalgarno reported the sequence of the 3' end of the *E. coli* 16S ribosomal RNA (rRNA), and observed that it was complementary to an A/G rich region in several 5' UTRS of mRNAs, directly upstream from the start codon[145]. They speculated that this complementarity may not be serendipitous, but that it may be required for proper interaction of the small ribosomal subunit on the mRNA to promote initiation of gene expression. Not long after this original hypothesis, it was shown experimentally through mutational analysis of the 16S rRNA that this interaction is imperative for translational efficiency[146,147]. Since this discovery, significant work has been done to identify the so-called "Shine-Dalgarno" (SD) sequence in prokaryotic mRNAs through bioinformatics searches to determine sequence similarities between different 5' UTRs of genes or thermodynamic free energy calculations[148].

The SD sequence is a purine rich 3-8 nucleotide stretch in the 5' UTRs of prokaryotic genes located generally 5-8 nucleotides directly upstream from the start codon[149]. The crystal structure of the *Thermus thermophilus* 30S subunit in complex with an mRNA mimic clearly shows the SD sequence of the mRNA interacting with the 3' end of the 16S rRNA[150]. This

[3] Arlie J. Rinaldi performed all experiments, with the exception of the *in vitro* translation assay, which was performed by Paul Lund.

SD:anti-SD interaction is imperative for proper placement of the start codon in the P site of the 30S subunit for subsequent translation initiation[63]. The optimal length of the SD sequence is variable, but usually hovers around 4-5 nucleotides. Longer SD sequences can actually inhibit translation efficiency since the SD:anti-SD interaction must be melted once the initiation complex is formed and the 50S subunit docks onto the small subunit to begin elongation. On the other hand, a stronger SD:anti-SD interaction does improve translational efficiency when an AUG start codon is not used[151], or when the start codon is sequestered in secondary structural elements[152,153]. While the region between the SD sequence and the start codon has no clear sequence requirements, it is generally A/U-rich, which may serve to minimize secondary structure formation[154].

Translational riboswitches are endogenous, structured elements in the 5' UTR of bacterial mRNAs. As described Chapter 2, the $preQ_1$ riboswitch from *Thermoanerobacter tengcongensis* is an exceptionally small RNA regulatory unit located in the 5' UTR of the COG1564 gene[115,155]. Previous structural studies on the aptamer domain have shown that the ligand-bound conformation forms a compact pseudoknot in which its A-rich 3' tail docks into the minor groove of the P1 stem[42] (**Figure 2.1B**). The crystal structure of the ligand-free conformation shows that it is structurally similar to the ligand-bound conformation and SAXS studies demonstrates that it is slightly more open in solution. Our smFRET studies (**Figure 2.2**) in combination with our TOPRNA coarse-grained RNA simulations (**Figure 2.6**) show that the ligand-free aptamer adopts a combination of conformations in which the 3' tail interacts segmentally with the P1-L1 stem-loop, and P2 is only partially formed.

In the ligand-bound and ligand-free crystal structures, the 3' nucleotide, G33 forms a Watson-Crick base pair with C9. G33 has been speculated to be the second nucleotide of the SD

sequence for the COG1564 gene. Since there is no structural information downstream of this nucleotide, the effect of ligand on the sequestration of the remaining SD sequence is unknown. Our smFRET results, however, indicate that P2 forms completely only in the presence of ligand. These results, together with the comparative SAXS data on the ligand-free and –bound aptamers[42] indicate that the crystal structure may not reveal the true conformation of the P2 helix in the absence of ligand, and that the 5' nucleotide of the SD sequence may be at least transiently exposed in the ligand-free state.

Up to this point, studies on translational riboswitches have primarily been focused on the aptamer domain, while few have included the downstream expression platform. Few to no studies have focused on the riboswitch in the presence of its mRNA. Studies in which translation is monitored as a function of ligand concentration in riboswitches usually involve a reporter gene fused to the 3' end of the riboswitch sequence. Since translational riboswitches may perform multiple rounds of genetic decisions per one mRNA transcript, which can have half-lives on the order of minutes, it seems imperative to study the effects of ligand-mediated regulation of translational riboswitches in the presence of the entire mRNA.

To this end, we have utilized the single molecule footprinting assay previously described for the ADD riboswitch in Chapter 3, and modified it to study the *Tte* preQ$_1$ riboswitch in its endogenous mRNA sequence, which contains two overlapping reading frames. We show through an *in vitro* translation assay, that both ORFs can be expressed, and that the expression of the upstream gene does not require an intact downstream gene. Using a fluorescently labeled RNA oligonucleotide with the exact sequence of the 16S rRNA from *Thermoanaerobacter tengcongensis*, we demonstrate a ligand-mediated occlusion of the SD sequence via a decrease in the binding rate constant, k_{on} of the probe. We attribute this as a direct effect of an increase in

preQ$_1$ concentration as the k_{on} of a control probe that hybridizes to a site within the ORF shows no discernible correlation with ligand concentration. Surprisingly, the rate of dissociation of the probe, k_{off}, increases with ligand concentration, indicating a competing effect between probe binding and P2 helix formation. These results provide, for the first time, a direct molecular explanation for the attenuation of translation initiation by single translational riboswitches, and demonstrate that the interplay between mRNA structure formation and the SD:anti-SD interaction between mRNA and 30S ribosomes is a delicate and carefully balanced system, more complex than a simple ON/OFF switch.

4.2 Materials and Methods

4.2.1 Cloning of *Tte* mRNA

Genomic sequences were downloaded from the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). The complete mRNA, including the COG1564 and COG1563 ORFs, and the 3' UTR as predicted from the FindTerm algorithm (SoftBerry), was amplified from the *Thermoanaerobacter tengcongensis* genomic DNA purchased from the NITE Biological Resource Center. The mRNA was cloned into pUC19 with an engineered T7 promoter at the 5' end between the BamHI and HindIII sites of pUC19. All primers are listed in **Table 4.1**.

4.2.2 mRNA transcription

The *Tte* pUC19 plasmid was linearized with HindIII (AflII or XbaI for *in vitro* translation assays) (New England Biolabs) for run-off transcription. Transcription reactions were performed in the presence of 120 mM HEPES-KOH, pH 7.5, 25 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 30 mM NTPs, 0.01% Triton X-100, 200 nM linearized plasmid, 0.01 U/µL

pyrophosphatase and 0.07 mg/mL T7 RNA polymerase in a total volume of 1 mL. Transcription reactions were incubated at 37 °C for four hours. Enzyme was removed by phenol/chloroform extraction and the resulting solution was spun in an Amicon 100 MWCO spin column to reduce the volume to ~100 µL. mRNA was purified by denaturing PAGE purification, detected using 254 nm UV radiation and gel eluted overnight. mRNAs were ethanol precipitated and resuspended in TE buffer at pH 7.0. The sequence of the *Tte* mRNA is listed in **Table 4.1**.

4.2.3 *In vitro* translation

in vitro translation experiments were performed using the PURExpress *In vitro* Protein Synthesis Kit (New England Biolabs) and the PURExpress Δ Ribosome Kit (New England Biolabs), in which house-purified ribosomes were added to the latter. Reactions were performed in the presence of 1X NEB solution A and Factor Mix, 4.2 µCi ^{35}S-Met (MP Biomedicals), 2.4 µM ribosomes and 50 nM mRNA in a total reaction volume of 10 µL. Reactions were incubated at 37 °C for 2 h. Following incubation, 1 µL of 1 M KOH was added to hydrolyze nucleic acids and terminate peptides. 50 µL of cold acetone was added to precipitate proteins, followed by centrifugation at 20.8k x *g* for 10 min. The supernatant was removed and the protein pellets were resuspended in SDS gel loading buffer. Protein products were resolved on a 16% Tris-Tricine SDS-PAGE gel and imaged using a Typhoon phosphorimager.

4.2.4 Single molecule footprinting of the *Tte* mRNA

10 nM *Tte* mRNA, Cy3-LNA and biotin capture strand were heat annealed at 70 °C for 2 min in the presence of 50 mM Tris-HCl, pH 7.5, 0.6 M NaCl and 20 mM $MgCl_2$, and allowed to slow cool to room temperature for 20 min in the presence or absence of $preQ_1$. Following slow cooling, the RNA mix was diluted to 100 pM in the same buffer in the presence or absence of $preQ_1$, with a 10X excess of Cy3-LNA and biotin capture strand to ensure the complex would

stay intact during dilution. All sequences of mRNA, capture strand, Cy3-LNA and Cy5-probes can be found in **Table 4.1**. The diluted complex was chilled on ice. The chilled solution was flowed over an assembled microfluidic channel on a quartz slide coated with biotinylated-BSA and streptavidin, as previously described[75,86]. 100 µL of the chilled, 100 pM RNA complex was flowed over the slide and allowed to equilibrate for 5 min. Excess RNA was washed off the slide with buffer +/- $preQ_1$. An oxygen scavenging system consisting of 5 mM protocatechuic acid and 50 nM protocatechuate-3,4-dioxygenase, +/- adenine, to slow photobleaching, 2 mM Trolox to reduce photoblinking[91] and 50 nM Cy5-probe was flowed over the slide and allowed to equilibrate for 5 min. Both Cy5 and Cy3 dyes were directly excited simultaneously using a 638 nm red diode laser and 532 nm green laser, respectively. Emission from both fluorophores was simultaneously recorded using an intensified CCD camera (I-Pentamax, Princeton Instruments) at 100 ms time resolution using Micro-Manager software. Fluorescence traces were extracted from the raw movie files using IDL (Research Systems) and analyzed using Matlab (The Math Works) scripts. Genuine traces exhibiting binding were manually selected using the following criteria: a single photobleaching step of the Cy3 signal, Cy3 fluorescence intensity of >200 intensity units and at least two binding events per trajectory with a signal to noise ratio of at least 3:1. A lower signal to noise ratio was utilized for these experiments than traditional smFRET experiments due to the higher than normal background noise in the Cy5 channel from the high concentration of probe in solution. Suitable traces were compiled and a nonlinear filter was applied to reduce the probability of false transitions in the data processing. Hidden Markov Modeling analysis was performed on the donor intensity using the segmental k-means algorithm in the QuB software suite as described[93]. A two-state model was used with an unbound and bound state to idealize the data. Transition density plots were constructed to extract the dwell

times in the bound and unbound states, as described[93]. Dwell times were fit to a single exponential in OriginLab 8.1 from which on and off rates were calculated.

4.2.5 Thermodynamic stability of SD-anti-SD interaction

Free energies of the interactions between the anti-SD probe and Shine-Dalgarno sequence were calculated using the bifold algorithm on the RNAstructure WebServer[156] (rna.urmc.rochester.edu) using the default parameters.

Table 4.1 Oligonucleotides used in Chapter 4.

Name	Sequence
5' Primer	5' GATCATCCATCCTAATACGACTCACTATAGGGGAACTCCTACTACAAGTTGCTAAGAGGC 3'
3' Primer	5' GATCATAAGCTTGCTTCCTCATCGTTCTCTGTAAACTC 3'
FL *Tte* mRNA	5'GGAACTCCTACTACAAGTTGCTAAGAGGCTATTTTTTAGTTCAAATTACTCATACAATCATGTTAAAAT TAATCGCAGTGAGCAACAAAATGCTCACCTGGGTCGCAGTAACCCCAGTTAACAAAACAAGGGAGGTA ATTTTGTGCCCAAAAAAAGAATAAAAGATTTAGCTGAAATTGCTCTTGTTGCAGCAATTTATTTCGCACT CACAATTATATTTTCGTCCATTTCGTTTTTACCCGTTCAATTTCGAATCGGGGAAATTACGAAATCCATTG TAGTATTCAATAAAAAATATGCTATTTCCATGATGATAGGAAATTTTTTTGCAAATTTGTTTAGCCCATTT GCTGGTGCAATGGAATTAATTTTTATGCCTCTTTCGAACTTAATAGGCTGTACAATTGGATACTACATTG GAAGACTTACTCACAAAGCGATAGGAGCTATATTCATAGCCCTTTGGATTGCAGCATCAGTTGCAATTA CTTTAAAGGTTTCTGCAGGCATACCATTTATTCCGACTTTCTTAAGCGTGGGAGTAGCGGAAACTGTACT TTTGGTAACTGGATATTTTTTGCTTTTCACAATTGAAAAGAAAGGAGTTGTGAAATTTGACAGATAAATA TAAAGAGAGAAGATTTGACATTTACGGTTACGAAAAAATTGACAAAGAAGTTCTAGAATCTATTGAATA TGAGTATCCTGAAAAAAATACTATCGTGGAGTATATTACCGATGAATTTTCTTCTGTTTGCCCTTGGACA GGATTACCTGACAATGCAAAACTTACTATAAGGTATATACCCCACAAAAAACTTGTAGAACTTAAATCC TTAAAATATTACCTTACATCTTATAGGAATGTAGGTATATTGCAAGAACATGCAATAAACAGAATTTTAG ATGATTTGGTGGAATTCCTGCAGCCAAAATTTATGGAAATAATAGGCGAATTTCAGGAAAGAGGAGGA ATAGCTACAAGAATTATAGCAAGGTATGAAAAAGAGGAGTATTAAACTTAAAAGGCTGCCTAAAATTTT GTAGGCAGCTTTTTTATTCATTTTAGTTTTTCTTCAAAATGAGTTTACAGAGAACGATGAGG 3'
Biotin-capture strand	5' GCCTCTTAGCAACTTGTAGTAGGAGTTCCAAAAAAAAAA-biotin 3'
LNA	5' TYE563-+GT+CAAATTT+CA+CAA+CT+C+CTTT+C 3'
Anti-SD Probe	5' Cy5-GAUCACCUCCUU 3'
Control probe	5' Cy5-GCAACAAGAGC 3'

Note: TYE563 is a Cy3 analog, and is referred to as Cy3 throughout the rest of the chapter.

“+” nucleotides refer to locked nucleic acids.

4.3 Results

4.3.1 The COG1564 mRNA contains two ORFs

The $preQ_1$ riboswitch from *Thermoanaerobacter tengcongensis* is located directly upstream from a putative $preQ_1$ transporter gene[115,155] that is 483 nucleotides long as predicted by the National Center for Biotechnology Information (NCBI). Interestingly, the 3' end of the COG1564 gene overlaps with a downstream gene, COG1563, by 11 nucleotides (**Figure 4.1**). Because of this, the 3' end of the upstream gene contains the putative SD sequence of the downstream gene. To determine the natural termination site of the mRNA, which would occur downstream of the COG1563 gene, the genomic sequence was entered into the FindTerm algorithm (Softberry), which identified a polyU terminator 86 nucleotides downstream from the COG1563 stop codon. For this reason, the mRNA was cloned to this point of the genomic DNA, with an engineered HindIII site (see **Materials and Methods**) to yield the 1,105 nucleotide mRNA product shown in **Figure 4.1**.

To determine whether both protein products can be expressed using the heterologous *E. coli* ribosome, or if translation of COG1564 requires an intact COG1563 gene, we performed *in vitro* translation using three different mRNA 3' truncations, as shown in **Figure 4.1**. The ΔHindIII construct is the full-length construct, and includes the complete COG1564 and COG1563 ORFs, as well as the 3' UTR of the full-length transcript. This construct should yield both full-length protein products since both ORFs are intact. The ΔXbaI construct contains an intact COG1564 ORF, but truncates the COG1563 ORF 63 nucleotides downstream from its start codon. Expression of the COG1563 protein product from this construct would yield a 2.9 kDa peptide that may not be visible on SDS-PAGE gels. Finally, the ΔAflII construct is truncated 85 nucleotides upstream of the COG1564 stop codon, eliminating the COG1563 ORF completely.

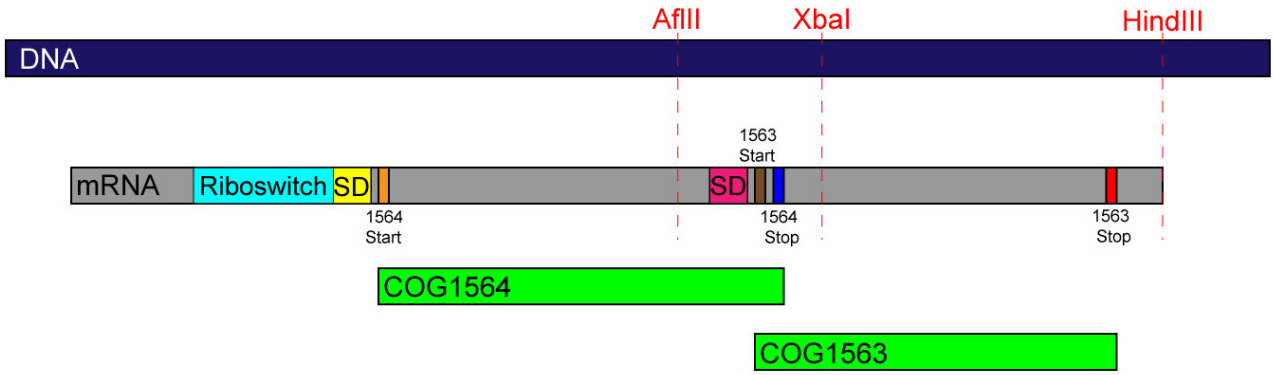

Figure 4.1 The downstream gene of the *Tte* preQ$_1$ riboswitch contains two overlapping reading frames.
Genomic DNA (blue) is transcribed to yield the mRNA transcript (gray). Three different 3' truncations are used in this study: AflII, XbaI and HindIII (red dashed lines). The HindIII truncation refers to the full-length transcript used in this study. XbaI truncations delete the majority of the downstream ORF, while keeping COG1564 intact. AflII truncates the COG1563 ORF and the COG1564 stop codon. The preQ$_1$ riboswitch is shown in cyan and the COG1564 and 1563 SD regions are shown in yellow and pink, respectively. The figure is not to scale.

As expected, *in vitro* translation of the full-length ΔHindIII construct yielded two protein products of the expected sizes, 17.6 and 16.1 kDa for COG1564 and COG1563, respectively (**Figure 4.2**, HindIII lane). Surprisingly, the ΔAflII truncation yielded no visible protein product at the expected size of 14.3 kDa (**Figure 4.2**, AflII lane). Since the truncation occurs upstream from the COG1564 stop codon, ribosomes could become stalled at the end of the RNA, and without a rescue system, such as tmRNP[157], a single round of translation may not be producing sufficient protein for visualization. Finally, the ΔXbaI truncation only yields one protein product corresponding to the full-length COG1564 gene (**Figure 4.2**, XbaI lane). The expected size of the COG1563 protein from this truncation is 2.9 kDa, and may either be too small to resolve, or, for reasons similar to the ΔAflII truncation, the amount of protein produced from one round of translation may not be sufficient. These results indicate that both intact ORFs can be expressed; and, since the COG1564 protein is produced even in the absence of COG1563 expression, COG1563 is not required for COG1564 expression.

4.3.2 *Tte* mRNA single molecule labeling scheme

To observe the occlusion of the SD sequence as a function of ligand concentration in the full-length *Tte* mRNA, we exploited the single molecule footprinting assay previously developed for the ADD riboswitch (see **Section 3.3.2** and **Figure 3.3**), with a few notable changes. Due to the much larger size of the full-length mRNA as compared to the ADD riboswitch (1,105 nucleotides compared to 111 nucleotides, respectively), we chose to non-covalently hybridize a biotinylated capture strand and a Cy3-labeled locked nucleic acid (LNA) probe to immobilize and localize the *Tte* mRNA to the slide, respectively. Since we chose to immobilize the mRNA to the slide by the 5' end, we extended the mRNA upstream from the

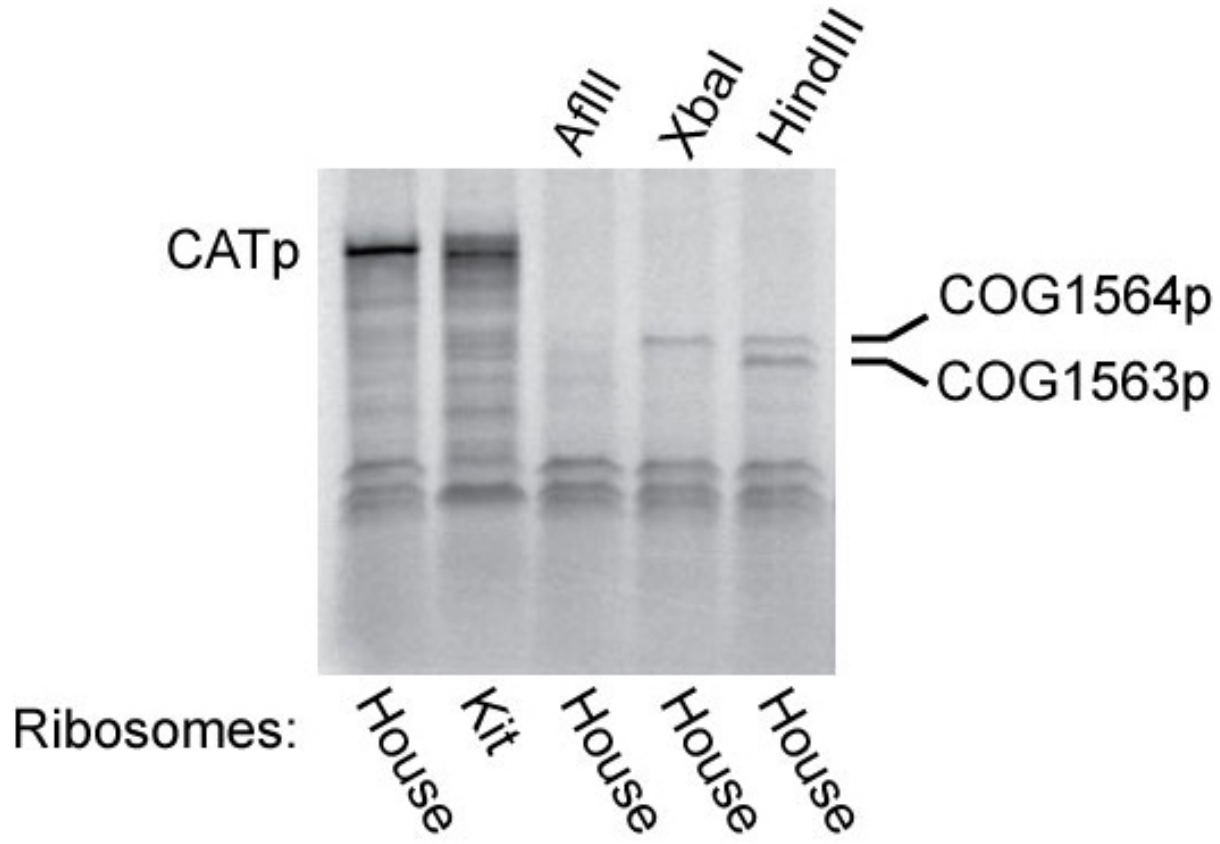

Figure 4.2 *in vitro* translation of truncated *Tte* RNA transcripts.
Chloramphenicol acetyl transferase (26 kDa) was included in the assay as a positive control. Ribosomes used were either provided by New England Biolabs (kit) or prepared in house (house). mRNA templates are indicated as described in **Figure 4.1**.

riboswitch sequence by 64 nucleotides, to ensure that hybridization of the capture strand would not interfere with any secondary and tertiary structure of the riboswitch sequence (**Figure 4.3** and **Table 4.1**). Additionally, since we have previously shown that the COG1563 gene can be initiated, and therefore, bind ribosomes (**Figure 4.2**), we chose to design a high melting temperature Cy3-labeled locked nucleic acid (LNA) probe that binds to its SD sequence and UUG start codon, thus occluding them from probe binding (**Figure 4.3** and **Table 4.1**). Therefore, any binding that we observe should be localized to the SD sequence of interest. Finally, we increased the concentration of monovalent and divalent cations in our buffer due to the much larger size of the *Tte* mRNA as compared to the ADD riboswitch. This buffer was required to observe longer lived binding events for observation at our 100 ms time resolution (data not shown).

To ensure that the mRNA:capture strand:LNA complex was capable of forming under our selected conditions at room temperature, the complex was formed and visualized on the slide. Indeed, a combination of these three components resulted in significant Cy3 signal over background as compared to buffer alone (**Figure 4.4**, right panel). When capture strand and mRNA were annealed together, the signal was comparable to background (**Figure 4.4**, middle panel), as well as when mRNA and Cy3-LNA were annealed together (**Figure 4.4**, right panel). These results indicate that the biotinylated capture strand and Cy3-LNA efficiently hybridize with the *Tte* mRNA under our selected buffer conditions, and that any signal that is observed originates from mRNA molecules.

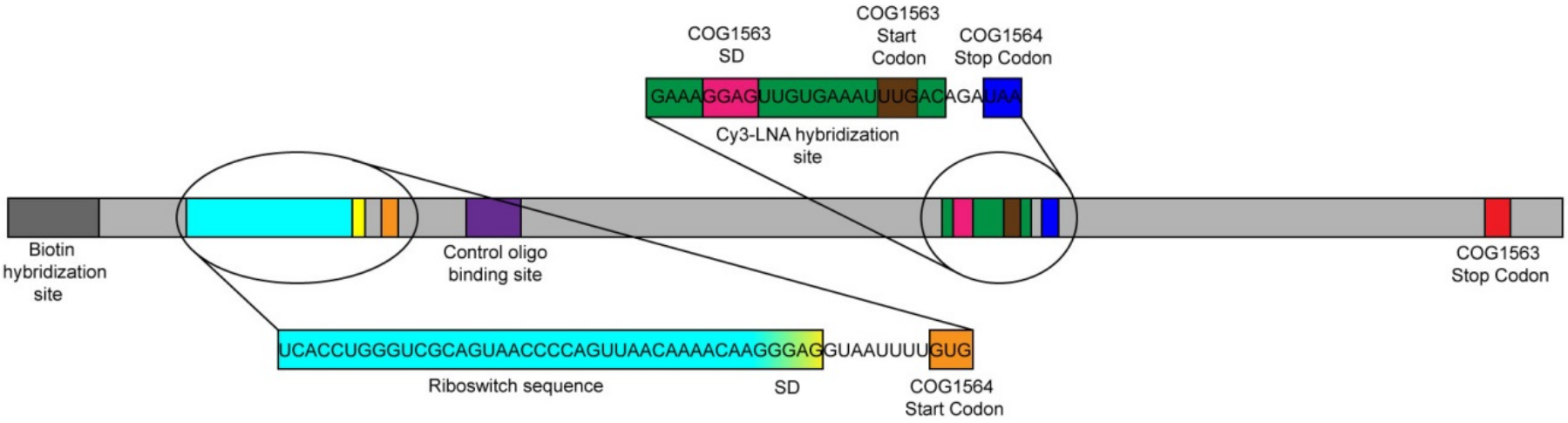

Figure 4.3 Map of key hybridization sites of the *Tte* mRNA relative to sequence elements.
Hybridization sites of key oligos used in this study to the full-length mRNA. Color code is as used in **Figure 4.1**. For details regarding exact sequences, refer to **Table 4.1**.

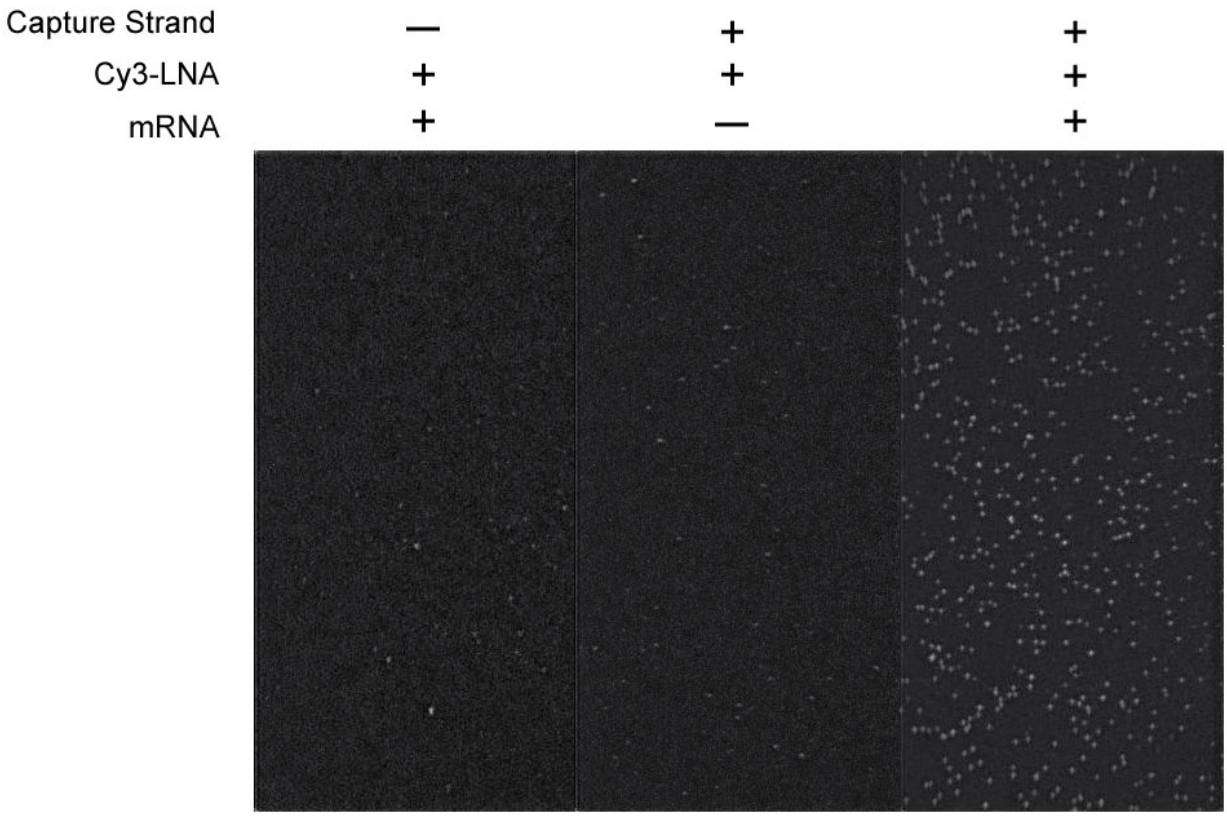

Figure 4.4 Capture strand:Cy3-LNA:mRNA complex forms on the slide surface.
Cy3 emission field of view. Background fluorescence is seen in panels 1 and 2 where the capture strand and mRNA were missing from the complex, respectively. Cy3 emission is only observed when all three components are annealed together (panel 3).

4.3.3 Single molecule footprinting of the *Tte* mRNA using an anti-SD oligo

According to our smFRET data on the *Tte* preQ$_1$ riboswitch aptamer, the riboswitch pseudoknot should be partially open in the absence of preQ$_1$ (**Figure 2.2** and **Figure 2.10**), leaving the SD sequence more exposed than in the presence of ligand. Additionally, SAXS data have shown that the ligand-free pseudoknot is more open than the ligand-bound pseudoknot in solution[42]. To test whether a Cy5-labeled RNA probe mimicking the 3' end of the *T. tengcongensis* 16S rRNA (**Table 4.1**) can transiently bind to the full-length *Tte* mRNA, we developed the experimental setup shown in **Figure 4.5**. Similarly to the ADD riboswitch, we immobilized the mRNA complex to the slide through a biotin-streptavidin interaction; however, we localized mRNA molecules via Cy3 emission. We then colocalized Cy5 emission from bound anti-SD probes with the Cy3 emission from the mRNA. To accomplish this, we flowed the probe over the mRNA complex in the absence of ligand using our TIRFM setup. Indeed, in the absence of preQ$_1$, over a thousand binding events were observed in >300 molecules (**Figure 4.6**). These binding events were compiled and the resulting dwell times were plotted and fit with a single exponential association curve (**Figure 4.7**). From this fit, the k_{on} was calculated to be 1.2×10^6 $M^{-1}s^{-1}$ whereas the k_{off} was found to be 1.31 s^{-1}. The presence of over a thousand binding events with residence times longer than our 100 ms time resolution indicate that our experimental conditions are ideal for observation of binding.

4.3.4 Occlusion of the SD sequence as a function of ligand concentration

To determine whether preQ$_1$ is capable of inducing a conformational change that results in occlusion of the SD sequence, we performed our single molecule footprinting assay with increasing concentrations of ligand. smFRET measurements on the *Tte* preQ$_1$ riboswitch aptamer demonstrated a conformational change from a mid-FRET state centering around ~0.7 to a high-

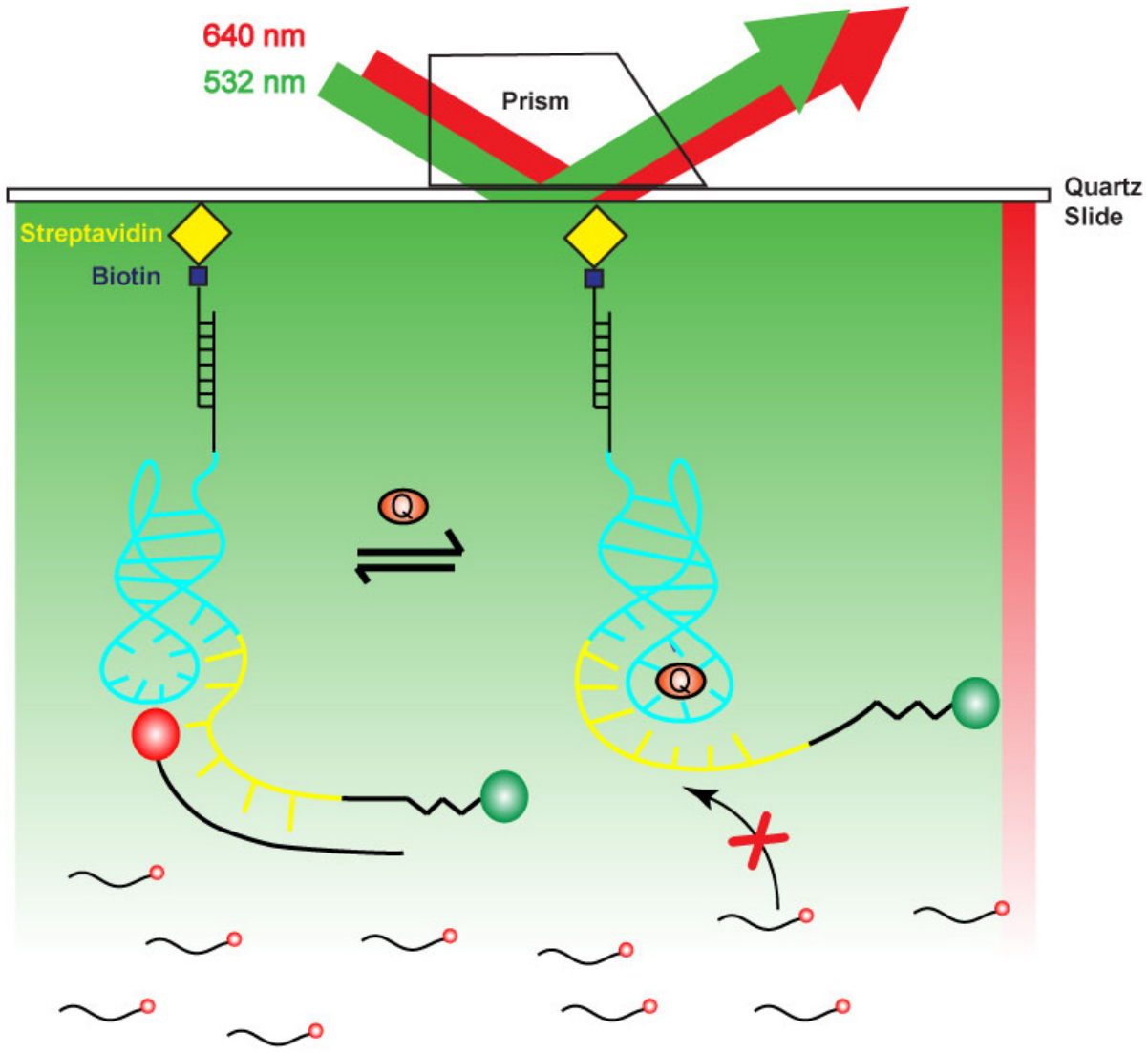

Figure 4.5 Single molecule footprinting experimental setup for the preQ$_1$ riboswitch.
The full mRNA and Cy3-LNA are not illustrated for simplicity. In the absence of preQ$_1$, the SD sequence should be exposed and therefore, available for probe binding. When the mRNA is annealed in the presence of preQ$_1$, helix P2 is stabilized, thus occluding the SD region from probe binding.

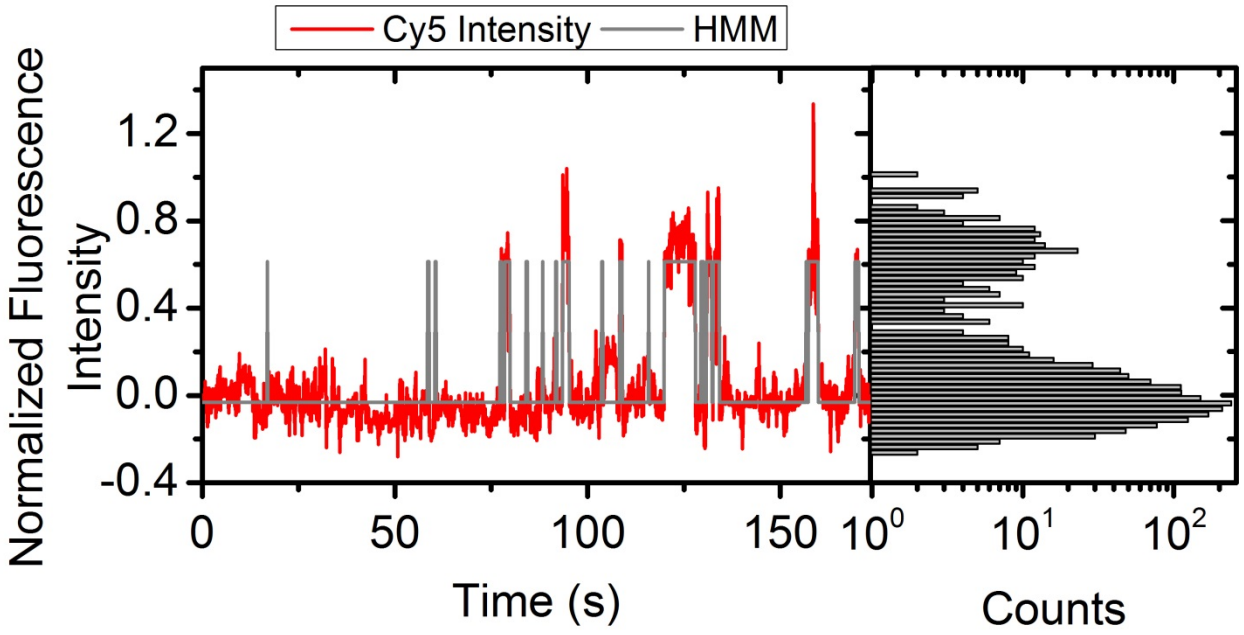

Figure 4.6 Sample binding trajectory of the anti-SD probe to the *Tte* mRNA.
Cy5 fluorescence (red) is reported as normalized to the entire dataset. The idealized Hidden Markov Model (HMM) is shown in gray. The number of counts of Cy5 fluorescence is shown as a histogram, which displays a clear shoulder.

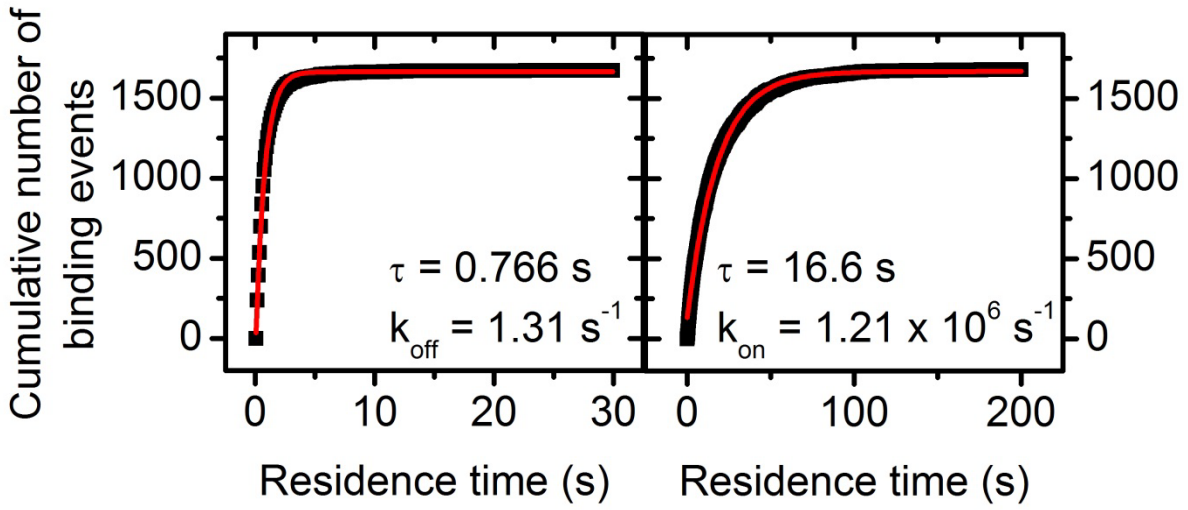

Figure 4.7 Exponential fits of dwell times in the absence of ligand.
Dwell times, as determined by HMM, in the bound and unbound states across an entire data set are compiled and plotted. Single exponential curves are fit to the data, yielding the rate constants for binding and dissociation as shown above.

FRET state centering around ~0.9 as $preQ_1$ concentration was increased (**Figure 2.10**), with a $K_{1/2}$ value of 69 nM (**Figure 2.11**), and fully saturated at 10 μM $preQ_1$. It is conceivable that the full-length mRNA would require a greater concentration of ligand in order to effect a conformational change to occlude the SD sequence to regulate gene expression. For this reason, we tested the effect of $preQ_1$ on the accessibility of the SD region in the context of its full-length mRNA up to 100 μM ligand. As expected, as $preQ_1$ concentration is increased, the k_{on} decreases from $1.2x10^6$ $M^{-1}s^{-1}$ in the absence of ligand to $9.2x10^5$ $M^{-1}s^{-1}$ at 100 μM ligand (**Figure 4.8**, right panel, black circles). The decrease in the value of the binding rate constant indicates that a slight occlusion of the probe binding site occurs as a function of $preQ_1$ concentration. However, the decrease is not as drastic as one would expect if the SD sequence was transitioning from completely open to completely closed. This indicates a more complex mechanism. Furthermore, the total number of binding and dissociation events per Cy3 localized molecule decreased as $preQ_1$ concentration increased, further attesting to the longer time between individual binding events (**Figure 4.9**). To ensure that the change in the binding rate constant is due to a conformational change localized to the riboswitch sequence, we performed the same experiment with a Cy5-labeled probe that is complementary to region within the COG1564 ORF, which should not be affected by $preQ_1$ concentration (**Figure 4.3**). As expected, there does not seem to be a trend in the rate constant of association of the control oligonucleotide (**Figure 4.8**, right panel, blue circles) suggesting that the occlusion of the anti-SD binding site is localized to the SD region in the 3' region of the *Tte* riboswitch sequence.

Unexpectedly, the rate constant for the dissociation of the anti-SD probe, k_{off}, increases with $preQ_1$ concentration (**Figure 4.8**, left panel, black circles), while the k_{off} of the control oligonucleotide shows no discernible trend (**Figure 4.8**, left panel, blue circles). This indicates

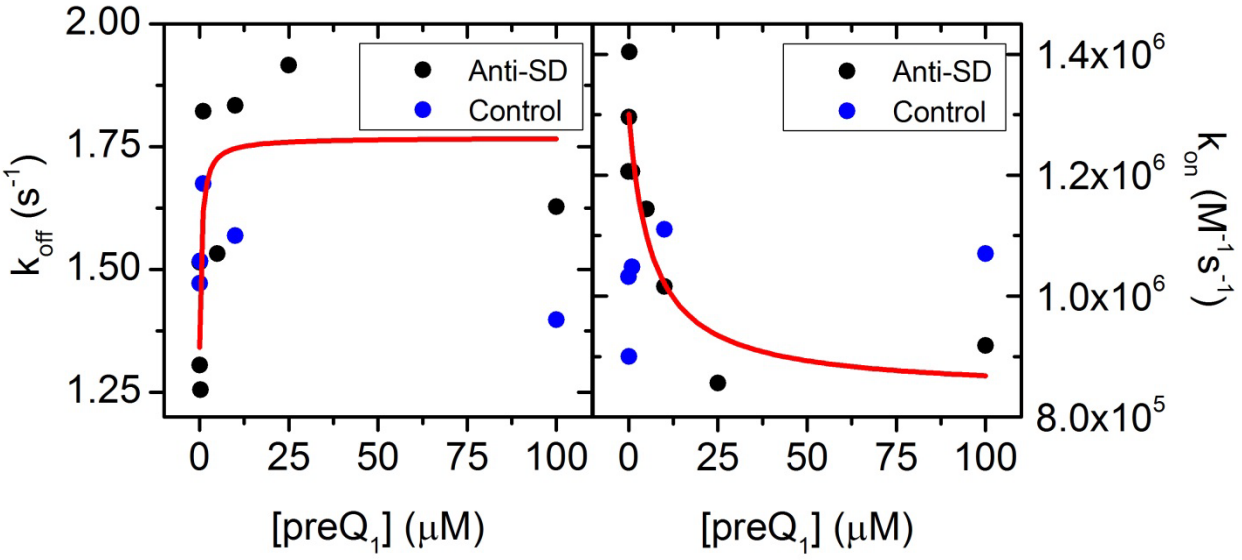

Figure 4.8 Rate constant analysis of anti-SD and control oligos as a function of preQ$_1$ concentration.

Each point represents the analysis described in Figure 4.7 for different concentrations of preQ$_1$ for the anti-SD oligo (black circles) and the control oligo (blue circles). Red line indicates the logistic fit to the anti-SD data.

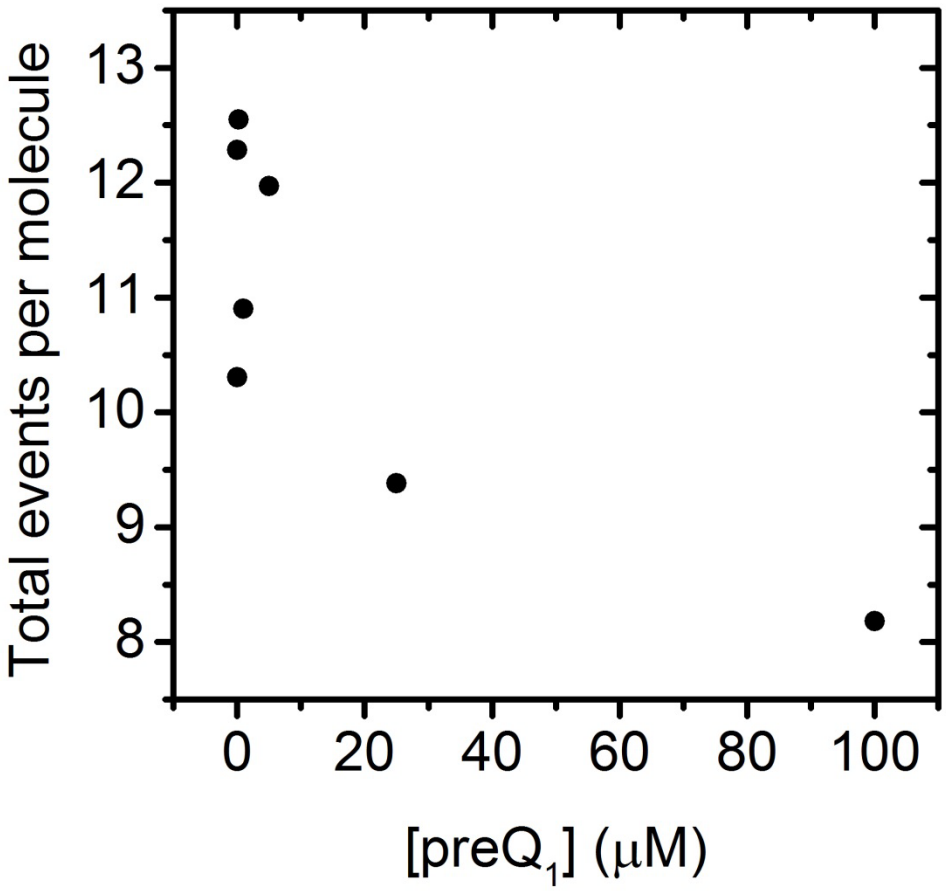

Figure 4.9 The number of total binding events per molecule decreases with $preQ_1$ concentration.
The number of binding and dissociation events per Cy3 labeled molecule were plotted as a function of $preQ_1$ concentration.

that along with an increase in the amount of time between individual binding events at high $preQ_1$ concentration, the residence time of the bound probe decreases. While there is a distinct change in the k_{off} of the anti-SD probe as a function of $preQ_1$ concentration, the change is only ~2-fold. The observation is consistent with the relatively small change in the value of k_{on}, supporting the hypothesis that a more complex mechanism is at work.

4.3.5 $preQ_1$ riboswitch can respond to fluctuations in physiological environment

It has been speculated that translational riboswitches have a longer time window in which to bind ligand in comparison to their transcriptional counterparts. In the latter, the RNA must bind ligand in a timely manner before RNA polymerase clears the expression platform to effect a genetic decision, while the former has the extra time before ribosomal subunit association to bind ligand. For this reason, it has been suggested that translational riboswitches can act as reversible regulators, responding to fluctuating changes in their physiological environments depending on the cellular milieu. Evidence supporting this hypothesis has been presented for the SAM-III riboswitch. Time-resolved 2-aminopurine fluorescence demonstrated a ligand-dependent conformational sampling between the ON and OFF states[59]. Further, it was shown that the translational ADD riboswitch could bind ligand in the context of its expression platform while the closely related transcriptional *pbuE* riboswitch could not[49]. These results support the notion that translational riboswitches can operate as reversible regulators.

To determine whether the $preQ_1$ translational riboswitch can act as a reversible regulator, we observed the changes in the SD accessibility using our single molecule footprinting assay. To do this, we assembled the mRNA with the LNA and capture strand in the absence of $preQ_1$. After imaging for a set amount of time, the lasers were turned off and either buffer and probe (**Figure 4.10A**) or buffer supplemented with $preQ_1$ and probe (**Figure 4.10B**) were flowed over the slide.

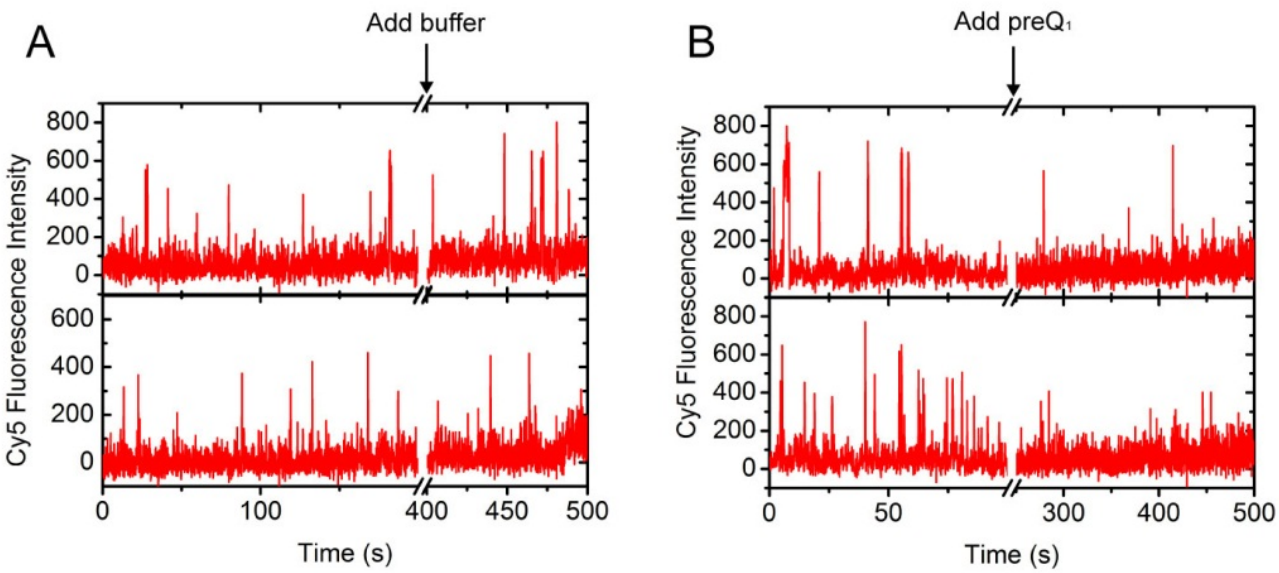

Figure 4.10 The $preQ_1$ riboswitch can respond to changing conditions.
mRNA complex was heat annealed in the absence of ligand in both panels. In panel (A), following 200 seconds, the lasers were turned off and buffer with anti-SD probe was flowed over the slide. Lasers were turned back on at 400 seconds and imaging resumed. In panel (B), the lasers were turned off at 100 seconds and buffer with $preQ_1$ and anti-SD probe was flowed over the slide. The lasers were turned back on at 250 seconds and imaging resumed.

The RNA was allowed to equilibrate for ~3 min before the lasers were turned back on and any changes in accessibility were observed. In the control, the probe is observed to bind relatively frequently before and after the dark period (**Figure 4.10A**). However, when $preQ_1$ is added to the RNA during the dark period, we see a marked decrease in the number of binding events when the lasers are turned back on (**Figure 4.10B**). These results indicate that the *Tte* $preQ_1$ riboswitch can sense and adapt to its physiological environment as a means of gene regulation.

4.4 Discussion

A large majority of studies conducted on riboswitches have focused on the aptamer domains, while only a few extended their analyses to the expression platform. The only studies that are extended to the mRNA open reading frame included the start codon, and perhaps a few nucleotides into the reading frame, or as translational fusions to non-endogenous reporter genes. Here, we have examined the *Tte* $preQ_1$ riboswitch in the context of its complete mRNA, which includes two overlapping open reading frames. We have shown, using an *in vitro* translation assay that both genes can be expressed, and that the presence of the downstream gene is not necessary for expression of the upstream gene. Using a novel single molecule footprinting assay, we demonstrated a direct effect on the accessibility of the SD sequence that overlaps the $preQ_1$ riboswitch aptamer as a function of ligand concentration. We show using a fluorescently labeled probe to mimic the 3' end of the 16S rRNA from *T. tengcongensis* that the binding rate of this probe decreases as a function of $preQ_1$ concentration, whereas the binding rate of a probe that hybridizes to a region within the ORF shows no discernible change. We interpret this effect as an indication that as $preQ_1$ concentration increases, the accessibility of the SD sequence decreases, which would inhibit subsequent translational expression of the downstream gene.

Unexpectedly, the decrease in the binding rate of the probe is accompanied by an increase in the rate of dissociation of the anti-SD probe. This could be due to higher thermodynamic stability of the P2 helix in the presence of high $preQ_1$ concentrations (**Figure 2.1C**), which competes with anti-SD probe hybridization. RNAstructure[156] predictions of the complex formed between the SD region of the 5' UTR of the COG1564 gene and the 3' end of the 16S rRNA from *T. tengcongensis* predicts the formation of seven Watson-Crick base-pairs and one G-U wobble, with a ΔG° of -12.0 kcal/mol (**Figure 4.11**). RNAstructure predicts that the last two nucleotides of the 16S rRNA, 3' UU, form a Watson-Crick base-pair and G-U wobble with A32 and G33 of the $preQ_1$ riboswitch, which were predicted to be the first two nucleotides of the SD sequence[42] (**Figure 2.1C** and **Figure 4.11**). However, this was derived from RNAstructure predictions between the 5' UTR and the *E. coli* 16S rRNA. The current study has extended this analysis to observe the potential base-pairing interactions with the native *T. tengcongensis* 16S rRNA. Nonetheless, these two nucleotides were shown to form intramolecular interactions; A33 forms a non-canonical base pair with A10 and G33 forms a Watson-Crick base-pair with C9 in both the ligand-free and ligand-bound structures. These interactions, however, may be an artifact of the crystallization process in the ligand-free structure, since SAXS data have shown the ligand-free conformation to be more open in solution than in the crystal lattice[42]. Furthermore, our smFRET data and TOPRNA simulations on the *Tte* aptamer domain indicate that the P2 helix is only partially formed in the absence of $preQ_1$ (**Figure 2.2** and **Figure 2.6**). Therefore, our results support the model shown in **Figure 4.12A**. In the absence of ligand, the 3' end of the SD sequence may be exposed and available for anti-SD probe binding, which begins a zippering effect along the remaining SD sequence. In the absence of ligand, the partially formed P2 helix is less thermodynamically stable than in the presence of

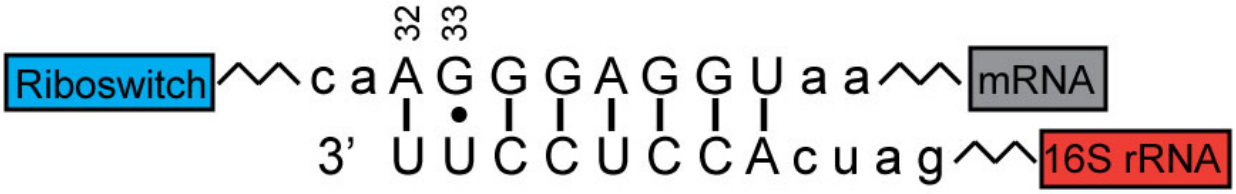

Figure 4.11 RNAstructure prediction of SD-anti-SD interaction.
A32 and G33 refer to the numbering as shown in **Figure 2.1C**. Lowercase letters indicate those that do not participate in base-pairing interactions with 16S rRNA.

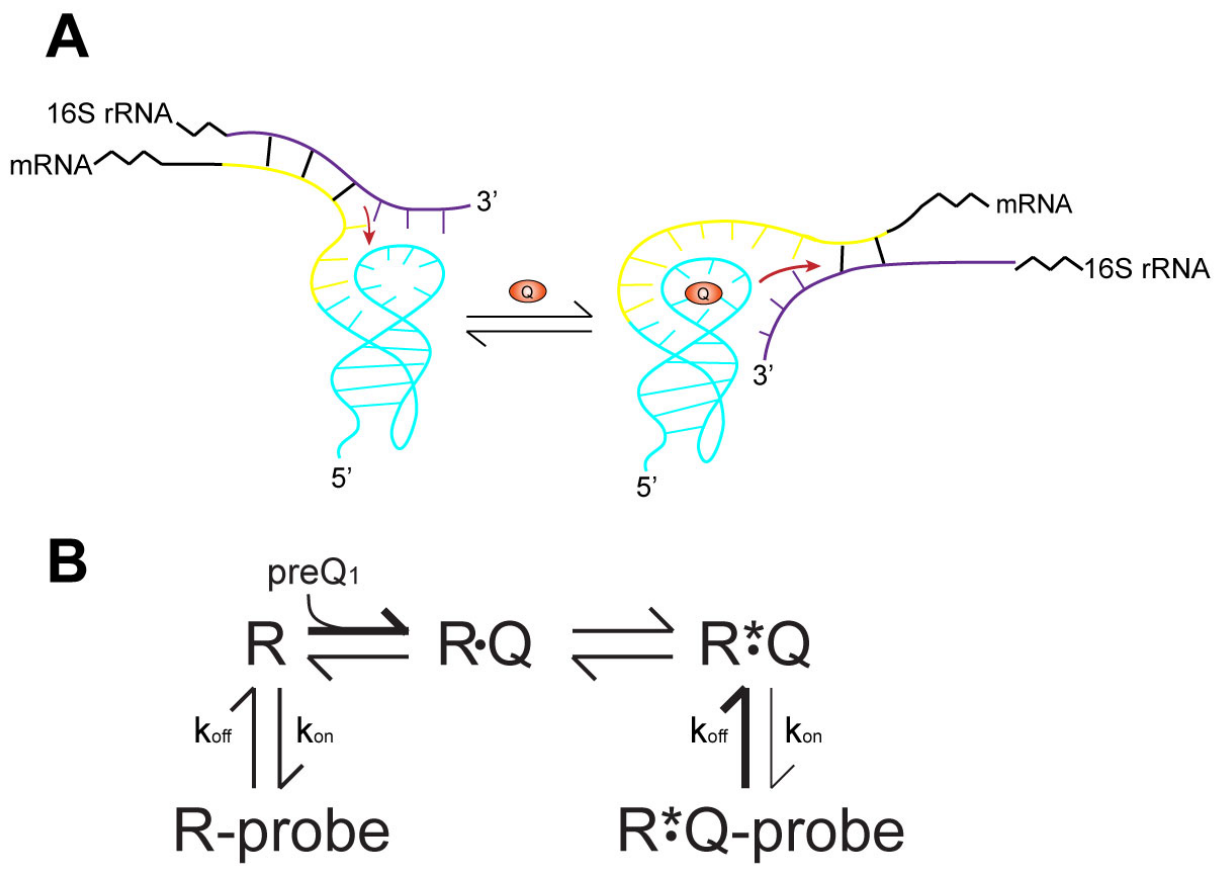

Figure 4.12 Model of the interaction between the preQ$_1$ riboswitch SD sequence and 16S rRNA.

(A) Cartoon model of the *Tte* preQ$_1$ riboswitch system. In the absence of ligand, P2 is only partially formed, as indicated by our smFRET data, leaving the SD sequence partially exposed. This allows the 5' region of the anti-SD portion of the 16S rRNA to nucleate onto the SD sequence, resulting in a zippering effect along the remaining SD sequence, as indicated by the red arrow. In the presence of ligand, however, the P2 helix is stabilized, exposing a smaller portion of the SD sequence. The 16S rRNA is able to associate only briefly to the SD sequence, as the formation of P2 helix competes with 16S rRNA hybridization, indicated by the opposite direction of the red arrow. (B) Kinetic model of the *Tte* preQ$_1$ riboswitch system. In the absence of ligand, probe is free to bind to the riboswitch (R). When preQ$_1$ is added, the ligand binds to the free riboswitch, inducing a conformational change, which orders the binding pocket (early ligand binding model, from Chapter 2) to form the ligand-bound state (R*Q) in which the P2 helix is completely formed. Addition of preQ$_1$ shifts the equilibrium towards formation of the ligand-bound state, increasing the rate of dissociation (k_{off}) and decreasing the rate of association (k_{on}) of the probe from the ligand-bound state.

ligand; therefore, the residence time of the anti-SD probe is longer-lived. However, in the presence of ligand, the P2 helix, which comprises the first two nucleotides of the SD sequence from previously described structural data[42], is stabilized, decreasing the probability of nucleation of probe contacts with the SD sequence, causing more transient binding to the 3' region of the SD sequence, resulting in shorter binding lifetimes and a faster k_{off}.

Interestingly, the half saturation points ($K_{1/2}$) for k_{on} is ~10-fold greater than that of k_{off} (6.45 μM versus 0.55 μM, **Figure 4.8**). Since the change in k_{off} as a function of ligand concentration can be thought of as a readout of P2 helix stability, this implies that P2 helix formation occurs early in the riboswitch's response to fluctuating $preQ_1$ concentrations. This hypothesis is also supported by our Gō-model simulations, which suggest that ligand binding occurs early in the folding trajectory (**Figure 2.14B**). Once the P2 helix is completely formed in the presence of ligand, the SD:anti-SD interaction becomes less favorable as reflected in the increase in k_{off} (**Figure 4.12B**). This cooperative interaction between P2 helix formation and SD:anti-SD hybridization could be important for the riboswitch as a reversible switch in that it would allow for multiple rounds of genetic decisions per single mRNA.

The extensive contacts predicted by RNAstructure to form between the 3' end of the *T. tengcongensis* 16S rRNA and the SD sequence of the COG1564 5' UTR results in an unusually high Gibbs free energy of -12.0 kcal/mol. This could be due to the use of an alternative start codon, GUG for the COG1564 gene. Studies where the canonical AUG start codon was replaced by a non-canonical GUG start showed an ~8-fold reduction in translation efficiency[158]. Despite this reduction, 12% of genes in *T. tengcongensis* with a thermodynamically defined SD region located upstream from the start codon use GUG[148]. The unusually high free energy of the SD:anti-SD interaction in the COG1564 5' UTR may be due to its usage of a non-canonical

GUG start codon, as studies have shown that translational efficiency is increased when the SD-anti-SD interaction is stabilized in the presence of non-AUG start codons[151].

The COG1564 gene is currently annotated as a putative $preQ_1$ transporter; in contrast, the COG1563 downstream gene, whose SD sequence and UUG start codon overlaps with the 3' end of the COG1564 gene, encodes a 7-cyano-7-deazaguanine reductase, which catalyzes the NADPH-dependent reduction of $preQ_0$ to $preQ_1$. Previous studies on the transcriptional $preQ_1$ riboswitch from *Bacillus subtilis* have indicated that this riboswitch regulates the *queCDEF* operon, which is directly involved in the biosynthesis of $preQ_1$[115]. While we have shown through *in vitro* translation assays that the presence of COG1563 is not required for the expression of COG1564, further studies will need to be performed to understand if both genes are controlled by the $preQ_1$ riboswitch directly upstream of COG1564. Preliminary *in vitro* translation assays in the presence of $preQ_1$ ligand show inconclusive evidence as to the effect of ligand on the expression of both genes (data not shown). Nonetheless, the presence of these overlapping genes could provide the basis for regulation of $preQ_1$ biosynthesis in *T. tengcongensis* by a remarkably small non-coding RNA in the 5' UTR.

In this study, we have added the *T. tengcongensis* $preQ_1$ riboswitch to the relatively short list of translational riboswitches that can respond to fluctuating physiological conditions to render a genetic decision. It has been previously shown that the SAM-III riboswitch can reversibly switch from the ON and OFF states in the presence of competitor RNA and SAM ligand, respectively[59]. We have shown using our single molecule footprinting assay that the accessibility of the SD sequence changes in the presence and absence of $preQ_1$ within a short time frame in the presence of its full-length mRNA. The question still remains, however, how $preQ_1$ can navigate its way into the binding pocket, since the aptamer encapsulates ~98% of the

surface area of the ligand in the bound crystal structure[42]. Further studies need to be performed to understand the role of A14 in the ligand-free to ligand-bound conformational switch, as A14 was shown to occupy the $preQ_1$ binding site in the ligand-free structure. For this reason, this residue may be important in proper ligand binding. Nonetheless, the possibility of a reversible translational riboswitch may be a thrifty commodity for the cell as a single mRNA can initiate repeated decisions, in a manner that the energy consumed by the transcription of an entire mRNA transcript is not wasted.

Altogether, we have shown using a novel single molecule footprinting assay that the accessibility of the SD sequence of the $preQ_1$ translational riboswitch in *T. tengcongensis* to a small reporter probe mimicking the 3' end of the 16S rRNA decreases as a direct result of $preQ_1$ concentration. In contrast to previous studies on the mode of regulation by translational riboswitches, we have expanded our analysis to include the full-length mRNA in which the riboswitch resides. Bioinformatic analyses have revealed that this mRNA contains two overlapping reading frames, which may both be implicated in the biosynthesis of $preQ_1$ in *T. tengcongensis*. Our novel single molecule assay has the advantage over bulk techniques in that both binding and dissociation rate constants for the 16S rRNA mimic have been elucidated as a function of ligand concentration, and suggest a "zippering" mechanism for its binding to the SD sequence of the mRNA. Finally, we have shown that the riboswitch can respond to changing physiological conditions, and alter the accessibility of the SD sequence to effect gene regulation. This assay can be exploited in many ways to study the folding and unfolding of other biologically relevant molecules, and RNA in general.

4.5 Acknowledgements

We thank Dr. George Garcia for his generous gift of $preQ_1$, Paul Lund for his work with the *in vitro* translation assay, and Mario Blanco and Matt Kahlscheuer for help with the Matlab scripts used to analyze the single molecule data.

CHAPTER 5

Conclusions and Future Directions

The relatively recent discovery of non-coding RNA and the subsequent explosion in non-coding RNA research has brought forth a new era in biology. It was found that the majority of the genome that was once thought to be junk DNA actually encodes biologically relevant RNA molecules, which perform a variety of essential functions throughout the cell. This newly sparked interest in RNA research has unveiled multiple classes of RNA molecules, previously unknown to scientists. The discovery by Fire and Mello that the introduction of homologous small RNAs can suppress the expression of endogenous genes[159,160], which later turned out to be the discovery of the RNA interference pathway, was the first major breakthrough in the non-coding RNA field. Since then, the variety of known non-coding RNAs has drastically increased, including other regulators of gene expression, such as microRNAs[161] and riboswitches[13].

Riboswitches regulate gene expression through a growing number of mechanisms[162], the most common being through transcriptional attenuation or inhibition of translation initiation, both as a result of a single ligand binding event. While there is a growing wealth of structural information regarding the ligand-free and ligand-bound conformations of riboswitches, a challenging question remaining in the field is how this ligand binding event can confer a sometimes large-scale downstream conformational change to effect gene regulation. In this dissertation, we have combined single molecule FRET with molecular dynamics simulations and NMR spectroscopy to address the ligand-mediated folding of two structurally similar

transcriptional and translational $preQ_1$ riboswitches from *Bacillus subtilis* and *Thermoanaerobacter tengcongensis,* respectively. We also describe a novel single molecule touchdown assay, which we have adapted to study the accessibility of the Shine-Dalgarno sequence of two translationally-acting riboswitches, the *V. vulnificus* ADD riboswitch and the *T. tengcongensis* $preQ_1$ riboswitch. We show that this technique can provide important kinetic information regarding the occlusion of the SD sequence of structured mRNAs, and provide insight towards the complex mechanism of translational regulation by riboswitches.

5.1 Ligand-free conformations of riboswitches

There have been several reports of riboswitch crystal and NMR structures in the ligand-bound state, which have provided valuable insights into the ligand-recognition mechanisms utilized by both transcriptional and translational riboswitches. Much less common, however, is structural information regarding their respective ligand-free states, where only four structures have been reported[42,44,45,142], and of those, the only translational riboswitch is the $preQ_1$ riboswitch from *T. tengcongensis*. To address this scarcity of structural information, significant work has been undertaken to understand the ligand-free conformation using ensemble methods, such as NMR spectroscopy, 2-aminopurine fluorescence assays, and chemical probing techniques. In Chapter 2, we describe the ligand-free conformational ensembles of two structurally similar transcriptional and translational $preQ_1$ riboswitches, which have previously been suggested to adopt a largely unfolded conformation[109] and a loose pseudoknot[42], respectively. Using a combination of single molecule FRET and coarse-grained RNA simulations, we show that these two riboswitches actually adopt similar ligand-free ensembles, in which the A-rich 3' tail interacts segmentally with the P1-L1 stem-loop. Additionally, we present evidence that buffer

conditions can affect the tertiary structure of the transcriptional riboswitch, particularly the low pH and ionic strength commonly used in NMR experiments, which provides an explanation for the unobserved tertiary interactions in previous NMR studies[110,118].

Since transcriptional riboswitches have a shorter time window in which to bind ligand, it was unclear as to why the translational riboswitch would adopt a more pre-folded state than the transcriptional counterpart. Our results ameliorate this discrepancy, as we have shown both riboswitches to adopt similar, relatively compact pseudoknot structures, which closely resemble the ligand-bound structure. We show that only upon ligand binding does the P2 helix form completely and complete the pseudoknot structure. While these results address the ligand-free conformation of the *B. subtilis* $preQ_1$ riboswitch in an indirect manner, it is still important to solve its structure, similar to that of *T. tengcongensis*. A common difficulty in this practice, however, is the common necessity to separate the aptamer domain from the expression platform in transcriptional riboswitches. This can complicate the crystallography process, and does not provide complete information regarding the mechanism for transcriptional regulation. Nonetheless, our results provide an important foundation towards understanding the ligand-free conformations of riboswitches.

5.2 Ligand-mediated folding in riboswitches

One of the most prevalent and challenging questions to address in the riboswitch field is how a single ligand binding event can elicit a large-scale conformational change to a downstream structural element to confer a genetic decision. As mentioned earlier, structural studies on the ligand-bound and ligand-free conformational ensembles adopted by riboswitches have provided minimal understanding for ligand recognition; however, since the aptamer domain and

expression platform are usually separated from one another in these studies, the mechanism of genetic regulation is not always well understood. In Chapter 2, we provided a framework for the subtle dynamic differences utilized by two structurally similar transcriptional and translational riboswitches. As outlined previously, the ligand-bound structures have been shown to be comprised of quite similar compact pseudoknots[42,110], and we show the ligand-free conformations to adopt similar pseudoknots in which the P2 helix is only partially formed. Using a combination of smFRET and Gō-model simulations, we show that the main distinction between the *Bsu* and *Tte* riboswitches lies in their relative binding of the $preQ_1$ ligand along their folding trajectories. We further demonstrate that the *Bsu* riboswitch tends to form its contacts with the ligand concomitantly with other secondary and tertiary structural elements, while the *Tte* riboswitch forms contacts with $preQ_1$ initially, followed by a nucleation of secondary and tertiary contacts. These results suggest that the *Bsu* and *Tte* riboswitches show relative tendencies to fold via the canonical conformational selection and induced fit mechanisms, respectively. We also show that the ligand binding properties of the two riboswitches can be manipulated through the rational design of a single nucleotide swap in the 3' tails, which has been shown to be deleterious to stacking in the *Bsu* riboswitch[118] and thought to increase stacking in the *Tte* riboswitch. Indeed, the deleterious mutation results in an ~80-fold decrease in ligand binding affinity, accompanied with a slight opening of the pseudoknot in solution.

While these results make it tempting to speculate as to the advantages of one mechanism over the other with respect to their overall genetic modes of regulation, a more thorough investigation of the kinetic rate constants of folding in response to ligand concentrations must be conducted. We believe that single molecule techniques are the optimal choice for this type of analysis, since they can provide rate constants for both the docking and undocking of the 3' tail

of the riboswitches, whereas only an observed rate constant can be derived from more classical ensemble techniques.

Riboswitches are endogenous, gene regulatory elements located in the 5' UTRs of bacterial mRNAs, and have been shown to regulate ~3% of all bacterial genes. Their ability to regulate the expression of essential genes without the need for proteins or exogenous factors (unlike RNAi therapeutics) makes them attractive drug targets. For example, not long ago, it was demonstrated that the antimicrobial roseoflavin, which inhibits the growth of several Gram-positive bacteria[163], directly targets the FMN riboswitch[164]. In these bacteria, all the genes involved in riboflavin biosynthesis are under the control of this single riboswitch[165,166]. Since riboswitches are almost exclusively found in bacteria, with the exception of the TPP riboswitch[167], it is easy to believe that they could represent appealing antimicrobial targets. By understanding the mechanisms by which riboswitches recognize their cognate ligands, the rational design of antibiotics could be accomplished.

5.3 Demonstrating translational regulation by riboswitches

Translational riboswitches regulate gene expression most commonly through the occlusion of the Shine-Dalgarno (SD) sequence via Watson-Crick base-pairing as a direct consequence of ligand binding (as in the *T. tengcongensis* $preQ_1$ riboswitch) or release (as in the *V. vulnificus* ADD riboswitch). Several structural studies have demonstrated the occlusion of the SD sequence in respective crystal structures[42,47,48,60], and some even show direct ligand recognition by nucleotides of the SD sequence[48]. However, direct evidence of an effect on ribosomal binding to riboswitch-embedded mRNA as a function of ligand concentration has only been demonstrated in the SAM-III riboswitch[62]. To this end, we have developed a single molecule assay that

directly probes the accessibility of the SD sequence as a function of ligand concentration. We utilize the ADD riboswitch from *V. vulnificus* as a model to demonstrate the feasibility of the assay, and finally, show that the accessibility of the SD sequence of the *T. tengcongensis* $preQ_1$ riboswitch decreases with ligand concentration. Importantly, we perform these assays in the context of the full-length mRNA, which could provide further information towards gene regulation by translational riboswitches. Similar experiments should be performed with truncated mRNAs to determine any artifacts that may exist from the large mRNA used in this study. In addition to probing the accessibility of the SD sequence, we have revealed a more complex mechanism for SD occlusion through analysis of the rate constants for dissociation of our oligonucleotide probe. Taken together with our results from Chapter 2, we show a competing effect between probe binding and P2 helix formation, where the P2 helix is stabilized in the presence of ligand, and only allows for more transient binding events to occur by the probe. This is demonstrated as an increase in k_{off} as a function of $preQ_1$ concentration.

Recently, we have developed a strategy in the lab to fluorescently label the 30S ribosomal protein S6 (rpS6) carrying a D41C mutation with Cy5-maleimide. The labeled protein is then reconstituted into 30S subunits which lack $rpS6^{168}$. We have preliminary data showing colocalization between reconstituted 30S subunits, initiation factors, tRNA and fluorescently labeled *Tte* mRNA truncations. The acquisition of these data suggests that single molecule footprinting of the *Tte* mRNA with reconstituted 30S subunits is not too far off. A final piece to the puzzle would be to show a $preQ_1$ dependent decrease in protein expression via our *in vitro* translation assay. These experiments, in combination with results laid out in this dissertation would provide an intricate and dynamic picture of the delicate genetic regulation mechanism by

the *T. tengcongensis* $preQ_1$ riboswitch that cannot be achieved through traditional ensemble measurements.

References

(1) Crick, F. H. *Symposia of the Society for Experimental Biology* **1958**, *12*, 138.

(2) Brachet, J.; Chantrenne, H. *Cold Spring Harbor symposia on quantitative biology* **1956**, *21*, 329.

(3) Brenner, S.; Jacob, F.; Meselson, M. *Nature* **1961**, *190*, 576.

(4) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell* **1981**, *27*, 487.

(5) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849.

(6) Hansen, J. L.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 11670.

(7) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 920.

(8) Hoagland, M. B.; Stephenson, M. L.; Scott, J. F.; Hecht, L. I.; Zamecnik, P. C. *The Journal of biological chemistry* **1958**, *231*, 241.

(9) Eddy, S. R. *Nature reviews. Genetics* **2001**, *2*, 919.

(10) Mattick, J. S.; Makunin, I. V. *Human molecular genetics* **2006**, *15 Spec No 1*, R17.

(11) Huttenhofer, A.; Schattner, P.; Polacek, N. *Trends in genetics : TIG* **2005**, *21*, 289.

(12) Brodersen, P.; Voinnet, O. *Nature reviews. Molecular cell biology* **2009**, *10*, 141.

(13) Roth, A.; Breaker, R. R. *Annual review of biochemistry* **2009**, *78*, 305.

(14) Steitz, T. A.; Moore, P. B. *Trends in biochemical sciences* **2003**, *28*, 411.

(15) Wahl, M. C.; Will, C. L.; Luhrmann, R. *Cell* **2009**, *136*, 701.

(16) Collins, K. *Nature reviews. Molecular cell biology* **2006**, *7*, 484.

(17) Thirumalai, D.; Lee, N.; Woodson, S. A.; Klimov, D. *Annual review of physical chemistry* **2001**, *52*, 751.

(18) Tinoco, I., Jr.; Bustamante, C. *Journal of molecular biology* **1999**, *293*, 271.

(19) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.

(20) Joyce, G. F. *Current opinion in structural biology* **1994**, *4*, 331.

(21) Grundy, F. J.; Rollins, S. M.; Henkin, T. M. *Journal of bacteriology* **1994**, *176*, 4518.

(22) Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952.

(23) Batey, R. T.; Gilbert, S. D.; Montange, R. K. *Nature* **2004**, *432*, 411.

(24) Mandal, M.; Breaker, R. R. *Nature structural & molecular biology* **2004**, *11*, 29.

(25) Blount, K. F.; Wang, J. X.; Lim, J.; Sudarsan, N.; Breaker, R. R. *Nature chemical biology* **2007**, *3*, 44.

(26) Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R. *Science* **2004**, *306*, 275.

(27) Barrick, J. E.; Breaker, R. R. *Genome biology* **2007**, *8*, R239.

(28) Sudarsan, N.; Hammond, M. C.; Block, K. F.; Welz, R.; Barrick, J. E.; Roth, A.; Breaker, R. R. *Science* **2006**, *314*, 300.

(29) Nechooshtan, G.; Elgrably-Weiss, M.; Sheaffer, A.; Westhof, E.; Altuvia, S. *Genes & development* **2009**, *23*, 2650.

(30) Kortmann, J.; Sczodrok, S.; Rinnenthal, J.; Schwalbe, H.; Narberhaus, F. *Nucleic acids research* **2011**, *39*, 2855.

(31) Dann, C. E., 3rd; Wakeman, C. A.; Sieling, C. L.; Baker, S. C.; Irnov, I.; Winkler, W. C. *Cell* **2007**, *130*, 878.

(32) Regulski, E. E.; Moy, R. H.; Weinberg, Z.; Barrick, J. E.; Yao, Z.; Ruzzo, W. L.; Breaker, R. R. *Molecular microbiology* **2008**, *68*, 918.

(33) Baker, J. L.; Sudarsan, N.; Weinberg, Z.; Roth, A.; Stockbridge, R. B.; Breaker, R. R. *Science* **2012**, *335*, 233.

(34) Wachter, A.; Tunc-Ozdemir, M.; Grove, B. C.; Green, P. J.; Shintani, D. K.; Breaker, R. R. *The Plant cell* **2007**, *19*, 3437.

(35) Cheah, M. T.; Wachter, A.; Sudarsan, N.; Breaker, R. R. *Nature* **2007**, *447*, 497.

(36) Thore, S.; Leibundgut, M.; Ban, N. *Science* **2006**, *312*, 1208.

(37) Caron, M. P.; Bastet, L.; Lussier, A.; Simoneau-Roy, M.; Masse, E.; Lafontaine, D. A. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, E3444.

(38) Collins, J. A.; Irnov, I.; Baker, S.; Winkler, W. C. *Genes & development* **2007**, *21*, 3356.

(39) Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. *Nature* **2004**, *428*, 281.

(40) Chen, A. G.; Sudarsan, N.; Breaker, R. R. *Rna* **2011**, *17*, 1967.

(41) Haller, A.; Rieder, U.; Aigner, M.; Blanchard, S. C.; Micura, R. *Nature chemical biology* **2011**, *7*, 393.

(42) Jenkins, J. L.; Krucinska, J.; McCarty, R. M.; Bandarian, V.; Wedekind, J. E. *The Journal of biological chemistry* **2011**, *286*, 24626.

(43) Santner, T.; Rieder, U.; Kreutz, C.; Micura, R. *Journal of the American Chemical Society* **2012**, *134*, 11928.

(44) Serganov, A.; Huang, L.; Patel, D. J. *Nature* **2008**, *455*, 1263.

(45) Stoddard, C. D.; Montange, R. K.; Hennelly, S. P.; Rambo, R. P.; Sanbonmatsu, K. Y.; Batey, R. T. *Structure* **2010**, *18*, 787.

(46) Vicens, Q.; Mondragon, E.; Batey, R. T. *Nucleic acids research* **2011**, *39*, 8586.

(47) Gilbert, S. D.; Rambo, R. P.; Van Tyne, D.; Batey, R. T. *Nature structural & molecular biology* **2008**, *15*, 177.

(48) Lu, C.; Smith, A. M.; Fuchs, R. T.; Ding, F.; Rajashankar, K.; Henkin, T. M.; Ke, A. *Nature structural & molecular biology* **2008**, *15*, 1076.

(49) Lemay, J. F.; Desnoyers, G.; Blouin, S.; Heppell, B.; Bastet, L.; St-Pierre, P.; Masse, E.; Lafontaine, D. A. *PLoS genetics* **2011**, *7*, e1001278.

(50) Rieder, R.; Lang, K.; Graber, D.; Micura, R. *Chembiochem : a European journal of chemical biology* **2007**, *8*, 896.

(51) Gallo, S.; Oberhuber, M.; Sigel, R. K.; Krautler, B. *Chembiochem : a European journal of chemical biology* **2008**, *9*, 1408.

(52) Lu, C.; Smith, A. M.; Ding, F.; Chowdhury, A.; Henkin, T. M.; Ke, A. *Journal of molecular biology* **2011**, *409*, 786.

(53) Wilson, R. C.; Smith, A. M.; Fuchs, R. T.; Kleckner, I. R.; Henkin, T. M.; Foster, M. P. *Journal of molecular biology* **2011**, *405*, 926.

(54) Chen, B.; Zuo, X.; Wang, Y. X.; Dayie, T. K. *Nucleic acids research* **2012**, *40*, 3117.

(55) Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J. *Chemistry & biology* **2004**, *11*, 1729.

(56) Anthony, P. C.; Perez, C. F.; Garcia-Garcia, C.; Block, S. M. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 1485.

(57) Neupane, K.; Yu, H.; Foster, D. A.; Wang, F.; Woodside, M. T. *Nucleic acids research* **2011**, *39*, 7677.

(58) Haller, A.; Altman, R. B.; Souliere, M. F.; Blanchard, S. C.; Micura, R. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 4188.

(59) Smith, A. M.; Fuchs, R. T.; Grundy, F. J.; Henkin, T. M. *Molecular microbiology* **2010**, *78*, 1393.

(60) Johnson, J. E., Jr.; Reyes, F. E.; Polaski, J. T.; Batey, R. T. *Nature* **2012**, *492*, 133.

(61) Fuchs, R. T.; Grundy, F. J.; Henkin, T. M. *Nature structural & molecular biology* **2006**, *13*, 226.

(62) Fuchs, R. T.; Grundy, F. J.; Henkin, T. M. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 4876.

(63) Fredrick, K.; Noller, H. F. *Molecular cell* **2002**, *9*, 1125.

(64) Joo, C.; Balci, H.; Ishitsuka, Y.; Buranachai, C.; Ha, T. *Annual review of biochemistry* **2008**, *77*, 51.

(65) Walter, N. G.; Huang, C. Y.; Manzo, A. J.; Sobhy, M. A. *Nature methods* **2008**, *5*, 475.

(66) Axelrod, D. *Methods in enzymology* **2003**, *361*, 1.

(67) Haustein, E.; Schwille, P. *Annual review of biophysics and biomolecular structure* **2007**, *36*, 151.

(68) Churchman, L. S.; Okten, Z.; Rock, R. S.; Dawson, J. F.; Spudich, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 1419.

(69) Lund, K.; Manzo, A. J.; Dabby, N.; Michelotti, N.; Johnson-Buck, A.; Nangreave, J.; Taylor, S.; Pei, R.; Stojanovic, M. N.; Walter, N. G.; Winfree, E.; Yan, H. *Nature* **2010**, *465*, 206.

(70) Manley, S.; Gillette, J. M.; Patterson, G. H.; Shroff, H.; Hess, H. F.; Betzig, E.; Lippincott-Schwartz, J. *Nature methods* **2008**, *5*, 155.

(71) Sharonov, A.; Hochstrasser, R. M. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 18911.

(72) Ashkin, A. *Physical Review Letters* **1970**, *24*, 156.

(73) Binnig, G.; Rohrer, H.; Gerber, C.; Weibel, E. *Physical Review Letters* **1982**, *49*, 57.

(74) Charvin, G.; Strick, T. R.; Bensimon, D.; Croquette, V. *Annual review of biophysics and biomolecular structure* **2005**, *34*, 201.

(75) Roy, R.; Hohng, S.; Ha, T. *Nature methods* **2008**, *5*, 507.

(76) Schuler, B.; Eaton, W. A. *Current opinion in structural biology* **2008**, *18*, 16.
(77) Aleman, E. A.; Lamichhane, R.; Rueda, D. *Current opinion in chemical biology* **2008**, *12*, 647.
(78) Bokinsky, G.; Rueda, D.; Misra, V. K.; Rhodes, M. M.; Gordus, A.; Babcock, H. P.; Walter, N. G.; Zhuang, X. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 9302.
(79) de Silva, C.; Walter, N. G. *Rna* **2009**, *15*, 76.
(80) Ditzler, M. A.; Rueda, D.; Mo, J.; Hakansson, K.; Walter, N. G. *Nucleic acids research* **2008**, *36*, 7088.
(81) Lemay, J. F.; Penedo, J. C.; Tremblay, R.; Lilley, D. M.; Lafontaine, D. A. *Chemistry & biology* **2006**, *13*, 857.
(82) Liu, S.; Bokinsky, G.; Walter, N. G.; Zhuang, X. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 12634.
(83) McDowell, S. E.; Jun, J. M.; Walter, N. G. *Rna* **2010**, *16*, 2414.
(84) Pereira, M. J.; Nikolova, E. N.; Hiley, S. L.; Jaikaran, D.; Collins, R. A.; Walter, N. G. *Journal of molecular biology* **2008**, *382*, 496.
(85) Pljevaljcic, G.; Robertson-Anderson, R.; van der Schans, E.; Millar, D. *Methods in molecular biology* **2012**, *875*, 271.
(86) Rueda, D.; Bokinsky, G.; Rhodes, M. M.; Rust, M. J.; Zhuang, X.; Walter, N. G. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 10066.
(87) Steiner, M.; Karunatilaka, K. S.; Sigel, R. K.; Rueda, D. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 13853.
(88) Tremblay, R.; Lemay, J. F.; Blouin, S.; Mulhbacher, J.; Bonneau, E.; Legault, P.; Dupont, P.; Penedo, J. C.; Lafontaine, D. A. *The Journal of biological chemistry* **2011**, *286*, 27406.
(89) Zhuang, X.; Kim, H.; Pereira, M. J.; Babcock, H. P.; Walter, N. G.; Chu, S. *Science* **2002**, *296*, 1473.
(90) Aitken, C. E.; Marshall, R. A.; Puglisi, J. D. *Biophysical journal* **2008**, *94*, 1826.
(91) Rasnik, I.; McKinney, S. A.; Ha, T. *Nature methods* **2006**, *3*, 891.
(92) Abelson, J.; Blanco, M.; Ditzler, M. A.; Fuller, F.; Aravamudhan, P.; Wood, M.; Villa, T.; Ryan, D. E.; Pleiss, J. A.; Maeder, C.; Guthrie, C.; Walter, N. G. *Nature structural & molecular biology* **2010**, *17*, 504.
(93) Blanco, M.; Walter, N. G. *Methods in enzymology* **2010**, *472*, 153.
(94) Breaker, R. R. *Molecular cell* **2011**, *43*, 867.
(95) Mandal, M.; Breaker, R. R. *Nature reviews. Molecular cell biology* **2004**, *5*, 451.
(96) Nudler, E.; Mironov, A. S. *Trends in biochemical sciences* **2004**, *29*, 11.
(97) Winkler, W. C.; Breaker, R. R. *Annual review of microbiology* **2005**, *59*, 487.
(98) Leontis, N. B.; Westhof, E. *Current opinion in structural biology* **2003**, *13*, 300.
(99) Kuchino, Y.; Kasai, H.; Nihei, K.; Nishimura, S. *Nucleic acids research* **1976**, *3*, 393.
(100) Okada, N.; Noguchi, S.; Nishimura, S.; Ohgi, T.; Goto, T.; Crain, P. F.; McCloskey, J. A. *Nucleic acids research* **1978**, *5*, 2289.
(101) Harada, F.; Nishimura, S. *Biochemistry* **1972**, *11*, 301.
(102) Bienz, M.; Kubli, E. **1981**.

(103) Meier, F.; Suter, B.; Grosjean, H.; Keith, G.; Kubli, E. *The EMBO journal* **1985**, *4*, 823.

(104) Urbonavicius, J.; Qian, Q.; Durand, J. M.; Hagervall, T. G.; Bjork, G. R. *The EMBO journal* **2001**, *20*, 4863.

(105) Durand, J. M.; Okada, N.; Tobe, T.; Watarai, M.; Fukuda, I.; Suzuki, T.; Nakata, N.; Komatsu, K.; Yoshikawa, M.; Sasakawa, C. *Journal of bacteriology* **1994**, *176*, 4627.

(106) Banas, P.; Walter, N. G.; Sponer, J.; Otyepka, M. *The journal of physical chemistry. B* **2010**, *114*, 8701.

(107) Feng, J.; Walter, N. G.; Brooks, C. L., 3rd *Journal of the American Chemical Society* **2011**, *133*, 4196.

(108) Gong, Z.; Zhao, Y.; Chen, C.; Xiao, Y. *PloS one* **2012**, *7*, e45239.

(109) Kang, M.; Peterson, R.; Feigon, J. *Molecular cell* **2009**, *33*, 784.

(110) Klein, D. J.; Edwards, T. E.; Ferre-D'Amare, A. R. *Nature structural & molecular biology* **2009**, *16*, 343.

(111) Petrone, P. M.; Dewhurst, J.; Tommasi, R.; Whitehead, L.; Pomerantz, A. K. *Journal of molecular graphics & modelling* **2011**, *30*, 179.

(112) Rieder, U.; Kreutz, C.; Micura, R. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 10804.

(113) Rieder, U.; Lang, K.; Kreutz, C.; Polacek, N.; Micura, R. *Chembiochem : a European journal of chemical biology* **2009**, *10*, 1141.

(114) Zhang, Q.; Kang, M.; Peterson, R. D.; Feigon, J. *Journal of the American Chemical Society* **2011**, *133*, 5190.

(115) Roth, A.; Winkler, W. C.; Regulski, E. E.; Lee, B. W.; Lim, J.; Jona, I.; Barrick, J. E.; Ritwik, A.; Kim, J. N.; Welz, R.; Iwata-Reuyl, D.; Breaker, R. R. *Nature structural & molecular biology* **2007**, *14*, 308.

(116) Spitale, R. C.; Torelli, A. T.; Krucinska, J.; Bandarian, V.; Wedekind, J. E. *The Journal of biological chemistry* **2009**, *284*, 11012.

(117) Capriotti, E.; Marti-Renom, M. A. *Bioinformatics* **2008**, *24*, i112.

(118) Eichhorn, C. D.; Feng, J.; Suddala, K. C.; Walter, N. G.; Brooks, C. L., 3rd; Al-Hashimi, H. M. *Nucleic acids research* **2012**, *40*, 1345.

(119) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *Journal of computational chemistry* **2009**, *30*, 1545.

(120) Feig, M.; Karanicolas, J.; Brooks, C. L., 3rd *Journal of molecular graphics & modelling* **2004**, *22*, 377.

(121) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 2754.

(122) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *Journal of computational chemistry* **2005**, *26*, 1701.

(123) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *Journal of biomolecular NMR* **1995**, *6*, 277.

(124) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. *Science* **2006**, *311*, 653.

(125) Ragunathan, K.; Liu, C.; Ha, T. *eLife* **2012**, *1*, e00067.
(126) Sekella, P. T.; Rueda, D.; Walter, N. G. *Rna* **2002**, *8*, 1242.
(127) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *Journal of molecular biology* **2000**, *298*, 937.
(128) Hills, R. D., Jr.; Brooks, C. L., 3rd *International journal of molecular sciences* **2009**, *10*, 889.
(129) Banas, P.; Sklenovsky, P.; Wedekind, J. E.; Sponer, J.; Otyepka, M. *The journal of physical chemistry. B* **2012**, *116*, 12721.
(130) Hammes, G. G.; Chang, Y. C.; Oas, T. G. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 13737.
(131) Hoops, G. C.; Townsend, L. B.; Garcia, G. A. *Biochemistry* **1995**, *34*, 15381.
(132) Nickson, A. A.; Clarke, J. *Methods* **2010**, *52*, 38.
(133) Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. *Cell* **2003**, *113*, 577.
(134) Gilbert, S. D.; Stoddard, C. D.; Wise, S. J.; Batey, R. T. *Journal of molecular biology* **2006**, *359*, 754.
(135) Wong, T. N.; Sosnick, T. R.; Pan, T. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 17995.
(136) Al-Hashimi, H. M.; Walter, N. G. *Current opinion in structural biology* **2008**, *18*, 321.
(137) Bhaskaran, H.; Russell, R. *Nature* **2007**, *449*, 1014.
(138) Russell, R. *Frontiers in bioscience : a journal and virtual library* **2008**, *13*, 1.
(139) Uhlenbeck, O. C. *Biopolymers* **2009**, *91*, 811.
(140) Johnson-Buck, A.; Nangreave, J.; Kim, D. N.; Bathe, M.; Yan, H.; Walter, N. G. *Nano letters* **2013**, *13*, 728.
(141) Jungmann, R.; Steinhauer, C.; Scheible, M.; Kuzyk, A.; Tinnefeld, P.; Simmel, F. C. *Nano letters* **2010**, *10*, 4756.
(142) Huang, L.; Serganov, A.; Patel, D. J. *Molecular cell* **2010**, *40*, 774.
(143) Nilsson, W. B.; Paranjype, R. N.; DePaola, A.; Strom, M. S. *Journal of clinical microbiology* **2003**, *41*, 442.
(144) Qin, P. Z.; Pyle, A. M. *Methods* **1999**, *18*, 60.
(145) Shine, J.; Dalgarno, L. *Proceedings of the National Academy of Sciences of the United States of America* **1974**, *71*, 1342.
(146) Hui, A.; de Boer, H. A. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 4762.
(147) Jacob, W. F.; Santer, M.; Dahlberg, A. E. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 4757.
(148) Starmer, J.; Stomp, A.; Vouk, M.; Bitzer, D. *PLoS computational biology* **2006**, *2*, e57.
(149) Chen, H.; Bjerknes, M.; Kumar, R.; Jay, E. *Nucleic acids research* **1994**, *22*, 4953.
(150) Kaminishi, T.; Wilson, D. N.; Takemoto, C.; Harms, J. M.; Kawazoe, M.; Schluenzen, F.; Hanawa-Suetsugu, K.; Shirouzu, M.; Fucini, P.; Yokoyama, S. *Structure* **2007**, *15*, 289.
(151) Weyens, G.; Charlier, D.; Roovers, M.; Pierard, A.; Glansdorff, N. *Journal of molecular biology* **1988**, *204*, 1045.

(152) de Smit, M. H.; van Duin, J. *Journal of molecular biology* **1994**, *235*, 173.
(153) Munson, L. M.; Stormo, G. D.; Niece, R. L.; Reznikoff, W. S. *Journal of molecular biology* **1984**, *177*, 663.
(154) Schauder, B.; McCarthy, J. E. *Gene* **1989**, *78*, 59.
(155) Meyer, M. M.; Roth, A.; Chervin, S. M.; Garcia, G. A.; Breaker, R. R. *Rna* **2008**, *14*, 685.
(156) Reuter, J. S.; Mathews, D. H. *BMC bioinformatics* **2010**, *11*, 129.
(157) Keiler, K. C. *Annual review of microbiology* **2008**, *62*, 133.
(158) Sussman, J. K.; Simons, E. L.; Simons, R. W. *Molecular microbiology* **1996**, *21*, 347.
(159) Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. *Nature* **1998**, *391*, 806.
(160) Timmons, L.; Fire, A. *Nature* **1998**, *395*, 854.
(161) Bartel, D. P. *Cell* **2009**, *136*, 215.
(162) Serganov, A.; Nudler, E. *Cell* **2013**, *152*, 17.
(163) Otani, S.; Takatsu, M.; Nakano, M.; Kasai, S.; Miura, R. *The Journal of antibiotics* **1974**, *27*, 86.
(164) Serganov, A.; Huang, L.; Patel, D. J. *Nature* **2009**, *458*, 233.
(165) Gelfand, M. S.; Mironov, A. A.; Jomantas, J.; Kozlov, Y. I.; Perumov, D. A. *Trends in genetics : TIG* **1999**, *15*, 439.
(166) Winkler, W. C.; Cohen-Chalamish, S.; Breaker, R. R. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 15908.
(167) Li, S.; Breaker, R. R. *Nucleic acids research* **2013**, *41*, 3022.
(168) Ermolenko, D. N.; Majumdar, Z. K.; Hickerson, R. P.; Spiegel, P. C.; Clegg, R. M.; Noller, H. F. *Journal of molecular biology* **2007**, *370*, 530.