A Tale of Two RNAs: Single Molecule Investigation of the Conformation, Dynamics and Ligand Binding to the $PreQ_1$ and T-box Riboswitches

by
Krishna Chaitanya Suddala

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in the University of Michigan
2014

Doctoral Committee:

Professor Nils G. Walter, Chair
Professor Hashim Al-Hashimi, Duke University
Professor Charles L. Brooks III
Assistant Professor Wei Cheng
Associate Professor Mark Saper

2014

Dedication

To my late grandfather who supported our family in many ways

Acknowledgements

Completing my doctoral research and writing up this thesis has been an exciting and beautiful journey of my career. I am indebted to the University of Michigan, the Biophysics program, the Walter lab and to the many people who have helped me achieve this milestone. I would like to first express my gratitude towards my advisor, Nils Walter, for his encouragement, support and guidance throughout the course of this dissertation. I am very much grateful to my parents, sister, my wife and all my family and friends for their love and encouragement all through my dissertation research. I am thankful to my committee members and collaborators for their guidance and for providing me with ideas and materials that went into this dissertation. I want to thank all the past and present members of the Walter lab who I interacted with, for creating a fun and stimulating atmosphere for doing science. I am grateful to all my teachers, lecturers and professors who taught me during my schooling, undergraduate education at Vellore Institute of Technology and in my Masters at Indian Institute of Technology, Kanpur. I cannot be but indebted to my wife Surma Talapatra, who made my time in grad school a memorable one. She is the biggest prize of my grad school, of course, along with all the knowledge that went into this dissertation ☺. I am also thankful to all the professors who gave me an opportunity to conduct research under them that helped keep my interest in science alive. I would specially like to thank Prof. Sankararamakrishnan, Prof. Kalluri Subba Rao and late Dr. Anupam Pal, who motivated and assisted me during my early years in research to pursue a PhD. I believe that their encouragement and recommendations helped me in getting admitted to the graduate school in this great University. Go Blue!

Table of Contents

CHAPTER 2 Single transcriptional and translational preQ$_1$ riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure

List of Figures

List of Tables

Abstract

Riboswitches are structured mRNA domains that can bind cellular metabolites and control gene expression of downstream genes mainly via transcription attenuation or inhibition of translation initiation. Although structures of many ligand-bound riboswitches are available, knowledge on their ligand-free conformations is scarce. Subsequently, the ligand-mediated folding process of riboswitches is poorly understood. In this dissertation, we used single molecule FRET to investigate the conformation and ligand binding properties of two very distinct riboswitches. We showed that, contrary to previous studies, the structurally similar but functionally different preQ$_1$ riboswitches from *B. subtilis* (*Bsu*) and *T. tencongensis* (*Tte*) have similar conformational ensemble in their ligand-free state with only subtle differences in their dynamics. Our smFRET data in combination with computational simulations suggested that both the riboswitches adopt ligand-free 'pre-folded' conformations and fold through distinct pathways that are similar to the conformational selection and induced fit mechanisms, respectively. We also demonstrated how remote mutations can affect the ligand binding affinities of riboswitches. Later, using smFRET, we probe the effect of various ligands on the kinetics of the *Bsu* riboswitch conformational dynamics with an aim to dissect its ligand binding mechanism. Our data suggest that the *Bsu* riboswitch can fold through both induced fit or conformational selection pathways, the relative extent of which is dependent on the presence of Mg^{2+}. The T-box riboswitch is one of the complex riboswitches that binds tRNA and controls gene expression by sensing the relative levels of charged and uncharged tRNA. The structure of a T-box riboswitch stem-I:tRNA complex was recently solved, but it lacks the important genetic regulatory domain. By using various designs of the *glyQS* T-box riboswitch, we have studied the global conformation of the full T-box riboswitch and estimated distances between different regions. We measured tRNA binding kinetics to different T-box variants and showed that the double T-loop motif only contributes modestly to decrease the tRNA dissociation rate. Further, we directly demonstrated that the presence of glycine increases the tRNA dissociation rate ~6-fold that forms the basis of T-box riboswitch mechanism. Based on our kinetic data, we propose an improved kinetic model of the T-box riboswitch function.

CHAPTER 1
Introduction

1.1 RNA is essential for life

1.1.1 The rise of RNA

The interest in understanding the chemical and physical basis of life led early chemists to the study of biological macromolecules, mainly proteins, for a major part of the 19th century until the early 20th century. Although DNA (known as 'nuclein') was discovered by Friedrich Miescher in 1868, its chemical composition and its connection to genes was not known for a long time until the early 1940s (1). At the same time, a few biochemists proposed the importance of RNA (then referred to as pentose nucleic acid) from their observation of large amounts of RNA in tissues containing high protein. However, the link between the two was unclear at that time. The main breakthrough came with the Watson-Crick structure of DNA in 1953 that later helped in establishing the link between DNA, RNA and protein in the 'central dogma of biology'. This led to the birth of molecular biology and a view of 'the molecular basis of life' - that life results from an intricate interaction network of chemically diverse macromolecules. Biological macromolecules, mainly protein and DNA, were interrogated with great interest to understand their functional roles and decipher their physical and chemical structures. However, studies on the function of RNA were rare and mainly limited to messenger RNA (mRNA). For a long time, the cellular functions of RNA were assumed to be confined to its roles in gene expression, primarily as mRNA, ribosomal RNA (rRNA) and transfer RNA (tRNA) (1). However, the discovery of RNAs with

---

Portions of this chapter will be part of a review Suddala, K. C and Walter, N. G. Riboswitch structure and dynamics by single molecule FRET, *Methods in Enzymology* (to be submitted)

enzymatic activity, known as ribozymes, by Thomas Cech and Sidney Altman in 1982 completely changed this existing view and added support to the RNA World hypothesis proposed independently by Carl Woese, Francis Crick andLeslie Orgel in the late 1960s (1). This hypothesis states that RNA was the genetic material and carried out essential enzymatic/housekeeping functions in early life forms (2). This idea further gained importance with the discovery of numerous other small ribozymes, as well as from the realization that the ribosome, the essential cellular machinery involved in protein synthesis, is also a ribozyme (2). In later years, a plethora of non-coding RNAs (ncRNAs) performing diverse functions in organisms ranging from simple archaea and eubacteria to complex multi-cellular animals were identified, further emphasizing the versatile nature of RNA in carrying out complex cellular functions (1,2).

1.1.2 The many roles of non-coding RNA (ncRNA)

Genome sequencing projects have revealed that only less than ~2% of the human genome codes for ~21,000 protein genes. We now know that the remaining over 98% of the human genome is transcribed into many classes of ncRNAs that are mainly involved in the regulation of gene expression at multiple levels. The high percentage of genome coding for ncRNAs has been suggested to be one of the main reasons for the complexity of higher organisms such as mammals as compared to smaller organisms, even though the number of protein coding genes are not very different (3). In eukaryotes, various kinds of ncRNAs such as miroRNAs (miRNAs), small interfering RNAs (siRNAs), PIWI interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs) and long ncRNAs are involved in gene regulation, and their malfunctioning is implicated in many diseases including cancer (4,5). Our knowledge of the various functions of ncRNAs in eukaryotes is expanding rapidly and newer classes of ncRNAs are being discovered at an increasing rate.

Bacteria and archaea are found in diverse environments and are extremely adaptable to frequent fluctuations in external physicochemical forces. By modulating their metabolic and physiological states, they can survive under environmental and nutritional stress conditions (6). This adaptability is achieved by regulation of gene expression in response to

both internal and external signals. Although less complex than eukaryotic organisms, bacteria also contain a large network of ncRNAs that perform diverse functions (6-9). The recent discovery of an RNA based viral defense system in bacteria, termed CRISPR (clustered regularly interspaced short palindromic repeats) system, highlights the many important roles played by ncRNAs even in the simplest forms of life (10,11). A majority of the ncRNAs in bacteria, such as the well-studied small regulatory RNAs (sRNAs), act in *trans*, where they bind to mRNAs and either cause RNase E mediated degradation or inhibit translation initiation by sequestering the Shine-Dalgarno (SD) sequence needed by the ribosome, with the help of accessory proteins such as Hfq (9). In contrast, bacteria also contain many *cis*-acting elements known as riboswitches that can sense intracellular metabolite concentrations and regulate expression of downstream genes without the need for protein cofactors (12,13).

1.1.3 Gene regulation by riboswitches

Riboswitches are structured domains, present usually in the 5'-untranslated region (5'-UTR) of certain mRNAs that regulate gene expression in response to a physiological signal (13-16). This physiological signal is generally a change in the concentration of a metabolite; but riboswitches that sense temperature, pH and metal ions have also been discovered (13,14). Many different classes of riboswitches have been identified that can bind nucleobases and their derivatives(17), amino acids, coenzymes, second messengers and metal ions (14,18). Almost all of these riboswitches are present in bacteria, although one class of riboswitches that binds to the essential coenzyme thiamine pyrophosphate (TPP) was identified in all forms of life, including archaea, fungi and plants (14). Riboswitches control gene expression generally by inhibiting transcription elongation or translation initiation, but certain TPP riboswitches in plants and fungi function through altering mRNA stability or by modulating mRNA splicing (13,14,19). Riboswitches consist of an aptamer domain that is involved in ligand sensing, followed by an expression platform (or gene regulatory element) that undergoes a structural change in response to ligand binding by the aptamer. Both domains share a common sequence referred to as the 'switching' sequence that communicates the ligand bound state of the aptamer domain to the expression

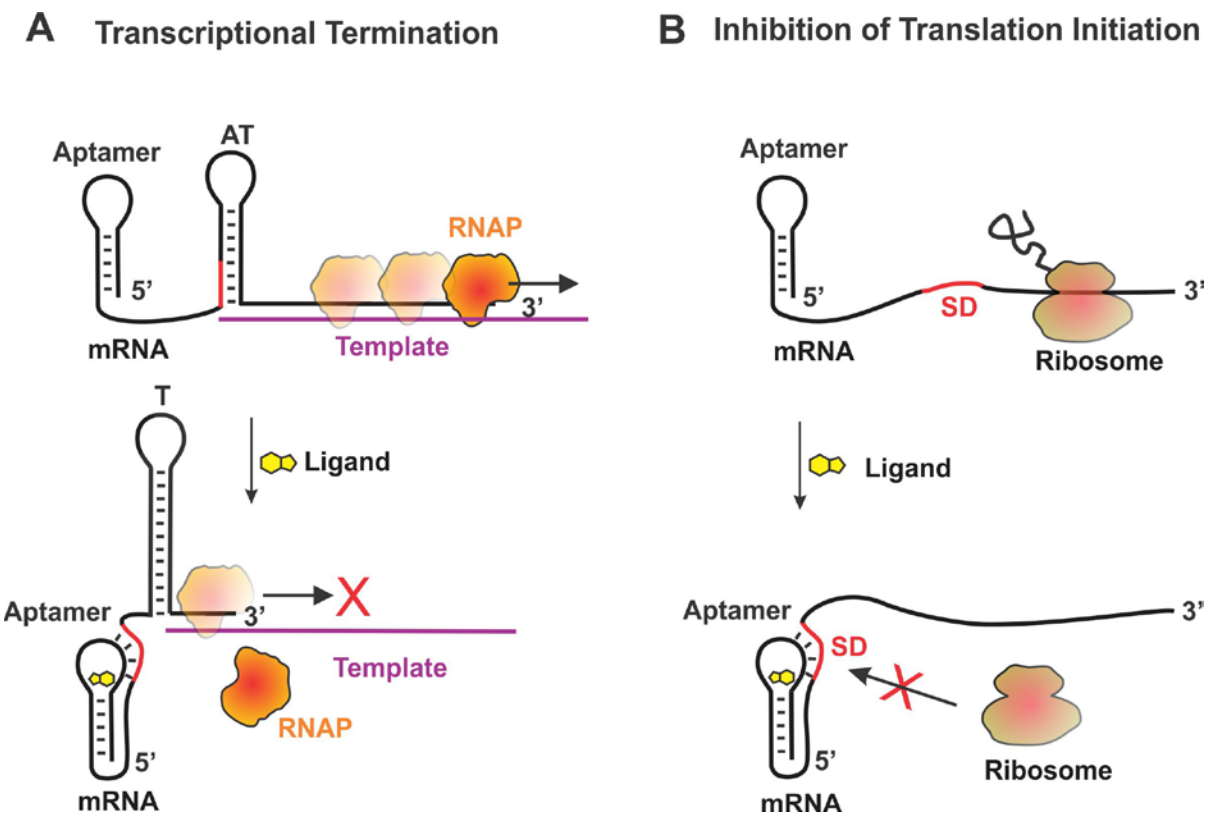

Figure 1.1 General mechanisms of riboswitch function.

(A) In the absence of ligand, the expression platform forms an anti-terminator (AT) hairpin that results in the transcription of downstream genes by RNA polymerase (RNAP). Stabilization of alternate conformation of the aptamer by the ligand leads to the formation of a terminator hairpin (T) thereby aborting transcription. (B) Folding of the aptamer in the presence of ligand sequesters the Shine-Dalgarno (SD) sequence leading to inhibition of translation initiation. The sequence shared between the aptamer and the expression platform, known as the 'switching' sequence is shown in red.

platform (16,20). In case of the transcriptionally acting riboswitches, the expression platform harbors two mutually exclusive hairpins – a Rho-independent terminator hairpin and an upstream anti-terminator hairpin, only one of which is specifically stabilized by ligand binding during transcription. Translationally acting riboswitches instead use ligand binding to decrease the accessibility of the SD sequence to the 30S ribosomal subunit, thereby inhibiting translation initiation (19). Translationally acting riboswitches can function as true 'switches' that make multiple gene regulatory decisions during the lifetime of the mRNA. For example, the SAM-III (S_{MK} box) riboswitch was shown to function in this manner where it can sense changing intracellular concentration of SAM (S-adenosylmethionine) over a long period and turn gene expression repeatedly ON or OFF. In contrast, transcriptionally acting riboswitches function more like 'fuses' since they make a single decision during transcription; once terminated they are eventually degraded (20,21).

Presently, more than 20 classes of riboswitches are known that bind chemically diverse ligands (16,18,22). In certain cases, multiple classes of riboswitches, with distinct secondary and tertiary structures, have been identified that recognize a common ligand. Examples include the more than five classes of riboswitches recognizing the coenzyme SAM and two classes of $preQ_1$ binding riboswitches (18,22). The aptamer domains of a given riboswitch class are highly conserved in structure across different species. Riboswitches utilize subtle sequence differences in the aptamer domain to 'fine-tune' their ligand binding affinity for optimal genetic control in different bacterial species (18,22). Additionally, the same class of aptamer domain in a particular bacterial species can have varied responses to a single ligand depending on the identity of the genes they regulate. For example, recent studies showed that the 11 SAM-I riboswitches of *B. subtilis* have very distinct ligand binding affinities (equilibrium dissociation constants, K_d) and show a wide range of effector mediated transcription termination efficiencies (23). In addition, riboswitches can also tune their optimal ligand responsiveness by using a mix-and-match approach, where different classes of aptamer domains can pair up with a wide variety of expression platforms (15,22,24). Furthermore, riboswitches can adopt complex architectures involving tandems of similar or different aptamer domains or tandems of entire riboswitches to achieve 'digital'-

like genetic control responses (25). Such tandem aptamer domains can exhibit cooperativity in ligand binding, as shown for the glycine riboswitch, which can help in achieving very high sensitivity to even small changes in ligand concentration (26,27). Interestingly, a riboswitch with tandem SAM-II/SAM-V aptamer domains has been identified with features similar to those of both transcriptional and translational genetic control, suggesting a sophisticated mechanism of action (15,28). Recently, a tandem arrangement of a self-splicing group I intron ribozyme with a c-di-GMP riboswitch was discovered in a pathogenic bacterium, further emphasizing the complex gene regulatory roles played by riboswitches (29). In addition, a riboswitch that recruits Rho-factor for transcription termination has been discovered (30). These examples highlight the intricacy of genetic regulatory mechanisms achieved by riboswitches. A large number of orphan riboswitches has been identified in bacteria whose cognate ligands are not yet known (24). Therefore, the functional repertoire of riboswitches will only increase in the future, as the roles of these riboswitches are unveiled.

1.1.4 Ligand recognition by riboswitches

Atomic resolution structures of riboswitch aptamer domains revealed many interesting aspects of riboswitches (16,22). First, the structures showed that similar to ribozymes and other non-coding RNAs, riboswitch aptamer domains can adopt structures of varying levels of size and complexity. Second, they also explained how RNA can recognize chemically diverse ligands despite having a limited arsenal of building blocks and functional groups. Riboswitches achieve tight ligand binding by mainly utilizing stacking interactions, hydrogen bonds and electrostatic interactions, sometimes mediated by co-ordinated metal ions and water molecules (18,22). For example, the c-di-GMP-II riboswitch has a K_d of only ~10 pM which is much lower than the K_d values of many proteins binding the same ligand (31). Most nucleotides of the ligand binding pocket are highly conserved while the residues surrounding it vary in sequence to fine-tune ligand binding affinity (22,23). A unique property of riboswitches is high selectivity to their cognate ligands. Closely related ligands differing by even a single function group can have large difference in their binding affinities

(18,32). In contrast, the aptamers of the *glmS*, $preQ_1$-I and THF riboswitches were shown to bind multiple closely related ligands relatively well (18). The discovery of cooperative ligand binding by some riboswitches showed that the molecular recognition properties of RNA extend beyond simple interactions. Cooperativity can be achieved using two interacting aptamer domains as seen in the glycine riboswitch or by using a single domain with two distant ligand binding sites, as in the THF (tetrahydrofolate) riboswitch (26,27,33). Interestingly, a recent study on the THF riboswitch showed that certain non-cognate ligands bound the aptamer with higher affinities, but could not elicit efficient gene regulatory responses (34). This shows a lack of correlation between the binding affinity and efficiency of gene regulation in riboswitches. These studies highlight the many tricks used by RNA to achieve sophisticated ligand binding and gene regulatory properties. However, our knowledge of the ligand binding mechanisms of riboswitches is still incomplete and therefore, studies need to be carried out in this aspect for a thorough understanding of riboswitch function.

1.1.5 Riboswitch structure and dynamics

Although riboswitches typically bind small molecules, the architectures of different aptamer domains vary vastly in their overall sizes (**Figure 1.2**) (22,35). While riboswitch aptamers adopt diverse configurations of RNA motifs, close inspection reveals that most of them are made of multihelical junctions and/or pseudoknot folds (22). For example, the crystal structures of purine, TPP, lysine and SAM-I binding aptamers all show multihelical junctions while the $preQ_1$, SAM-II and fluoride riboswitches exist as pseudoknots (**Figure 1.2**) (22). Among all riboswitches, the $preQ_1$-I riboswitch has the smallest aptamer domain (~33-nt) while the T-box riboswitch that binds tRNA and senses its aminoacylation state is ~200-nt long (36,37). The large size of T-box riboswitch enables it to recognize the overall geometry of tRNA and make sequence specific contacts with different ends of the tRNA. In contrast, the lysine riboswitch has one of the largest aptamers (~ 170-nt) among the metabolite binding riboswitches, even though it recognizes only a small molecule (38). Artificially selected RNA aptamers are also known to bind ligands with high affinity, yet

they are smaller in size than the naturally-occurring aptamers (39). However, synthetic aptamers display comparably poor ligand specificity (40). Therefore, the larger size of riboswitch aptamers may have evolved to provide extensive intramolecular tertiary interactions that form complex architectures required for achieving high ligand specificity and fine-tuned gene regulatory function (16).

For RNA, like all other biological macromolecules, structural dynamics are crucial for proper biological function (41). The conformational dynamics of RNA can be local or global and spans a wide range of timescales (42). Due to a rugged free energy landscape and inherent flexibility, RNA structures adopt multiple conformations that interconvert on a range of timescales (43,44). Therefore, RNA molecules exist in an ensemble of functionally active conformations and the idea of a single native structure is generally not applicable (43). As the name suggests, riboswitch RNAs are dynamic in nature. Structural dynamics play a critical role in ligand-mediated folding and gene regulation by riboswitches (45). Although the crystal structures of riboswitches show atomic detail of the ligand recognition mode, they do not provide information on riboswitch dynamics or the ligand-mediated folding pathwaysthat are critical to understanding riboswitch function (46). Furthermore, the current scarcity of available ligand-free aptamer structures limits our understanding of the folding process that forms the basis of gene regulation by riboswitches (46).

A number of biophysical studies using nuclear magnetic resonance (NMR) spectroscopy, molecular dynamics (MD) simulations, small angle x-ray scattering (SAXS) and single molecule experiments have provided details of the conformation, dynamics and effect of ligand on the folding of a number of riboswitches (47-62). These studies point to a general mechanism where ligand-free riboswitches exist in multiple interconverting conformations including ‘folded-like’ states that are stabilized by ligand binding. Divalent metal (mainly Mg^{2+}) ions are known to be crucial for RNA folding by stabilizing tertiary interactions (63,64). In the case of many riboswitches, Mg^{2+} was shown to be essential for the RNAs to sample ‘folded’-like conformations in the absence of ligand. Although Mg^{2+} is not required for ligand binding, it was shown to generally accelerate ligand dependent

folding and slow down the unfolding rate of the aptamer (49,54,62,65). In particular, single molecule fluorescence resonance energy transfer (smFRET) has been extensively used to study the dynamics and ligand dependent folding of riboswitches.

1.2 Single molecule fluorescence microscopy

1.2.1 Single molecule methods

Conventional experiments on biological macromolecules are done in bulk where a large number of molecules (typically ~10^{10}–10^{15}) provides an average signal for an observable parameter of interest. This parameter may be the catalytic rate constant k_{cat} for enzymes, or the dimensions or the diffusion constant of a molecular species. While such ensemble methods are valuable in providing information on the general behavior of a given target molecule and will continue to be useful, they suffer from an important problem – ensemble averaging. Ensemble methods do not provide a distribution of the observable parameter but rather only a single average value, thus leading to a loss of valuable information (66). For example, when ensemble methods are used to study a biomolecule existing in equal populations of two distinct states, they will report an average state that is in fact not a real conformation. In addition, the presence of any lowly populated transient conformations (sometimes – ambiguously – referred to as 'excited' states) can be extremely challenging to detect using ensemble methods (66). Therefore, single molecule methods are ideally suited for studying biomolecular systems such as RNA that generally exist in multiple distinct conformations (67,68). In addition, along with static heterogeneity of the entire population, the ability to observe a single molecule for a long period of time will enable studies of dynamic heterogeneity. Systems can be studied under equilibrium and non-equilibrium conditions and the rate constants for conversion between different conformations can be obtained (66). Furthermore, single molecule methods are usually done at very low concentrations and therefore require little material to work with. This is especially advantageous for studying systems that aggregate or form multimers at high concentrations. Due to the many advantages, over the past two decades single molecule microscopy has been

applied to study the structure, folding and dynamics of biomolecules such as DNA, RNA, and proteins, as well as for investigating large macromolecular assemblies (44,66,69-79).

A wide variety of single molecule methods has been used for studying biomolecules *in vitro* and *in vivo* (66). In general, single molecule methods can be classified into fluorescence and non-fluorescence (often mechanical) methods (80,81). Fluorescence based single molecule methods require the labeling of biomolecules with typically organic dyes (fluorophores). Fluorophore labeling enables the study of various molecular properties such as localization, movement (diffusion), binding to other molecules, intra- or inter-molecular distances, dynamics and oligomerization state under different conditions. Methods such as fluorescence correlation spectroscopy (FCS) using confocal microscopy and wide-field microscopy using prism- or objective-based total internal reflection fluorescence (TIRF) are widely used (80,82). Mechanical or force based methods do not require the labeling of biomolecules with fluorophores. However, recently, hybrid methods combining both fluorescence and force (e.g., "fleezers") have enabled the investigation of both the mechanical and conformational properties of biomolecules in a single experiment (83). Among the force-based methods, atomic force microscopy (AFM), as well as optical and magnetic tweezers are commonly utilized to study the folding and dynamics of biomolecules (82).

1.2.2 Fluorescence resonance energy transfer (FRET)

Fluorescence or Förster resonance energy transfer (FRET) refers to the non-radiative energy transfer between a donor and an acceptor fluorophore that are spatially proximal to each other (84). To be suitable as a FRET pair, the emission spectrum of the donor needs to overlap with the excitation spectrum of the acceptor. Energy transfer via FRET occurs only when the fluorophores are present within a certain distance that depends on the identity of the FRET pair and is generally <100 Å. Therefore, FRET can be used as a sensitive spectroscopic ruler to measure very short intra- or inter-molecular distances in the nm range (**Figure 1.3A**) (84). The excitation energy of the donor is then transferred to the acceptor

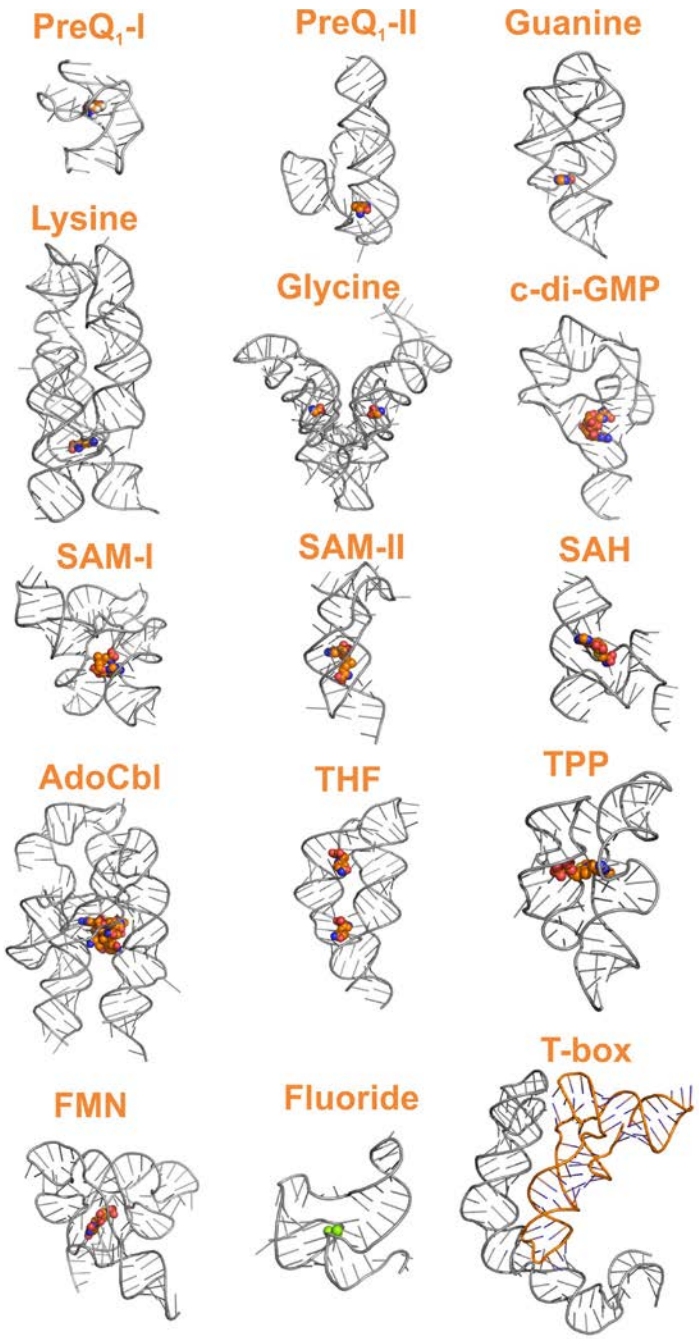

Figure 1.2 Crystal structures of various riboswitch classes showing diverse architectures.

The riboswitches and their ligands are shown in grey and colored, respectively. The structures shown are not drawn to scale.

through a dipole-dipole coupling interaction that is distance dependent. The FRET efficiency, *E*, is given by the equation:

$$E = \frac{1}{1 + (R/R_0)^6}$$

where *R* is the distance between the donor and the acceptor fluorophores. $R_{0,}$ known as the Förster radius, refers to the distance between the fluorophores when the energy transfer efficiency is 50% (**Figure 1.3A**). The value of R_0 depends on the properties of the donor and acceptor fluorophores and the local environment. Generally, R_0 is constant for a given FRET pair under similar buffer conditions. The R_0 values for commonly used FRET pairs range from 40–60 Å. For example, the R_0 value for the Cy3 (donor) - Cy5 (acceptor) pair of fluorophores is ~54 Å, with a roughly linear dependency of the FRET efficiency from ~30–75 Å (84). Therefore, distance changes below 30 Å and above 75 Å cannot easily be distinguished and will yield FRET values close to 1 and 0, respectively.

FRET results in the decrease in the intensity of the donor fluorophore with a simultaneous increase in the acceptor fluorophore intensity. Combining FRET with single molecule detection results in a powerful technique known as single molecule FRET (smFRET), which has been used to study the conformational dynamics of a variety of molecules in real-time (84). In smFRET experiments, one generally monitors the donor and acceptor fluorophore intensities of individual molecules as the FRET efficiency, calculated as:

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

where I_D and I_A are the background corrected donor and acceptor intensities, respectively. Structural dynamics in molecules labeled with a FRET pair manifest as anti-correlated changes in the intensities of the donor and acceptor fluorophores (**Figure 1.3C**). By immobilizing molecules to permit long observation times of their FRET changes, conformational dynamics and folding of biomolecules can be studied (44).

1.2.3 Single molecule FRET using TIR microscopy

Single molecule FRET microscopy for studying conformational dynamics is generally done using wide-field illumination (**Figure 1.3B**) (84). In this method, intensities from hundreds of molecules can be recorded simultaneously using a CCD camera. The molecules are generally immobilized onto the surface of a microfluidic channel for enabling longer observation times. The immobilized molecules can be excited either by epi-illumination or by using an evanescent field. The evanescent field is generated using total internal reflection (TIR) of the excitation laser at the quartz-water interface. The intensity of the evanescent filed decays exponentially with distance in the z-direction and therefore only excites immobilized molecules that are within ~100 nm of the surface (**Figure 1.3B**). The advantage of using TIR microscopy thus is that it eliminates the background autofluorescence from solvent and any free labeled molecules in solution, resulting in a high signal-to-noise ratio that is critical for single molecule detection. A quartz prism or an objective with high numerical aperture can be used to achieve TIR. The prism-based TIRF microscopy setup is relatively simple and has been widely used for studying conformational dynamics of many non-coding RNAs such as ribozymes, telomere RNA, the ribosome, the spliceosome, and many different riboswitches (44,49,50,66,68,70,75,85).

1.2.4 Fluorophore labeling and purification of riboswitches for smFRET

For studying riboswitch structure and dynamics using smFRET, molecules have to be labeled with both donor and acceptor fluorophores (69,84). In addition, a biotin moiety needs to be present on the molecule for immobilization on the quartz slide surface, typically using biotin-streptavidin chemistry. A key aspect of obtaining a doubly labeled RNA is the selection of nucleotides for fluorophore labeling. Depending on the structure of the aptamer, the fluorophores need to be positioned such that they report large-scale distance changes between functionally important regions forming the key tertiary interactions involved in ‘switching’. For example, in the purine riboswitches, fluorophores were placed near the hairpin loops P2 and P3 that form a critical tertiary interaction stabilized by ligand binding (54,62). Similarly, for riboswitches adopting a pseudoknot fold the fluorophores should

report formation of the thermodynamically less stable helix (generally P2 or P3, closest to the 3'-end) that is stabilized by the ligand (49,51). In this regard, the availability of a high-resolution crystal or NMR structure will aid in choosing the nucleotides to label. This is particularly useful for labeling riboswitches with complex architectures. For simple riboswitches, RNA secondary structure along with nucleotide conservation data may be sufficient for the selection of suitable labeling positions. In the absence of a high-resolution structure for the RNA, chemical structure probing using methods such as selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) or in-line probing can reveal the identity of nucleotides that form intramolecular interactions (or inter-molecular interactions with the ligand), as well as those exposed to solvent (86). As a rule of thumb, nucleotides that are evolutionarily less conserved, solvent exposed and not involved in any intra-molecular interactions should be chosen for fluorophores labeling. This strategy will ensure that the presence of the bulky fluorophores will not affect the structure, folding or ligand binding of the riboswitch. Furthermore, the labeling sites should be within the linear FRET range of the fluorophores for maximum sensitivity. The larger the difference between the FRET levels, the easier it is to distinguish them. Ideally, the labeling sites should be chosen such that a minimum difference of 0.2 is obtained between two FRET values.

There are many ways of achieving site-specific labeling of RNA with fluorophores for smFRET experiments (87). A common way of internally labeling riboswitches is to conjugate N-hydroxysuccinimide (NHS) ester derivatives of fluorophores to free primary amine functional groups of modified nucleotides. Commonly, aminoallyl uridine can be incorporated into the sequence in place of uridine during chemical synthesis of oligoribonucleotide (RNA) and labeled using an NHS ester fluorophore. In the recent years, due to an increase in the efficiency of RNA chemical synthesis, ordering custom designed singly or doubly fluorophore labeled short RNAs with additional modifications, such as 3' or 5' biotin, has become affordable. However, this is only feasible for aptamers that are <80 nt in length. For larger riboswitches, alternatively, one or more short fluorophore labeled RNA oligos can be convalently linked to *in vitro* transcribed larger RNA using enzymatic ligation (87) or simply hybridized. For labeling of RNA at the 3' or 5' end, free amine group can be

introduced during chemical synthesis, which can be labeled using a fluorophore NHS ester. Labeling is often non-trivial and requires multiple strategies of purification to remove unlabeled and singly labeled RNAs. Excess unreacted dye is removed by, for example, gel filtration and ethanol precipitation or reversed-phase HPLC separation. The labeling efficiency can be assessed by using their extinction coefficients and measuring the absorbances of RNA and the two fluorophores using a UV-Vis spectrophotometer. If the labeling efficiency is high enough (>80%), the RNA can be directly used for smFRET experiments without further purification. However, purification of doubly labeled RNA using denaturing gel electrophoresis or HPLC is recommended when the labeling yield is low. This will enable imaging of hundreds of doubly labeled RNA molecules during smFRET experiments.

1.2.5 Preparation of quartz slides for smFRET

SmFRET experiments are routinely performed using quartz slides with a microfluidic channel (84). For single molecule experiments, the slide surface needs to be thoroughly cleaned using a multi-step protocol to remove any organic impurities (70). Presence of significant fluorescent impurities will impede observation of true single molecules and contribute to false background signals. Impurities may also affect the local environment of molecules that may result in heterogeneities in their behavior. The cleaning process first involves boiling the used slides to remove the cover slips and then scrubbing the surface with a thick paste of detergent (Alconox, Inc), followed by sonication in the detergent solution. The slides are then sonicated in various organic and inorganic solvents such as acetone, methanol or ethanol, 1 M KOH and a 5:1:1 volume ratio mixture of water:ammonium hydroxide:hydrogen peroxide. Each sonication step in the various solvents is done for 30 min, and after every step, the slides are sonicated in double-distilled water for 10 min. In the end, slide surfaces are flamed off using a propane torch to destroy any remaining fluorescent organic impurities. Once the slides are clean, the microfluidic channel can be made by sandwiching two strips of double-sided sticky tape ~3-5 mm apart between the quartz slide and a clean glass coverslip (84). The channel thus made has a low volume of ~30 – 50 µL. A

pair of holes drilled into the quartz slide will act as inlet and outlet ports for the flow of buffers containing different buffers and samples into the channel. in one standard protocol, the microfluidic channel is coated first with biotinylated-BSA (bBSA) that adsorbs onto the slide surface. After washing the excess unbound bBSA using a 1x smFRET buffer of choice, a streptavidin solution is flowed onto the surface. After incubation for a few min, unbound streptavidin is washed off with the 1x buffer. Next, 50-100 pM of doubly labeled RNA is flowed onto the slide until an optimal density is obtained (~300–400 molecules per field of view). Unbound molecules are washed away and molecules are imaged in 1x buffer containing an oxygen scavenging system (OSS) that will increase the life-time of the fluorophores permitting long observation times. A widely used OSS consists of protocatechuic acid (PCA), protocatechuate-3,4-dioxygenase (PCD) enzyme and Trolox to slow down fluorophore photobeaching and blinking, respectively (70).

1.2.6 Heat-annealing of riboswitch RNA

Before immobilization onto the slide, riboswitch RNAs need to be folded into their native conformations. RNA is generally stored at low temperatures (-20 °C or -80 °C) in autoclaved double-distilled water or in low pH buffers without any divalent cations to prevent its degradation over time. However, under such low-temperature, low-ionic strength conditions and due to repeated freeze-thawing, RNA molecules can misfold or form non-specific aggregates. Therefore, heating will unfold the misfolded RNAs and break the aggregates, allowing the RNA to fold into its native conformation. There is no standard protocol for folding of RNAs and diverse heat-annealing protocols can be found in the literature. Commonly, RNA is heated in a buffer with monovalent ions such as Na^+ or K^+, but without Mg^{2+}, to 75 °C for 2 min. The RNA is cooled at either rapidly or slowly to allow folding into its secondary structure in the presence of monovalent ions. This is followed by the addition of Mg^{2+} to the required concentration after cooling down to 37 °C or RT. Addition of Mg^{2+} will induce formation of tertiary interactions and complete the RNA folding. For smFRET experiments, heat annealing after diluting the RNA to the required low concentrations of 50 - 100 pM will prevent the formation of dimers/aggregates. Higher

temperatures such as 90 °C should be tried in cases where heating at 75 °C is not effective in breaking the aggregates. As a control, a few different temperature and time regimes should be tried to optimize the folding protocol and to test if the exact folding protocol has any significant effect on the smFRET results.

1.2.7 Analysis of smFRET data

Using wide-field smFRET microscopy, time traces of the donor and acceptor fluorophore intensities can be obtained for hundreds of molecules. These smFRET time traces provide direct information on the conformational states sampled by individual molecules and the times spent in each of these states before transitioning to other FRET states. The smFRET traces can be analyzed using a number of methods to yield data on the structure and dynamics of the RNA (88). The time traces of a few hundreds of molecules typically are first plotted into a population FRET histogram by sampling the first 50 or 100 frames of each molecule. The FRET histogram monitors the ensemble behavior of all the molecules. By fitting the histogram as a sum of Gaussian functions, the number of conformations sampled by the riboswitch and their apparent equilibrium distribution can be obtained under a given condition. Generally, for riboswitches the FRET histogram is first obtained in the absence of both Mg^{2+} and ligand. Later, Mg^{2+} or ligand titrations can be performed in the same experiment to identify their effect(s), either individually or together, on the conformational dynamics of the riboswitch. The occupancy (as seen from the FRET histogram) of one of the conformations, corresponding to the ligand-bound folded state is expected to increase as a function of ligand concentration, from which a half-saturation value ($K_{1/2}$) value is estimated that relates to a K_d obtained from biochemical methods. The mean and width of the individual Gaussian peaks will provide additional information on the compactness and dynamic nature of the underlying conformations. Following the changes in these fitting parameters at varying ligand and/or Mg^{2+} concentrations can provide details on the folding of the RNA. However, the most important data from smFRET experiments are the dwell times of individual molecules in different FRET states before transitioning to other states. Statistical methods such as Hidden Markov Modeling (HMM) are used to idealize the

smFRET traces in an unbiased manner, detect transitions between FRET states, and obtain the exact dwell times of individual molecules in different states (88,89). From the idealized traces, transition density plots (TDPs) can be made that show the number of transitions of a given type in all the molecules as heat maps (**Figure 1.3E**) (88). Dwell times of all molecules in each FRET state before transitioning to a different FRET state can be extracted from the TDP. The dwell times are then plotted as a cumulative distribution plot and fit with exponential functions to extract the rate constants of conformational dynamics between the different states (**Figure 1.3F**) (88). The transition rates of individual molecules will reveal the extent of conformational heterogeneity present in riboswitch molecules. However, in certain cases where the conformational dynamics are fast and close to the time resolution of the camera (typically tens to hundreds of milliseconds), using HMM is not ideal. In such cases, cross-correlation analysis can be performed on the smFRET traces to quantify the anti-correlation between the donor and acceptor intensities. The inverse of the lifetime of the cross-correlation function yields the sum of transitions rates between the FRET states (90,91). By analyzing how increasing concentrations of ligand affect the rates of conformational dynamics between different FRET states, one can potentially reveal the mechanism of ligand binding, which is extremely difficult to investigate using ensemble methods (92).

The aim of this thesis was to study the conformation, dynamics and ligand binding of two very distinct riboswitches – the $preQ_1$ riboswitch (32,36) as the smallest known riboswitch, and the T-box RNA (37) as one of the largest and most complex riboswitches (Figure 1.2). To this end, in Chapter 2, using smFRET, computational simulations and NMR, we have investigated the conformation and ligand-mediated folding of each one transcriptionally and translationally acting $preQ_1$ riboswitch. We showed that in contrast to previous studies, both riboswitches have similar ligand-free conformations with only subtle differences in their dynamics. We also demonstrated that remote mutations away from the

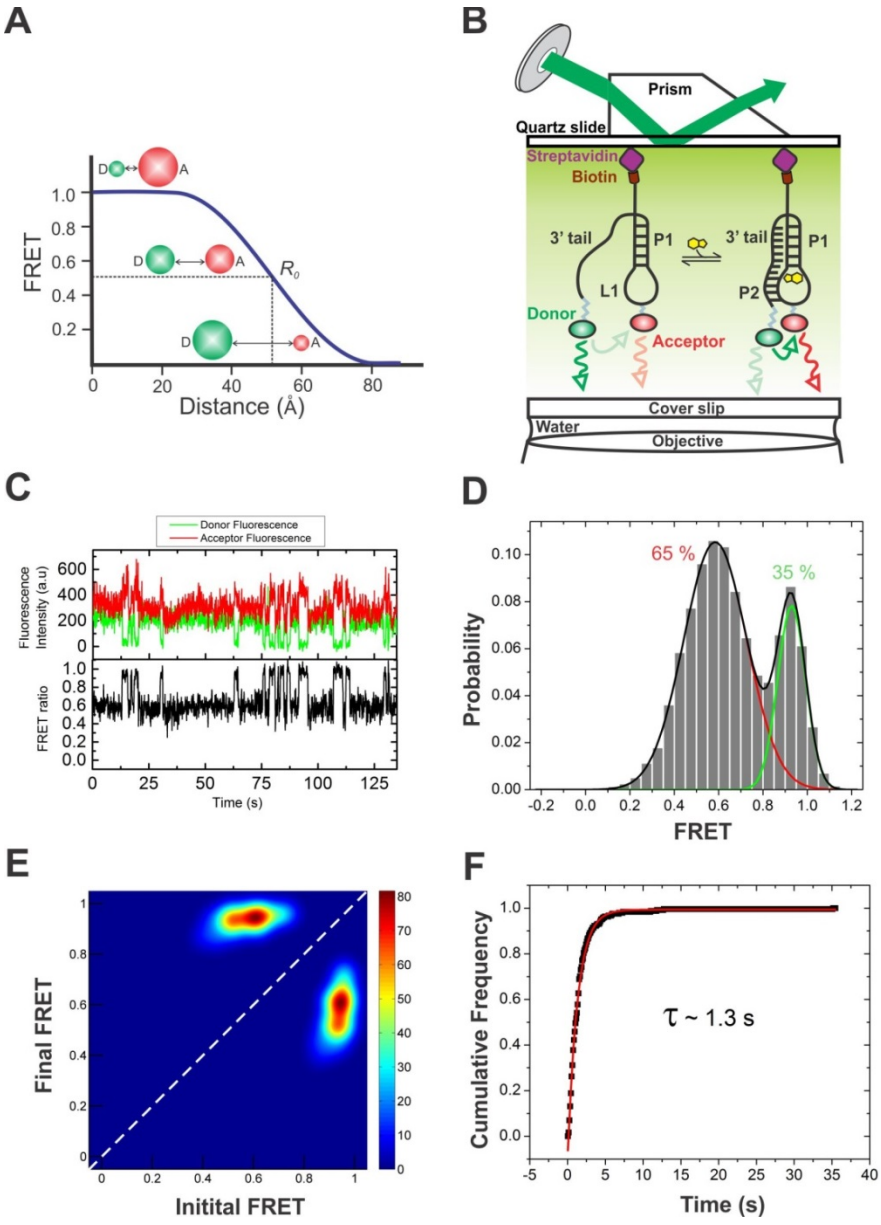

Figure 1.3 Studying riboswitch structure and dynamics using single molecule FRET.

(A) FRET efficiency vs distance plot (B) Schematic of smFRET setup using TIRF microscopy (C) Example smFRET trace showing dynamics between two FRET states (D) Population FRET histogram showing the distribution of two conformations (E) A transition density plot (TDP) that shows the total number of transitions between the different FRET states (two in this example) in all molecules (F) Cumulative dwell time distribution fit to a single exponential function to extract kinetic information.

ligand binding pocket can affect riboswitch conformation and ligand binding, especially for the transcriptional $preQ_1$ riboswitch. Based on our smFRET data and computational folding simulations, we proposed that both the RNAs follow distinct ligand-mediated folding pathways that resemble the classical models of 'induced-fit' and 'conformational selection'. In Chapter 3, we further investigated the effect of ligand on the structural dynamics of the transcriptionally acting Bsu $preQ_1$ riboswitch at faster time resolutions to further delineate its folding mechanism. In addition, the effect of different ligands on the conformation and dynamics of the *Bsu* riboswitch were probed. Finally, in **chapter 4**, we have probed the conformation and tRNA binding kinetics of various *glyQS* T-box riboswitch designs using smFRET. We showed that the T-box riboswitch exists in a stable conformation with minimum global conformational dynamics. We also showed that tRNA binding does not cause major structural changes, suggesting a pre-organized conformation. Our kinetic binding data directly demonstrate that the presence of a small amino acid such as glycine on the 3'-end of the tRNA significantly increases its dissociation rate by ~6-fold, which explains why a charged tRNA does not result in transcription antitermination. Using our kinetic data, we present an improved quantitative kinetic model for the T-box riboswitch mechanism of action.

Studying riboswitch structure, dynamics and ligand binding mechanism using ensemble methods has been challenging. In this dissertation, we have used smFRET to shed light on the conformational behavior of two contrasting riboswitches. In addition, our single molecule kinetic assays used to probe T-box-tRNA binding are applicable to the study of many other RNA-RNA interactions. The smFRET methods used in this dissertation generally can be applied to more deeply probe the molecular structure and conformational properties of the diverse classes of riboswitches that range in size between the two RNAs investigated here to expand our so far limited knowledge of ligand-mediated riboswitch folding.

1.3 References

1. Darnell, J.E. (2011) *RNA : life's indispensable molecule*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
2. Gesteland, R.F., Cech, T. and Atkins, J.F. (2006) *The RNA world : the nature of modern RNA suggests a prebiotic RNA world*. 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
3. Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep*, **2**, 986-991.
4. Esteller, M. (2011) Non-coding RNAs in human disease. *Nat Rev Genet*, **12**, 861-874.
5. Taft, R.J., Pang, K.C., Mercer, T.R., Dinger, M. and Mattick, J.S. (2010) Non-coding RNAs: regulators of disease. *J Pathol*, **220**, 126-139.
6. Wassarman, K.M. (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell*, **109**, 141-144.
7. Gottesman, S. and Storz, G. (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol*, **3**.
8. Repoila, F. and Darfeuille, F. (2009) Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biology of the Cell*, **101**, 117-131.
9. Storz, G., Vogel, J. and Wassarman, K.M. (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell*, **43**, 880-891.
10. Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167-170.
11. Karginov, F.V. and Hannon, G.J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell*, **37**, 7-19.
12. Winkler, W.C. and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol*, **59**, 487-517.
13. Bastet, L., Dube, A., Masse, E. and Lafontaine, D.A. (2011) New insights into riboswitch regulation mechanisms. *Mol Microbiol*, **80**, 1148-1154.
14. Breaker, R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol*, **4**.
15. Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867-879.
16. Peselis, A. and Serganov, A. (2014) Themes and variations in riboswitch structure and function. *Biochim Biophys Acta*.
17. Batey, R.T. (2012) Structure and mechanism of purine-binding riboswitches. *Q Rev Biophys*, **45**, 345-381.
18. Serganov, A. and Patel, D.J. (2012) Molecular recognition and function of riboswitches. *Curr Opin Struct Biol*, **22**, 279-286.
19. Garst, A.D., Edwards, A.L. and Batey, R.T. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol*, **3**.
20. Garst, A.D. and Batey, R.T. (2009) A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta*, **1789**, 584-591.
21. Smith, A.M., Fuchs, R.T., Grundy, F.J. and Henkin, T.M. (2010) The SAM-responsive S(MK) box is a reversible riboswitch. *Mol Microbiol*, **78**, 1393-1402.

22. Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, **152**, 17-24.
23. Stoddard, C.D., Widmann, J., Trausch, J.J., Marcano-Velazquez, J.G., Knight, R. and Batey, R.T. (2013) Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch. *J Mol Biol*, **425**, 1596-1611.
24. Barrick, J.E. and Breaker, R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol*, **8**, R239.
25. Sudarsan, N., Hammond, M.C., Block, K.F., Welz, R., Barrick, J.E., Roth, A. and Breaker, R.R. (2006) Tandem riboswitch architectures exhibit complex gene control functions. *Science*, **314**, 300-304.
26. Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L. and Breaker, R.R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science*, **306**, 275-279.
27. Huang, L., Serganov, A. and Patel, D.J. (2010) Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. *Mol Cell*, **40**, 774-786.
28. Poiata, E., Meyer, M.M., Ames, T.D. and Breaker, R.R. (2009) A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA*, **15**, 2046-2056.
29. Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N. and Breaker, R.R. (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science*, **329**, 845-848.
30. Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E. and Groisman, E.A. (2012) Riboswitch control of Rho-dependent transcription termination. *Proc Natl Acad Sci U S A*, **109**, 5376-5381.
31. Smith, K.D., Shanahan, C.A., Moore, E.L., Simon, A.C. and Strobel, S.A. (2011) Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proc Natl Acad Sci U S A*, **108**, 7757-7762.
32. 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. *et al.* (2007) A riboswitch selective for the queuosine precursor preQ$_1$ contains an unusually small aptamer domain. *Nat Struct Mol Biol*, **14**, 308-317.
33. Trausch, J.J., Ceres, P., Reyes, F.E. and Batey, R.T. (2011) The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. *Structure*, **19**, 1413-1423.
34. Trausch, J.J. and Batey, R.T. (2014) A Disconnect between High-Affinity Binding and Efficient Regulation by Antifolates and Purines in the Tetrahydrofolate Riboswitch. *Chem Biol*, **21**, 205-216.
35. Serganov, A. and Patel, D.J. (2012) Metabolite recognition principles and molecular mechanisms underlying riboswitch function. *Annu Rev Biophys*, **41**, 343-370.
36. Klein, D.J., Edwards, T.E. and Ferre-D'Amare, A.R. (2009) Cocrystal structure of a class I preQ$_1$ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol*, **16**, 343-344.

37. Zhang, J. and Ferre-D'Amare, A.R. (2013) Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature*, **500**, 363-366.
38. Garst, A.D., Heroux, A., Rambo, R.P. and Batey, R.T. (2008) Crystal structure of the lysine riboswitch regulatory mRNA element. *J Biol Chem*, **283**, 22347-22351.
39. Edwards, T.E., Klein, D.J. and Ferre-D'Amare, A.R. (2007) Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr Opin Struct Biol*, **17**, 273-279.
40. Carothers, J.M., Oestreich, S.C. and Szostak, J.W. (2006) Aptamers selected for higher-affinity binding are not more specific for the target ligand. *J Am Chem Soc*, **128**, 7929-7937.
41. Dethoff, E.A., Chugh, J., Mustoe, A.M. and Al-Hashimi, H.M. (2012) Functional complexity and regulation through RNA dynamics. *Nature*, **482**, 322-330.
42. Al-Hashimi, H.M. and Walter, N.G. (2008) RNA dynamics: it is about time. *Curr Opin Struct Biol*, **18**, 321-329.
43. Solomatin, S.V., Greenfeld, M., Chu, S. and Herschlag, D. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*, **463**, 681-684.
44. Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science*, **296**, 1473-1476.
45. Haller, A., Souliere, M.F. and Micura, R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc Chem Res*, **44**, 1339-1348.
46. Liberman, J.A. and Wedekind, J.E. (2012) Riboswitch structure in the ligand-free state. *Wiley Interdiscip Rev RNA*, **3**, 369-384.
47. Reining, A., Nozinovic, S., Schlepckow, K., Buhr, F., Furtig, B. and Schwalbe, H. (2013) Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature*, **499**, 355-359.
48. Rieder, U., Kreutz, C. and Micura, R. (2010) Folding of a transcriptionally acting preQ$_1$ riboswitch. *Proc Natl Acad Sci U S A*, **107**, 10804-10809.
49. Haller, A., Rieder, U., Aigner, M., Blanchard, S.C. and Micura, R. (2011) Conformational capture of the SAM-II riboswitch. *Nat Chem Biol*, **7**, 393-400.
50. Haller, A., Altman, R.B., Souliere, M.F., Blanchard, S.C. and Micura, R. (2013) Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc Natl Acad Sci U S A*, **110**, 4188-4193.
51. Souliere, M.F., Altman, R.B., Schwarz, V., Haller, A., Blanchard, S.C. and Micura, R. (2013) Tuning a riboswitch response through structural extension of a pseudoknot. *Proc Natl Acad Sci U S A*, **110**, E3256-3264.
52. Fiegland, L.R., Garst, A.D., Batey, R.T. and Nesbitt, D.J. (2012) Single-molecule studies of the lysine riboswitch reveal effector-dependent conformational dynamics of the aptamer domain. *Biochemistry*, **51**, 9223-9233.
53. Wood, S., Ferre-D'Amare, A.R. and Rueda, D. (2012) Allosteric tertiary interactions preorganize the c-di-GMP riboswitch and accelerate ligand binding. *ACS Chem Biol*, **7**, 920-927.

54. Brenner, M.D., Scanlan, M.S., Nahas, M.K., Ha, T. and Silverman, S.K. (2010) Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine. *Biochemistry*, **49**, 1596-1605.
55. Chen, B., Zuo, X., Wang, Y.X. and Dayie, T.K. (2012) Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res*, **40**, 3117-3130.
56. Noeske, J., Buck, J., Furtig, B., Nasiri, H.R., Schwalbe, H. and Wohnert, J. (2007) Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res*, **35**, 572-583.
57. Buck, J., Wacker, A., Warkentin, E., Wohnert, J., Wirmer-Bartoschek, J. and Schwalbe, H. (2011) Influence of ground-state structure and Mg2+ binding on folding kinetics of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res*, **39**, 9768-9778.
58. Wacker, A., Buck, J., Mathieu, D., Richter, C., Wohnert, J. and Schwalbe, H. (2011) Structure and dynamics of the deoxyguanosine-sensing riboswitch studied by NMR-spectroscopy. *Nucleic Acids Res*, **39**, 6802-6812.
59. Frieda, K.L. and Block, S.M. (2012) Direct observation of cotranscriptional folding in an adenine riboswitch. *Science*, **338**, 397-400.
60. Feng, J., Walter, N.G. and Brooks, C.L., 3rd. (2011) Cooperative and directional folding of the $preQ_1$ riboswitch aptamer domain. *J Am Chem Soc*, **133**, 4196-4199.
61. Neupane, K., Yu, H., Foster, D.A., Wang, F. and Woodside, M.T. (2011) Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res*, **39**, 7677-7687.
62. Lemay, J.F., Penedo, J.C., Tremblay, R., Lilley, D.M. and Lafontaine, D.A. (2006) Folding of the adenine riboswitch. *Chem Biol*, **13**, 857-868.
63. Misra, V.K. and Draper, D.E. (1998) On the role of magnesium ions in RNA stability. *Biopolymers*, **48**, 113-135.
64. Misra, V.K. and Draper, D.E. (2002) The linkage between magnesium binding and RNA folding. *J Mol Biol*, **317**, 507-521.
65. Santner, T., Rieder, U., Kreutz, C. and Micura, R. (2012) Pseudoknot preorganization of the $preQ_1$ class I riboswitch. *J Am Chem Soc*, **134**, 11928-11931.
66. Tinoco, I., Jr. and Gonzalez, R.L., Jr. (2011) Biological mechanisms, one molecule at a time. *Genes Dev*, **25**, 1205-1231.
67. Zhuang, X., Bartley, L.E., Babcock, H.P., Russell, R., Ha, T., Herschlag, D. and Chu, S. (2000) A single-molecule study of RNA catalysis and folding. *Science*, **288**, 2048-2051.
68. Zhuang, X. (2005) Single-molecule RNA science. *Annu Rev Biophys Biomol Struct*, **34**, 399-414.
69. Abelson, J., Blanco, M., Ditzler, M.A., Fuller, F., Aravamudhan, P., Wood, M., Villa, T., Ryan, D.E., Pleiss, J.A., Maeder, C. *et al.* (2010) Conformational dynamics of single pre-mRNA molecules during in vitro splicing. *Nat Struct Mol Biol*, **17**, 504-512.

70. Krishnan, R., Blanco, M.R., Kahlscheuer, M.L., Abelson, J., Guthrie, C. and Walter, N.G. (2013) Biased Brownian ratcheting leads to pre-mRNA remodeling and capture prior to first-step splicing. *Nat Struct Mol Biol*, **20**, 1450-1457.
71. Schuler, B., Lipman, E.A. and Eaton, W.A. (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, **419**, 743-747.
72. Lipman, E.A., Schuler, B., Bakajin, O. and Eaton, W.A. (2003) Single-molecule measurement of protein folding kinetics. *Science*, **301**, 1233-1235.
73. Borgia, A., Williams, P.M. and Clarke, J. (2008) Single-molecule studies of protein folding. *Annu Rev Biochem*, **77**, 101-125.
74. Blanchard, S.C. (2009) Single-molecule observations of ribosome function. *Curr Opin Struct Biol*, **19**, 103-109.
75. Brakmann, S. (2010) Single-molecule analysis: A ribosome in action. *Nature*, **464**, 987-988.
76. Wen, J.D., Lancaster, L., Hodges, C., Zeri, A.C., Yoshimura, S.H., Noller, H.F., Bustamante, C. and Tinoco, I. (2008) Following translation by single ribosomes one codon at a time. *Nature*, **452**, 598-603.
77. Bai, L., Santangelo, T.J. and Wang, M.D. (2006) Single-molecule analysis of RNA polymerase transcription. *Annu Rev Biophys Biomol Struct*, **35**, 343-360.
78. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B. *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133-138.
79. McKinney, S.A., Declais, A.C., Lilley, D.M. and Ha, T. (2003) Structural dynamics of individual Holliday junctions. *Nat Struct Biol*, **10**, 93-97.
80. Walter, N.G., Huang, C.Y., Manzo, A.J. and Sobhy, M.A. (2008) Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nat Methods*, **5**, 475-489.
81. Deniz, A.A., Mukhopadhyay, S. and Lemke, E.A. (2008) Single-molecule biophysics: at the interface of biology, physics and chemistry. *J R Soc Interface*, **5**, 15-45.
82. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. and Ha, T. (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annu Rev Biochem*, **77**, 51-76.
83. Hohng, S., Zhou, R., Nahas, M.K., Yu, J., Schulten, K., Lilley, D.M. and Ha, T. (2007) Fluorescence-force spectroscopy maps two-dimensional reaction landscape of the holliday junction. *Science*, **318**, 279-283.
84. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single-molecule FRET. *Nat Methods*, **5**, 507-516.
85. Hengesbach, M., Kim, N.K., Feigon, J. and Stone, M.D. (2012) Single-molecule FRET reveals the folding dynamics of the human telomerase RNA pseudoknot domain. *Angew Chem Int Ed Engl*, **51**, 5876-5879.
86. Weeks, K.M. (2010) Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol*, **20**, 295-304.
87. Solomatin, S. and Herschlag, D. (2009) Methods of site-specific labeling of RNA with fluorescent dyes. *Methods Enzymol*, **469**, 47-68.

88. Blanco, M. and Walter, N.G. (2010) Analysis of complex single-molecule FRET time trajectories. *Methods Enzymol*, **472**, 153-178.
89. Qin, F. and Li, L. (2004) Model-based fitting of single-channel dwell-time distributions. *Biophys J*, **87**, 1657-1671.
90. Kim, H.D., Nienhaus, G.U., Ha, T., Orr, J.W., Williamson, J.R. and Chu, S. (2002) Mg2+-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules. *Proc Natl Acad Sci U S A*, **99**, 4284-4289.
91. Ragunathan, K., Liu, C. and Ha, T. (2012) RecA filament sliding on DNA facilitates homology search. *Elife*, **1**, e00067.
92. Kim, E., Lee, S., Jeon, A., Choi, J.M., Lee, H.S., Hohng, S. and Kim, H.S. (2013) A single-molecule dissection of ligand binding to a protein with intrinsic dynamics. *Nat Chem Biol*, **9**, 313-318.

CHAPTER 2

Single transcriptional and translational $preQ_1$ riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure

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 (1-5). 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 (6,7), amino acids (8,9), cofactors of metabolic enzymes (10,11) and metal ions (12,13), and are found in a multitude of bacterial species (5). 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 bacteria (14,15). Queuosine is found in bacteria and eukaryotes

---

This work is reproduced in part from Suddala, K.C. *et al*. Single transcriptional and translational $preQ_1$ riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure. *Nucleic Acids Res* **41**, 10462-75 (2013). Krishna C. Suddala performed labeling and smFRET experiments on the *Bsu* riboswitch. Arlie J. Rinaldi performed labeling and smFRET experiments on the *Tte* riboswitch. Jun Feng performed the Gō-model simulations in **Figures 2.13 and 2.14**, Anthony M. Mustoe performed the TOPRNA simulations in **Figure 2.9**, Catherine D. Eichhorn performed the NMR experiments in **Figure 2.12** and Joseph A. Liberman carried out the ITC measurements in **Figure 2.7**.

at the wobble position of tRNAs for His, Tyr, Asn, and Asp (16), where it is thought to be essential for translational fidelity (17-19) as well as bacterial virulence (20). 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 (21-30) and are known to regulate the *queCDEF* operon through transcription termination (31). 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 (32,33).

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* (21) and *Tte* (33) aptamer domains (referred to henceforth simply as "riboswitches") overlay closely with a backbone unit-vector RMSD (URMSD; using an all-C3' atom trace) of only 1.8 Å, as based on the RNA structure alignment program SARA (34) (**Figure 2.1B**). This value is comparable to a URMSD of 1.6 Å between the lowest-energy NMR (25) and crystal (21) 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 (35).

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. NMR studies of the *Bsu* riboswitch (25,36) have 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. Such an extended conformation lacks a pre-organized binding site, leaving unresolved the question of how this 'open' conformation senses ligand. Further, the *Bsu* riboswitch must bind ligand efficiently during the narrow time window of ~1-2 s during which RNA polymerase proceeds from the 3' end of the aptamer domain to the 3' end of the intrinsic terminator hairpin and therefore, an unfolded ligand-free conformation is ill-suited for such a transcriptionally acting riboswitch. In contrast, X-ray crystallography and SAXS studies on the *Tte* riboswitch (33) suggest that it does form a pre-folded tertiary conformation poised for recognition in the absence of ligand. Such pre-folding of the *Tte* riboswitch at room temperature may not be surprising given its origin from a thermophilic bacterium that grows optimally at 75 °C where the riboswitch may be less folded. Such contrasting behavior of these two structurally similar riboswitches prompted us to investigate their conformational distribution and dynamics under similar physiologically relevant buffer conditions. In addition, even though some structural data are available, how the two riboswitches compare in their 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) (37-39) 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

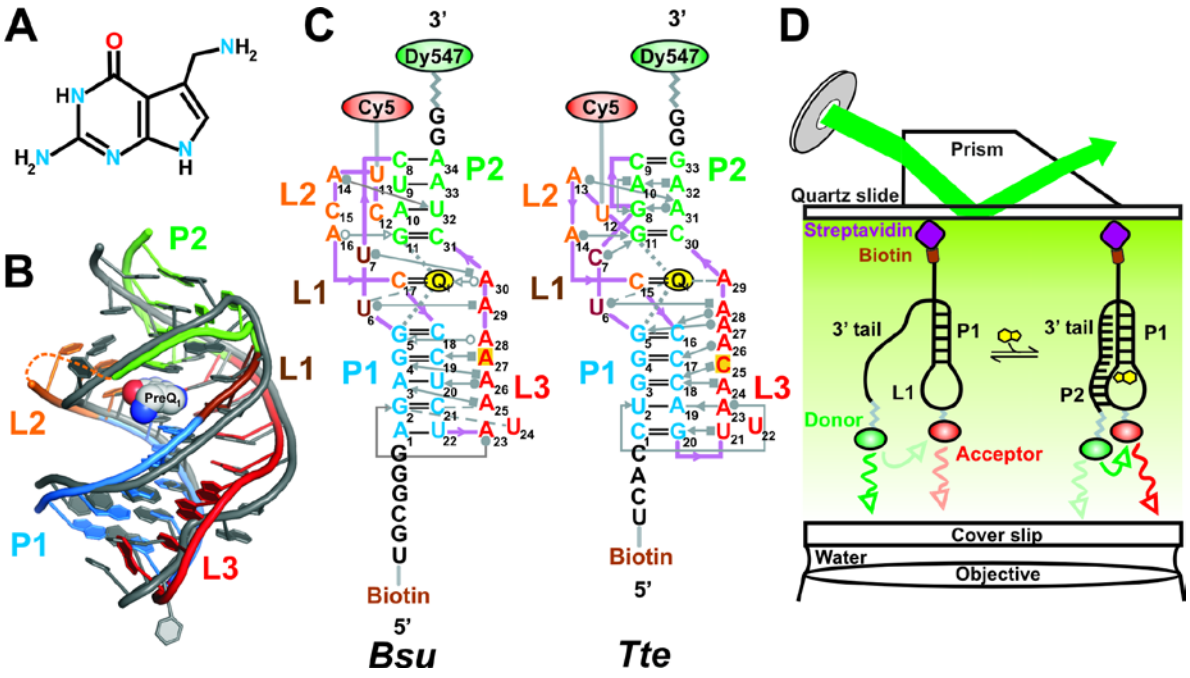

Figure 2.1 Structural comparison of the *Bsu* and *Tte* preQ$_1$ riboswitches

(A) Structure of preQ$_1$ (7-aminomethyl-7-deazaguanine). (B) Structural overlay of the *Bsu* (colored, PDB ID 3FU2, chain A) and *Tte* (grey, 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 (35). Individual secondary structures are color-coded as in B and the locations of fluorophores and biotin are indicated. (D) Prism-based TIRFM setup for smFRET.

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 (36), which leads to differences in long-range transient interactions. Together with results from structure-based 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 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 (38,40). 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 (39). 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 (39-44). 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 (45). 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 (45).

2.2.3 Cross-correlation analysis

Cross-correlation analysis between the donor and acceptor signal intensities of individual smFRET traces was performed similar to previous studies (46) using custom written Matlab programs. Cross-correlation functions were fit with single-exponentials to obtain the reported transition time constants (that are the inverse of the sum of the forward and backward rate constants).

2.2.4 Isothermal Titration Calorimetry

The calorimetric methods were conducted as described (47) with minor modifications. *T. tengcongensis* $preQ_1$-I riboswitch 33-mer was chemically synthesized (Dharmacon or Fidelity Systems Inc) and purified by reverse phase HPLC (48). The $preQ_1$ ligand was prepared by chemical synthesis (LeadGen Labs, LLC). Lyophilized RNA samples were dissolved in 10 mM HEPES-NaOH, pH 7.0, containing 100 mM NaCl and the solution was heated to 65 °C for 5 min. $MgCl_2$ was added slowly to a final concentration of 6 mM, followed by slow cooling to 24 °C. The folded RNA sample was dialyzed overnight at 4 °C against 4 L of 100 mM NaCl with 6 mM $MgCl_2$ buffered at pH 7.0 by 50 mM HEPES-NaOH, pH 7.0. Following dialysis, riboswitch samples were diluted with dialysis buffer to: 5 μM for the samples analyzed at 25 °C, and 3 μM for the samples analyzed at 60 °C. $PreQ_1$ was dissolved in dialysis buffer to a concentration 7- to 12-fold higher than the RNA. ITC measurements were conducted by use of a VP-ITC calorimeter (MicroCal Inc). Titrations were conducted by injecting $preQ_1$ from the syringe into the riboswitch using 28 or 29 injections of 10 μL each – except for the first injection of 3 μL; an interval of 260 or 300 sec was used between injections. The resulting thermograms were analyzed with Origin 7.0 (MicroCal) using a 1:1 binding model. Titrations were performed in duplicate at each temperature. Representative titrations and curve fits are shown in **Figure 2.7**.

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 (49) and Sparky 3. Uniformly labeled $^{13}C/^{15}N$ samples were prepared by *in vitro* transcription using T7 RNA polymerase as described previously (50). 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.

2.2.6 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 (51) 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 **Tables 2.1**

and 2.2 for restraint details). Temperature replica exchange simulations were then performed in CHARMM (51) with the MMTSB replica exchange server (52) 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 of the *Bsu* preQ$_1$ 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 (21) 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 of the *Tte* preQ$_1$ 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 2.2**. Constraint centers were derived from the 3Q50 crystal structure (33).

2.2.7 Gō-Model RNA Simulations

Gō model RNA simulations were performed essentially as described before (24). 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 (51) 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 (53) 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.

2.3 Results

2.3.1 smFRET detects two conformational states in both ligand-free riboswitches with different transition dynamics

To exploit smFRET (37-39) 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 (21,25,33). 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 (37,39,54), 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 **Figures 2.2A,B**), which survey conformational sampling of the entire population (45), 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 (**Figures 2.2A,B**). 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 the former, in particular, 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 (**Figures 2.2C,D** and **Figures 2.3, 2.5 and 2.8**). Additionally, some molecules appear to stably persist in the high-FRET state with dwell times of up to tens of seconds (**Figure 2.3, panel IV**), similar to observations for other riboswitches (55).

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 on 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**, 100 ms), suggesting that we may miss a significant number of even faster transitions. We therefore performed cross-correlation analysis (46) 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** and **Figure 2.4A**). 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 2.4B**). Even at this faster time resolution, however, 26% of all *Bsu* molecules in the absence of $preQ_1$ reveal no detectable anti-correlation (**Figure 2.4B**), suggesting that at least this

fraction of molecules undergoes transitions that are faster still, similar to observations on the *Fnu* preQ$_1$ riboswitch (28) .

2.3.2 Ligand titrations together with coarse-grained simulations identify the two FRET states as pre-folded and folded

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) (56). The distance of ≤30 Å associated with a FRET value of 0.9 agrees well with the crystal structures of the ligand-bound *Bsu* (21) and *Tte* (32,33) riboswitches, as well as with that of the ligand-free *Tte* riboswitch (33). 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 3'-tail were truly random.

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 (**Figures 2.2A,B**). We fitted each FRET histogram with a sum of two Gaussian functions (**Figures 2.2A,B**) and plotted the fraction of the high-FRET state as a function of ligand concentration (**Figure 2.6A**). 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 2.6A**). 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 2.2 and 2.6**. 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 (31) and 2 nM in 10 mM

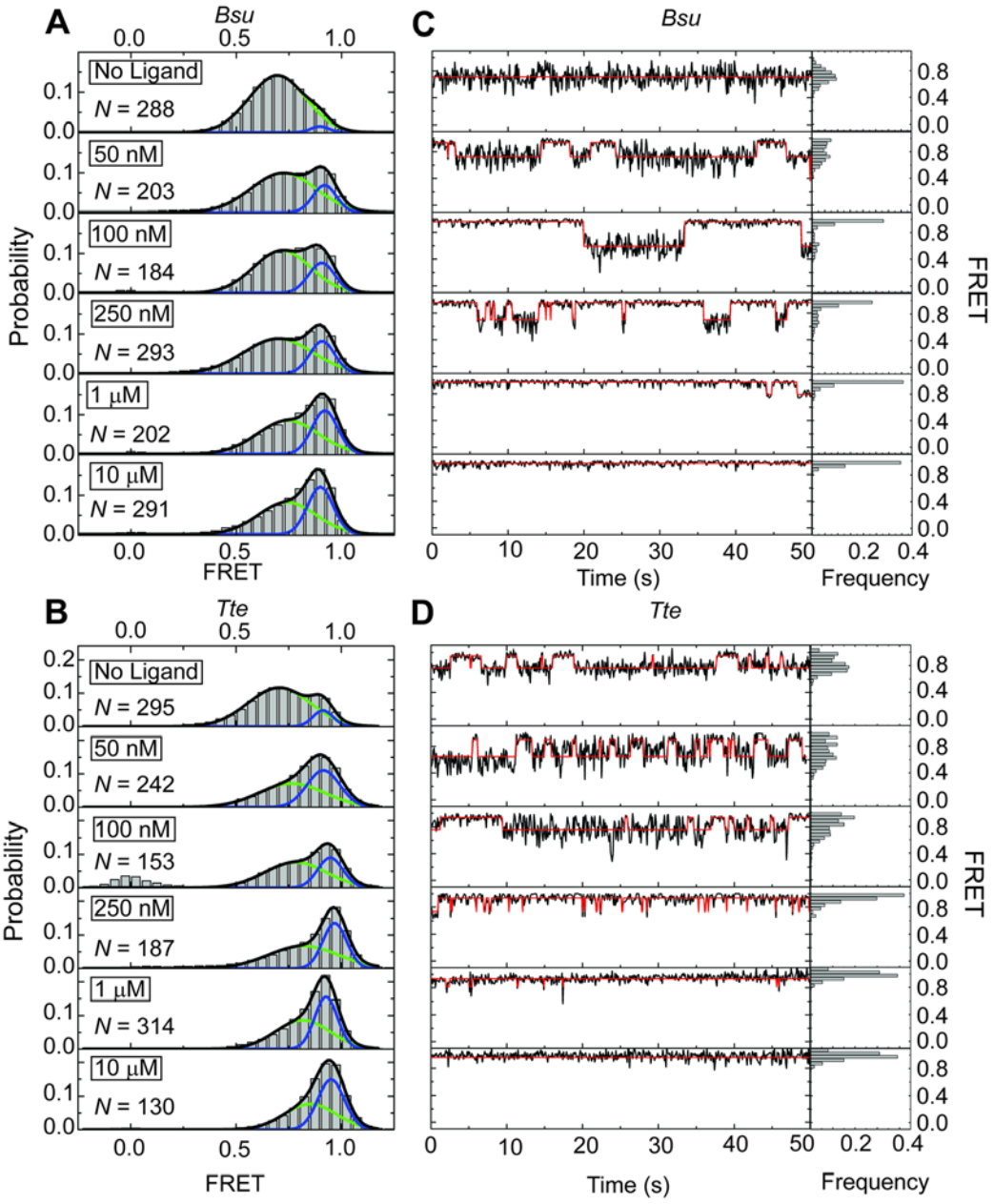

Figure 2.2 smFRET characterization of single preQ$_1$ riboswitch molecules

(A) smFRET histograms of the *Bsu* riboswitch 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. (C) Exemplary FRET time traces of the *Bsu* riboswitch for each condition. Idealized HMM fits are shown as red line. The population of each FRET state is shown as a frequency bar graph to the right. (D) Same as in C, but for the *Tte* riboswitch.

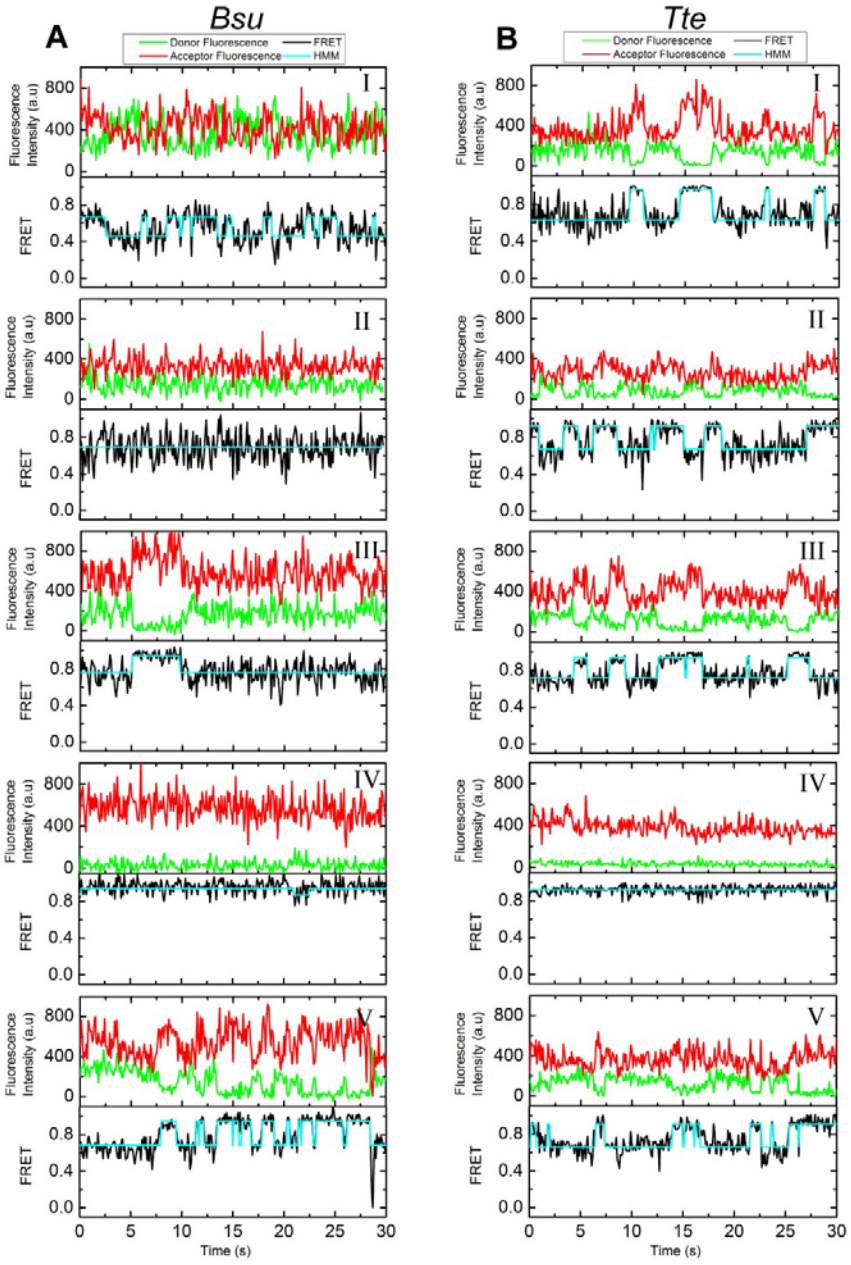

Figure 2.3 Raw smFRET traces of the *Bsu* and *Tte* preQ$_1$ riboswitches in the absence of ligand

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

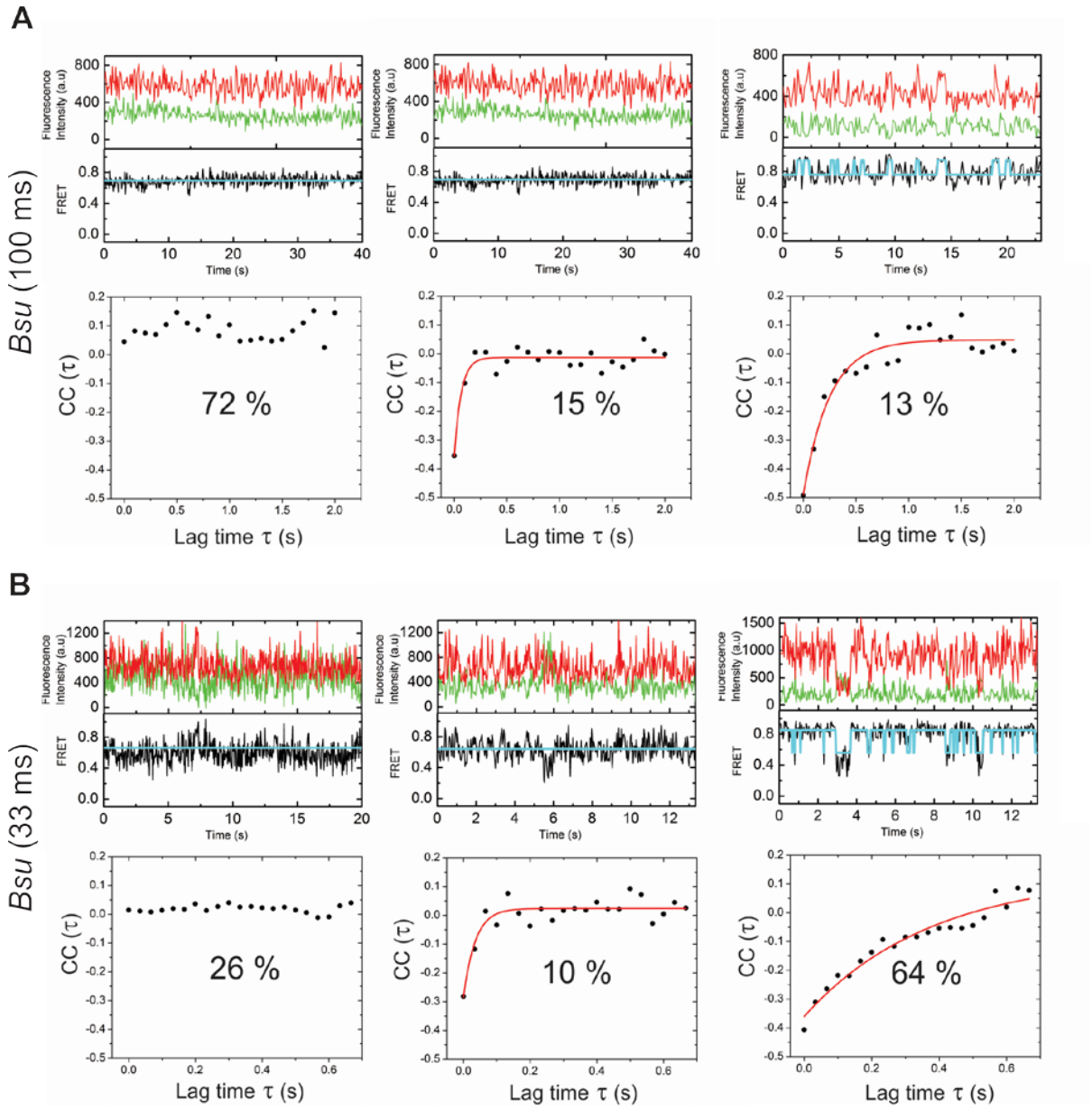

Figure 2.4 Donor-acceptor cross-correlation analysis of exemplary smFRET traces of the *Bsu* riboswitch in the absence of preQ$_1$

(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.

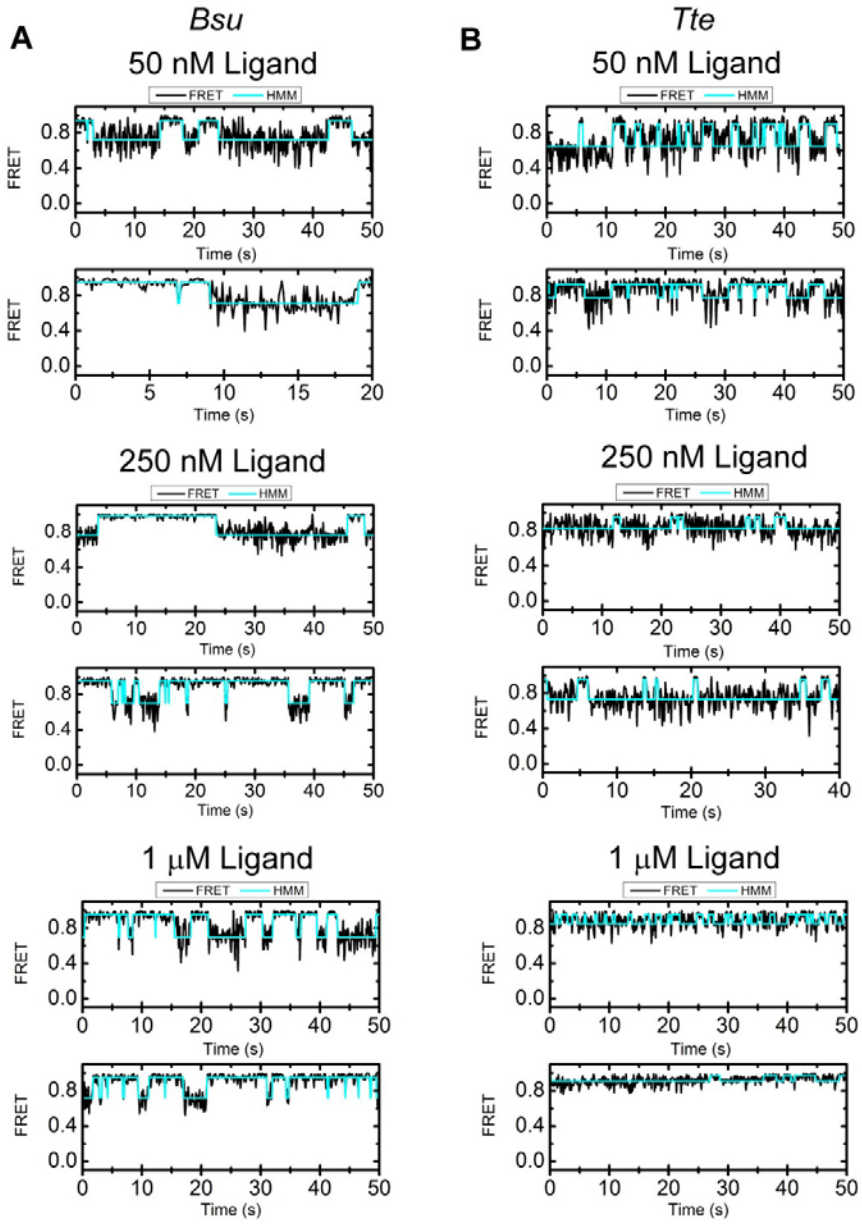

Figure 2.5 Exemplary smFRET traces of the *Bsu* and *Tte* riboswitches in the presence of preQ$_1$ ligand, showing differences in dynamics

(**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.

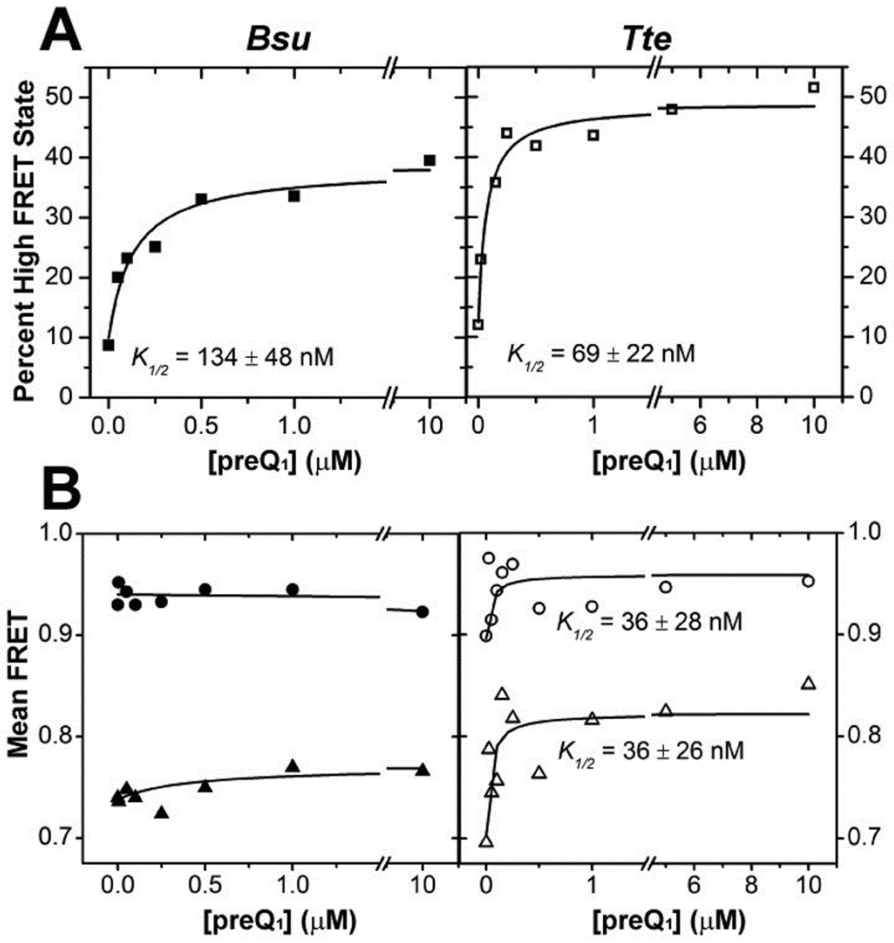

Figure 2.6 Effect of ligand on the distribution of the mid- and high-FRET states

(A) The FRET histograms of **Figure 2.2** 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 2.2** 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.

sodium cacodylate, pH 7.0, 3 mM $MgCl_2$ (33) for the *Bsu* and *Tte* riboswitches, respectively). This may be due to differences in technique (in-line probing and surface immobilization, respectively) and/or buffer conditions, especially pH and Mg^{2+} concentration, which we later show to have a significant effect on the compactness of the *Bsu* riboswitch. Of note, the ligand binding affinity of the *Tte* riboswitch measured here by smFRET and previous work by surface plasmon resonance (SPR) (33) were obtained at room temperature. However, the affinity of the *Tte* riboswitch at physiological growth temperatures of *Thermoanaerobacter tengcongensis* (50 - 80 °C) may be well lower. We therefore used isothermal titration calorimetry (ITC) to investigate the effect of temperature on the ligand binding affinity of the *Tte* riboswitch. The K_D measured at 25 °C is 7.3 ± 2.3 nM (with a reaction stoichiometry N = 0.98), close to the K_D measured using SPR (33). At a temperature of 60 °C, the ligand binding affinity decreased ~60-fold with a K_D of 430 ± 60 nM (**Figure 2.7 and Table 2.3**). The lower N value of 0.75 indicates that not all RNA is competent to bind the ligand at 60 °C. (Please note that the K_D value could not be measured at the optimum *Tte* growth temperature of 75 °C due to instability of the instrument). Our ITC data show that, while the affinity is reduced at high temperature, the *Tte* riboswitch still binds ligand with significant affinity.

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, in part perhaps due to the existence of alternative RNA conformations incapable of binding ligand (25,37). In both riboswitches, ligand addition also causes changes in the transition kinetics between the mid- and high-FRET states (**Figures 2.5 and 2.8**). 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.

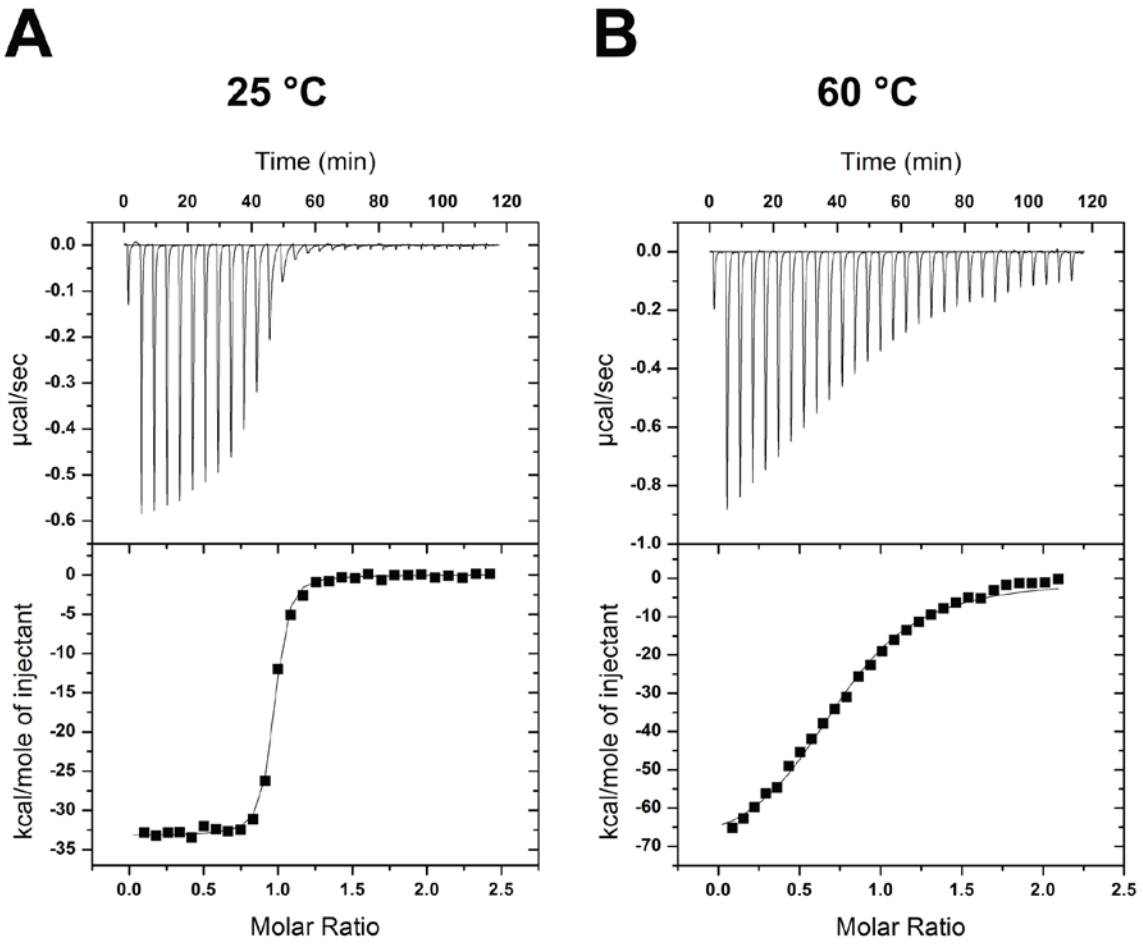

Figure 2.7 Representative isothermal titration calorimetry (ITC) data to measure preQ$_1$ binding to the *Tte* riboswitch

(A) ITC thermogram of preQ$_1$ binding to the *Tte* riboswitch (top) and resulting binding isotherm fitted with a single-site binding model (bottom) at 25 °C. **(B)** Same as in (A), but at 60 °C.

Table 2.3 Average Isothermal titration calorimetry (ITC) binding data for the *Tte* riboswitch[a]

Temp (°C)	K_D (nM)	N	ΔH (kcal mol^{-1})	-TΔS (kcal mol^{-1})	ΔG (kcal mol^{-1})
25	7.4 ± 2.3	0.98 ± 0.07	-41.5 ± 11.7	30.3 ± 11.5	-11.2 ± 0.3
60	425 ± 60	0.75 ± 0.01	-97.9 ± 34.7	88.1 ± 34.6	-9.7 ± 0.1

[a]Values represent the average of two independent titration experiments recorded at each temperature (i.e. n = 2)

To further evaluate the conformations 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 2.9 and Table 2.4**). 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 2.9 and Table 2.4**). 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 (**Figures 2.2A,B**). 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 2.9 and Table 2.4**). Altogether, we have strong evidence that the mid-FRET state in both riboswitches consists of a pre-folded conformational ensemble in which the 3'-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).

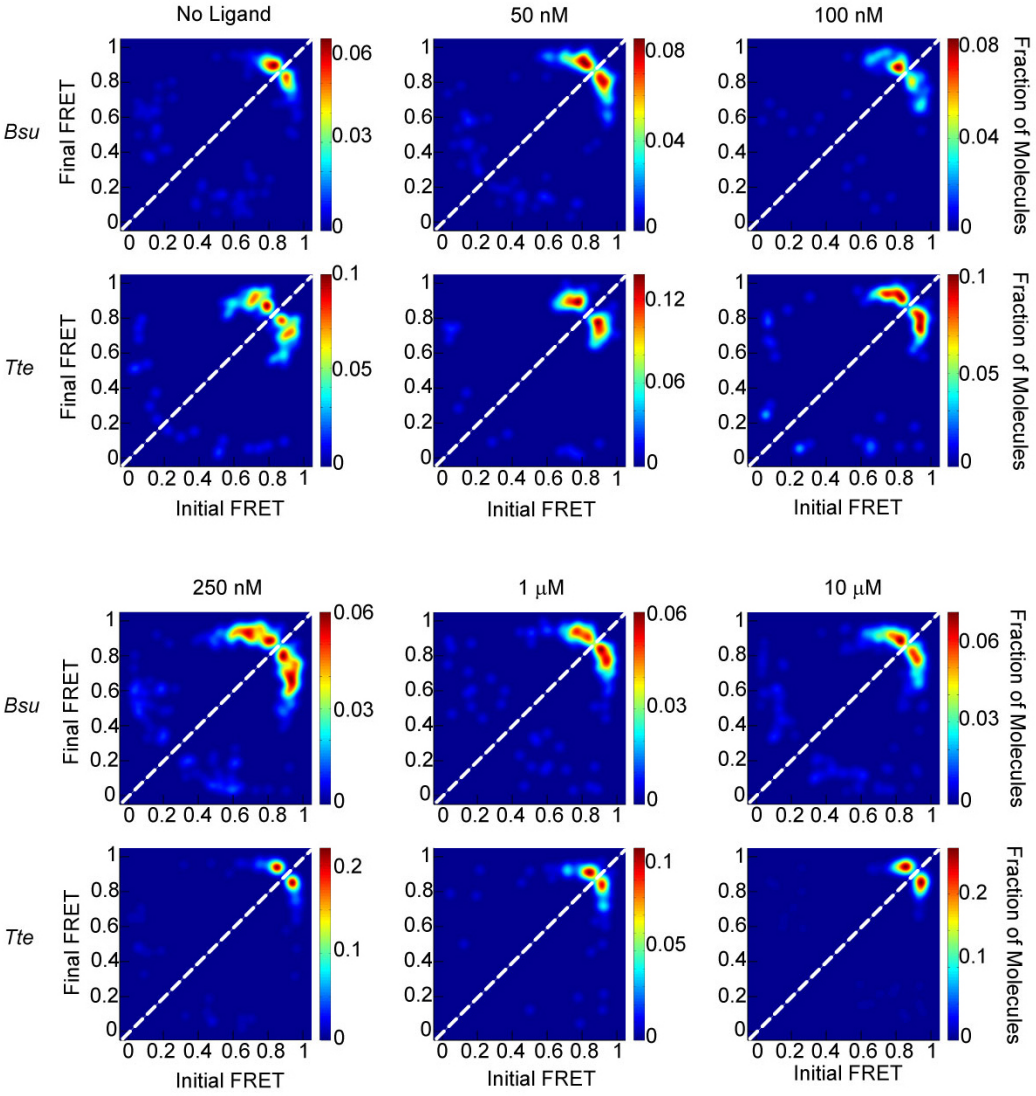

Figure 2.8 Transition occupancy density plots (TODPs) of the preQ$_1$ riboswitches at varying ligand concentrations.

TODPs (45) 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, for the ligand-free *Bsu* riboswitch, the TODP with no ligand displays contours close to the diagonal that move farther away with increasing ligand concentration (until ~250 nM ligand).

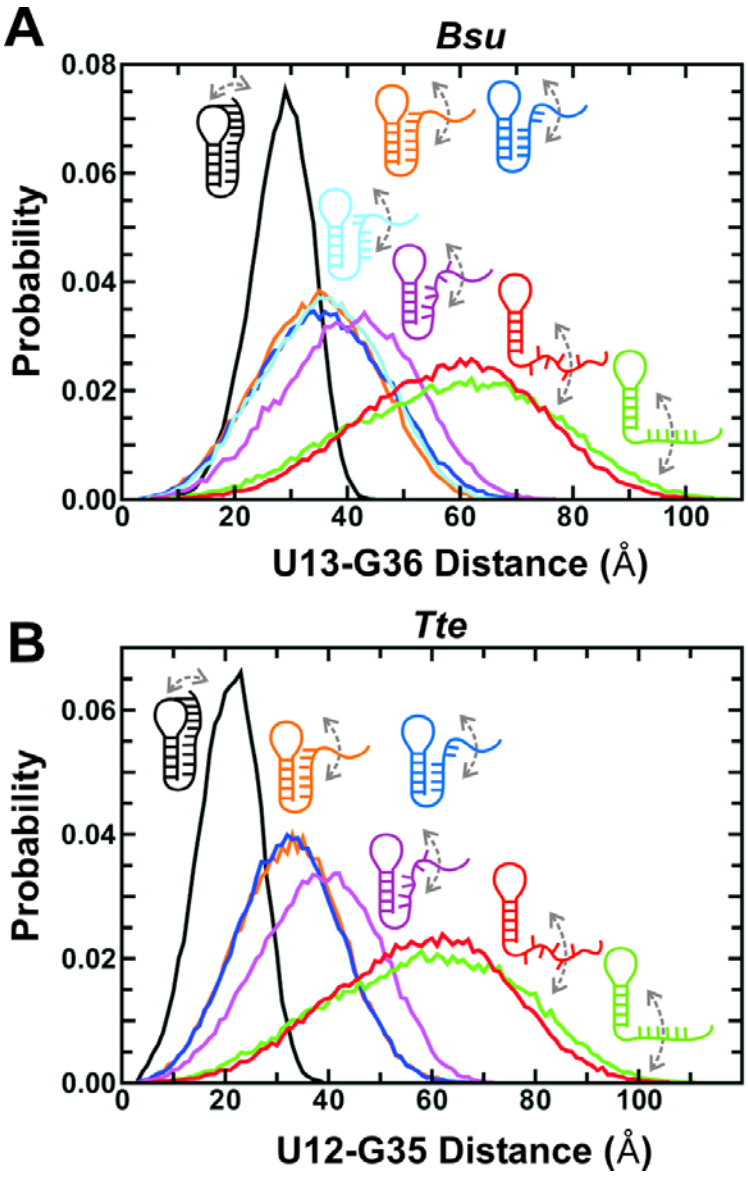

Figure 2.9 TOPRNA simulations of the *Bsu* and *Tte* riboswitches

Coarse grained TOPRNA simulations predicting distance distributions between the fluorophore labeled residues as a function of specific interactions in the *Bsu* (A) and *Tte* (B) riboswitches. Color code is as follows: green, stacked 3' tail; red, unstacked 3' tail; purple, blue, orange and cyan (A, 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 (see **Tables 2.1, 2.2 and 2.4 for details**).

Table 2.4 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 (57):

Here, r is the end-to-end 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.

$$E = \left\langle 1 \Big/ 1 + \left(\frac{r + d_l}{R_0} \right)^6 \right\rangle$$

2.3.3 Differences to previous NMR studies are explained by dimerization and variations in Mg^{2+} concentration and pH

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 (25,36), 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 (33). To resolve this apparent discrepancy, we studied the buffer 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 2.10**). 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 (25,36). 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 2.10**). 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 (26), 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 (33), 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 2.11A**). Similarly, the *Tte* riboswitch

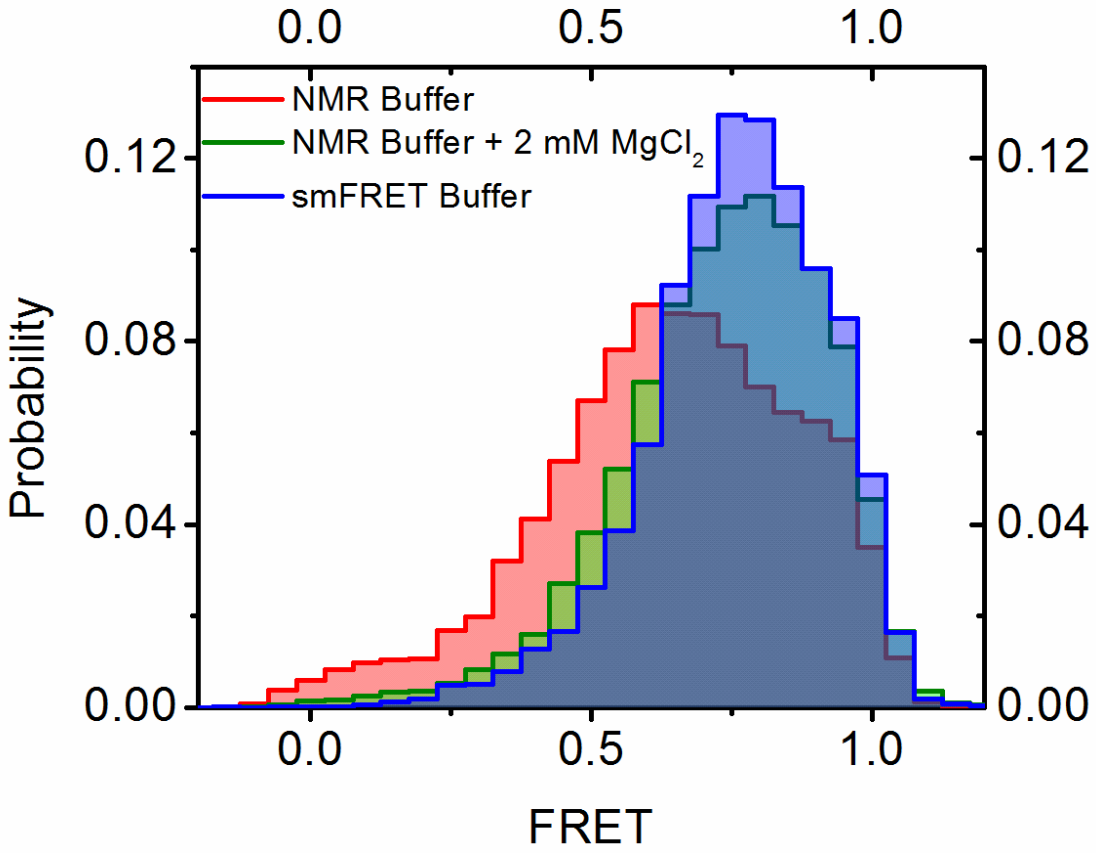

Figure 2.10 Buffer dependence of the ligand-free *Bsu* riboswitch analyzedby smFRET

Low ionic strength NMR buffer (15 mM Na_iPO_4, 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).

shows a FRET peak around 0.71 ± 0.17, accompanied by a peak at 0.94 ± 0.06 (F**igure 2.11B**). The lower mean FRET value and larger width of the *Bsu* mid-FRET state is consistent with our observations in NMR buffer (**Figure 2.10**) and shows that the ligand-free *Tte* riboswitch is more compact than the *Bsu* 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 2.11**), 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 (58,59).

To further reconcile our smFRET data with those from the previous NMR study of the ligand-free *Bsu* riboswitch (25,36), 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**) (36), as observed also for other $preQ_1$ riboswitches (23). A double C12U/C15U mutant eliminates dimer formation at the high concentrations used for NMR (25,36). 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 2.12**), 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 2.12**), 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 (**Figures 2.10 and 2.11**) and together provide strong evidence that the capacity to form

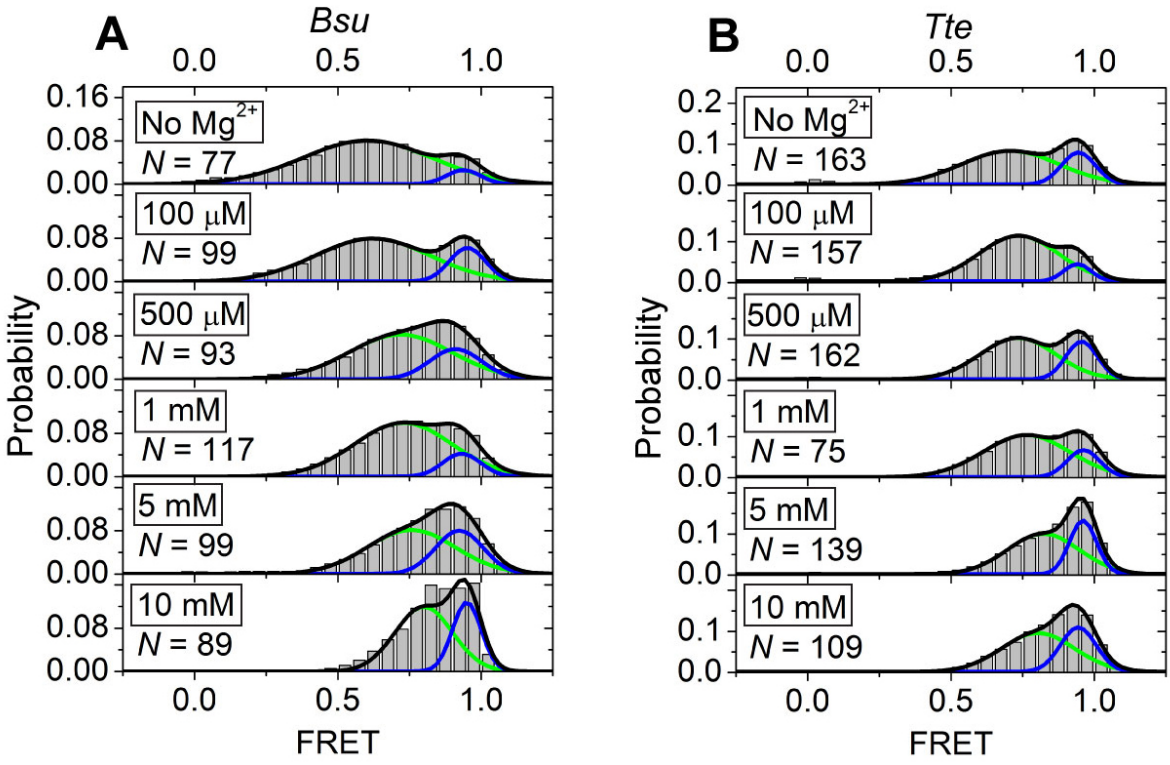

Figure 2.11 Mg^{2+} titration of the *Bsu* (A) and *Tte* (B) riboswitches.

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.

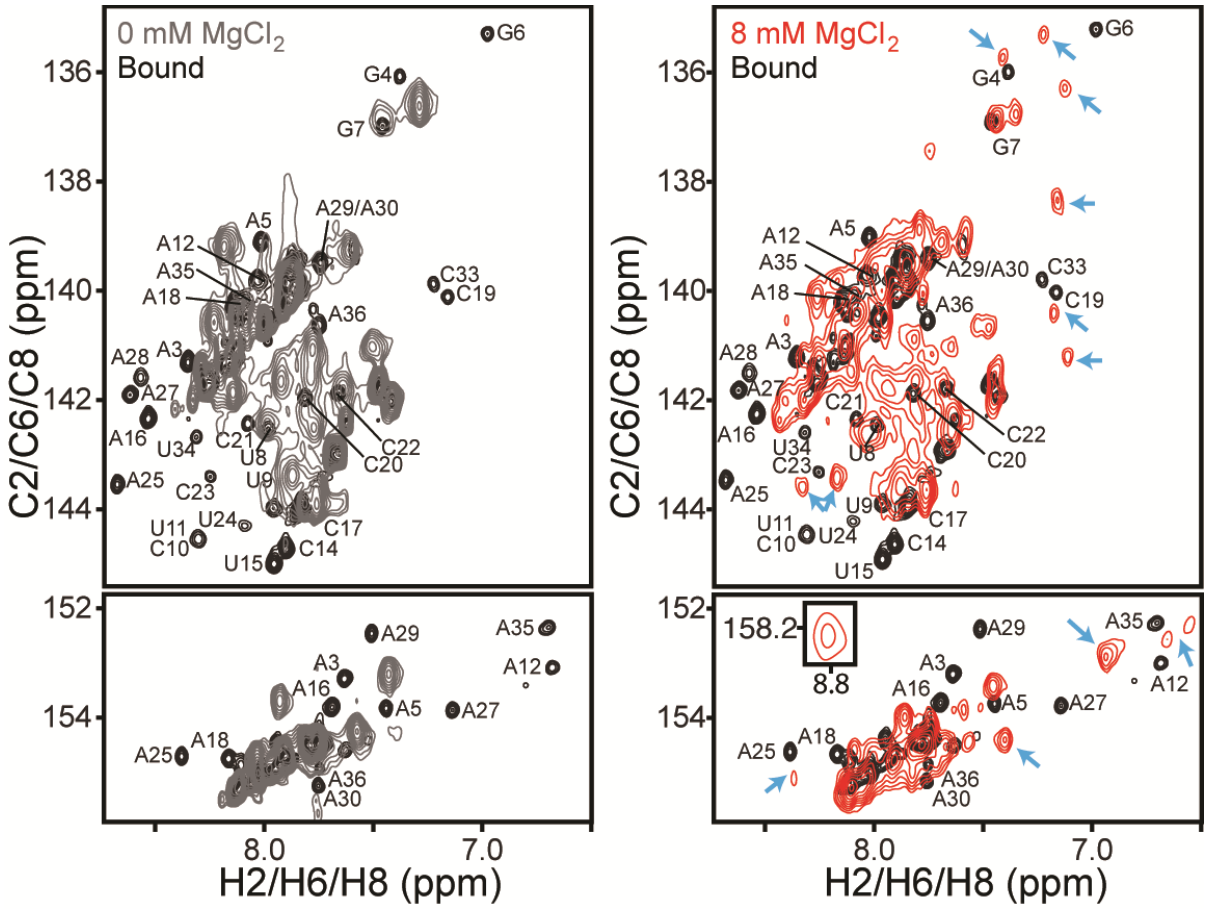

Figure 2.12 NMR characterization of the *Bsu* preQ$_1$ riboswitch – effect of Mg^{2+} on the ligand-free conformation.

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 (25,36).

dimers at the high 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 (28). Therefore, we show that buffer differences, in particular sub-saturating concentrations of Mg^{2+} (**Figures 2.10, 2.11 and 2.12**) and low pH, can have significant effects on the pre-folded state of the *Bsu* riboswitch and account for discrepancies with previous studies (25,36). Finally, we find that saturating Mg^{2+} concentrations have a compacting effect on the riboswitches similar to that of high ligand concentrations (**Figure 2.11**), further highlighting the importance of buffer conditions for RNA folding and tertiary structure.

2.3.4 smFRET provides evidence that the pre-folded *Tte* riboswitch senses $preQ_1$

When fitting the FRET histograms from the $preQ_1$ titration in **Figure 2.2** 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 2.6B**). 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 2.6B**). 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 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 2.6B**). 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 or induced fit, 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 reveal tendencies of the *Bsu* and *Tte* riboswitches towards ligand binding by conformational selection and induced fit, respectively

To better elucidate the differences in *Bsu* and *Tte* riboswitch folding, we utilized Gō-model simulations to compare the folding pathways of the *Bsu* and *Tte* riboswitches (**Figure 2.13**), essentially as described (24). Gō models are biased toward the fully folded state, leading to smooth folding free energy landscapes and have been used to specifically study the folding pathways of biomolecules including RNA (24,60,61). Starting each simulation from a fully unfolded random conformation, the RNA is allowed to fold to its crystallographically defined native pseudoknot structure (as defined by an appropriate potential function) to probe its most likely folding pathway(s). 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 2.13 and 2.14**). 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 of formation of the remaining contacts, however, significantly differs. In the transcriptional *Bsu* riboswitch, ligand binding, stable docking of the 3'-tail, and folding of the P2 stem all occur late and almost concomitantly (**Figure 2.13A**), whereas in the translational *Tte* riboswitch, ligand binding to the top of P1 occurs early, trailed by 3'-tail docking just prior to folding of the P2 stem (**Figure 2.13B**). 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 2.14B**). 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 2.14A**). 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.

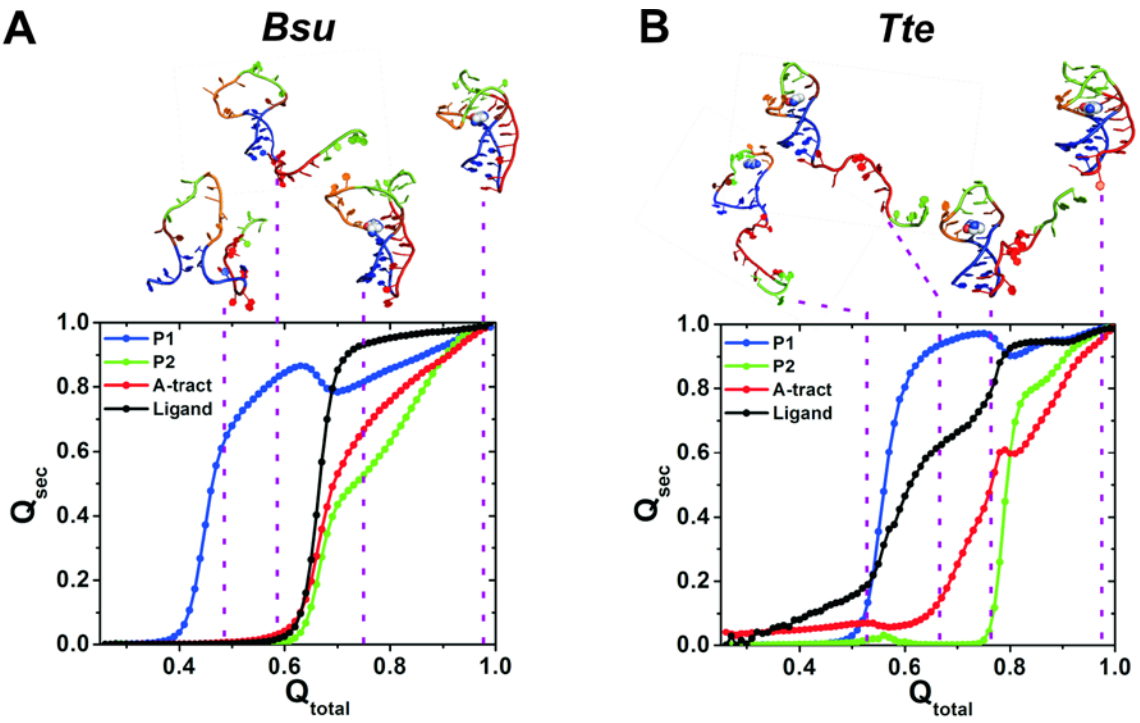

Figure 2.13 Gō model simulations of single *Bsu* (A) and *Tte* (B) riboswitch molecules Fraction of native contacts for each structural component, Q_{sec} (P1, blue; P2, green; A-tract, red; ligand, black), averaged over each 51 simulations and plotted as a function of the fraction of total contacts observed in the native folded structure, Q_{total}. Above, characteristic points along the folding pathway are illustrated with each one representative conformation.

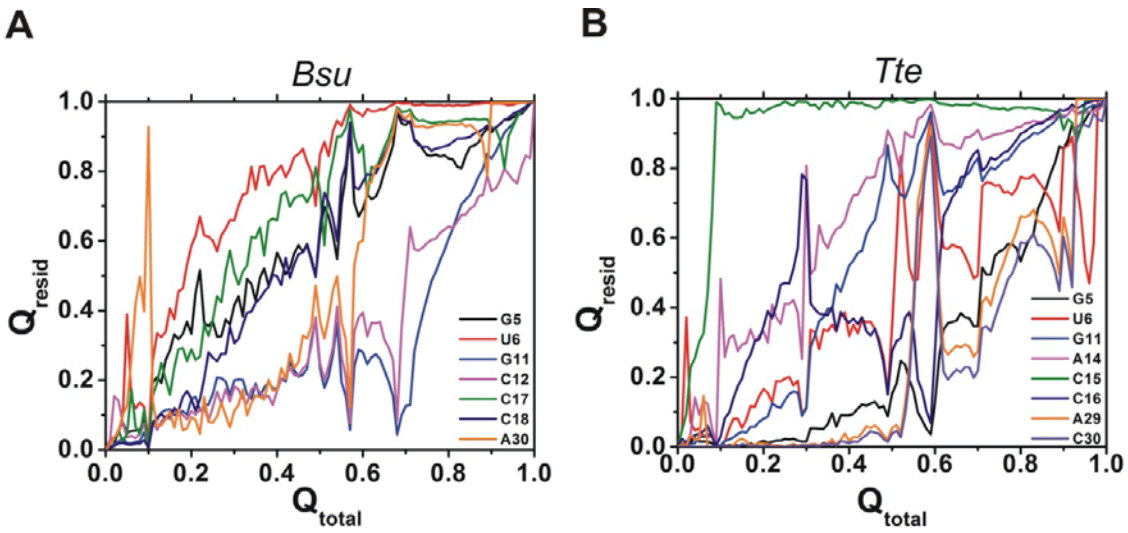

Figure 2.14 Gō model simulations of ligand binding to the *Bsu* (A) and *Tte* (B) riboswitches. The fraction of native contacts, Q_{resid}, formed with preQ$_1$ by each nucleotide, as indicated, is plotted as a function of the fraction of total ligand contacts, Q_{total}. Nucleotides are numbered as in **Figure 2.1C**.

2.3.6 A mutation distal from the binding site impacts ligand binding as predicted by the pre-folded state model

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 (36). 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 positionsof 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 2.15A, 2.16A and 2.6A**), 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 dueto a (partial) loss of 3'-tail rigidity as demonstrated by NMR (36), which our TOPRNA simulations predict will lead to an increase in the inter-fluorophore distance monitored by smFRET (**purple line in Figure 2.9A**). 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, 2.6A** and **Figure 2.15B**), 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,

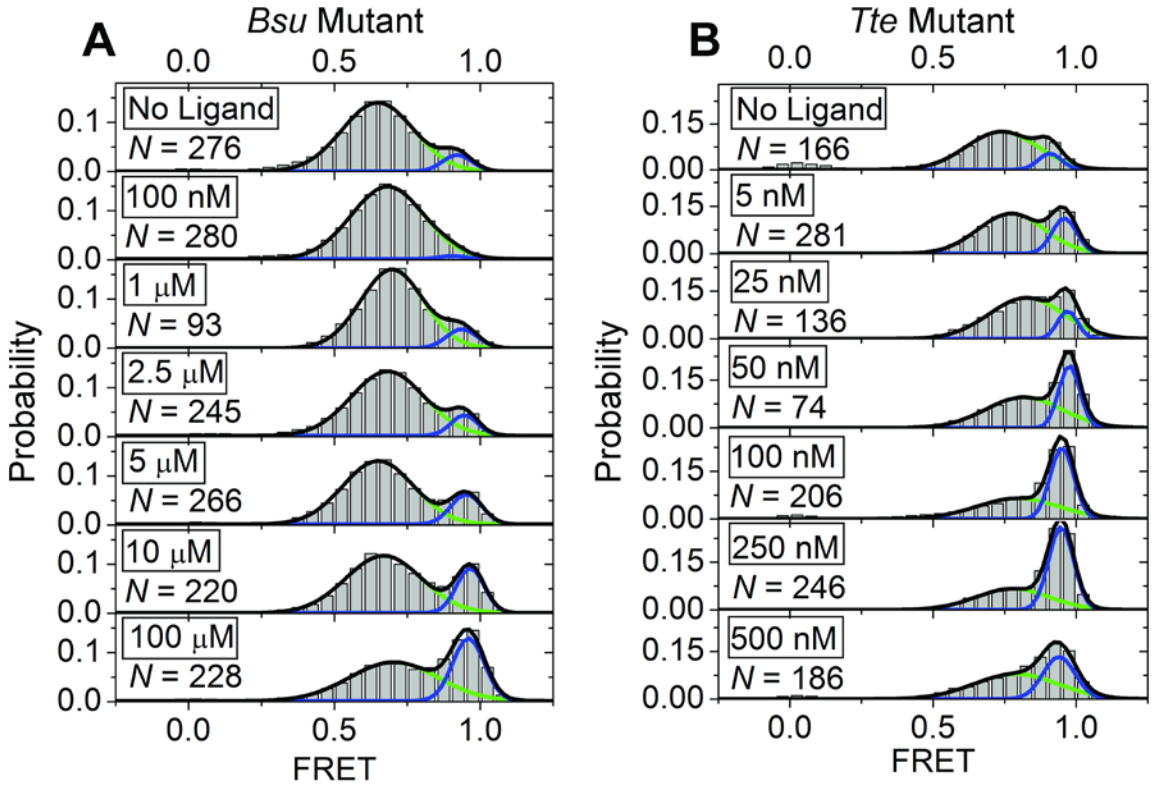

Figure 2.15 smFRET characterization of preQ$_1$ binding to the *Bsu* (A) and *Tte* (B) riboswitch mutants

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.

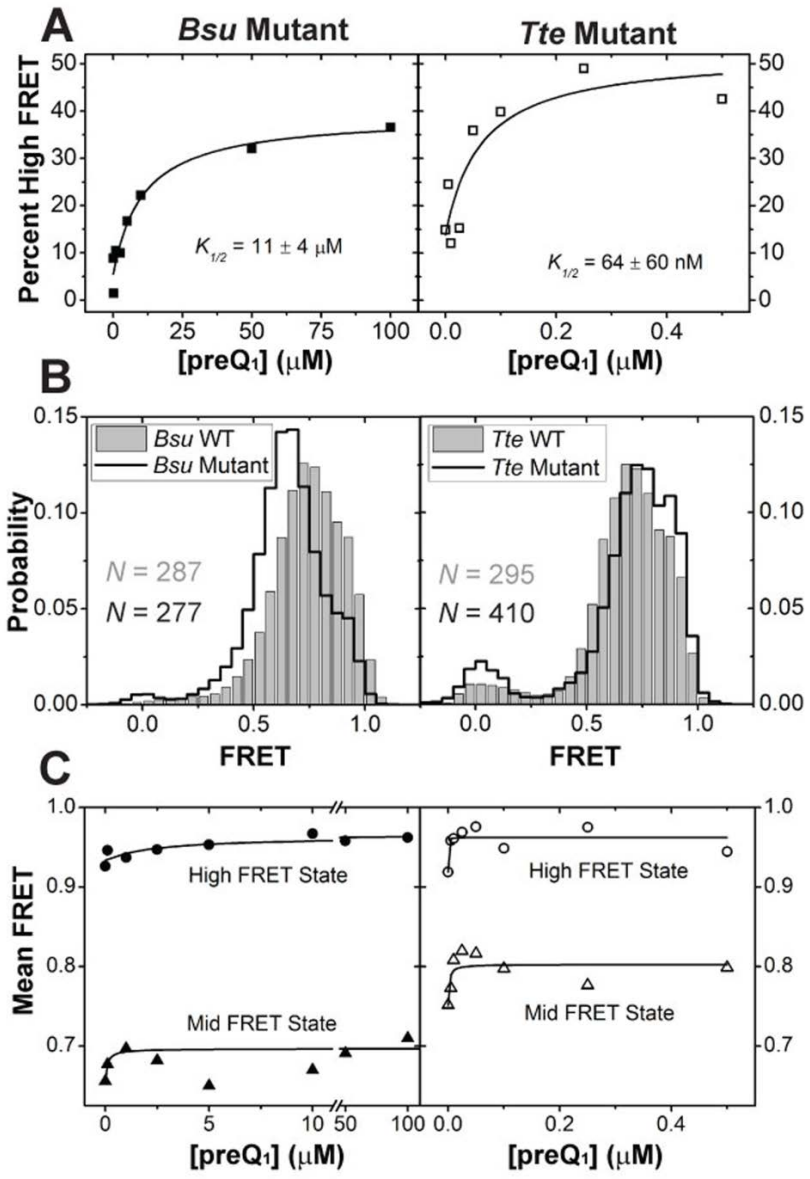

Figure 2.16 smFRET characterization of riboswitch mutants

(A) The Gaussian distributions from **Figure 2.15** 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 2.15** were plotted as a function of ligand concentration for both the *Bsu* (closed symbols) and *Tte* (open symbols) riboswitches.

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. By 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 (25,33). Here, we have used smFRET and both coarse-grained and 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 buffer 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 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 2.17**, where the differences between the

structurally similar transcriptional and translational preQ$_1$ riboswitches reduce to subtle, yet significant, relative shifts in their conformational sampling upon ligand binding. This model finds further support as we show that remote mutations in the 3' A-rich tail, which we previously have shown to diminish or enhance stacking of the A's (36), 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 (25) and the closely related *Fnu* (*Fusobacterium nucleatum*) riboswitch (22,23) 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 in the absence of ligand by NMR. This apparent discrepancy turns out to be due to competing RNA dimerization facilitated by kissing-loop interactions (between U_9AGCUA_{14} in L1 loop of the *Bsu* riboswitch) at the high RNA concentrations used for NMR, as recent studies attest (28,36), as well as differences in buffer conditions as we show by decreasing the Mg^{2+} concentration and pH (**Figures 2.10, 2.11 and 2.12**). 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 (28). 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 (27,29,30).

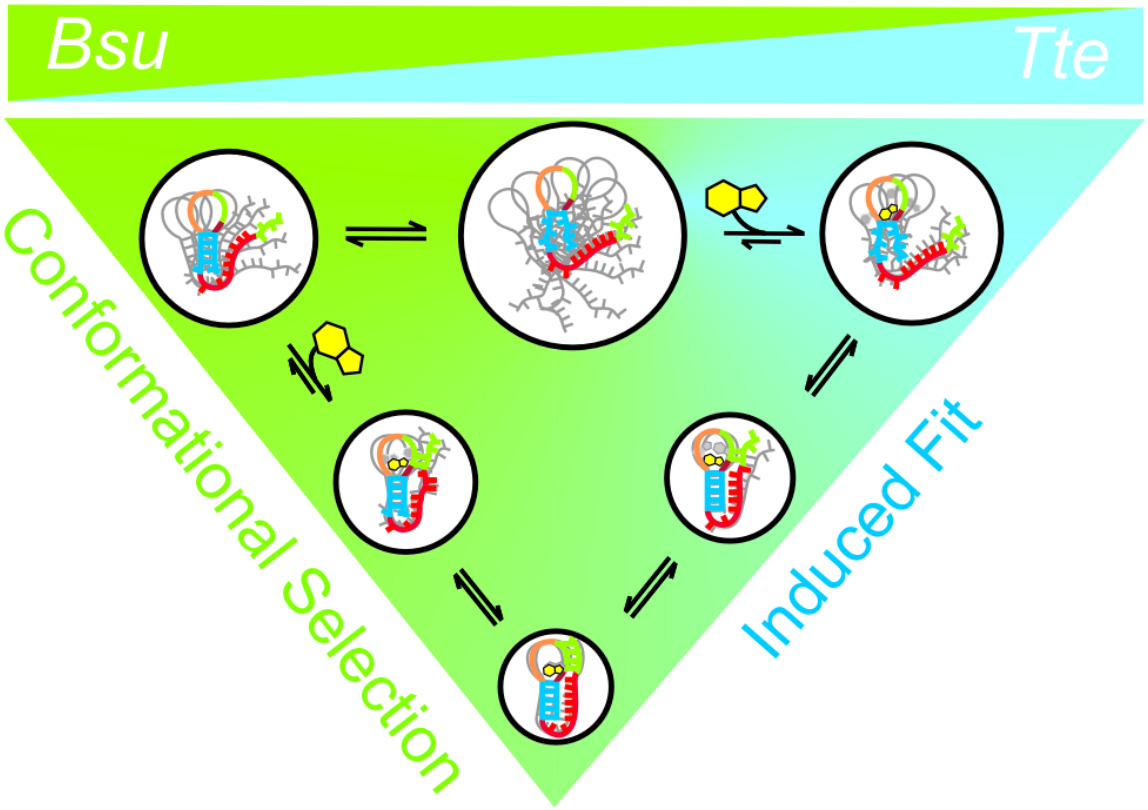

Figure 2.17 Parsimonious folding model of the *Bsu* and *Tte* preQ$_1$ riboswitches

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). 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 (62). The size of the white circle and the gray outlines describe the extent of conformational heterogeneity of each state.

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 (33). 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 transientinteractions 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 2.17**). By contrast, conformational selection (capture) was proposed for the translational SAM-II riboswitch that also folds into an H-type pseudoknot (58). 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 (58), 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 2.17**).

Our model in **Figure 2.17** 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 ~80-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 2.17**), making the two mechanisms notoriously difficult to distinguish (62). 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 (62). In general, a combination of both mechanisms is often utilized in complex biomolecular binding processes and the major mechanism followed depends on many factors including ligand and RNA concentrations (62). Moreover, 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 has to still close to entirely envelop the ligand. Caution is therefore warranted when assigning one or the other mechanism to a specific riboswitch.

A previous SPR study on the ligand binding affinity of the *Tte* riboswitch showed a very tight binding interaction at 25 °C with a K_D of 2 nM (33). Our ITC data at 25 °C show a comparable affinity (K_D = 7.3 nM) when the riboswitch is free in solution, which is similar to the phylogenetically unrelated class II $preQ_1$ riboswitch (K_D = 17.9 nM) analyzed by ITC under the same conditions. The *Tte* riboswitch loses significant affinity (K_D = 425 nM) at temperatures as high as 60 °C (**Figure 2.7 & Table 2.3**) and the affinity will be further decreased at the optimum growth temperature of 75 °C . Despite this decrease, it still binds $preQ_1$ surprisingly well with an affinity comparable to that of the class-I $preQ_1$ riboswitch (K_D = 283 nM) from *Fusobacterium nucleatum* at 25 °C (23). Interestingly, the *Tte*

riboswitch has a shorter L2 loop and lacks two nucleotides that are unresolved in the *Bsu* crystal structure (21), 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, although genetically closely related (63) 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 *add* 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 (64). 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 ligand-free, pre-folded conformation on 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 (65). 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

The authors 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, Dr. Jermaine Jenkins for help collecting ITC data, and Mario Blanco for assistance in the HMM analysis of smFRET data.

Note

The concentrations of $preQ_1$ reported in this chapter and in the published paper are incorrect. The actual concentrations of $preQ_1$ are ~ 6-fold lower than the values mentioned. This does not affect the conclusions of this study in any way and in fact, results in better agreement between the $K_{1/2}$ values obtained in our smFRET measurements and the K_D values measured in the previous study (31).

2.6 References

1. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. and Breaker, R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, **428**, 281-286.
2. Nudler, E. and Mironov, A.S. (2004) The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.*, **29**, 11-17.
3. Mandal, M. and Breaker, R.R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, **5**, 451-463.
4. Winkler, W.C. and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annual review of microbiology*, **59**, 487-517.
5. Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867-879.
6. Mandal, M. and Breaker, R.R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.*, **11**, 29-35.
7. Batey, R.T., Gilbert, S.D. and Montange, R.K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411-415.

8. Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L. and Breaker, R.R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science*, **306**, 275-279.
9. Blount, K.F., Wang, J.X., Lim, J., Sudarsan, N. and Breaker, R.R. (2007) Antibacterial lysine analogs that target lysine riboswitches. *Nat. Chem. Biol.*, **3**, 44-49.
10. Sudarsan, N., Hammond, M.C., Block, K.F., Welz, R., Barrick, J.E., Roth, A. and Breaker, R.R. (2006) Tandem riboswitch architectures exhibit complex gene control functions. *Science*, **314**, 300-304.
11. Barrick, J.E. and Breaker, R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome biology*, **8**, R239.
12. Dann, C.E., 3rd, Wakeman, C.A., Sieling, C.L., Baker, S.C., Irnov, I. and Winkler, W.C. (2007) Structure and mechanism of a metal-sensing regulatory RNA. *Cell*, **130**, 878-892.
13. Regulski, E.E., Moy, R.H., Weinberg, Z., Barrick, J.E., Yao, Z., Ruzzo, W.L. and Breaker, R.R. (2008) A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol. Microbiol.*, **68**, 918-932.
14. Kuchino, Y., Kasai, H., Nihei, K. and Nishimura, S. (1976) Biosynthesis of the modified nucleoside Q in transfer RNA. *Nucleic Acids Res.*, **3**, 393-398.
15. Okada, N., Noguchi, S., Nishimura, S., Ohgi, T., Goto, T., Crain, P.F. and McCloskey, J.A. (1978) Structure determination of a nucleoside Q precursor isolated from E. coli tRNA: 7-(aminomethyl)-7-deazaguanosine. *Nucleic Acids Res.*, **5**, 2289-2296.
16. Harada, F. and Nishimura, S. (1972) Possible anticodon sequences of tRNA His , tRNA Asm , and tRNA Asp from Escherichia coli B. Universal presence of nucleoside Q in the first postion of the anticondons of these transfer ribonucleic acids. *Biochemistry*, **11**, 301-308.
17. Bienz, M. and Kubli, E. (1981) Wild-Type Transfer-Rna Gtyr Reads the Tmv Rna Stop Codon, but Q-Base-Modified Transfer-Rna Gtyr Does Not. *Nature*, **294**, 188-190.
18. Meier, F., Suter, B., Grosjean, H., Keith, G. and Kubli, E. (1985) Queuosine Modification of the Wobble Base in Transfer Rnahis Influences Invivo Decoding Properties. *EMBO J.*, **4**, 823-827.
19. Urbonavicius, J., Qian, O., Durand, J.M.B., Hagervall, T.G. and Bjork, G.R. (2001) Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.*, **20**, 4863-4873.
20. Durand, J.M., Okada, N., Tobe, T., Watarai, M., Fukuda, I., Suzuki, T., Nakata, N., Komatsu, K., Yoshikawa, M. and Sasakawa, C. (1994) vacC, a virulence-associated chromosomal locus of Shigella flexneri, is homologous to tgt, a gene encoding tRNA-guanine transglycosylase (Tgt) of Escherichia coli K-12. *J. Bacteriol.*, **176**, 4627-4634.
21. Klein, D.J., Edwards, T.E. and Ferre-D'Amare, A.R. (2009) Cocrystal structure of a class I preQ$_1$ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.*, **16**, 343-344.

22. Rieder, U., Lang, K., Kreutz, C., Polacek, N. and Micura, R. (2009) Evidence for pseudoknot formation of class I preQ1 riboswitch aptamers. *ChemBioChem*, **10**, 1141-1144.
23. Rieder, U., Kreutz, C. and Micura, R. (2010) Folding of a transcriptionally acting preQ1 riboswitch. *Proc. Natl Acad. Sci. USA*, **107**, 10804-10809.
24. Feng, J., Walter, N.G. and Brooks, C.L., III. (2011) Cooperative and directional folding of the $preQ_1$ riboswitch aptamer domain. *J. Am. Chem. Soc.*, **133**, 4196-4199.
25. Kang, M., Peterson, R. and Feigon, J. (2009) Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol. Cell*, **33**, 784-790.
26. Zhang, Q., Kang, M., Peterson, R.D. and Feigon, J. (2011) Comparison of solution and crystal structures of preQ1 riboswitch reveals calcium-induced changes in conformation and dynamics. *J. Am. Chem. Soc.*, **133**, 5190-5193.
27. Petrone, P.M., Dewhurst, J., Tommasi, R., Whitehead, L. and Pomerantz, A.K. (2011) Atomic-scale characterization of conformational changes in the preQ(1) riboswitch aptamer upon ligand binding. *J. Mol. Graph. Model.*, **30**, 179-185.
28. Santner, T., Rieder, U., Kreutz, C. and Micura, R. (2012) Pseudoknot preorganization of the $preQ_1$ class I riboswitch. *J. Am. Chem. Soc.*, **134**, 11928-11931.
29. Gong, Z., Zhao, Y., Chen, C. and Xiao, Y. (2012) Computational study of unfolding and regulation mechanism of preQ1 riboswitches. *PLoS One*, **7**, e45239.
30. Banas, P., Sklenovsky, P., Wedekind, J.E., Sponer, J. and Otyepka, M. (2012) Molecular mechanism of $preQ_1$ riboswitch action: a molecular dynamics study. *J. Phys. Chem. B*, **116**, 12721-12734.
31. 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. *et al.* (2007) A riboswitch selective for the queuosine precursor $preQ_1$ contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.*, **14**, 308-317.
32. Spitale, R.C., Torelli, A.T., Krucinska, J., Bandarian, V. and Wedekind, J.E. (2009) The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain. *J. Biol. Chem.*, **284**, 11012-11016.
33. Jenkins, J.L., Krucinska, J., McCarty, R.M., Bandarian, V. and Wedekind, J.E. (2011) Comparison of a $preQ_1$ riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.*, **286**, 24626-24637.
34. Capriotti, E. and Marti-Renom, M.A. (2008) RNA structure alignment by a unit-vector approach. *Bioinformatics*, **24**, I112-I118.
35. Leontis, N.B. and Westhof, E. (2003) Analysis of RNA motifs. *Curr. Opin. Struct. Biol.*, **13**, 300-308.
36. Eichhorn, C.D., Feng, J., Suddala, K.C., Walter, N.G., Brooks, C.L., III and Al-Hashimi, H.M. (2012) Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch. *Nucleic Acids Res.*, **40**, 1345-1355.
37. Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science*, **296**, 1473-1476.

38. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single-molecule FRET. *Nature methods*, **5**, 507-516.
39. Abelson, J., Blanco, M., Ditzler, M.A., Fuller, F., Aravamudhan, P., Wood, M., Villa, T., Ryan, D.E., Pleiss, J.A., Maeder, C. *et al.* (2010) Conformational dynamics of single pre-mRNA molecules during in vitro splicing. *Nat Struct Mol Biol*, **17**, 504-512.
40. Rueda, D., Bokinsky, G., Rhodes, M.M., Rust, M.J., Zhuang, X. and Walter, N.G. (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc. Natl Acad. Sci. USA*, **101**, 10066-10071.
41. Ditzler, M.A., Rueda, D., Mo, J., Hakansson, K. and Walter, N.G. (2008) A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res.*, **36**, 7088-7099.
42. Pereira, M.J., Nikolova, E.N., Hiley, S.L., Jaikaran, D., Collins, R.A. and Walter, N.G. (2008) Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis. *J. Mol. Biol.*, **382**, 496-509.
43. de Silva, C. and Walter, N.G. (2009) Leakage and slow allostery limit performance of single drug-sensing aptazyme molecules based on the hammerhead ribozyme. *RNA*, **15**, 76-84.
44. McDowell, S.E., Jun, J.M. and Walter, N.G. (2010) Long-range tertiary interactions in single hammerhead ribozymes bias motional sampling toward catalytically active conformations. *RNA*, **16**, 2414-2426.
45. Blanco, M. and Walter, N.G. (2010) Analysis of complex single-molecule FRET time trajectories. *Methods Enzymol*, **472**, 153-178.
46. Ragunathan, K., Liu, C. and Ha, T. (2012) RecA filament sliding on DNA facilitates homology search. *Elife*, **1**, e00067.
47. Liberman, J.A., Salim, M., Krucinska, J. and Wedekind, J.E. (2013) Structure of a class II preQ1 riboswitch reveals ligand recognition by a new fold. *Nat. Chem. Biol.*, **9**, 353-355.
48. Wedekind, J.E. and McKay, D.B. (2000) Purification, crystallization, and X-ray diffraction analysis of small ribozymes. *Methods Enzymol.*, **317**, 149-168.
49. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.
50. Zhang, Q., Sun, X., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, **311**, 653-656.
51. Brooks, B.R., Brooks, C.L., III, Mackerell, A.D., Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545-1614.
52. Feig, M., Karanicolas, J. and Brooks, C.L., III. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.*, **22**, 377-395.
53. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.*, **26**, 1701-1718.

54. Lund, K., Manzo, A.J., Dabby, N., Michelotti, N., Johnson-Buck, A., Nangreave, J., Taylor, S., Pei, R., Stojanovic, M.N., Walter, N.G. *et al.* (2010) Molecular robots guided by prescriptive landscapes. *Nature*, **465**, 206-210.
55. Wood, S., Ferre-D'Amare, A.R. and Rueda, D. (2012) Allosteric tertiary interactions preorganize the c-di-GMP riboswitch and accelerate ligand binding. *ACS Chem. Biol.*, **7**, 920-927.
56. Sekella, P.T., Rueda, D. and Walter, N.G. (2002) A biosensor for theophylline based on fluorescence detection of ligand-induced hammerhead ribozyme cleavage. *RNA*, **8**, 1242-1252.
57. Schuler, B., Lipman, E.A., Steinbach, P.J., Kumke, M. and Eaton, W.A. (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA*, **102**, 2754-2759.
58. Haller, A., Rieder, U., Aigner, M., Blanchard, S.C. and Micura, R. (2011) Conformational capture of the SAM-II riboswitch. *Nat. Chem. Biol.*, **7**, 393-400.
59. Chen, B., Zuo, X., Wang, Y.X. and Dayie, T.K. (2012) Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res.*, **40**, 3117-3130.
60. Clementi, C., Nymeyer, H. and Onuchic, J.N. (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, **298**, 937-953.
61. Hills, R.D., Jr. and Brooks, C.L., 3rd. (2009) Insights from coarse-grained go models for protein folding and dynamics. *Int. J. Mol. Sci.*, **10**, 889-905.
62. Hammes, G.G., Chang, Y.C. and Oas, T.G. (2009) Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl Acad. Sci. USA*, **106**, 13737-13741.
63. Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y. *et al.* (2002) A complete sequence of the T. tengcongensis genome. *Genome Res.*, **12**, 689-700.
64. Neupane, K., Yu, H., Foster, D.A., Wang, F. and Woodside, M.T. (2011) Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res.*, **39**, 7677-7687.
65. Nickson, A.A. and Clarke, J. (2010) What lessons can be learned from studying the folding of homologous proteins? *Methods*, **52**, 38-50.

CHAPTER 3

Probing the ligand binding mechanism of the $preQ_1$-I riboswitch using single molecule FRET

3.1 Introduction

Riboswitches are metabolite binding motifs in messenger RNAs (mRNA) that control gene expression (1-3). They are generally present in the 5'-untranslated regions of mRNAs and control expression of proteins associated with a given metabolite, by sensing the intracellular concentration changes of the metabolite. Riboswitches consist of an aptamer domain that directly binds the ligand and a downstream expression platform that controls gene expression. Ligand binding stabilizes one of the multiple conformations of the aptamer, leading to structural rearrangement of the expression platform and modulation of gene expression (1). Modulation of gene expression generally occurs via Rho-independent transcription termination or repression of translation inhibition. Riboswitches affect gene expression by coupling ligand binding to RNA conformational changes (4). Crystal structures of a number of ligand bound aptamer domains have been determined that showed the principles of ligand recognition by RNA (2). In addition, high-resolution structures of a few ligand-free aptamer domains showed that, in a crystal, they closely resemble the ligand-bound structures with only local conformational differences. However, due to the lack of ligand-free structures for many riboswitches and a general uncertainty concerning their conformational ensemble in solution, the mechanism of ligand-mediated folding is not well understood (7).

Krishna C. Suddala performed the smFRET experiments and data analysis of the *Bsu* riboswitch $preQ_1$ titration in the presence and absence of Mg^{2+}. Jiarui Wang performed the smFRET experiments and data analysis of ligand ($preQ_1$, $preQ_0$ and guanine) comparisons on the *Bsu* riboswitch conformational dynamics.

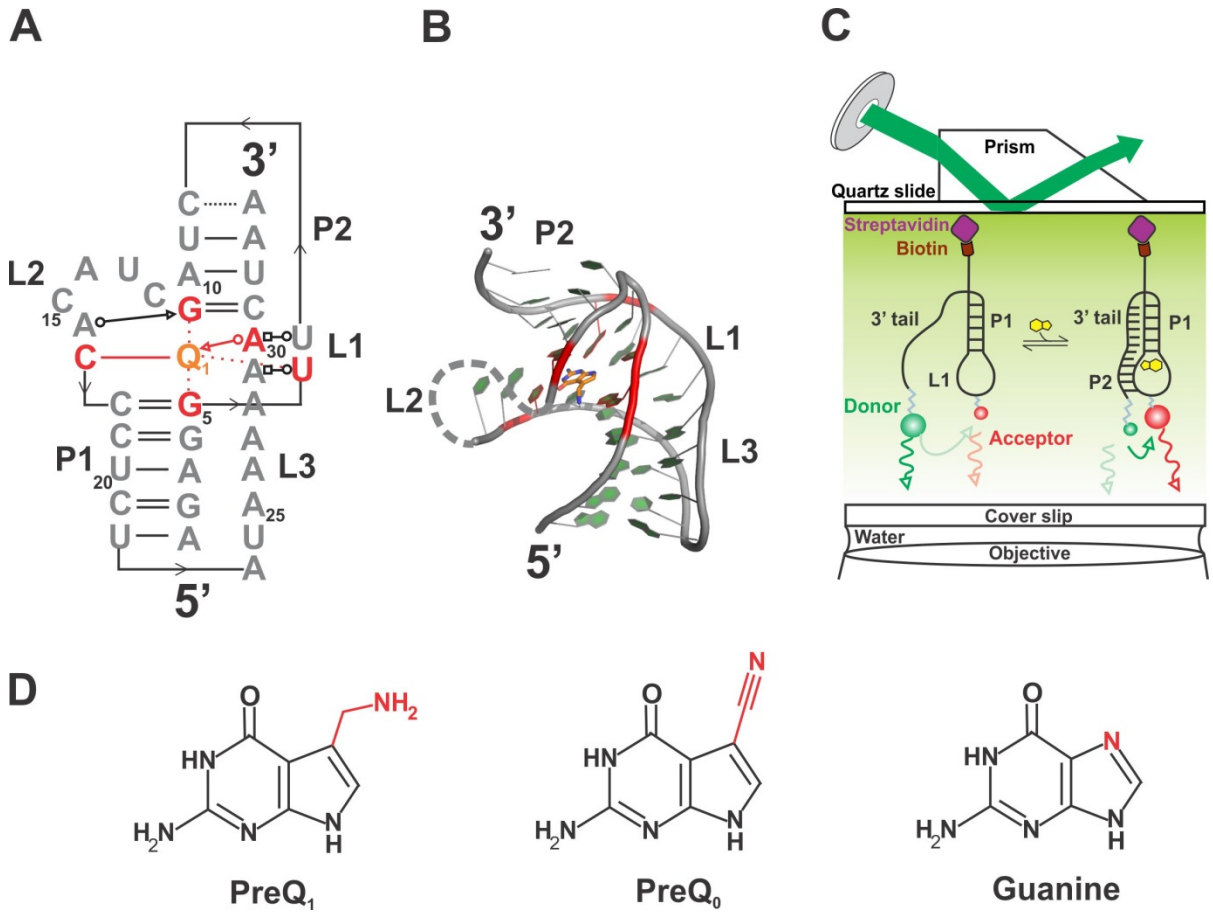

Figure 3.1 Sequence and structure of the *Bsu* preQ1 riboswitch

(A) Sequence and secondary structure of the *Bsu* $preQ_1$ riboswitch showing the important intra-molecular tertiary interactions and inter-molecular interactions with ligand using Leontis-Westhof nomenclature (5). (B) Crystal structure of the *Bsu* $preQ_1$ riboswitch (grey, PDB ID: 3FU2, chain A) (6). The ligand, $preQ_1$ is shown in stick model and the nucleotides interacting with the ligand are shown in red. (C) Prism-based TIRFM setup for performing smFRET experiments. (D) Chemical structures of $preQ_1$ and the related ligands $preQ_0$ and guanine, with differences highlighted in red.

Recently, studies of ligand-free conformations using nuclear magnetic resonance (NMR) spectroscopy (8-11), small-angle X-ray scattering (SAXS) (12-15), computational simulations (16-19) and single molecule techniques (10,17,20-23) have provided valuable insight into the structural and dynamic properties of riboswitches. In particular, due to their ability to detect transient and lowly populated conformations, single molecule methods are being applied extensively to decipher the conformational properties of riboswitches (17). In addition, the equilibrium distribution of different states, their lifetimes and interconversion between different conformations can be directly observed using single molecule techniques (23-28). Riboswitches that act through transcription termination have a small time window during which they need to recognize ligand and fold into the right conformation to affect gene regulation (29,30). Therefore, studying the kinetics of ligand-dependent conformational dynamics of transcriptional riboswitches is important. In addition, probing the effect(s) of ligand on the folding rate can reveal the mechanism of ligand binding by riboswitches, a phenomenon that is not well understood (30,31).

The $preQ_1$ (7-aminomethyl-7-deaza-guanine) riboswitch from *B. subtilis* (henceforth referred to as *Bsu* riboswitch) has the smallest known aptamer domain, containing only 34 nucleotides (nt), and controls the expression of proteins involved in the biosynthesis of queuosine through a transcriptional attenuation mechanism (32). The crystal and NMR structures of the aptamer domain bound to $preQ_1$ show that it forms a compact H-type pseudoknot fold with a stable 5-nt stem P1, loops L1-L3 and a 4-nt stem P2 that is stabilized by the ligand (**Figure 3.1A**) (6,33). $PreQ_1$ is similar to guanine and, therefore, the ligand is recognized through Watson-Crick base pairing by conserved nucleotide C17 in loop L1. In addition, the ligand stacks in between two guanines, G5 and G11, that form the top and bottom base-pairs of stems P1 and P2, respectively. In the crystal structure, the aminomethyl group of $preQ_1$ (**Figure 3.1D**) forms three hydrogen bonds with G5, G11 and a hydration water. In addition, residues from loops L1 and L3 are involved in hydrogen bonding with different functional groups of the $preQ_1$. Therefore, the binding pocket is formed jointly by L2, L1, the top and bottom of P1 and P2 and the A-rich L3, enclosing ~92 % of the solvent accessible surface area of the ligand (**Figure 3.1B**) (6). The exocyclic aminomethyl group of

the ligand is the only solvent accessible region seen in the solution structure. The alternative ligand $preQ_0$, a precursor of the biosynthesis of $preQ_1$, has a linear nitrile group (**Figure 3.1D**) that cannot donate the same hydrogen bond and sticks out the binding site (33). These features probably explain why $preQ_0$ has a 5-fold lower affinity than $preQ_1$ (32).

Initial studies on the ligand-free *Bsu* riboswitch suggested that it exists in an extended conformation where the 3'-tail (which forms L3 and the 3' portion of P2 in the ligand bound structure) does not form any tertiary interactions with stem-loop P1-L1 (33). However, using single molecule fluorescence energy transfer (smFRET), NMR spectroscopy and computational simulations we have previously shown that the ligand-free *Bsu* riboswitch exists in a pre-folded conformation where the 3'-tail already forms transient interactions with P1 and/or L1 (17). We also have demonstrated the important role of Mg^{2+} ions in promoting compact folded-like conformations, even in the absence of ligand. In the presence of ligand, we directly observed that the *Bsu* riboswitch shows transitions between the 'pre-folded' and folded states. Furthermore, based on our smFRET data and Gō-model simulations, we proposed that $preQ_1$ binds to folded-like conformations of the *Bsu* riboswitch through a pathway resembling a 'conformational selection' mechanism (16,17)

In general, biomolecular recognition (ligand binding) can be characterized into two major contrasting mechanisms – induced fit and conformational selection (34-37) (**Figure 3.2A**). The induced fit mechanism posits that ligand binding to an 'open' conformation drives the receptor (riboswitch) into the folded conformation. In contrast, the conformational selection mechanism states that the apo (ligand-free) receptor exists in multiple conformations that are in equilibrium with one another, including a minor population of 'native' or folded-like conformations. The ligand 'captures' specifically these native-like conformations, thereby shifting the equilibrium to the final folded state. However, these mechanisms are not mutually exclusive and a given ligand-mediated folding event can proceed through both mechanisms to different extents depending on many factors, such as

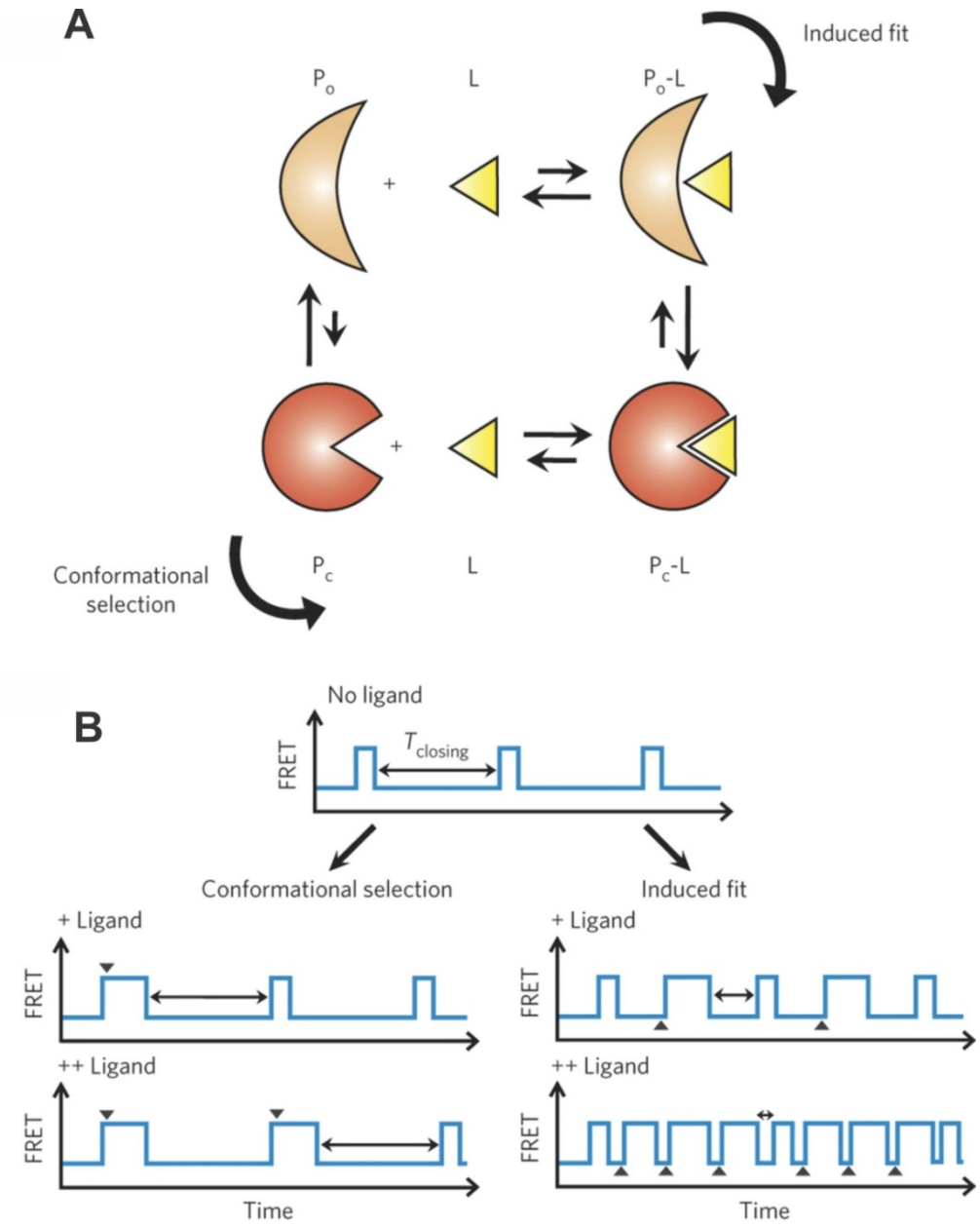

Figure 3.2 Molecular mechanisms of ligand binding coupled to folding

A) Thermodynamic cycles for conformational selection and induced-fit mechanisms. Po, Pc and L indicate the receptor's open conformation, closed conformation and a ligand, respectively. (B) The expected time traces of the FRET efficiency in smFRET measurements for the respective recognition mechanisms. $T_{closing}$ indicates the time for a receptor to undergo a conformational change from an open to a closed state in the absence or presence of varying ligand concentrations. Plus symbols represent the ligand concentration. Arrowheads indicate the binding of a ligand. Figure reproduced with permission from Kim *et al.* (31).

the relative concentrations of the receptor, ligand and rate constants of ligand binding (35,38). In addition, it is important to note that the mere presence of a ligand-free 'folded-like' conformation does not necessarily confirm the conformational selection mechanism; it only suggests the possibility of conformational selection (35).

Our understanding of the principles of biomolecular recognition processes has been majorly expanded by powerful NMR spectroscopic studies on a number of proteins and nucleic acid systems (34,39). Recent studies suggested that to accurately classify a binding interaction into one of these two mechanisms, calculations of flux through each pathway have to be carried out, a task that is extremely difficult and therefore realized only very recently (35,38). However, the major mechanism under a given set of conditions can be dissected by studying the influence of ligand on the kinetic signature of conformational dynamics (35,40,41). Kinetic assays have been used to study the mechanism of binding, mainly in proteins (40). Generally, in ensemble binding assays an increasing observed rate, k_{obs}, as a function of ligand concentration suggests an induced-fit like mechanism while a decreasing k_{obs} supports a conformational selection pathway (41). More recently, smFRET microscopy was applied to study the kinetics of ligand-dependent conformational dynamics of the maltose binding protein (MBP) as a model system (31). The study showed that ligand binding proceeds mainly through an induced-fit like pathway. In addition, using a fluorophore labeled ligand for three-color smFRET, the authors directly observed that ligand binds to both 'open' and 'closed' conformations of MBP and demonstrated that ~80 % of binding events occur in the open state, further supporting an induced-fit like mechanism (31). This study shows the great potential of smFRET microscopy for studying ligand-dependent folding mechanisms of biomolecules that are difficult to characterize using ensemble methods (42).

Here, we have characterized the kinetics of *Bsu* riboswitch conformational dynamics and investigated the effect of ligand on them, with the aim to decipher the mechanism of ligand binding. Using smFRET, we have studied the ligand dependent conformational dynamics of the transcriptionally acting *Bsu* $preQ_1$ riboswitch. In addition, we have also

directly monitored the efficacy of non-cognate ligands that are similar to $preQ_1$ in promoting folding of the riboswitch. SmFRET using total internal reflection fluorescence microscopy (TIRFM) enables the observation of surface immobilized riboswitches for long periods to visualize their dynamics in real-time and calculate the individual rate constants (24,26,27,43). Our results suggest contrasting ligand-mediated folding behavior of the *Bsu* riboswitch in the presence and absence of Mg^{2+}. In the absence of Mg^{2+}, the folding rate (k_{dock}) increases almost linearly with increasing $preQ_1$ concentration, whereas the unfolding rate (k_{undock}) decreases only slightly. In contrast, in the presence of Mg^{2+} k_{dock} increases only modestly, while k_{undock} significantly decreases with increasing $preQ_1$ concentration. These data suggest that in the presence of Mg^{2+}, the *Bsu* riboswitch folds through a conformational selection mechanism, while an induced-fit like mechanism dominates in the absence of Mg^{2+}. Our smFRET kinetic studies comparing the effect of the chemically similar ligands $preQ_1$, $preQ_0$ and guanine (**Figure 3.1D**) on the *Bsu* riboswitch dynamics show that the relative stabilizing effect of the ligands on the riboswitch folded conformation varies greatly with Mg^{2+}. Furthermore, we show that the affinities of the ligands $preQ_0$ and guanine do not correlate well with their ability to stabilize the folded conformation.

3.2 Materials and Methods

3.2.1 Labeling and purification of RNAs for smFRET

The doubly labeled *Bsu* riboswitch aptamers were generated as described in our previous study (17). RNA with a 5'-biotin, internal 5-aminoallyl-uridine (5 NU) at position U13, and a 3'- DY547 fluorophore was chemically synthesized by Dharmacon Inc. (Fayette, CO). The RNA was first deprotected as per the manufacturer's protocol and then ~3.4 nmol of it was used for labeling with 100 µg Cy5-NHS ester (GE Healthcare) in a 50 µL reaction volume containing 30 µL DMSO and 0.1 M sodium bicarbonate buffer, pH 8.7. The reaction was incubated with constant tumbling in the dark at RT for 4 h. Excess free dye was removed by gel filtration using Nap-5 columns (GE Healthcare) and the fractions were ethanol precipitated. The pellet obtained was dried and suspended in autoclaved deionized water for later use.

3.2.2 Single molecule FRET microscopy

SmFRET experiments were performed using prism-based TIRFM as described in (17,43). Quartz slides with microfluidic channel were coated with biotinylated BSA and streptavidin for immobilization of RNA with 5'-biotin. RNA was first folded by heating at 90 ºC for 1 min in the absence of Mg^{2+} and Mg^{2+} added to a final concentration of 1 mM after cooling the RNA to 37 ºC. 50-100 pM of the heat-annealed doubly labeled *Bsu* aptamer was immobilized on the slide and excess unbound RNA was washed out with 1x smFRET buffer (50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM Mg^{2+}). Experiments were carried out in 1x smFRET buffer with an oxygen scavenging system containing 5 mM protocatechuic acid (PCA) and 50 nM protocatechuate-3,4-dioxygenase (PCD) to slow photobleaching and 2 mM Trolox to decrease photoblinking of the dyes (43). Kinetic experiments with ligand titration were performed in the presence or absence of Mg^{2+} on the same slide. For smFRET experiments of ligand comparison, molecules were imaged using an intensified CCD camera (I-Pentamax, Princeton Instruments) at a time resolution of ~60 ms using the full CCD chip. Ligand titration experiments in the absence of Mg^{2+} were performed at a faster time resolution of ~35 ms, obtained by using half (512 x 256 pixels) of the CCD chip. Ligand titration experiments in the presence of Mg^{2+} were performed using an EMCCD camera (iXon, Andor Technology) on a similar prism-based TIRF microscopy setup. The raw movies were recorded using a custom written Matlab (The Math Works) program and were processed using IDL (Research Systems) to extract smFRET time traces. FRET ratio was calculated as $I_A/(I_A + I_D)$, where I_A and I_D represent the background corrected acceptor (Cy5) and donor (DY547) intensities, respectively. Individual traces displaying FRET with a minimum total ($I_A + I_D$) intensity of 300 and single step photobleaching were selected for further analysis. The traces were idealized by hidden Markov modeling (HMM) using segmental k-means algorithm in QuB program (44). From the idealized smFRET traces with dynamics, dwell times in each state were obtained which were fit with exponential functions to extract the rate constants of transitions under a given condition.

3.3 Results

3.3.1 $PreQ_1$ dependent kinetics of the *Bsu* riboswitch in the absence of Mg^{2+} suggests an induced-fit mechanism of ligand binding

We first measured the kinetics of the conformational dynamics of the *Bsu* riboswitch in the absence of Mg^{2+} and in the presence of $preQ_1$. Mg^{2+} was omitted from the buffer to decouple its stabilizing effect on the folded conformations and to determine the effect of ligand by itself on the dynamics. In the absence of Mg^{2+}, individual smFRET traces were largely static with stable FRET value of ~0.6. However, in the presence of 50 nM $preQ_1$, the smFRET traces displayed multiple transitions between a pre-folded ~0.6 FRET state and a folded ~0.9 FRET state (**Figure 3.3A**). The cumulative distributions of dwell times in the pre-folded and folded states were plotted and fit with single-exponential functions to obtain k_{dock} and k_{undock}, respectively (**Figure 3.3B**). Under these conditions, the folding rate, k_{dock} of the *Bsu* riboswitch measured at ~ 35 ms time resolution was ~0.3 s^{-1} (**Figure 3.3C**). With increasing ligand concentration, the value of k_{dock} increased almost linearly, to ~ 3.7 s^{-1} at 1 µM $preQ_1$. Least square fitting with a straight line gave a slope of ~3.5 $\mu M^{-1} s^{-1}$ for the k_{dock}. In contrast, the value of k_{undock} remained almost constant, with only a very modest decrease from 1.18 s^{-1} at 50 nM $preQ_1$ to 1.01 s^{-1} at 1 µM $preQ_1$ (**Figure 3.3C**). The positive dependence of k_{dock} on the ligand concentration suggests that $preQ_1$ binds to the ligand-free open conformation and 'induces' folding of the *Bsu* riboswitch, similar to an induced-fit like mechanism (41).

3.3.2 $PreQ_1$ dependent kinetics in the presence of Mg^{2+} suggests folding via a major conformational selection mechanism

Mg^{2+} is important for RNA folding in general and for riboswitches in particular, where Mg^{2+} has been shown to promote folded-like compact conformations and enhance ligand binding (17,21,45). We therefore probed the dynamics of the *Bsu* riboswitch in the presence of a physiologically relevant concentration of Mg^{2+} (1 mM). **Figure 3.4A** shows smFRET traces at different ligand concentrations. In the absence of ligand, smFRET traces of single

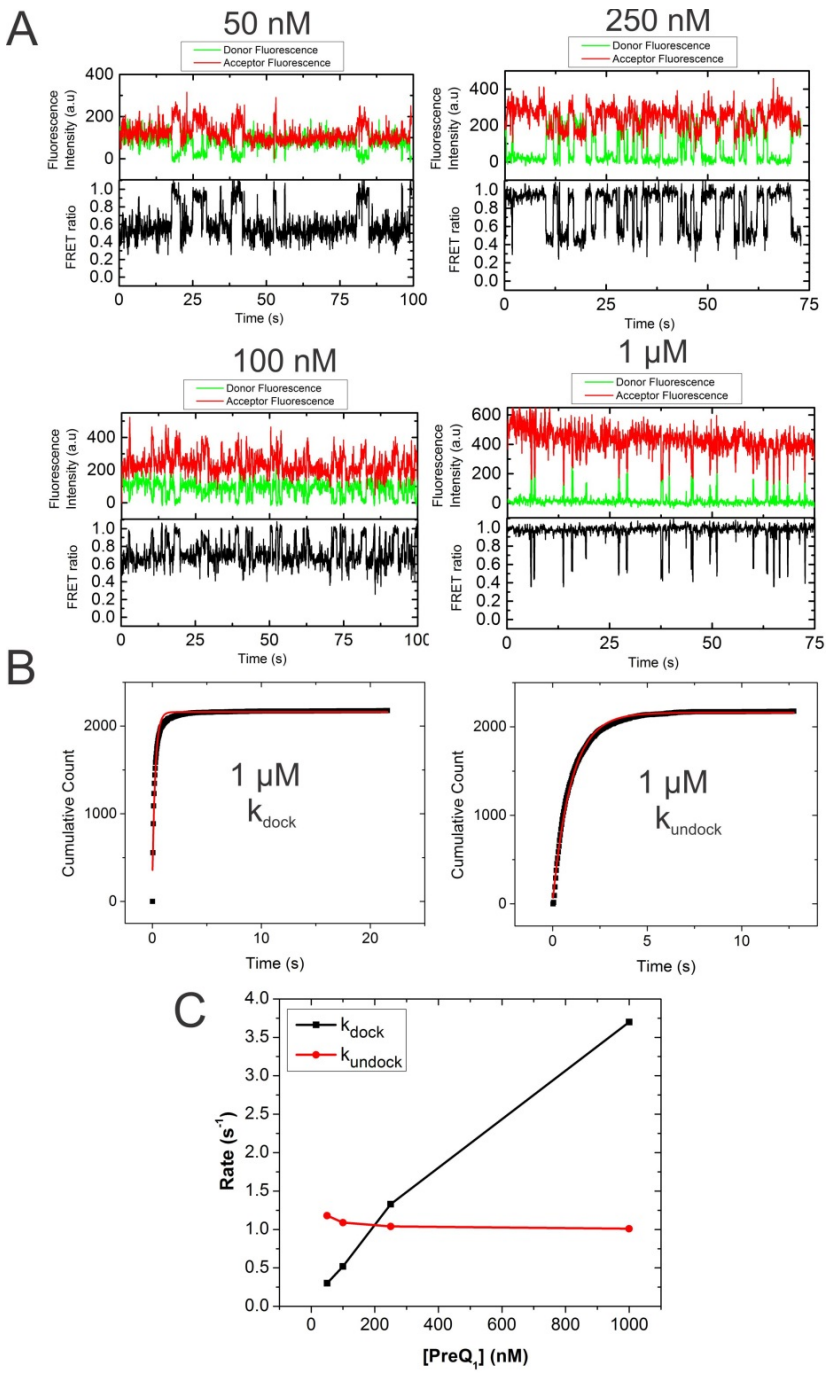

Figure 3.3 $PreQ_1$ dependent conformational dynamics in the absence of Mg^{2+}

(A) Representative smFRET traces at different $preQ_1$ concentrations. (B) Cumulative dwell time distributions with single exponential fits (in red) to obtain k_{dock} and k_{undock} for the 1 μM $preQ_1$ condition. (C) Plot showing the kinetics of conformational dynamics (k_{dock} and k_{undock}) as a function of $preQ_1$ concentration in the absence of Mg^{2+}.

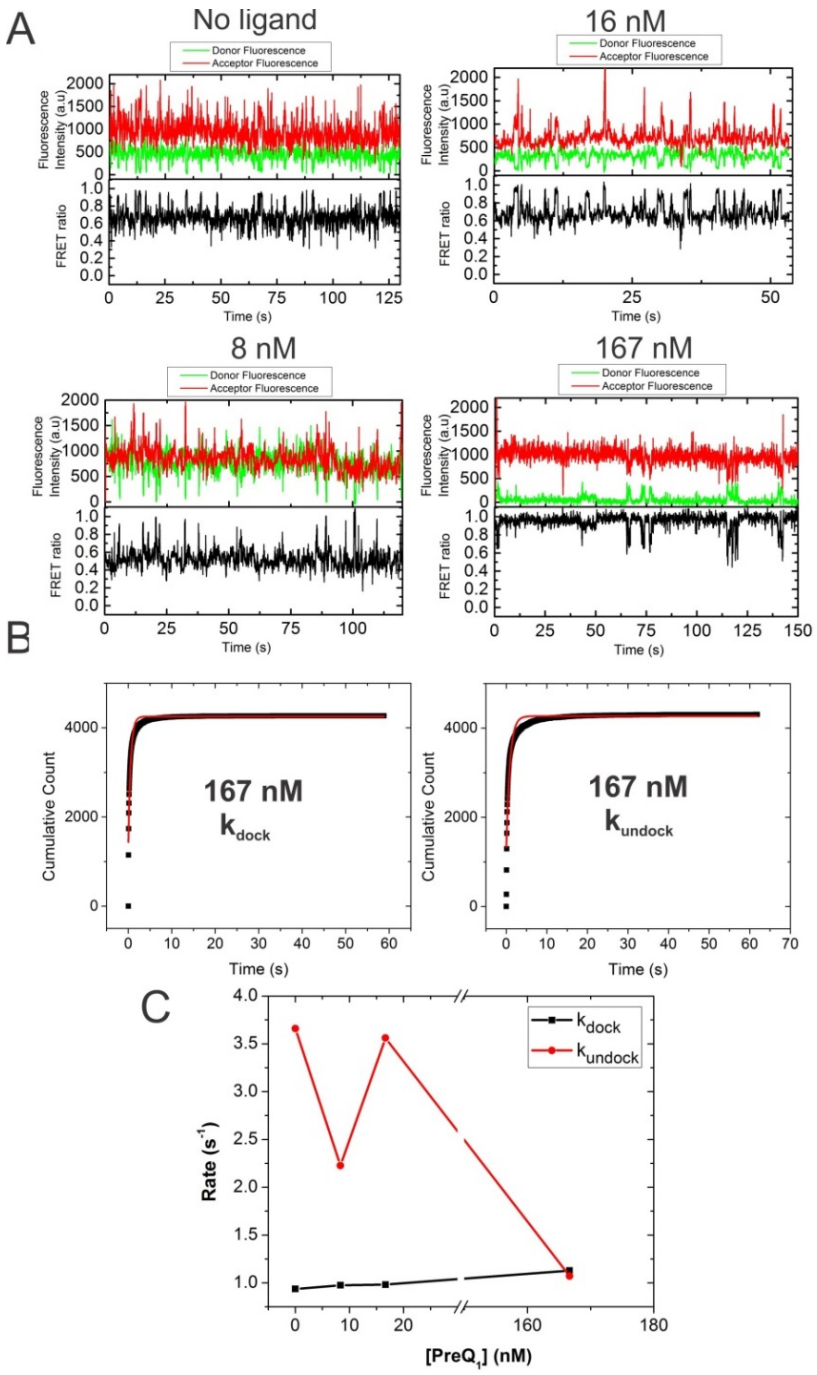

Figure 3.4 PreQ$_1$ dependent conformational dynamics in the presence of Mg^{2+}

(A) Representative smFRET traces at different preQ$_1$ concentrations. (B) Cumulative dwell time distributions with single exponential fits (in red) to obtain k_{dock} and k_{undock} for the 167 nM preQ$_1$ condition. (C) Plot showing the kinetics of conformational dynamics (k_{dock} and k_{undock}) as a function of preQ$_1$ concentration in the absence of Mg^{2+}.

molecules imaged at 33 ms time resolution showed that the riboswitch exists mainly in an ~0.7 pre-folded state, making frequent transitions with short dwell times into pseudoknot like conformations with a FRET value of ~0.91 (**Figures 3.4A**). Dwell time distributions fit with single-exponential functions gave average k_{fold} and k_{unfold} values of 0.94 s^{-1} and 3.7 s^{-1}, respectively (**Figures 3.4B**). In stark contrast to the no Mg^{2+} condition described above, the value of k_{dock} only increased marginally to ~ 1.12 s^{-1} at 167 nM $preQ_1$ (**Figure 3.4C**). On the other hand, the unfolding rate, k_{undock} decreased significantly from ~3.7 s^{-1} in the absence of ligand to ~1.1 s^{-1} at 167 nM $preQ_1$. These smFRET data directly show that in the presence of Mg^{2+} alone, the *Bsu* riboswitch has fast intrinsic dynamics and samples the ligand-bound like conformations at a rate that is not majorly influenced by $preQ_1$ concentration. In addition, the data suggest that $preQ_1$ binding stabilizes the folded conformation and decreases its unfolding rate. The observation of folded-like conformations in the absence ligand and a constant k_{dock} value suggests that, in the presence of Mg^{2+}, the *Bsu* riboswitch binds $preQ_1$ through a major conformational selection like mechanism (35,37,41).

3.3.3 Comparison of the effect of different ligands on the *Bsu* riboswitch conformational dynamics

The *Bsu* $preQ_1$ riboswitch was shown to bind the closely related non-cognate ligands $preQ_0$ and guanine with high affinities (32). Therefore, we investigated the effects of both of these ligands in comparison with $preQ_1$ on the conformational dynamics of the *Bsu* riboswitch at 60 ms time resolution. The K_d values for $preQ_1$, $preQ_0$ and guanine are 20 nM, 100 nM and ~320 nM, respectively as measured in previous studies using ensemble in-line probing assays that monitor increasing solvent protection of the RNA upon ligand addition (32). Due to the different K_d values, we measured the riboswitch dynamics under a constant and saturating ligand concentration of 10 µM. Examplary smFRET traces for each ligand are shown in **Figure 3.5A**. We performed three independent measurements for each ligand and the rates are shown as mean ± standard deviation. In the absence of Mg^{2+}, $preQ_1$ has the highest k_{dock} rate followed by $preQ_0$ and guanine, as expected (**Figure 3.6A**). The k_{dock} values for $preQ_1$, $preQ_0$ and guanine are 0.96 ± 0.6, 0.83 ± 0.18 and 0.5 ± 0.17, respectively. In contrast, we observed that the k_{undock} does not follow the same trend. The k_{undock} values for $preQ_1$, $preQ_0$

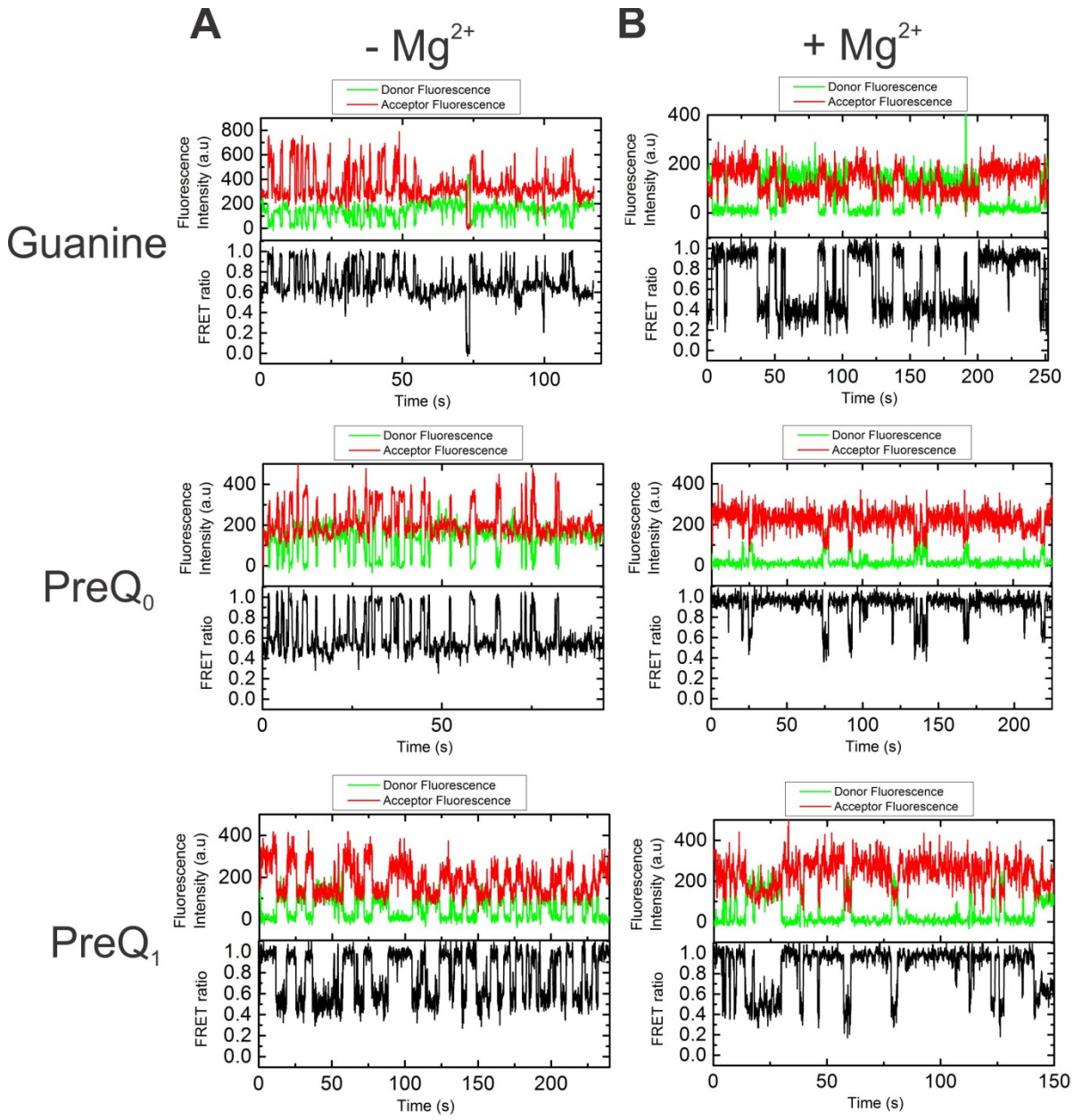

Figure 3.5 Effect of different ligands on the *Bsu* riboswitch conformational dynamics

Representative smFRET traces for guanine, preQ$_0$ and preQ$_1$in the (A) absence and (B) presence of Mg^{2+}.

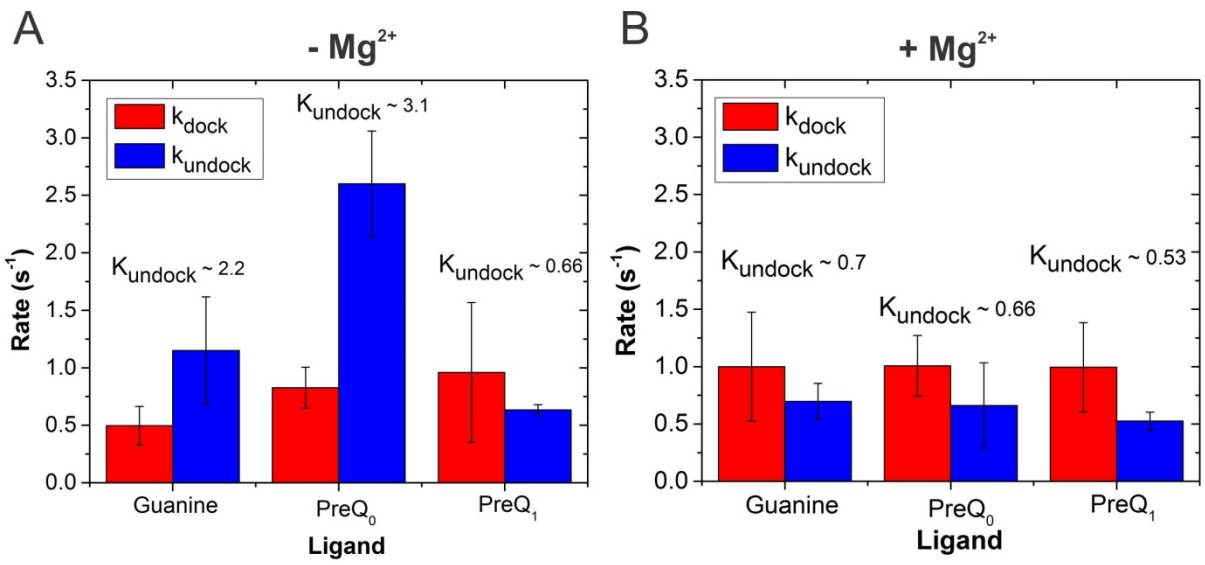

Figure 3.6 Rates of the *Bsu* riboswitch conformational dynamics in the presence of different ligands

Rates of ligand dependent conformational dynamics in the (A) absence and (B) presence of Mg^{2+}.

process is important for a deeper understanding of the molecular mechanism of riboswitches (30). SmFRET microscopy is ideally suited for studying riboswitch conformational dynamics and guanine are 0.63 ± 0.05, 2.6 ± 0.46 and 1.1 ± 0.47, respectively (**Figure 3.6A**). These data show that, while $preQ_1$ has the smallest k_{undock} among the three ligands, guanine has a lower k_{undock} value than $preQ_0$. We calculated the equilibrium undocking constant, K_{undock} as the ratio k_{undock}/k_{dock} for each ligand. The K_{undock} for $preQ_1$, $preQ_0$ and guanine are ~0.66, ~3.1 and ~2.2, respectively. A lower K_{undock} value implies a higher stabilized folded state and therefore, as expected, the cognate ligand $preQ_1$ stabilizes the folded state most effectively of all three ligands. Interestingly, despite its lower affinity guanine's K_{undock} is smaller than that of $preQ_0$. This suggests that K_{undock} does not directly correspond to the ability of the ligand to stabilize folded conformations.

Next, we tested the effect of the three ligands on the docking and undocking rates of the *Bsu* riboswitch in the presence of 10 μM ligand and 1 mM Mg^{2+}. The rates were obtained from three independent measurements for $preQ_0$, guanine and two measurements for $preQ_1$. Example smFRET traces for each ligand are shown in **Figure 3.5B.** Under these conditions, the ligands display subtle differences in their effect on k_{dock} and k_{undock} as compared to the no Mg^{2+} condition described above. Interestingly, all three ligands showed the same average k_{dock} value of ~1.0 s^{-1} (**Figure 3.6B**). In addition, k_{undock} is similar for $preQ_1$, $preQ_0$ and guanine with values of 0.52 ± 0.08, 0.66 ± 0.37 and 0.70 ± 0.16, respectively (**Figure 3.6B**). These results show that in the presence of Mg^{2+}, all three ligands have similar efficiencies in stabilizing the pseudoknot conformation of the *Bsu* riboswitch. Comparison of the K_{undock} values with and without Mg^{2+} shows that Mg^{2+} decreases the K_{undock} value significantly by ~3-fold and ~5-fold for guanine and $preQ_0$, respectively, whereas only a modest decrease was seen for $preQ_1$. The Mg^{2+}-promoted stabilization of the folded state by $preQ_0$ and guanine is mainly due to a decrease in k_{undock} and a minor increase in k_{dock}.

3.4 Discussion

Riboswitches utilize ligand dependent RNA folding to modulate gene expression (4). Although structures of a number of riboswitches are available, yielding atomic level insight into the interactions between RNA and small molecule ligands, due to the scarcity of ligand-

free structures our knowledge of how the ligand-free conformation folds into the ligand-bound conformation is limited (2,7,46,47). Therefore, studying the ligand dependent folding and observing minor functionally important conformations that are generally difficult to detect using ensemble averaging methods (17,24). SmFRET has been extensively used for studying riboswitch dynamics and RNA folding in general (25,30). In this study, we have used smFRET to investigate the effects of ligand on the conformational dynamics of the *Bsu* preQ$_1$ riboswitch.

Ligand binding coupled folding processes of biological macromolecules are mainly classified into induced-fit and conformational selection mechanisms (34,35,48) (**Figure 3.2A**). These mechanisms are not necessarily mutually exclusive as shown by flux calculations through each pathway and directly measured using smFRET experiments (31,35). Distinguishing the two mechanisms has been difficult using traditional ensemble methods and many studies have incorrectly attributed folding via the conformational selection mechanism merely by the observation of folded-like conformations in the absence of ligand (35,38). One way of differentiating the two mechanisms is to study the effect of increasing ligand concentration on the kinetics of (un)folding (k_{obs}) of a biomolecule, measured as a sum of folding and unfolding rate constants during ensemble experiments (41). By contrast, smFRET microscopy can be used to study the individual rate constants of folding and unfolding to directly pinpoint the most-populated ligand binding mechanism, as recently demonstrated by dissecting the ligand binding mechanism of the maltose binding protein (31). The kinetic signature of induced-fit and conformational selection mechanisms manifest directly in smFRET traces as an increasing folding rate constant (or decreasing $t_{closing}$, as shown in **Figure 3.2B**) and a constant folding rate constant, respectively.

Accordingly, our single molecule kinetic data suggest that in the absence of Mg^{2+}, the *Bsu* riboswitch folds via an induced-fit mechanism (**Figure 3.2A**) where the rate of folding (k_{dock}) increases as a function of ligand concentration. This is due to a decrease in the dwell-time ($t_{closing}$) in the open state as shown in **Figure 3.2B**. In contrast, the ligand-mediated folding pathway in the presence of Mg^{2+} favors a conformational selection mechanism where the folding rate (k_{dock}) only increases slightly, remaining constant with increasing preQ$_1$

concentration (**Figure 3.4C**). These data of folding rate versus ligand concentration agree well with the kinetic signatures of induced-fit and conformational selection as shown in **Figure 3.2B** (31). This shift in the folding mechanism can be explained by the effect of Mg^{2+} on promoting pseudoknot-like conformations in the absence of ligand, as shown in our previous study and also demonstrated using NMR on a different class-I $preQ_1$ riboswitch (11,17). In the absence of Mg^{2+}, the folded-like conformations are either not populated or only exist very transiently (with dwell times <<33 ms). Therefore ligand presumably binds to the dominant 'open' conformation that probably contains stem-loop P1-L1 with or without transient L3 interactions with P1. However, in the presence of Mg^{2+}, the riboswitch adopts a higher population (~9 %) of loose pseudoknot-like conformations (0.9 FRET state) (17). This is seen clearly as short transitions into the 0.9 FRET state in the smFRET traces (at 33 ms time resolution) in the presence of Mg^{2+} alone (**Figure 3.4A**, No ligand trace). The ligand can then bind to these loose pseudoknot-like states and fold the RNA into the native pseudoknot after minor local conformational adjustments (6,33). It is worth noting that since the ligand binding pocket is mostly solvent inaccessible (and therefore closed) in the final ligand-bound structure, $preQ_1$ cannot bind to this exact conformation. An NMR study on the ligand-bound *Bsu* riboswitch has observed fast (µs - ms) dynamics in a few residues in the loop L1 and the authors proposed that these residues act as a 'lid' on the binding pocket (49). Therefore, the ligand binding competent conformation can be likely described as a loose pseudoknot with a dynamic loop L1, through which ligand can access the binding pocket. According to a recently emerging nomenclature, this mechanism of binding to a state that resembles the final structure followed by locally induced conformational adjustments has been referred to as 'extended conformational selection' (34,36). It is interesting that the riboswitch does not fold majorly through an induced-fit mechanism in the presence of Mg^{2+}, even though only ~9 % population adopts a folded-like conformation while the remaining ~90 % is only partially pre-folded. Although the reason for this observation is not clear, we speculate that this is due to the more unfolded nature of the conformational ensemble in the absence of Mg^{2+} as shown in our previous study (17). In the absence of Mg^{2+}, the ligand-free (or open) state has a lower mean FRET value of 0.62 (as compared to ~0.72 for the + Mg^{2+} condition) and has a larger

width with a relatively higher population of molecules showing lower FRET states (≤0.3 FRET states). Furthermore, NMR data on the ligand-free *Bsu* riboswitch showed that the addition of Mg^{2+} causes changes in the structure of loop L1, resulting in the appearance of chemical shifts that correspond to tertiary interactions indicative of a pre-formed ligand binding pocket. Therefore, the ligand-free conformational ensemble in the presence of Mg^{2+} is more compact with a partial organization of the binding pocket. This is distinct from the conformations in the absence of Mg^{2+} that are of a more unfolded nature, lacking a pre-formed binding pocket and therefore binding ligand through an induced fit mechanism.

Many riboswitches exhibit high specificity for their cognate ligands and can discriminate closely related metabolites with affinities differing by an order of magnitude or more (46,47). However, a few riboswitches such as the $preQ_1$ riboswitch bind multiple related ligands with strong affinities (32). The *Bsu* riboswitch in particular binds $preQ_1$ with a K_d of ~ 20 nM; but it can also bind $preQ_0$ and guanine relatively well with K_d values in the presence of Mg^{2+} of ~100 nM and ~320 nM, respectively (32). This suggests that *in vivo* both $preQ_0$ and guanine can act as effectors for the *Bsu* riboswitch to modulate gene expression depending on their concentration. Therefore, we used smFRET to compare the ability of the non-cognate ligands $preQ_0$ and guanine to promote folding of the *Bsu* riboswitch. Our results show differential, Mg^{2+} dependent effects of the three ligands. In the absence of Mg^{2+}, the cognate ligand $preQ_1$ has the lowest equilibrium undocking constant K_{undock} whereas $preQ_0$ has a higher K_{undock} than guanine by virtue of its ~2.5-fold higher k_{undock} (**Figure 3.6A**). In contrast, all three ligands show largely similar K_{undock} values in the presence of Mg^{2+} (**Figure 3.6A**). Our smFRET data show that Mg^{2+} decreases the K_{undock} value mainly by slowing down k_{undock} and increasing k_{dock}. Similar effects of Mg^{2+} were also observed in other RNAs, such as the purine riboswitches (21,50). Of note, the only subtle differences between the ligands as observed in the presence of Mg^{2+} may in part be due to the high (10 μM, >> K_d) ligand concentrations used in the current study. SmFRET experiments performed under concentrations closer to the K_d values may thus provide better estimates of the differences between the ligands and will be carried out in the future.

Our observation of a lower K_{undock} value for guanine (relative to $preQ_0$) is surprising since it has been reported to have a weaker affinity (~3-fold) than $preQ_0$ (32). This suggests that the equilibrium binding affinity (K_d) values do not necessarily correspond to the ability of ligand to stabilize a folded conformation. Similar observations were made in studies of the lysine and tetrahydrofolate (THF) riboswitches, which showed that a disconnect exists between the affinity of a ligand and its ability to modulate gene expression (51,52). Ligand analogs that bound with higher affinities had lower efficiencies of promoting transcription termination by the THF riboswitch than THF itself. This may be due to the inability of the non-cognate ligands to strongly stabilize the regions of RNA that form key long-range tertiary interactions incorporating the 'switching' sequence needed to form downstream transcription terminator hairpin. Alternatively, it may be due to differences in the ligand binding kinetics (k_{on} and k_{off}) between the ligands. The K_d values of ~100 nM and ~320 nM for $preQ_0$ and guanine, respectively, were measured in the presence of high Mg^{2+} concentrations and, therefore, it is possible that they display different relative affinities in the absence of Mg^{2+}. Interestingly, our observation of increasing k_{dock} with increasing ligand affinities (k_{dock} of guanine < $preQ_0$ < $preQ_1$) in the absence of Mg^{2+} and a constant k_{dock} in the presence of Mg^{2+} mirrors the behavior of k_{dock} versus $preQ_1$ concentration in the absence and presence of Mg^{2+} (**Figures 3.3, 3.4 and 3.6**). This observation also points to differential induced-fit and conformational selection mechanisms in the absence and presence of Mg^{2+}, respectively, for the *Bsu* $preQ_1$ riboswitch.

In summary, using smFRET kinetic measurements of conformational dynamics, we have demonstrated that the *Bsu* riboswitch folds through contrasting ligand-mediated mechanisms that depend on the presence of Mg^{2+}. In addition, we have shown that the binding affinities of ligands do not always translate into their effect on stabilizing folded conformations. In light of our smFRET-based ligand comparison and the recently emerging results on the disconnect between ligand binding affinity and gene regulatory effects, future studies should investigate the link between riboswitch structural dynamics and how they are fine-tuned in different organisms to promote efficient ligand-mediated genetic regulation (51). Molecular recognition plays an important role in riboswitch function and, therefore, the

smFRET approach used in this study can be applied to other riboswitch RNAs to elucidate their ligand-mediated folding mechanisms.

3.5 Acknowledgements

We thank Dr. George Garcia for a generous gift of the ligands $preQ_1$ and $preQ_0$, Dr. David Rueda for the use of his prism-based TIRF microscope for the acquiring data at 33 ms time resolution and Kyle Vrtis for assistance using it, and Mario Blanco for help in the analysis of smFRET data at 35 ms time resolution.

3.6 References

1. Garst, A.D., Edwards, A.L. and Batey, R.T. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol*, 3.
2. Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, 152, 17-24.
3. Peselis, A. and Serganov, A. (2014) Themes and variations in riboswitch structure and function. *Biochim Biophys Acta*.
4. Mandal, M. and Breaker, R.R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, 5, 451-463.
5. Leontis, N.B. and Westhof, E. (2003) Analysis of RNA motifs. *Curr. Opin. Struct. Biol.*, 13, 300-308.
6. Klein, D.J., Edwards, T.E. and Ferre-D'Amare, A.R. (2009) Cocrystal structure of a class I $preQ_1$ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol*, 16, 343-344.
7. Liberman, J.A. and Wedekind, J.E. (2012) Riboswitch structure in the ligand-free state. *Wiley Interdiscip Rev RNA*, 3, 369-384.
8. Rieder, U., Kreutz, C. and Micura, R. (2010) Folding of a transcriptionally acting preQ1 riboswitch. *Proc Natl Acad Sci U S A*, 107, 10804-10809.
9. Wilson, R.C., Smith, A.M., Fuchs, R.T., Kleckner, I.R., Henkin, T.M. and Foster, M.P. (2011) Tuning riboswitch regulation through conformational selection. *J Mol Biol*, 405, 926-938.
10. Haller, A., Rieder, U., Aigner, M., Blanchard, S.C. and Micura, R. (2011) Conformational capture of the SAM-II riboswitch. *Nat Chem Biol*, 7, 393-400.
11. Santner, T., Rieder, U., Kreutz, C. and Micura, R. (2012) Pseudoknot preorganization of the $preQ_1$ class I riboswitch. *J Am Chem Soc*, 134, 11928-11931.
12. Chen, B., Zuo, X., Wang, Y.X. and Dayie, T.K. (2012) Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res*, 40, 3117-3130.
13. Baird, N.J. and Ferre-D'Amare, A.R. (2010) Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *RNA*, 16, 598-609.

14. Lipfert, J., Herschlag, D. and Doniach, S. (2009) Riboswitch conformations revealed by small-angle X-ray scattering. *Methods Mol Biol*, 540, 141-159.
15. Baird, N.J. and Ferre-D'Amare, A.R. (2014) Analysis of riboswitch structure and ligand binding using small-angle X-ray scattering (SAXS). *Methods Mol Biol*, 1103, 211-225.
16. Feng, J., Walter, N.G. and Brooks, C.L., 3rd. (2011) Cooperative and directional folding of the preQ$_1$ riboswitch aptamer domain. *J Am Chem Soc*, 133, 4196-4199.
17. Suddala, K.C., Rinaldi, A.J., Feng, J., Mustoe, A.M., Eichhorn, C.D., Liberman, J.A., Wedekind, J.E., Al-Hashimi, H.M., Brooks, C.L., 3rd and Walter, N.G. (2013) Single transcriptional and translational preQ1 riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure. *Nucleic Acids Res*, 41, 10462-10475.
18. Lin, J.C. and Thirumalai, D. (2008) Relative stability of helices determines the folding landscape of adenine riboswitch aptamers. *J Am Chem Soc*, 130, 14080-14081.
19. Lin, J.C. and Thirumalai, D. (2013) Kinetics of allosteric transitions in S-adenosylmethionine riboswitch are accurately predicted from the folding landscape. *J Am Chem Soc*, 135, 16641-16650.
20. Haller, A., Altman, R.B., Souliere, M.F., Blanchard, S.C. and Micura, R. (2013) Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc Natl Acad Sci U S A*, 110, 4188-4193.
21. Lemay, J.F., Penedo, J.C., Tremblay, R., Lilley, D.M. and Lafontaine, D.A. (2006) Folding of the adenine riboswitch. *Chem Biol*, 13, 857-868.
22. Greenleaf, W.J., Frieda, K.L., Foster, D.A., Woodside, M.T. and Block, S.M. (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science*, 319, 630-633.
23. Frieda, K.L. and Block, S.M. (2012) Direct observation of cotranscriptional folding in an adenine riboswitch. *Science*, 338, 397-400.
24. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single-molecule FRET. *Nat Methods*, 5, 507-516.
25. Zhuang, X. (2005) Single-molecule RNA science. *Annu Rev Biophys Biomol Struct*, 34, 399-414.
26. Deniz, A.A., Mukhopadhyay, S. and Lemke, E.A. (2008) Single-molecule biophysics: at the interface of biology, physics and chemistry. *J R Soc Interface*, 5, 15-45.
27. Walter, N.G., Huang, C.Y., Manzo, A.J. and Sobhy, M.A. (2008) Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nat Methods*, 5, 475-489.
28. Tinoco, I., Jr. and Gonzalez, R.L., Jr. (2011) Biological mechanisms, one molecule at a time. *Genes Dev*, 25, 1205-1231.
29. Garst, A.D. and Batey, R.T. (2009) A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta*, 1789, 584-591.
30. Haller, A., Souliere, M.F. and Micura, R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc Chem Res*, 44, 1339-1348.

31. Kim, E., Lee, S., Jeon, A., Choi, J.M., Lee, H.S., Hohng, S. and Kim, H.S. (2013) A single-molecule dissection of ligand binding to a protein with intrinsic dynamics. *Nat Chem Biol*, 9, 313-318.
32. 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. *et al.* (2007) A riboswitch selective for the queuosine precursor preQ$_1$ contains an unusually small aptamer domain. *Nat Struct Mol Biol*, 14, 308-317.
33. Kang, M., Peterson, R. and Feigon, J. (2009) Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol Cell*, 33, 784-790.
34. Boehr, D.D., Nussinov, R. and Wright, P.E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol*, 5, 789-796.
35. Hammes, G.G., Chang, Y.C. and Oas, T.G. (2009) Conformational selection or induced fit: a flux description of reaction mechanism. *Proc Natl Acad Sci U S A*, 106, 13737-13741.
36. Csermely, P., Palotai, R. and Nussinov, R. (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci*, 35, 539-546.
37. Vogt, A.D. and Di Cera, E. (2013) Conformational selection is a dominant mechanism of ligand binding. *Biochemistry*, 52, 5723-5729.
38. Daniels, K.G., Tonthat, N.K., McClure, D.R., Chang, Y.C., Liu, X., Schumacher, M.A., Fierke, C.A., Schmidler, S.C. and Oas, T.G. (2014) Ligand concentration regulates the pathways of coupled protein folding and binding. *J Am Chem Soc*, 136, 822-825.
39. Zhang, Q., Stelzer, A.C., Fisher, C.K. and Al-Hashimi, H.M. (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, 450, 1263-1267.
40. Weikl, T.R. and von Deuster, C. (2009) Selected-fit versus induced-fit protein binding: Kinetic differences and mutational analysis. *Proteins: Structure, Function, and Bioinformatics*, 75, 104-110.
41. Vogt, A.D. and Di Cera, E. (2012) Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. *Biochemistry*, 51, 5894-5902.
42. Hatzakis, N.S. (2014) Single molecule insights on conformational selection and induced fit mechanism. *Biophysical Chemistry*, 186, 46-54.
43. Abelson, J., Blanco, M., Ditzler, M.A., Fuller, F., Aravamudhan, P., Wood, M., Villa, T., Ryan, D.E., Pleiss, J.A., Maeder, C. *et al.* (2010) Conformational dynamics of single pre-mRNA molecules during in vitro splicing. *Nat Struct Mol Biol*, 17, 504-512.
44. Blanco, M. and Walter, N.G. (2010) Analysis of complex single-molecule FRET time trajectories. *Methods in enzymology*, 472, 153-178.
45. Misra, V.K. and Draper, D.E. (1998) On the role of magnesium ions in RNA stability. *Biopolymers*, 48, 113-135.
46. Serganov, A. and Patel, D.J. (2012) Metabolite recognition principles and molecular mechanisms underlying riboswitch function. *Annu Rev Biophys*, 41, 343-370.

47. Serganov, A. and Patel, D.J. (2012) Molecular recognition and function of riboswitches. *Curr Opin Struct Biol*, 22, 279-286.
48. Changeux, J.P. and Edelstein, S. (2011) Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol Rep*, 3, 19.
49. Zhang, Q., Kang, M., Peterson, R.D. and Feigon, J. (2011) Comparison of solution and crystal structures of preQ$_1$ riboswitch reveals calcium-induced changes in conformation and dynamics. *J Am Chem Soc*, 133, 5190-5193.
50. Brenner, M.D., Scanlan, M.S., Nahas, M.K., Ha, T. and Silverman, S.K. (2010) Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine. *Biochemistry*, 49, 1596-1605.
51. Trausch, J.J. and Batey, R.T. (2014) A Disconnect between High-Affinity Binding and Efficient Regulation by Antifolates and Purines in the Tetrahydrofolate Riboswitch. *Chem Biol*, 21, 205-216.
52. Garst, A.D., Porter, E.B. and Batey, R.T. (2012) Insights into the regulatory landscape of the lysine riboswitch. *J Mol Biol*, 423, 17-33.

CHAPTER 4

Single molecule investigation of the conformation and tRNA binding kinetics of the *glyQS* T-box riboswitch

4.1 Introduction

Riboswitches are structured mRNA motifs usually present in 5'-untranslated regions (5'-UTRs) that can sense the intracellular concentration of metabolites and accordingly regulate the expression of proteins involved in their synthesis and transport (1-4). Riboswitches are ubiquitous in bacteria where they regulate at least ~4 % of all genes mainly through inhibition of transcription elongation or translation initiation (2,5). A large number of structurally diverse riboswitches recognizing amino acids, nucleobases, nucleotides, coenzymes, metal ions, temperature have been discovered (6,7). The T-box riboswitch embodies one of the earliest discovered RNA based gene regulation mechanisms in bacteria (8,9). In Gram-positive bacteria, the T-box mechanism is used to control the expression of proteins related to transfer RNA (tRNA) charging, amino acid metabolism and transport (10). Regulation occurs through a transcription attenuation mechanism. In the absence of tRNA, transcription of the 5'-leader results in the formation of terminator hairpin and destabilization of the RNA polymerase complex. Binding of an uncharged cognate tRNA to the 5'-leader region of the nascent transcript stabilizes an anti-terminator hairpin, preventing formation of a terminator hairpin and leads to transcription of the downstream genes (10) (**Figure 4.1**). T-box riboswitches have also been identified in Gram-negative bacteria and in some high G+C-rich Gram-positive bacteria where they are predicted to function at the level of translation Initiation (11,12). In this mechanism, binding of uncharged tRNA stabilizes an anti-sequester hairpin, similar to the anti-terminator hairpin in the transcriptionally acting T-box systems.

Krishna C. Suddala performed all the experiments and the data analysis. Dr. Malgorzata Michnicka made the T-box and tRNA constructs used for the single molecule experiments.

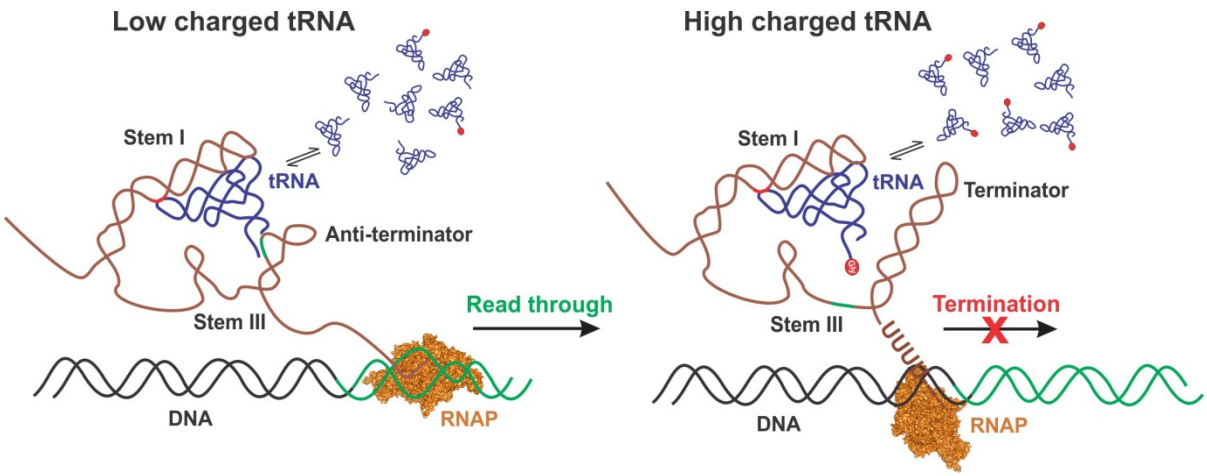

Figure 4.1 Mechanism of the T-box riboswitch

tRNA binding to the T-box riboswitch under low charging conditions stabilizes the anti-terminator (AT) hairpin, leading to transcription of the downstream genes (green). Under high concentration of charged tRNA, presence of the amino acid (indicated by a small red sphere at the 3'-end of tRNA) prevents the formation of tRNA with the AT hairpin, leading to transcription termination.

This event prevents the sequestering of the Shine-Dalgarno (SD) sequence and instead leads to the translation of the mRNA (11,12). Due to its similarities in the mechanism of action to small molecule binding riboswitches, the T-box system has been referred to as a riboswitch. Therefore, the T-box system can be classified as a unique riboswitch recognizing the aminoacylation state of a structurally complex macromolecular ligand, tRNA. Although T-box riboswitches bind tRNA as the ligand, it is important to note that the genetic switch is dependent on the presence or absence of amino acid on its 3'-end. In this respect, the T-box system is similar to a small molecule (amino acid) binding riboswitch.

The T-box riboswitch was first identified as a conserved sequence in the leader region of the *tyrS* mRNA from *Bacillus subtilis* coding for tyrosyl tRNA synthetase(13). In a subsequent breakthrough study, Grundy *et al* (8) showed that uncharged $tRNA^{tyr}$ alone, in the absence of any protein cofactors, can act as a positive effector to enhance transcription of the downstream genes. T-box riboswitches were later discovered to regulate the expression of genes associated with several other amino acids (9,10). Analyses of bacterial genomes have revealed putative T-box riboswitches controlling the expression of aminoacyl tRNA synthetase genes (*aaRS*) for almost every amino acid (12,14), highlighting its importance as a conserved gene regulatory element in Gram-positive bacteria. Among these, the *tyrS* (8,15,16) and the *glyQS* (17-19) T-box riboswitch systems have been well studied using various biochemical techniques that identified the conserved sequence and structural motifs required for tRNA recognition.

T-box riboswitches have a conserved organization of secondary structural elements comprising a long stem-I, stem-II, stem-IIA/stem-IIB pseudoknot and stem-III, followed by the expression platform (regulatory domain) that can form mutually exclusive anti-terminator (AT) and terminator hairpins (10,14) (**Figure 4.2A**). Initial studies established that the cognate tRNA recognition by T-box riboswitches is mediated by a codon-anticodon like interaction with the tRNA using a highly conserved 3-nt 'specifier' sequence within an internal loop of stem-I (8). In addition, the 3'-NCCA acceptor end of tRNA was shown to interact with a highly conserved and complementary 4-nt UGGN sequence within a bulge of

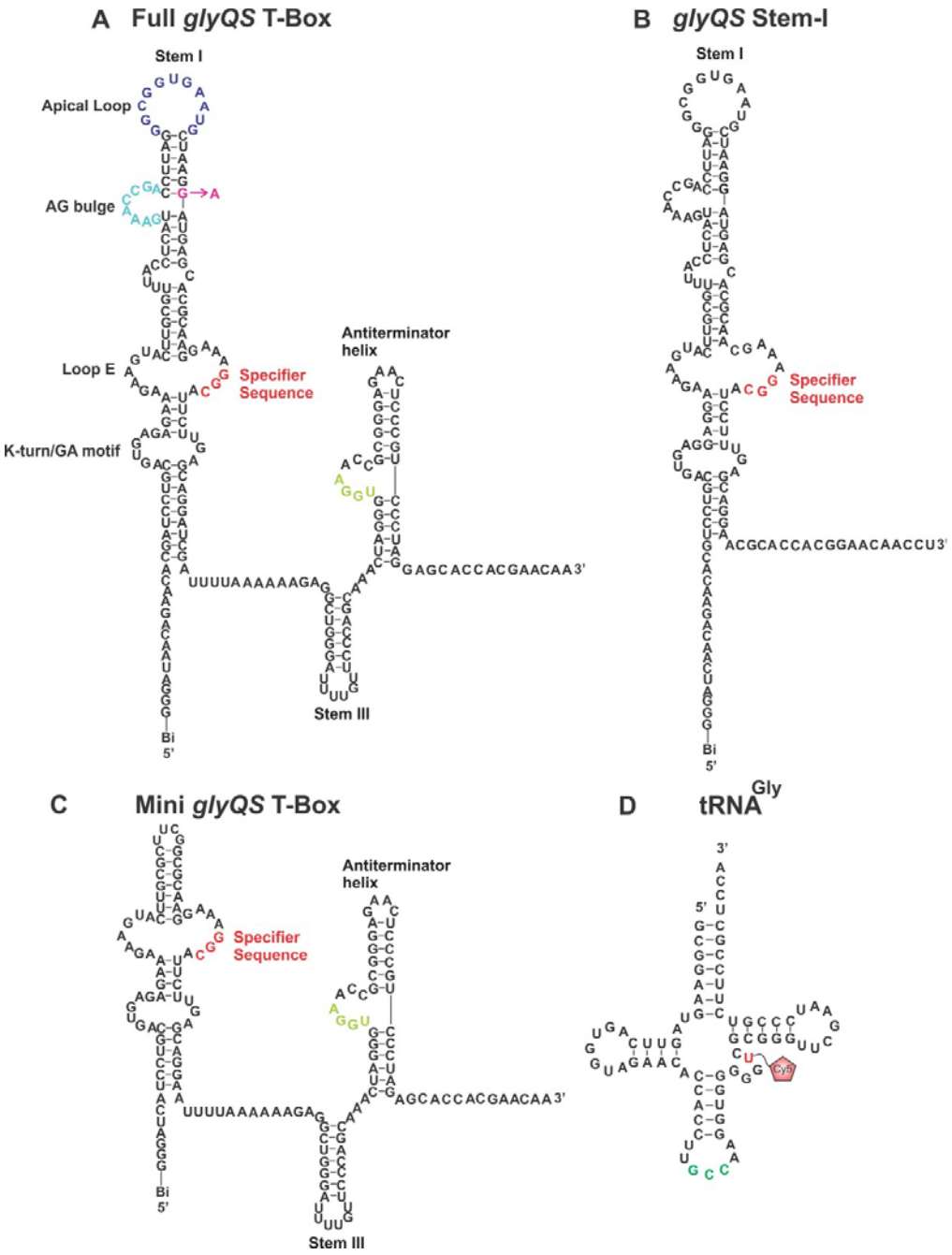

Figure 4.2 Sequences of the different T-box riboswitch designs used in this study

Sequence and secondary structures of (A) The WT full T-box riboswitch (B) Stem-I only construct (C) Mini T-box riboswitch and (D) $tRNA^{Gly}$. The specifier sequence (red), AG bulge (cyan), apical loop (blue), the conserved sequence in the AT bulge (green) are shown. The G81A mutation in the mutant riboswitch is shown in magenta. The tRNA anticodon sequence and the labeling position of Cy5 are shown in green and red, respectively in (D).

the anti-terminator hairpin. This variable 'N' nucleotide in the complementary AT bulge sequence is complementary to the discriminator base of the tRNA and therefore acts as another checkpoint for cognate tRNA recognition, in addition to the specifier sequence (20). Interestingly, in some cases, the specificity of a T-box riboswitch to its cognate tRNA can be switched to a different tRNA by simply mutating its specifier sequence to a differentcodon and by changing the variable 'N' position in the AT bulge (10). However, in all cases the maximal extent of transcriptional anti-termination achieved by the mutant T-box riboswitches was always found to be less than the extent achieved by the wild-type riboswitches with the cognate tRNA (10). This shows that the interaction between the T-box riboswitch and tRNA is more complex and extends beyond the two major known contacts, possibly involving tRNA interaction with sequences within the single-stranded linker region. This observation was also supported by structure probing studies of the *glyQS* riboswitch (19). *In vivo* studies on the *tyrS* T-box system demonstrated that uncharged tRNA acts as the effector and also showed that the riboswitch senses the relative concentrationsof both the uncharged and charged tRNA (20). The tertiary structure of tRNA was also shown to be very important for tRNA mediated anti-termination *in vitro* (21,22).

The *glyQS* riboswitch is one of the best-studied T-box riboswitch systems due to its simple structure lacking stem-II and the stem IIA/B pseudoknot elements that are present in larger systems (10) (**Figure 4.2A**). Furthermore, it was shown to function well in *in vitro* transcription anti-termination assays using purified RNA polymerase (RNAP) and *in vitro* transcribed $tRNA^{Gly}$ (**Figure 4.2D**) without the need for other cellular components (23). Therefore, the *glyQS* T-box riboswitch has been extensively used as a model system to study tRNA mediated anti-termination (17-19,22-24). In the *glyQS* riboswitch, stem-I contains highly conserved structural motifs such as a GA/K-turn motif, an S-turn/Loop E motif and the AG bulge (10) . Also present are a 11-nt apical loop, a 2-4 nt bulge below the AG bulge and an internal loop (Specifier loop) that contains the conserved 3-nt "specifier" sequence that is complementary to the tRNA anticodon (10). The downstream AT hairpin contains a 7-nt bulge with a 5'-**UGGA** sequence complementary to the **UCCA-**3' sequence in the tRNA acceptor stem (22) (**Figure 4.2A**).

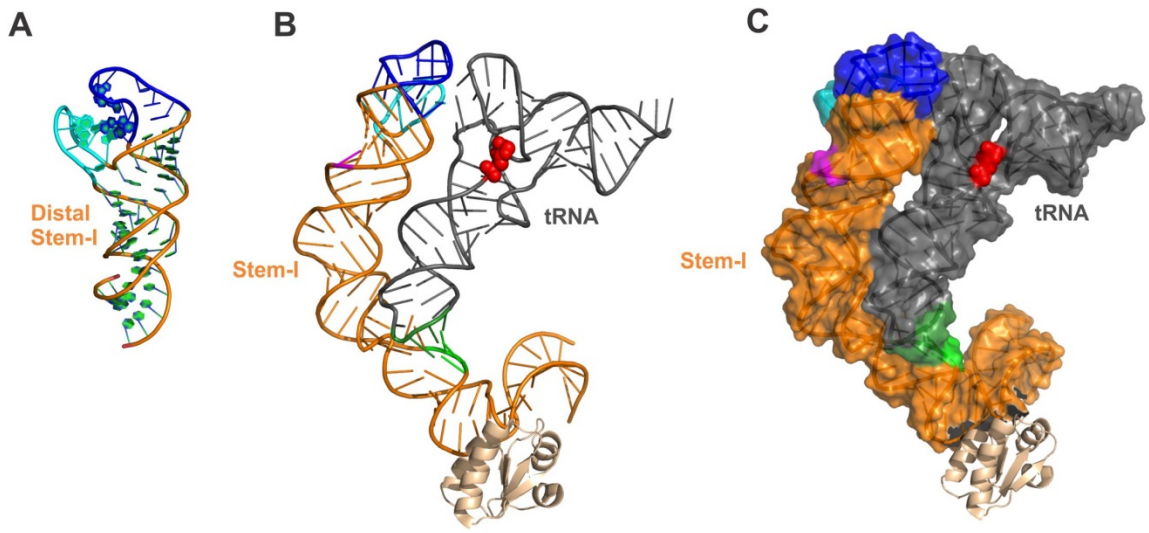

Figure 4.3 Structures of the T-box riboswitch

Crystal structures of (A) the distal stem-I region of the *Gkau glyQS* T-box riboswitch (25). Stem-I:tRNA complex of *glyQ* T-box riboswitch (26) in (B) cartoon and (C) surface representations. The T-loop 1 and T-loop 2 that form the interdigitated T-loop motif are colored as cyan and blue, respectively. Stem-I is shown in orange. The $tRNA^{Gly}$, the specifier sequence, anticodon, the G81 nucleotide and the position of Cy5 labeling (U46) are shown in grey, light green, dark green, magenta and red, respectively. The extensive non-specific helix-helix contact between the tRNA and stem-I is clearly visible in (C).

In spite of the wealth of biochemical knowledge available on the sequence and structural features of both T-box riboswitch and tRNA required for efficient anti-termination, a global structural model of the interaction between the two is not yet available. Biophysical studies have used NMR to derive the structures of different motifs in the T-box riboswitch such as the specifier loop, proximal region of stem-I, GA/k-turn motif and the AT bulge (27)and have offered insight into the local conformation and dynamics of conserved nucleotides in these motifs that are important for tRNA binding (27-29). Bioinformatics analysis of T-box riboswitch sequences predicted the existence of an interlocking T-loop interaction motif in the upper region of stem-I, which is known to dock against the tRNA elbow region in ribosomal L1 stalk and RNase P (30). A crystal structure of the upper portion of stem-I and small angle x-ray scattering (SAXS) studies of stem-I:tRNA complexes showed that the AG bulge and the conserved apical loop indeed form an interlocking T-loop motif that interacts with the tRNA elbow containing the D/T-loops (25) (**Figure 4.3A**). However, a major advance in our understanding of the tRNA interaction with the T-box riboswitch only emerged recently with the availability of two high-resolution crystal structures of the stem-I:tRNA complex, which have provided insight into the important role of stem-I in the molecular recognition of tRNA (26,31). The structures showed that stem-I arches into a "C" shaped structure to precisely recognize the geometry of tRNA (**Figure 4.3B**). The necessary kink in the stem-I backbone is introduced by a 2-4 nt bulge, situated below the AG bulge, that is hypothesized to impart the flexibility required for efficient recognition of the tRNA structure. The S-turn motif present in the upper specifier loop region further increases this kink in the backbone, which facilitates tRNA contact with the specifier sequence. The structure also shows that the GA/k-turn motif present in the proximal region of stem-I may be important for orienting the downstream expression platform of the T-box riboswitch for interaction with the tRNA 3'-end (26). These structural studies have provided valuable insights into the complex architecture of stem-I, involving a series of conserved motifs to precisely recognize not only the conserved sequence and structural motifs of tRNA, but also its overall dimensions (26,31).

Although the crystal structures (26,31) of the stem-I:tRNA complex (**Figure 4.3B**) provided details about the role of stem-I in tRNA binding, the conformations of the full T-box riboswitch and the T-box:tRNA complex are still unknown. It is also not known if the tRNA conformation when bound to the full T-box riboswitch is similar to the one observed in the crystal structure, bound to stem-I. Therefore, in the absence of a high-resolution structure of the full T-box riboswitch showing the interaction between tRNA 3'-CCA end and the AT bulge, critical molecular details of how the presence of even a small amino acid such as glycine can block this contact to affect gene regulation are still lacking. More importantly, the kinetics of tRNA binding to the T-box riboswitch have not been studied yet. Since the T-box riboswitch mainly functions at the level of premature transcription termination (10), the kinetics of tRNA binding, rate of transcription by RNAP and the presence of pause sites all play an important role in the final outcome of gene regulation. Therefore, studying the kinetics of tRNA binding and how aminoacylation affects the kinetics is important for an understanding of the molecular mechanism underlying T-box riboswitch function. Here, we have used single molecule FRET microscopy (32,33) to study the conformation and dynamics of the full *glyQS* T-box riboswitch either alone or bound to $tRNA^{Gly}$. Furthermore, we have monitored the kinetics of tRNA binding to the *glyQS* T-box riboswitch and to the isolated stem-I domain. To probe the importance of the newly discovered double T-loop motif (25,30), we have investigated two T-box riboswitch designs that we will refer to as the 'mini' and 'mutant' T-box riboswitches, in which the motif is completely deleted and mutated to destabilize it, respectively. Specifically, in the 'mini' construct, the stem-I is made shorter by deleting the 4-nt (5'-UUAC) bulge, the AG bulge, the apical loop and by capping it with an inert UUCG tetraloop (**Figure 4.2C**). By contrast, the 'mutant' T-box riboswitch construct contains a single G81A point-mutation (**Figure 4.2A**). This mutation is present at the base-pair immediately 3' to the AG bulge that forms the interlocking T-loop motif, and therefore is expected to destabilize it. Using both these complementary constructs, we tested the importance of the stability and presence of the

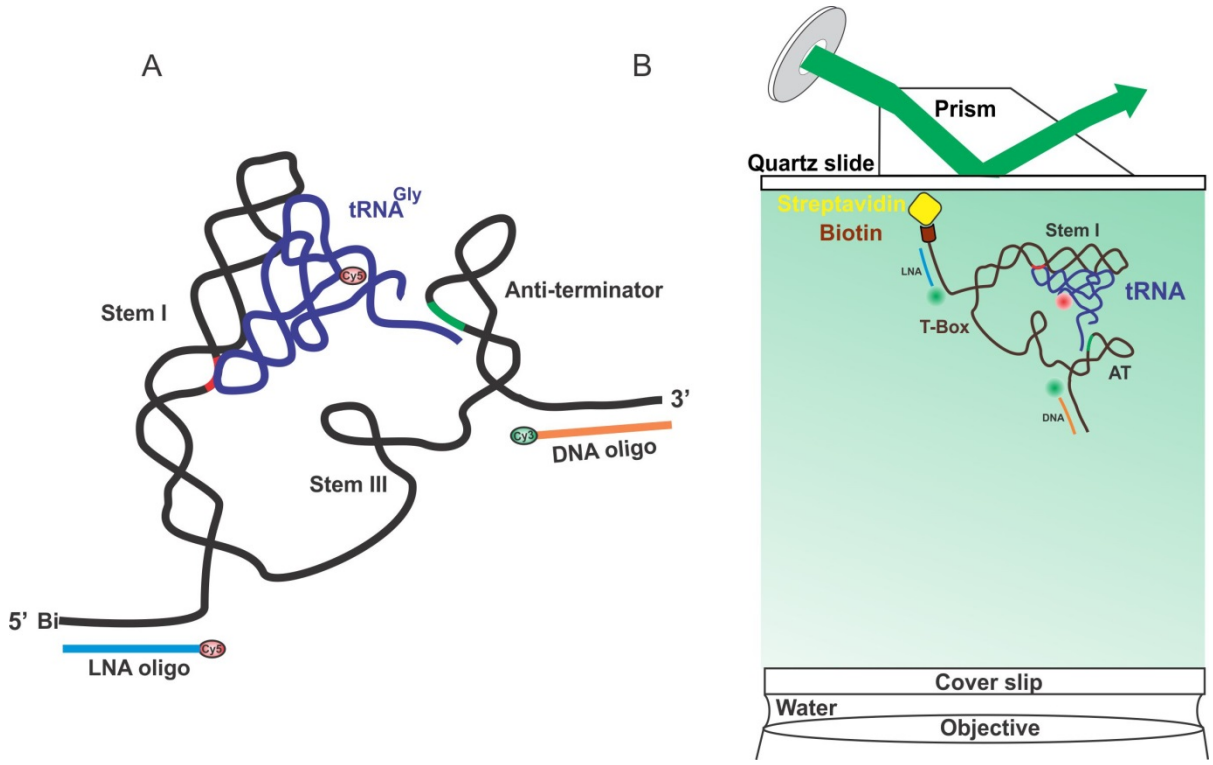

Figure 4.4 Fluorophore labeling strategy and single molecule FRET setup for studying T-box riboswitch conformation

(A) Model of the full T-box riboswitch showing the different fluorophore labeling sites for distance measurements (B) Prism TIRF microscopy setup used for the smFRET experiments.

double T-loop motif in tRNA binding. The ‘full’ T-box riboswitch constructs used in this work only include the 3’-end of the AT hairpin and do not include regions that can form the thermodynamically more stable long transcription terminator hairpin. This ensures that the AT conformation is always present in these riboswitch constructs, as required to study the conformation of the functionally relevant full T-box:tRNA complex immediately upon its exit from the RNAP. A stem-I only containing variant was also used for our smFRET and kinetic studies (**Figure 4.2D**).

Using multiple fluorophore labeling positions (**Figure 4.4A**), we here have estimated the distances between different regions of the full (WT and mutant) and mini T-box riboswitches and monitored changes under different conditions using smFRET (**Figure 4.4B**). We have also measured the kinetics of tRNA binding to the stem-I (WT), mini and full T-box riboswitches. Further, we addressed the important question of how tRNA aminoacylation affects its binding kinetics using the WT and mutant full T-box riboswitches. Our results show that all the three T-box riboswitch variants, either alone or in complex with tRNA, adopt a stable conformation devoid of rapid, large-scale structural changes. Further, we show that tRNA binding does not induce global structural changes in the full T-box riboswitch, suggesting a pre-organized conformation. In addition, we demonstrate that stem-I by itself cannot ‘anchor’ the tRNA and show that the downstream sequences containing AT hairpin are necessary for stable complex formation, required to effect gene regulation. Comparison of tRNA binding kinetics for the different T-box riboswitch constructs suggests that presence of the double T-loop motif in stem-I slows down the dissociation rate of tRNA. Further, our kinetic data clearly elucidate that the presence of amino acid on the tRNA accelerates its dissociation rate from the T-box riboswitch, thus weakening the binding of charged tRNA as compared to uncharged tRNA. Based on our smFRET and kinetic data, we propose an improved kinetic model for the molecular mechanism of the T-box riboswitch.

4.2 Materials and Methods

4.2.1 Transcription and native purification of *glyQS* T-box riboswitch

The *glyQS* T-box riboswitch RNAs were made by *in vitro* transcription of PCR generated DNA templates using AmpliScribe™ T7 High Yield Transcription Kit (Illumina Inc.). Transcription reactions were done using 3 mM 5'-biotin-GMP (Trilink BioTechnologies), 2 mM GTP and 5 mM each of ATP, CTP and UTP and incubated for 4-6 h (34). At the end of the reaction, precipitate in the transcription reaction was removed by centrifugation. The transcribed RNA was purified from the excess free nucleotides and salts by filtration (using 10 K filter), concentrated and reconstituted with 10 mM Tris, pH 7.4, 50 mM KCl, 2 mM $MgCl_2$. The concentration of RNA was measured by UV/Vis spectroscopy and its purity was analyzed using denaturing PAGE.

4.2.2 Single molecule FRET

Single molecule FRET experiments were performed using a prism-based TIRF (total internal reflection fluorescence) microscope (35-37) (**Figure 4.4B**). All measurements were taken at 10 frames per second using an intensified CCD camera (I-Pentamax, Princeton Instruments) (36,37). Quartz slides with a microfluidic channel containing inlet and outlet ports for buffer exchange were assembled as in previous work(36). In short, the surface of the microfluidic channel was coated with biotinylated BSA followed by streptavidin to immobilize biotinylated T-box RNA molecules. We flowed 50-100 pM of fluorophore labeled RNA molecules in 1x smFRET buffer (10 mM Tris-HCl, pH 6.1, 50 mM KCl, 10 mM Mg^{2+}) into the channel for surface immobilization. Any unbound molecules were washed out with the same buffer. Doubly labeled T-box riboswitch molecules or T-box riboswitch:tRNA complexes were excited with 532 nm green laser for measuring FRET and later with 640 nm red laser to check for the presence of Cy5 (acceptor) in order to distinguish low/zero FRET molecules from the ones missing an acceptor fluorophore. An enzymatic oxygen scavenging system consisting of 5 mM protocatechuic acid (PCD) and 50 mM protocatechuate-3,4-dioxygenase (PCA) along with 2 mM Trolox was included in the smFRET buffer to extend

the lifetime of fluorophores and to prevent photoblinking of the dyes(38). The raw movies were processed using IDL (Research Systems) to extract smFRET time traces and later analyzed using custom written Matlab (The Math Works) scripts. Individual smFRET time traces displaying 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 were manually selected. These selection parameters ensure that only single molecules and not aggregates or background impurities are analyzed. The ratio $I_A/(I_A + I_D)$, where I_A and I_D represent the background corrected fluorescence intensities of the acceptor (Cy5) and donor (Cy3) fluorophores, respectively was used to calculated the FRET value at every time point in smFRET traces. The first 50 frames from all the smFRET traces in every condition were used to plot population FRET ratio histograms using OriginLab. 8.1 and fit with sum of Gaussian functions that gave the number of conformations and their mean FRET values. Distances (*R*) between the fluorophores were then calculated from the equation below using the mean FRET (*E*) value of the major non-zero FRET peak and an R_0 value of 54 Å for Cy3 - Cy5 pair (35)

$$E = \frac{1}{1 + (R/R_0)^6}$$

4.2.3 Single molecule tRNA binding kinetic assays

Single molecule binding assays were done similarly as above using prism-based TIRF microscopy. Experiments were also done either by directly exciting only Cy3 (FRET mode) or by exciting both Cy3 and Cy5 with green and red lasers (direct excitation mode). In the direct mode, both the green and red lasers were used to excite the T-box and tRNA molecules respectively, all the time. Alternatively, in some cases, the green laser was turned off initially after a short time and then red laser was turned on immediately and left for the entire length of the movies. Due to a higher background, traces with Cy5 S/N ratio of >3:1 where the binding events were clearly discernible were manually selected. Multiple movies with observation times of at least 400 s were taken at 100 ms time resolution. Traces with Cy3

intensity and showing atleast one binding event were only taken for further analysis. The Cy5 intensity in the traces (in both FRET and direct excitation mode) was fit to a two-state model (bound and unbound) using segmental k-means algorithm in the QuB(39,40). From the idealized traces, dwell times of tRNA in the bound (t_{bound}) and unbound ($t_{unbound}$) states were obtained. Cumulative bound and unbound dwell time distributions were plotted and fitted to exponential functions (single or double) to obtain life-times in the bound (τ_{bound}) and unbound ($\tau_{unbound}$) states, respectively. The tRNA dissociation rate constant (k_{off}) was calculated as the inverse of τ_{bound} whereas the association rate constant was calculated by dividing the inverse of $\tau_{unbound}$ by the concentration of free tRNA-Cy5 used.

4.2.4 Labeling of DNA/LNA oligonucleotides with fluorophores

The 15-nt LNA/DNA oligo with 5'-amine linker (/5AmMC6/+T+GTTCT+GT+TGATC+C+C, RNA T_m ~ 63 °C, where '+N' represents an LNA nucleotide) was chemically synthesized from Exiqon and labeled with Cy3/Cy5. One dye pack of Cy3/Cy5 monofunctional NHS ester (GE Healthcare) was dissolved in 30 µL DMSO and used to label ~5 nmol of the LNA oligo in a 50 µL reaction volume containing 0.1 M sodium bicarbonate buffer, pH 8.7. The reaction was incubated at RT for 4 hrs with tumbling in the dark. Excess unlabeled dye was removed by precipitation with 3 volumes of 100 % ethanol and 300 mM sodium acetate, pH 5.2. After centrifugation, the pellet was dried in a speedvac and suspended in autoclaved milliQ for further use. The 14-nt DNA oligo (5'-TGTTCGTGGTGCTC-Cy3-3') complementary to the 3'-end of the T-box was chemically synthesized with a 3'-Cy3 from IDT (Integrated DNA Technologies).

4.2.5 Creation of Cy5 labeled $tRNA^{Gly}$

The Cy5-labeled $tRNA^{Gly}$ RNA was made using a two-piece ligation strategy (41). A 35-nt 5'-half of the tRNA and the 40-nt 3'-half with an internal Cy5 label on U46 were chemically synthesized from Integrated DNA Technologies (IDT). An equimolar mixture of the two RNAs was annealed by heating at 95 °C for 1 min followed by cooling on ice. The RNAs were then ligated by using 1 U/µL concentration of T4 RNA Ligase 1 (New England

Biolabs) in the supplied 1x buffer for 1 hr at 37 °C. RNA concentration was ~ 20 – 70 µM in the ligation reaction. The final ligated product corresponding to the full-length tRNA was then purified using denaturing urea PAGE. The purified tRNA was folded by heating at 65 ºC for 3 min followed by slow cooling at RT for 20 min.

4.2.6 Formation of T-box riboswitch:tRNA complexes for smFRET

1 µM of the T-box riboswitch RNA was first incubated with 10 µM of DNA or 5 µM of LNA oligo for 15 min at room temperature in 1x smFRET buffer contaning 10 mM Mg^{2+}. This was then used to make a stock of 0.5 µM T-box:tRNA complex by incubating with 1 µM of Cy5 labeled tRNA at RT for at least 15 min. Required dilution of this was made in 1x smFRET buffer to 50-100 pM for immobilization onto the slide for smFRET experiments.

4.2.7 Aminoacylation of tRNA and acid urea PAGE

Aminoacylation of Cy5 labeled $tRNA^{Gly}$ was done using *E. coli* S100 extracts in a 25 µL reaction volume following a protocol previously described (42) . In short, ~ 4 µM of tRNA-Cy5 was charged with 0.5 µL of *E. coli* S100 extracts in a 25 µL reaction containing 100 mM HEPES-KOH, pH 7.6, 200 µM Glycine, 10 mM ATP, 1 mM DTT, 10 mM KCl, 20 mM $MgCl_2$ and incubated at 37 °C for 30 min. The reaction was stopped by adding 3 M sodium acetate, pH 5.2, to a final concentration of 100 mM in 100 µL volume. Acid urea page (42) was used for checking the aminoacylation efficiency of the reaction. We added 2 µL of the 25 µL charging reaction to 98 µL of the gel loading dye (0.1 M sodium acetate, pH 5.0, 8 M urea, 0.05 % bromophenol blue, 0.05 % xylene cyanol FF). 10 µL of this was run on a 10 % acid urea (8 M) gel using 10 W power at 4 °C for ~ 20-25 hrs in 30 cm long gel plates. The gel was cast using 0.1 M sodium acetate, pH 5.0 and the same buffer was used for gel electrophoresis. The buffer was constantly circulated between the top and bottom chambers using a peristaltic pump to maintain the pH at 5.0. The gel was scanned with Typhoon scanner (GE Healthcare Life Sciences) using 640 nm red laser for imaging the Cy5 labeled tRNA. Aminoacylated tRNA band will move slightly slower than uncharged tRNA enabling the estimation of charging efficiency. Under our reaction conditions, the aminoacylation

efficiency was close to ~ 80-85 % and reached saturation quickly in < 10 min (**Figure 4.12A**). For the binding experiments, charged tRNA from the 100 µL stopped reaction stock was diluted to 200 uL using 100 mM sodium acetate, pH 5.2 and the protein components were removed using phenol-chloroform extraction. Concentration of this phenol-chloroform extracted charged tRNA is estimated and diluted to required concentration in final 1x smFRET buffer for single molecule binding experiments.

4.3 Results

To study the conformation and dynamics of the *glyQS* T-box riboswitch by smFRET, we used each a Cy3 and a Cy5 fluorophore labeled DNA and LNA oligonucleotides complementary to the 3' and 5' ends, respectively, of the RNA (**Figure 4.4A**). This hybridization labeling strategy for labeling was employed since it allowed for the use of a natively purified T-box riboswitch RNA. Furthermore, internal covalent labeling of large RNAs is challenging and requires purification methods that denature the RNA, possibly leading to a heterogeneously folded population of molecules (43). Therefore the *glyQS* T-box riboswitch RNA for the smFRET experiments was natively purified to avoid any potential misfolded conformations as a result of denaturing protocols. Unmodified synthetic $tRNA^{Gly}$ labeled with Cy5 at position U46 (**Figure 4.2D**) in the variable loop was used for all smFRET studies. Since this position is not conserved and in the crystal structure does not form any interactions with stem-I (**Figure 4.3B**), labeling at this position is not expected to affect tRNA binding to the T-box riboswitch. Using this labeling strategy and prism-based smFRET microscopy, we estimated the distances between the bound $tRNA^{Gly}$ (from position U46 in the variable loop) and base of stem-I and the base of AT hairpin as illustrated in **Figure 4.4B**. Additionally, the distance between the bases of stem-I and AT hairpin was measured in the presence and absence of tRNA to investigate any tRNA induced conformational changes in this region. The tRNA used for all smFRET-based distance measurements was uncharged and Cy5-labeled unless otherwise specified. Since the mini T-box riboswitch lacked a long 5' single stranded region to which DNA/LNA oligo could be

hybridized, a distance estimate could only be obtained between its 3'-end and the bound $tRNA^{Gly}$.

4.3.1 $tRNA^{Gly}$ is bound in a similar orientation to both the mini and full T-box riboswitches

First, we measured FRET between the base of AT hairpin and the bound tRNA-Cy5 to investigate the position of the tRNA relative to the AT hairpin in the complex (**Figure 4.5A**). This will also reveal the presence of any large-scale motions of the $tRNA^{Gly}$ bound to the T-box riboswitch. In 10 mM Mg^{2+}, the population FRET histograms of the WT full T-box:tRNA complexes displayed a single FRET state with a mean FRET value of 0.25±0.13 (**Figure 4.5B**). A few smFRET traces displayed zero FRET state, but their population was negligible (**Figure 4.5C**). In comparison, the mutant riboswitch showed a bimodal distribution with a major (81 %) 0.28 ± 0.13 and a minor (19 %) ~0 FRET state (**Figure 4.5D**). The FRET histograms for the mutant and WT full T-box riboswitches are very similar to each other with only subtle differences. The two FRET states of ~0.25 and 0 correspond to distances between the fluorophores of ~65 Å and >75 Å respectively (using a Förster radius of 54 Å between Cy3 and Cy5 (44)). Visual inspection of individual smFRET traces showed that transitions between the two states were rare (<5 % molecules, **Figures 4.5C,E**) and a majority of molecules displayed stable FRET states for extended periods of time. It is interesting to note that the tRNA was very stably bound (dwell times >100 s) under these conditions and some complexes were visible even after an hour of immobilization on the slide surface.

Previous *in vitro* binding experiments using ITC showed that deleting the interlocking T-loop motif (**Figure 4.3A**) from a stem-I only construct resulted in weakened tRNA binding, which suggests a possible role for the motif in tRNA 'docking' during transcription (26). We therefore investigated the mini T-box lacking the interlocking T-loop motif to test its tRNA binding ability and to measure distance between the bound tRNA and base of AT hairpin (**Figure 4.5F**). Single molecule FRET traces at 10 mM Mg^{2+} showed that the mini T-box riboswitch binds tRNA relatively well. However, higher (50 mM) Mg^{2+} was required to

extend the tRNA bound times to enable longer observation of single complexes using smFRET. These observations suggest that while the interlocking T-loop motif enhances the affinity for tRNA by decreasing its dissociation rate, it is not required for tRNA binding in the presence of the downstream RNA containing stem-III and AT hairpin. At 50 mM Mg^{2+}, the FRET histogram showed a major (~78 %) 0.31 ± 0.11 FRET state and a minor (~22 %) ~0 FRET state (**Figure 4.5F**). These FRET populations correspond to distances of ~62 Å and >75 Å, respectively. The ~ 0 FRET state appears to be a genuine conformation with large separation between the fluorophores (>75 Å) and not due to the absence of a FRET acceptor, as confirmed by the observation of Cy5 intensity when excited directly with 640 nm red laser (**Figure 4.7C**). SmFRET traces of the mini T-box-tRNA complexes displayed rare transitions between the two FRET states at a time resolution of 100 ms, suggesting that the two conformations are stable and not readily in exchange with each other, similar to the WT and mutant full T-box riboswitch:tRNA complexes (**Figure 4.5G**). Importantly, the population FRET histogram of the mini T-box:tRNA complexes was highly similar to the FRET histograms of both the full T-box:tRNA complexes. This observation suggests that the relative orientation of the tRNA bound to the mini and full T-box riboswitches is very similar.

4.3.2 Conformation of the linker region is not affected by tRNA binding

In T-box riboswitches, the single stranded 'linker' region connecting the stem-I and AT hairpin is variable in size, sometimes lacking stem II and the stem IIA/stem IIB pseudoknot, and thus is not highly conserved (10). In the *glyQS* riboswitch, this region contains a single-stranded 14-nt stretch followed by stem-III (**Figure 4.2A**). The role of this linker region and its importance in tRNA binding are not well understood. Previous studies have shown that the linker region does not contribute to tRNA binding and RNA structure probing data have suggested that it exists largely as a single-stranded region (19). However, upon tRNA binding, a majority of the bases in the linker region showed protection against chemical footprinting. This prompted the authors to propose that this region undergoes a structural change, presumably by interacting with the tRNA directly or by forming new

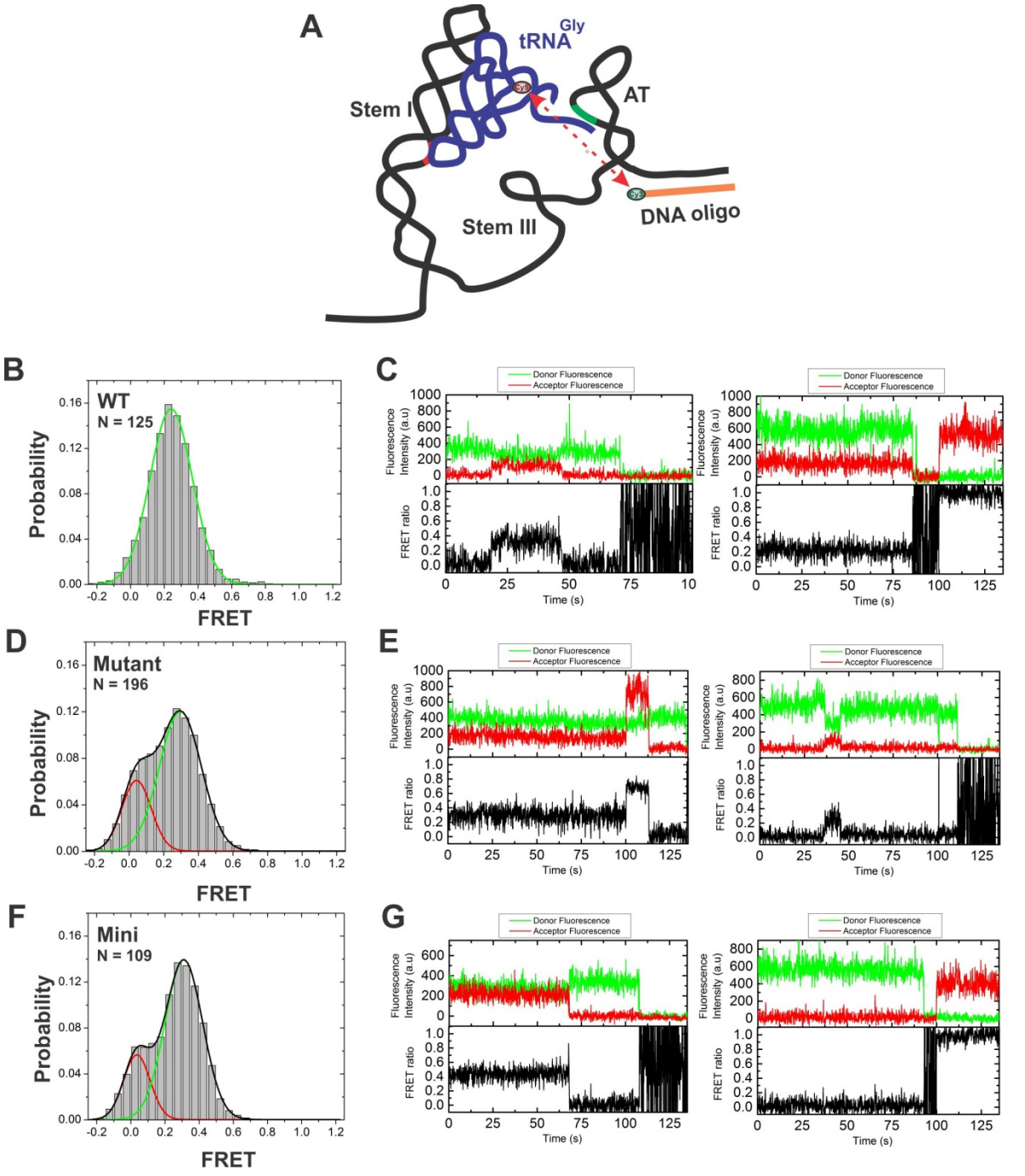

Figure 4.5 Determination of the distance between the bound tRNA and the base of the AT hairpin

(A) Schematic of the full T-box riboswitch showing the positions of fluorophores for FRET. Population FRET histograms and example smFRET traces for the (B, C) WT, (D, E) Mutant and (F, G) Mini T-box riboswitches.

intra-molecular tertiary interactions (19). Therefore, to investigate the conformation of the linker region and to check if tRNA binding induces large-scale structural changes, we labeled the full (WT and mutant) T-box riboswitches at the bases of stem-I and the AT hairpin using 5'-Cy5-LNA and3'-Cy3 DNA oligonucleotides, respectively, for smFRET studies (**Figure 4.6A**). This labeling scheme permits the measurement of end-to-end distance changes in the linker region using smFRET.

In the absence of free unlabeled $tRNA^{Gly}$, the FRET histogram for the doubly labeled full WT T-box riboswitch at 10 mM Mg^{2+} displayed two states with mean values of ~0.52 ± 0.12 (91 %) and ~ 0.09 ± 0.12 (9 %) (**Figure 4.6B**). Addition of 2 μM unlabeled (K_d ~ 50 - 100 nM, as measured from our smFRET experiments and previous studies) $tRNA^{Gly}$ only slightly increased the mean value of the major FRET state to ~ 0.57 ± 0.12, but did not affect the relative distribution (~ 10:1 ratio) between the two states (**Figure 4.6C**) (26). Similarly, in the absence of tRNA, the full mutant T-box riboswitch showed two FRET states with mean values of 0.65 ± 0.14 (78 %) and ~ 0 (22 %) (**Figure 4.6E**). In the presence of 1 μM tRNA, the mean value of the major FRET state modestly shifted to 0.71 ± 0.12 (83 %) with a 5 % increase in its population (**Figure 4.6F**). The increase in FRET value by ~0.05 only corresponds to distance changes by only a few Å, which is within error. Therefore, these results show that tRNA binding does not induce large-scale conformational changes in the linker region, at least as detected by our specific FRET labeling scheme. Individual smFRET traces of the WT and mutant riboswitches showed very stable FRET values with only few (in <5 % of all traces) transitions between the two FRET states, in both the presence and absence of tRNA (**Figures 4.6D,G**).

The FRET histogram of the mutant T-box riboswitch under low (1 mM) Mg^{2+} condition also showed a major ~ 0.46 ± 0.13 (87 %) and a minor ~ 0 (13 %) FRET state (**Figure 4.6H**). A significant decrease in the mean FRET value of the major state from ~0.65 to ~ 0.46 indicates that the linker region adopts an extended conformation under lower, more physiologically relevant Mg^{2+}. Again, the addition of 1 μM unlabeled $tRNA^{Gly}$ had no significant effect on the FRET histogram (**Figure 4.6I**). However, it should be noted that

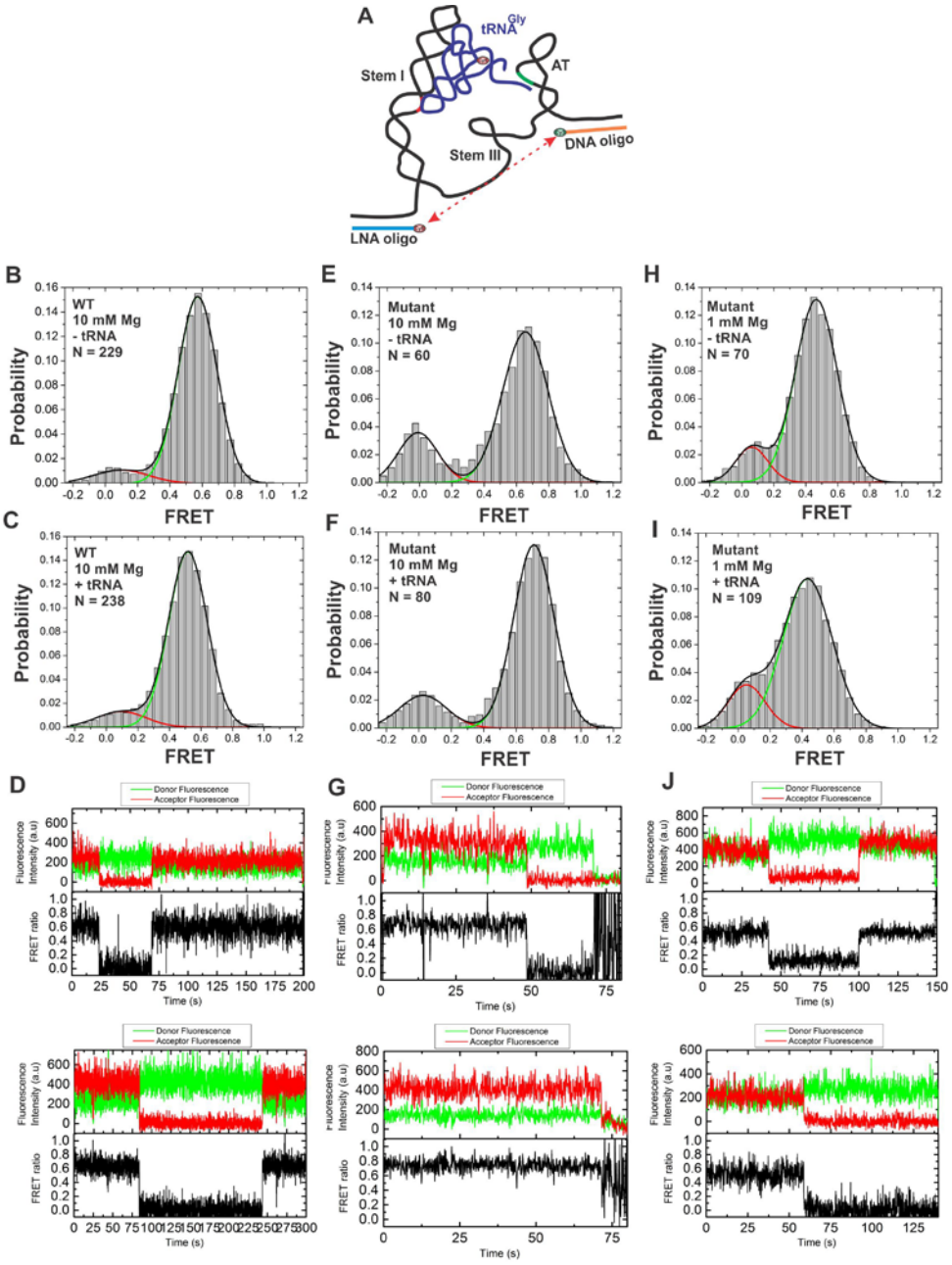

Figure 4.6 smFRET between the bases of stem-I and the AT hairpin

(A) Schematic of the full T-box riboswitch showing the positions of fluorophores for measuring FRET between the bases of stem-I and the AT hairpin. Population FRET histogram for (B) WT in 10 mM Mg^{2+}, - tRNA, (C) WT in 10 mM Mg^{2+}, + tRNA, (E) Mutant in 10 mM Mg^{2+}, - tRNA, (F) Mutant in 10 mM Mg^{2+}, + tRNA, (H) Mutant in 10 mM Mg^{2+}, - tRNA, (I) Mutant in 10 mM Mg^{2+}, + tRNA. Examplary smFRET traces for (D) the WT in 10 mM Mg^{2+}, (G) Mutant in 10 mM Mg^{2+} and (J) Mutant in 1 mM Mg^{2+}.

our smFRET experiments where doubly labeled complexes could not be observed with Mg^{2+} <5 mM (19). These data show that tRNA binding does not induce detectable large-scale conformational changes, at least between the bases of stem-I and AT hairpin, suggesting apre-organized conformation of the full T-box riboswitch.

4.3.3 The distance between the base of stem-I and the bound $tRNA^{Gly}$ shows a two-state distribution that is sensitive to Mg^{2+}

We hybridized a 5'-Cy3-labeled LNA oligonucleotide to the single stranded region upstream of stem-I to measure FRET between the bound tRNA and the base of stem-I in the full T-box riboswitch (**Figure 4.7A**). In the WT riboswitch, FRET histogram at 10 mM Mg^{2+} showed a major (61 %) 0.38 ± 0.12 and a minor (39 %) ~0 FRET state (**Figure 4.7D**). The mutant T-box riboswitch showed a very similar two-state distribution with 0.42 ± 0.11 (59 %) and ~0 (41 %) FRET states (**Figure 4.7B**). In both the WT and mutant riboswitches, individual smFRET traces showed that these states are very stable with only a few (~4 %) traces showing slow transitions between them (**Figure 4.7C, E**). These FRET transitions may be either due to the motion of tRNA in the complex or may be a result of different conformations of the stem-I domain. The major ~0.4 FRET state corresponds to a distance estimate of ~58 Å between the base of stem-I and position U46 of tRNA. The corresponding distance between U46 of tRNA and the base of stem-I as measured from the crystal structure (26) is ~ 64 Å. Since the structure was solved in complex with a k-turn binding protein that orients the 3'-end of stem-I towards the tRNA, this distance is expected to increase in the absence of the protein. In addition, the full T-box RNA constructs used in our studies have a 9-bp stem below the k-turn/GA motif as compared to a 6-bp stem in the crystal structure, resulting in a slightly longer stem-I in our constructs. Therefore, the presence of a longer stem-I and the absence of a k-turn stabilizing protein in our smFRET experiments is expected to yield a distance estimate in excess of the ~64 Å seen in the crystal structure (26). This suggests that in solution, the tRNA is slightly closer to the base of stem-I when bound to the full T-box riboswitch. Consequently, we speculated that the ~ 0 FRET state with distance >75 Å between the base of stem-I and the tRNA corresponds to molecules in which the

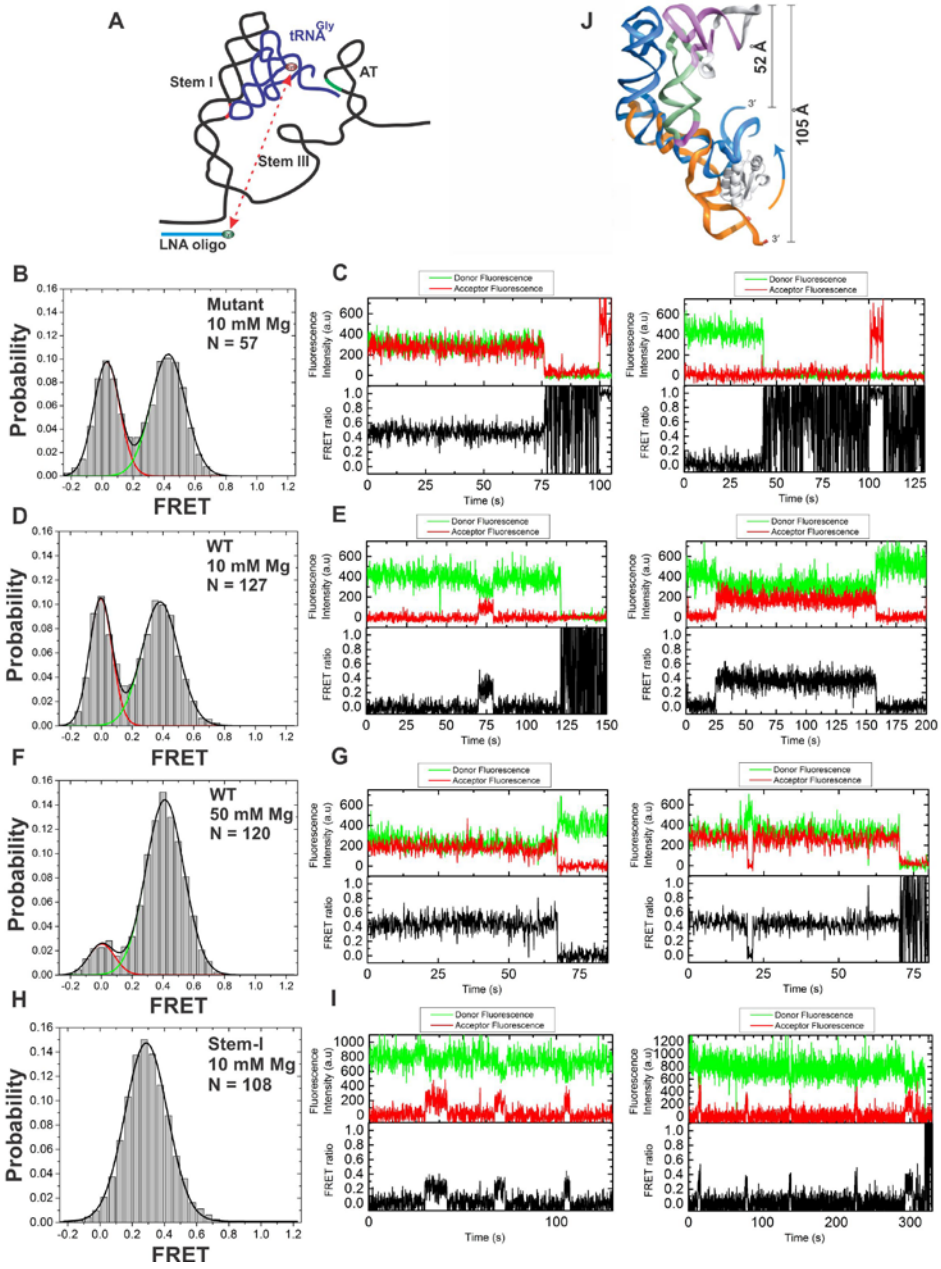

Figure 4.7: smFRET between the bound tRNA and the base of stem-I

(A) Schematic of the full T-box riboswitch showing the positions of fluorophores for FRET between tRNA and base of stem-I. Population FRET histograms and example smFRET traces for (B, C) Mutant T-box riboswitch in 10 mM Mg^{2+}, (D, E) WT in 10 mM Mg^{2+}, (F, G) WT in 50 mM Mg^{2+} and (H, I) stem-I in 10 mM Mg^{2+}. (J) Structural comparison of the crystal structure of stem-I (blue) and the NMR structure of the proximal region of stem-I (orange) (33). Distances between the 3'-end of stem-I and the tRNA are shown.

k-turn motif is in an extended conformation. The GA/k-turn motif in RNA is a well characterized structural motif (45,46). Various k-turn motifs are identified in structurally diverse RNAs and are known to have differential requirements for Mg^{2+} in the absence of stabilizing proteins (29,45,47). We therefore hypothesized that the addition of higher Mg^{2+} concentrations in the smFRET experiments should induce more T-box riboswitch molecules to adopt kinked conformations in stem-I. Indeed, as expected, when the Mg^{2+} was increased to 50 mM, the population of the major FRET state, now with a slightly higher mean FRET value of 0.41 ± 0.12, increased from ~61 % to ~90 % (**Figure 4.7F**). This observation shows that the conformation of stem-I is Mg^{2+} dependent and also explains the need for a high (at least 5 mM) Mg^{2+} for $tRNA^{Gly}$ binding and for high anti- termination efficiency in *in vitro* transcription assays for the *glyQS* riboswitch (19). This high Mg^{2+} dependency of the T-box riboswitch could be a result of a cumulative requirement of the various bulges and motifs present in the stem-I region (26). A comparison of the 'kinked' conformation of the stem-I in the crystal structure (33) and the 'extended' conformation of the NMR structure of the stem-I proximal region (PDB ID: 2KZL) is shown in **Figure 4.7J**. The distances between the 3'-end of stem-I and the bound tRNA in these structures support our smFRET data and our hypothesis that the ~0.4 and ~0.0 FRET states correspond to the kinked and extended conformations of stem-I, respectively.

Figure 4.8 summarizes the distance estimates obtained for the WT full T-box riboswitch at 10 mM Mg^{2+} from our smFRET measurements. Among the three estimated distances, the largest distance of ~62 Å is between the base of AT hairpin and the bound tRNA. We are working on using these distance estimates to generate a solution model of the full T-box:tRNA complex using SAXS, which would provide valuable insight into the global organization of the different domains of full T-box riboswitch. In this direction, the distances we measured using smFRET will be important in refining the SAXS model using the triangulation method.

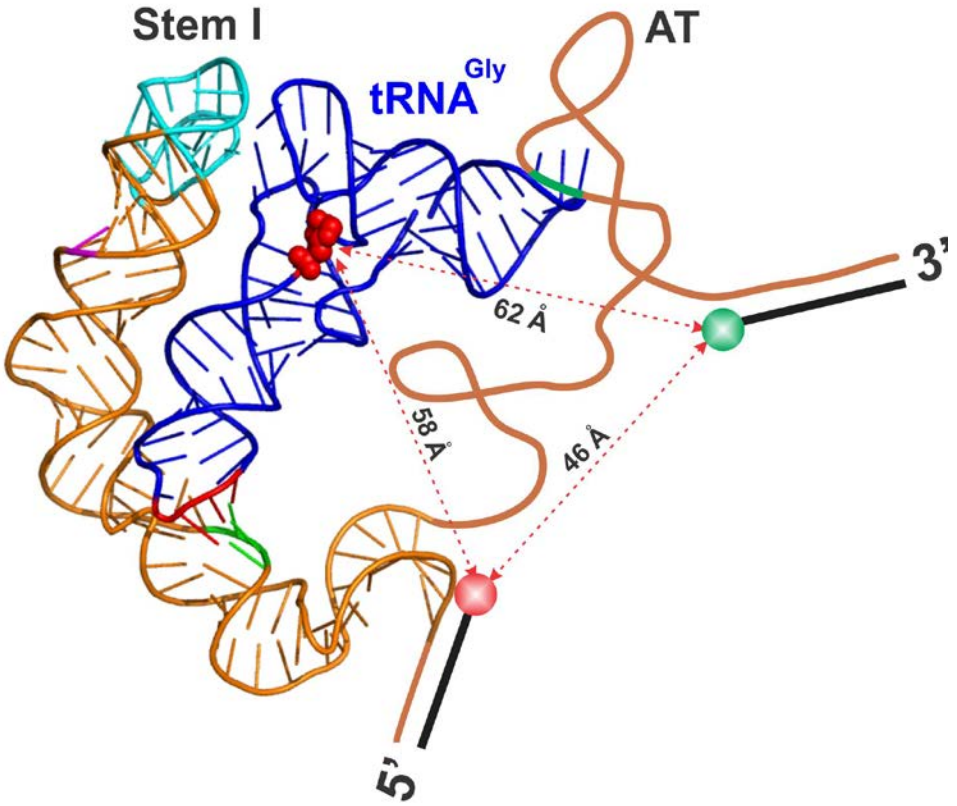

Figure 4.8 Schematic showing the distance estimates between different regions of the full WT T-box riboswitch

4.3.4 The anti-terminator hairpin is necessary to stably anchor the tRNA

Next, we performed smFRET experiments on the stem-I only construct (**Figure 4.2B**) with an aim to deduce conformations of the two FRET states observed for the full T-box riboswitches (**Figure 4.7B-G**). Intriguingly, we could not observe stable stem-I:tRNA complexes in our smFRET experiments, even at high (50 mM) Mg^{2+} concentration. This was surprising since previous ITC measurements indicated that stem-I alone can bind tRNA equally well or even stronger than the full T-box riboswitch (26). Our failure to observe stem-I:tRNA complexes in smFRET experiments (without free tRNA in the microfluidic channel) suggested that the tRNA binds only weakly to the stem-I only RNA and dissociates into the microfluidic channel containing buffer, where their concentration would be too low to allow them to rebind stem-I. This is contrary to previous studies and the recently modified model for the T-box riboswitch function in which stem-I was proposed to be important for 'anchoring' the tRNA during transcription (25,26,48). Therefore, we performed the experiment with 25 nM free Cy5 labeled tRNA in the microfluidic channel. Transient binding events of tRNA to stem-I were detected by the appearance of FRET (**Figure 4.7I**). The multiple short binding events were pooled together to construct a FRET histogram to obtain the average distance estimate between tRNA and the base of stem-I. The FRET histogram of stem-I:tRNA complexes at 10 mM Mg^{2+} shows a distribution that can be fit with a Gaussian function to give a mean of 0.28 ± 0.13 (**Figure 4.7H**). This corresponds to a distance if ~63 Å, similar to the crystal structureand larger than the distance obtained for the major conformation using the full T-box riboswitch (~58 Å) (**Figure 4.7D**). This shows that the tRNA is slightly closer to the base of stem-I when bound to the full T-box riboswitch as compared to stem-I only construct. We speculate that this may be due to a sharper kink in the stem-I of the full T-box riboswitch or due to a different tRNA conformation when bound to the full T-box riboswitch. The fraction of the ~0 FRET state, if at all it is sampled in the stem-I:tRNA complex, could not be probed in this experiment, since tRNA binds only transiently and therefore the histogram only shows a single conformation (**Figure 4.7I**). Our smFRET data showing only transient binding of tRNA to the stem-I but stable binding to the

full T-box riboswitches (WT, mutant and mini) demonstrate that the sequences downstream to stem-I including the AT hairpin are required for stable tRNA binding.

tRNA binding kinetics of the T-box riboswitch

The T-box riboswitch has to bind uncharged tRNA and stabilize the AT hairpin during active transcription elongation by RNAP in real-time to prevent termination (10,17). Therefore, studying the kinetics of tRNA binding is important for understanding the temporal details of its mechanism (48). To this end, we have used smFRET microscopy (33,35,49) to study the tRNA binding kinetics of different T-box riboswitch constructs and to probe the effect of aminoacylation and destabilization of the double T-loop motif on the binding kinetics. These are different from the distance measurement experiments where only green laser was used to excite the donor fluorophore (Cy3) to measure FRET. In these kinetic experiments, red laser was also used to directly excite the Cy5 on tRNA to detect its binding events and measure the individual dwell times (**Figure 4.9**). This has the advantage of detecting all tRNA binding events, including the ones with ~0 FRET values, which cannot be detected when using only green laser.

4.3.5 Kinetics of stem-I vs full WT T-box riboswitch: The interlocking T-loop motif increases the affinity for tRNA by lowering the dissociation rate

The T-box riboswitch functions during transcription by RNAP and therefore, the kinetics of tRNA (ligand) binding play an important role for achieving proper gene regulation (17,50). Using smFRET, we measured the kinetics of tRNA binding to different T-box variants (**Figure 4.9A**). SmFRET is powerful in detecting very transient binding events (dwell times as short as 100 ms or even lower) in real-time and therefore can be used to probe even weak binding interactions at concentrations significantly below the K_d, as opposed to other ensemble methods requiring high RNA concentrations (51). Binding events can be detected as an increase in the Cy5 intensity over the background. The dwell times in the bound and unbound states can be obtained for a large number of molecules to obtain the rates (**Figure 4.9B, also see section 4.2.3**).

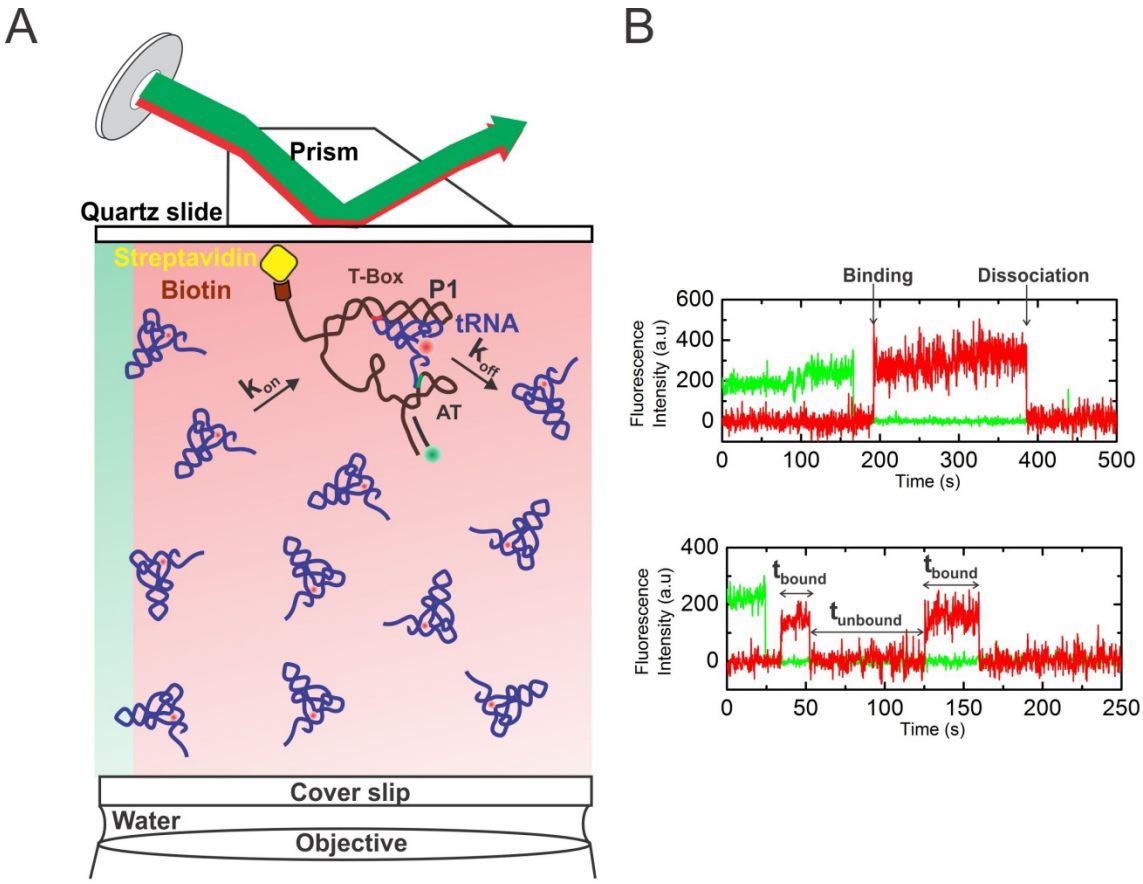

Figure 4.9 Strategy for measuring the tRNA binding kinetics at the single molecule level

(A) Prism based TIRF microscopy setup for measuring tRNA binding rate constants (k_{on} and k_{off}) to different T-box riboswitch constructs. (B) Example single molecule traces showing tRNA binding and dissociation events as sharp appearance and loss of Cy5 signal. The dwell times in the bound (t_{bound}) and unbound ($t_{unbound}$) states are also shown from which rate constants of dissociation and association can be estimated.

Intrigued by our failure to observe stem-I:tRNA complexes on the slide, we first probed the tRNA binding kinetics of stem-I using 25 nM free tRNA-Cy5 using both FRETand direct excitation with red laser. As mentioned previously, investigation of the smFRET traces showed that tRNA bound only transiently to stem-I as expected (**Figure 4.7I, Figure 4.11A**) with values of t_{bound}. In contrast, the full WT T-box riboswitch bound tRNA very well with long dwell times extending even up to 400 s (**Figure 4.11B)**. The cumulative distribution of dwell times in the bound state (t_{bound}) fit well with a single exponential function giving a tRNA dissociation rate constant (k_{off}) of ~0.23 s^{-1} for stem-I (**Figure 4.10B and Table 4.1**). The distribution of dwell times in the unbound state ($t_{unbound}$) also fit fairly well to a single exponential function with a life time of ~51 s, giving an association rate constant, k_{on} value of ~0.8 x 10^6 M^{-1} s^{-1} for the stem-I only construct (**Figure 4.10A**).

In comparison, the kinetics of the full WT T-box riboswitch were more heterogeneous and fit better with double-exponential functions. The cumulative distribution of $t_{unbound}$ had a slower 132.7 s (46 %) and a faster 29.9 s (54 %) component, with a weighted average of ~72.5 s, from which we calculated a k_{on} value of ~0.7 x 10^6 M^{-1} s^{-1} (**Figure 4.10C, Figure 4.11B and Table 4.1**). In a similar way, an average k_{off} value of 0.038 s^{-1} was obtained from the cumulative distribution of t_{bound} with a slower 96.5 s (24 %) and a faster 4.1 s (76 %) lifetime (**Figure 4.10D, Figure 4.11B and Table 4.1**). This shows that, while the average k_{on} values are similar for both RNAs, the k_{off} for stem-I is ~6 fold higher than the k_{off} value for the full WT T-box riboswitch. This clearly demonstrates that the sequence downstream of stem-I, including the AT hairpin, is required to stably bind the tRNA for longer time.

4.3.6 Kinetics of the full mutant versus WT riboswitches: Destabilization of the double T-loop motif increases the tRNA dissociation rate

The full mutant T-box riboswitch also showed heterogeneous kinetics with double-exponential distributions of both the t_{bound} and $t_{unbound}$ dwell times, similar to the WT T-box riboswitch. It has an average k_{on} value of ~0.8 x 10^6 M^{-1} s^{-1} with a slower 74.4 s (79 %) and a

faster 15.8 s (21 %) lifetime (**Figure 4.10E, Figure 4.11C and Table 4.1**). We calculated an average k_{off} value of 0.08 s^{-1} from the slower 24.2 s (47 %) and faster 1.1 s (53 %) components (**Figure 4.10F, Figure 4.11C and Table 4.1**). The average k_{on} value is similar to the value for WT riboswitch whereas the average k_{off} is faster by ~2-fold. However, the individual lifetimes for $t_{unbound}$ and t_{bound} are roughly 2- and 4-fold shorter, respectively, than those of the WT T-box riboswitch and display different distributions (**Figure 4.10, Figure 4.11C and Table 4.1**). These data show that the G81A substitution proximal to the double T-loop motif (**Figures 4.2A, 4.3)** in the mutant riboswitch has a modest destabilization effect of increasing the dissociation rate by ~2-fold. While the average association rate was not significantly affected, the individual values of $t_{unbound}$ were shorter for the mutant T-box riboswitch.

4.3.7 Kinetics of tRNA binding to the mini and mutant T-box riboswitches

Single molecule binding experiments performed at 20 nM tRNA show that the mini T-box riboswitch binds uncharged tRNA relatively well with dwell times in the bound state (t_{bound}) ranging from 0.1 s to 75 s (**Figure 4.11D**). The cumulative distribution of $t_{unbound}$ fit well with a single-exponential function with a lifetime of ~42.5 s from which we estimate a k_{on} value of ~1.2 x 10^6 M^{-1} s^{-1} (**Figure 4.10G and Table 4.1**). The cumulative distribution of t_{bound} fit well with a double-exponential function with a major (65 %) slow (~18.4 s) component and a minor (35 %) fast (~0.95 s) component. From these, an average lifetime of tRNA bound to the mini T-box riboswitch was calculated as ~12.3 s giving a k_{off} value of ~0.08 s^{-1} (**Figure 4.10H and Table 4.1**). These kinetics are similar to those of the mutant full T-box riboswitch with some subtle differences. The cumulative distributions of t_{bound} showed double exponential behavior in both constructs with similar short and long lifetimes, but their relative contributions are different. In particular, the value of the slower component is higher by ~6 s for the mutant T-box riboswitch. Both constructs also have similar average k_{off} values of ~0.08 s^{-1}. This shows that the mini and the full mutant T-box riboswitches bind tRNA equally well with only subtle differences in their kinetics (**Figures 4.10E-H**).

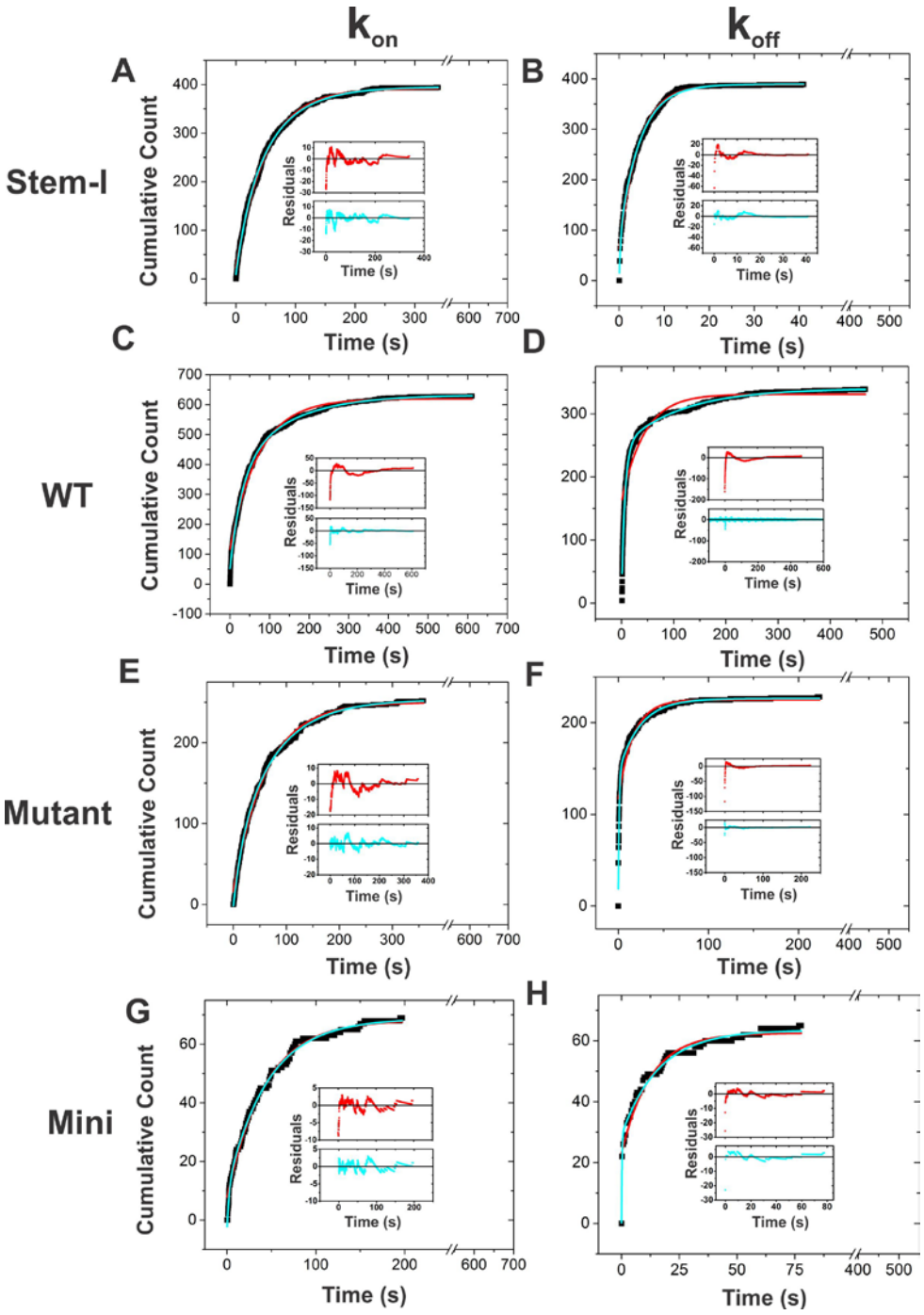

Figure 4.10: tRNA binding kinetics of the different T-box riboswitch constructs

Cumulative dwell time distributions fit to single (red) and double (cyan) exponential functions for calculating k_{on} and k_{off} values for stem-I (A, B), WT (C, D), Mutant (E, F) and Mini (G, H) T-box riboswitch constructs. The residual plots for single (red) and double (cyan) exponential fits are shown in the inset.

	Single			Double		
	$\tau_{unbound}$ (s)	τ_{bound} (s)		$\tau_{unbound}$ (s)		τ_{bound} (s)	
Stem-I	51	4.3		22.4 (26 %)	60.6 (74 %)	0.3 (21 %)	4.6 (79 %)
WT	72.5	33.7		29.9 (54 %)	132.7 (46 %)	4.1 (76 %)	96.5 (24 %)
Mutant	61.6	16.6		15.8 (21 %)	74.4 (79 %)	1.1 (53 %)	24.2 (47 %)
Mini	42.5	12.4		2.3 (19 %)	45.7 (81 %)	0.95 (35 %)	18.4 (65 %)

Table 4.1: Kinetics of tRNA binding to the different T-box riboswitch constructs

The life times (τ) of the single and double exponential fits to the cumulative distributions of the unbound and bound dwell times for the different T-box variants are shown.

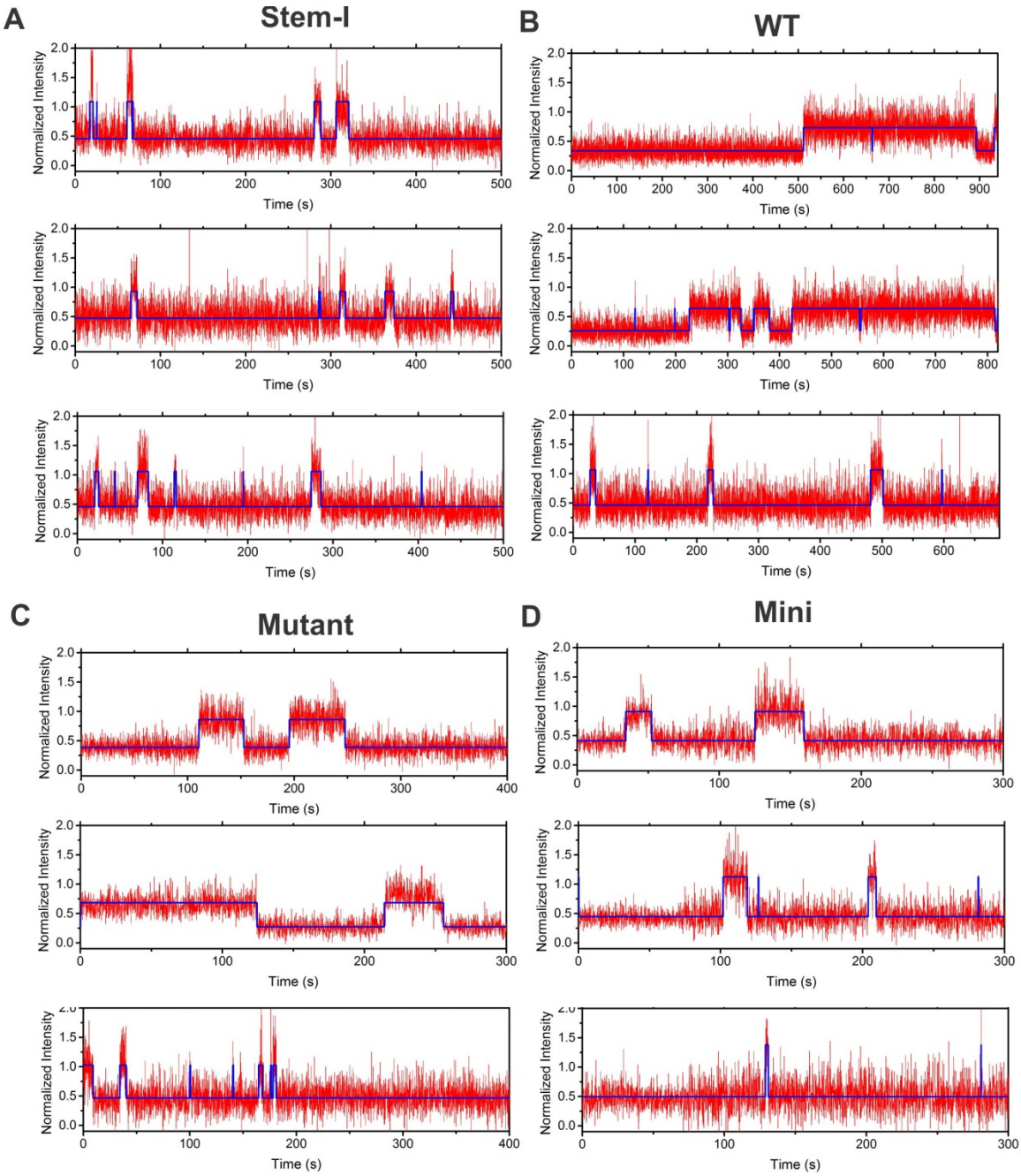

Figure 4.11 Example single molecule binding traces

Single molecule traces showing multiple tRNA binding events to the (A) Stem-I (B) WT (C) Mutant and (D) Mini T-box riboswitch constructs. The HMM (blue) fits for the traces are shown overlaid onto the raw Cy5 intensity (red).

4.3.8 Aminoacylation of tRNA increases its dissociation rate from the T-box riboswitch

In the T-box riboswitch, the presence of an amino acid coupled to the tRNA acceptor end prevents its interaction with the AT hairpin bulge leading to the termination of transcription elongation (20). For the *glyQS* riboswitch, addition of a charged tRNA mimic with an extra base at the 3'-end or by including glycyl-tRNA synthetase in the *in vitro* transcription assays was shown to decrease anti-termination (18,22). *In vivo* studies also confirmed that charged tRNA decreases anti-termination efficiency (10,52). However, the effect of charging on tRNA binding to the T-box riboswitch as a critical event during riboswitching has not been probed. To this end, we measured the binding kinetics of charged $tRNA^{Gly}$ to the mutant full T-box riboswitch. Specifically, we have charged $tRNA^{Gly}$ using *E. coli* S100 extract. The efficiency of charging was at least ~80 % as quantified from acid urea gel (**Figure 4.12A**) and reached saturation within <10 min of incubation at 37 °C. We used this aminoacylated tRNA for binding experiments with the mutant full T-box riboswitch. The acidic pH of 6.1for the buffer used for our binding experiments and the fact that Gly-$tRNA^{Gly}$ has a longer half-life as compared to other aa-tRNAs ensures that the tRNA should be fairly stable (53).

Visual inspection of individual traces for charged $tRNA^{Gly}$ displayed binding events with very short dwell times (<1–10 s) in the bound state (t_{bound}) (**Figure 4.12B**). A cumulative distribution of $t_{unbound}$ fit well with a double-exponential function with slower and faster lifetimes of 31.3 s (64 %) and 7.3 s (36 %) giving a weighted average k_{on} value of ~2 x 10^6 M^{-1} s^{-1} at ~20 nM charged tRNA. Furthermore, the cumulative distribution of t_{bound} also fit well with a double-exponential function with slower and faster lifetimes of 7.4 s (37 %) and 0.35 s (73 %), respectively, giving an average k_{off} value of ~ 0.45 s^{-1} (average lifetime of ~2.2 s). Accounting for the presence of ~15 % uncharged tRNA (due to <100 % charging efficiency), we corrected the rates obtained to give the k_{on} and k_{off} values of ~2.2 x 10^6 M^{-1} s^{-1} and 0.51 s^{-1}, respectively for the charged $tRNA^{Gly}$. Therefore, the rate of tRNA dissociation from the full mutant T-box riboswitch increases by ~6-fold due to the presence of amino acid on the tRNA 3'-end. Our kinetic data thus clearly demonstrate that charged tRNA does not remain bound to the T-box as long as the uncharged tRNA.

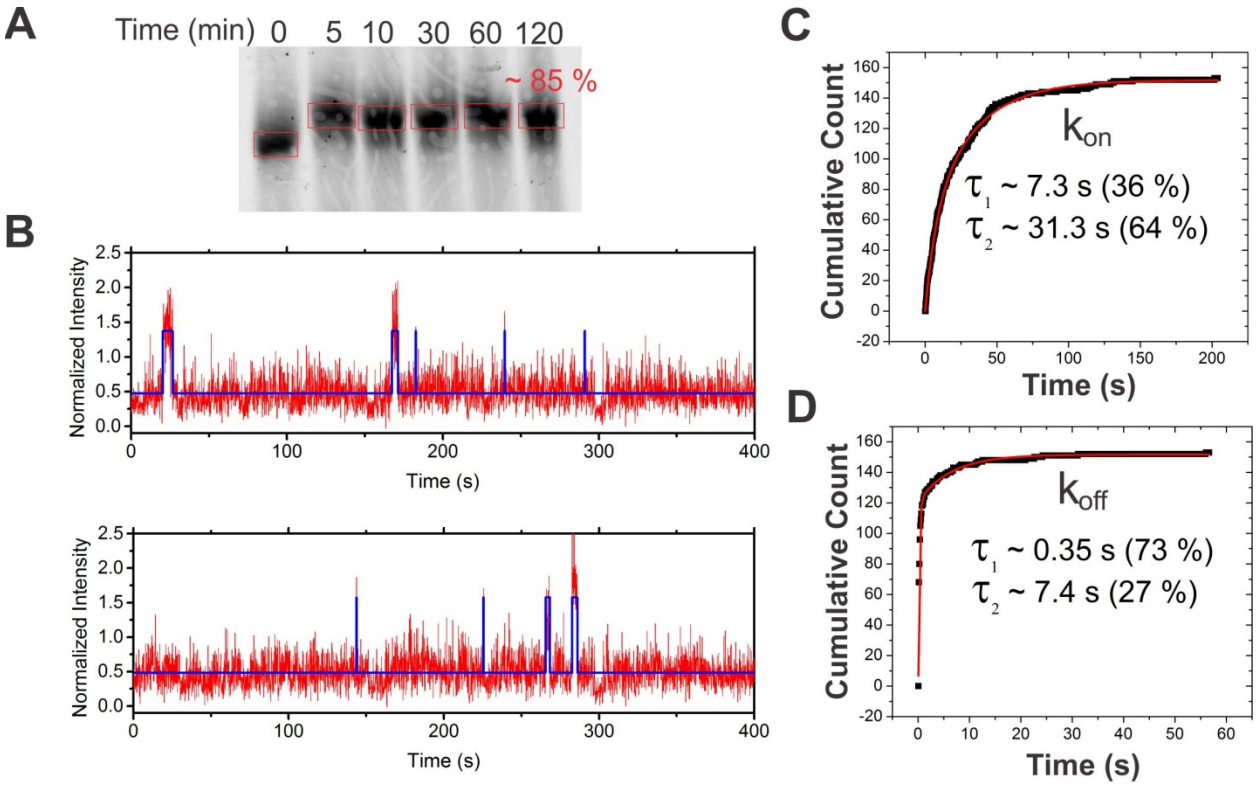

Figure 4.12 Effect of aminoacylation on the tRNA binding kinetics

(A) Acid urea gel showing time course of the tRNA aminoacylation reaction stained with SYBR Gold (Life Technologies). (B) Example binding traces for the charged $tRNA^{Gly}$ fit to two state HMM showing very short dwell times in the bound state. Cumulative dwell time distributions of charged tRNA binding to the full mutant T-box riboswitch for obtaining (C) k_{on} and (D) k_{off}. Individual life times of the exponential distributions are also shown.

4.4 Discussion

Despite its discovery almost 20 years ago (8,13), details of the T-box riboswitch molecular mechanism are still not completely understood. A number of biochemical studies have established the roles of various conserved sequence and structural motifs of the T-box riboswitch in tRNA mediated transcription anti-termination (8,15,18,20,22,25). However, biophysical studies on the structure and kinetic behavior of the riboswitch are still rare. Initial studies used NMR to solve the structures of different motifs in the RNA that provided atomic details of the architectural features of the T-box riboswitch (27-29). The recently solved crystal structure of the *glyQS* stem-I revealed the presence of an interdigitated T-loop motif at its distal end that is known to bind the tRNA elbow (25) (**Figure 4.3A**). Additional major insight into the interaction between the tRNA and T-box arose with the recently solved crystal structures of the *glyQ* stem-I:tRNA complex (26,31) (**Figure 4.3B,C**). However, knowledge about the conformation of the full T- box riboswitch including the downstream expression platform is still lacking. Perhaps even more importantly, the kinetics of tRNA binding to the T-box riboswitch and the role of different motifs in tRNA recognition has not been directly investigated. In this work, we have studied the conformation and tRNA binding kinetics of the *glyQS* T-box riboswitch. Our work used different T-box RNA constructs to delineate the function of the newly discovered double T-loop motif (25) and to probe the role of the AT hairpin in tRNA binding kinetics. Using smFRET, we have estimated distances between the different regions of the RNA constructs and monitored their changes in real-time. Our distance estimates on the T-box riboswitch will be useful in generating a global model for the full T-box riboswitch:tRNA complex using SAXS.

Our smFRET data show that the histograms for distance measurements between the bound tRNA and the base of AT hairpin for the mini and full (WT and mutant) riboswitches are similar with only subtle differences (**Figure 4.5**). This shows that the conformation of the tRNA bound to all three constructs is possibly the same. There are two plausible explanations for this observation. First, it posits that the interaction between the tRNA elbow and the T-loop motif as seen in the crystal structure of the stem-I:tRNA complex is absent in the final

complex involving the full T-box riboswitch, where the tRNA 3'-CCA end makes contact with its complementary sequence in the AT bulge. If this model is correct, then the crystal structure of the stem-I:tRNA complex (26,31) could be an intermediate conformation en route to the formation of a final complex involving the downstream expression platform containing the AT hairpin. We have preliminary evidence from SAXS studies (unpublished, data not shown) of the full WT T-box riboswitch:tRNA complexes supporting this model. The other plausible explanation, that is supported by our smFRET data is that the tRNA retains the same orientation as seen in the crystal structure even when bound to the mini T-box riboswitch, by virtue of its interaction with the specifier sequence and the extensive non-specific helix-helix contact between the tRNA anti-codon stem and the stem-I backbone (**Figure 4.3C**).

Our work also shows that the conformation of the single-stranded linker region including stem-III, connecting stem-I and the AT hairpin is sensitive to Mg^{2+} (**Figure 4.6**). Under low Mg^{2+} (1 mM), a FRET value of ~0.46 corresponding to ~56 Å shows that this ~19-nt long (assuming stem-III width as a 2-nt sequence) is in an extended single-stranded conformation that undergoes compaction at higher Mg^{2+} concentrations. This conclusion agrees well with previous structure probing studies which showed lack of secondary structure in a majority of the linker region residues in the absence of tRNA (19). However, we did not observe any major FRET changes upon addition of excess $tRNA^{Gly}$, suggesting that the conformation is not affected by tRNA binding. This shows that the full T-box riboswitch is in a largely pre-organized conformation prior to tRNA binding. This result is in contrast to a previous study that proposed a structural transition in this region upon tRNA binding, based on strcture probing of RNA (19). The observation of a pre-folded conformation in the full T-box riboswitch may not be of functional significance *in vivo* for transcriptionally acting T-box riboswitches like the *glyQS* riboswitch, since tRNA binding may occur before the AT hairpin is fully transcribed (17). Nevertheless, since T-box riboswitches that can modulate translation initiation have been identified (11,12), it is possible that these RNAs can act as thermodynamic 'switches' in fully transcribed mRNA that can continually sense tRNA levels during its lifetime. In fact, the S-box (or S_{MK} box) system, which is a translationally acting

riboswitch, was shown to act as a true switch capable of making multiple translation ON/OFF regulatory decisions during the lifetime of its mRNA transcript, in response to changes in concentration of its ligand, SAM (54). In this context, studying the conformation and tRNA binding kinetics of the full T-box riboswitch assumes significance.

We have also demonstrated that stem-I in the full T-box riboswitch adopts two conformations corresponding to extended and kinked (as seen in the crystal structure (26,31)) structures and Mg^{2+} affects the equilibrium between these two states (**Figure 4.7**). Our FRET measurements suggest that this could be due to the k-turn motif which becomes close to fully kinked at high (50 mM) Mg^{2+}. This also might explain why very high Mg^{2+} is required for efficient $tRNA^{Gly}$ mediated anti-termination in *in vitro* assays on the *glyQS* riboswitch (19). The k-turn motif functions to position the downstream AT hairpin close to the tRNA acceptor end so that the key interaction with the conserved bulge sequence can be established to result in transcriptional anti-termination (26).

Riboswitches that work at the level of transcription termination are typically kinetically controlled (55,56). This means that the rate of ligand binding is a very important factor that decides the outcome of gene regulation. In case of the T-box riboswitch, subsequent to the formation of stem-I, tRNA binding mediated stabilization of the AT hairpin must occur before the RNAP moves past the 3' end of the terminator hairpin (**Figure 4.1**). Therefore, tRNA binding has to occur within this time window, the duration of which depends on the rate of transcription by RNAP and the presence of any pause sites. Investigating the kinetics of tRNA binding is thus important for a complete understanding of the T-box riboswitch function. Our kinetic studies measured a k_{on} value of ~0.7×10^6 M^{-1} s^{-1} for both the stem-I only and the full WT T-box riboswitches (**Figure 4.10A, C**). This suggests that the regions 3' to stem-I including the linker region and the AT hairpin have no effect on k_{on} and therefore may not participate in the initial tRNA recognition event. In contrast, a 6-fold faster k_{off} for stem-I only construct shows that the AT hairpin is required to 'anchor' the cognate tRNA in place to effect transcriptional read-through (**Figure 4.10B, D**). The transient binding of tRNA by stem-I should be an important feature for proper

functioning of the T-box riboswitches since it ensures that stochastic binding events of charged tRNA to stem-I when their concentrations are low do not last long and affect the outcome of the regulatory switch during the critical decision making point. Our kinetic data on the mutant and mini T-box riboswitches show that destabilizing or deleting the double T-loop motif did not change the k_{on} significantly but only accelerated the k_{off} value by ~2-fold (**Figure 4.10E-H**).This suggests that the double T-loop motif functions to stabilize the bound tRNA but does not contribute to the tRNA association rate. In certain bacteria, T-box riboswitches have been identified that presumably work at the level of translation initiation (12). Interestingly, the stem-I domain in these T-box systems is shorter and lacks the sequences that can form the double T-loop motif. Therefore, these T-box riboswitches will probably have a faster k_{off} as compared to their transcriptional counterparts. A faster k_{off} is a general feature of translational riboswitches that work under thermodynamic control (56).

The value of k_{on} obtained for the T-box riboswitch is in line with the ligand binding rates for metabolite binding riboswitches, which have k_{on} in the range of ~10^4 – 10^5 M^{-1} s^{-1} (56).The k_{off} value for stem-I is an order of magnitude faster whereas the k_{off} value the full T-box riboswitch is similar to those of ligand dissociation rates of riboswitches (10^{-2} – 10^{-3} s^{-1}) (56). The faster k_{off} for stem-I will enable it to achieve equilibrium with the intracellular tRNA concentration before the genetic decision point. Our tRNA binding affinities obtained from the single molecule binding experiments (K_d ~54 nM and ~290 nM for the full WT riboswitch and stem-I, respectively) match relatively well with previous tRNA binding measurements done on various *glyQS* T-box riboswitch truncation variants (19). The full WT riboswitch has an ~6-fold lower K_d than the stem-I only construct, which agrees well with a ~5-6 fold efficient binding of tRNA by the full T-box riboswitch as compared to the stem-I only construct, measured in a previous study using filter retention assays (19). It is interesting to note that the tRNA binding kinetics to the different T-box variants are biphasic. A possible explanation for such biphasic kinetics of binding and dissociation are the presence of two conformational sub-states of the T-box riboswitch with different tRNA binding properties. While we see some correlation in the FRET populations

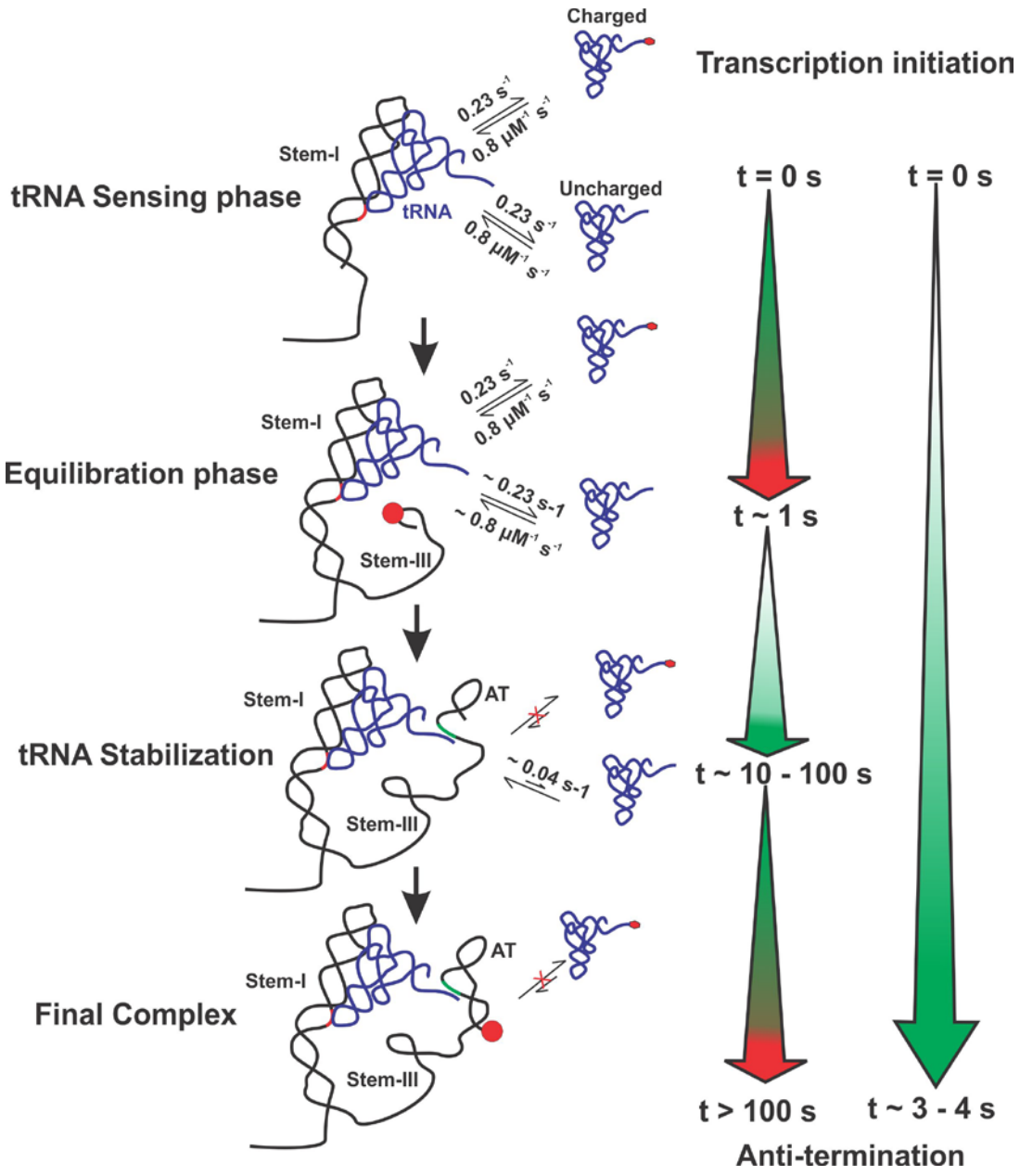

Figure 4.13 Kinetic model of the T-box riboswitch mechanism

Kinetic model showing the different stages of tRNA binding during transcription by the *glyQS* T-box riboswitch. Rates estimated from our single molecule kinetic measurements are shown. Major RNAP pause sites are indicated as red circles and charged tRNA is shown with a small red ellipse on its end. A reference time of t = 0 is when the tRNA binding capable portion of stem-I is transcribed. The broken and continuous arrows represent the thermodynamic and kinetic control pathways of the T-box riboswitch action, defined as having a commitment time window >> $1/k_{off}$ or ~ $1/k_{off}$, respectively (56).

and the biphasic tRNA binding kinetics for certain T-box variants, in general, we do not observe a strong correlation between the two. Further experiments will need to be carried out to correlate the conformational distributions of the T-box riboswitch variants to their tRNA binding kinetics. In addition, kinetic experiments can be done under different Mg^{2+} concentrations (> 10 mM Mg^{2+}) to probe if the amplitudes of the slow and fast kinetic phases change similar to the change in the distribution of the FRET states (**Figure 4.7 D-G**), which would establish the dependence of tRNA binding kinetics on the conformational state of the T-box riboswitch.

Our work also provides direct evidence that aminoacylation of tRNA leads to increased dissociation rate from the T-box riboswitch (**Figure 4.12**). The ~2.5 fold higher k_{on} value for charged tRNA compared to uncharged tRNA may be due to inaccurate measurement of the tRNA concentrations under dilute conditions. Phenol-chloroform extraction was done to remove the proteins components after the charging reaction, which diluted the tRNA to ~0.3 – 0.4 µM. At this concentration, measuring the concentration of Cy5 using Nanodrop with absorbance < 0.1 is difficult. This needs to be further verified. Due to the presence of amino acid on the 3'-end, the charged tRNA probably makes interaction mainly with the stem-I domain. Measuring the kinetics of charged tRNA binding to stem-I only will further support this. In this respect, the interaction of charged tRNA with the mutant T-box riboswitch is probably equivalent to the interaction with uncharged tRNA with a mutant stem-I only construct. Interestingly, our observation of ~6-fold faster k_{off} for the charged tRNA (as compared to the uncharged tRNA) binding to the full mutant T-box riboswitch matches well with the ~6-fold difference in k_{off} values observed for uncharged tRNA binding to the stem-I only and full WT T-box riboswitches. This suggests that charging the tRNA has a similar effect as deleting/mutating the AT hairpin in terms of dissociation rate from the T-box riboswitch. Studying the binding kinetics of tRNA to a mutant stem-I only construct may support this result further and will also help in quantifying the exact effect of the G81A mutation on tRNA binding.

Figure 4.13 shows a kinetic model of the *glyQS* T-box riboswitch mechanism incorporating temporal details obtained from our tRNA binding kinetic studies and also fromprevious bulk kinetic studies (17,19). The 'sensing' of tRNA by the T-box riboswitch can start as soon as the portion of stem-I up to the 3' end of the specifier loop is transcribed, since the downstream k-turn motif is shown to be not required for tRNA binding (25,26). At this stage, the stem-I can make multiple binding interactions with both the charged and uncharged $tRNA^{Gly}$. The Glycyl-tRNA synthetases have K_M values in the low µM range, which is also close to the intracellular concentration of $tRNA^{Gly}$. At these concentrations, the rate of tRNA binding by stem-I is ~ 1-3 s^{-1} using our k_{on} value of ~0.8 μM^{-1} s^{-1}. Three major pause sites were detected by previous *in vitro* kinetic studies on the ***glyQS*** riboswitch (17). One of them included a long pause site in the loop of stem-III with half-life of few minutes under low NTP concentrations. Pausing of RNAP at this point may allow sufficient time for stem-I to equilibrate with the relative concentrations of charged and uncharged tRNA. In the next phase, the formation of a stable complex occurs when the 5' portion of the AT hairpin is transcribed (19) as sequences that form the 3' half of the AT hairpin are not required for binding. At this stage, the charged tRNA probably cannot compete with the uncharged tRNA, as suggested by previous *in vitro* kinetic assays (17). Another pause site identified near the 3' end of the AT hairpin (the decision making point) allows more time for the uncharged tRNA to form a stable complex required to prevent transcription termination. An approximate timeline is shown assuming a transcription rate of ~25 nt/s (55,57) by bacterial RNAP and accounting for the effect of pause sites (17). This predicts that stem-I has sufficient (~100 s) time for stable tRNA binding in case of RNAP pausing. However, it is not certain if these pause sites are functional under intracellular conditions where the NTP concentration is much higher and in the presence of other cellular factors. Therefore, in the absence of RNAP pausing, only a short ~3-4 s (for transcribing ~80 nt up to the 3'-end of AT hairpin) time window is available for the T-box riboswitch to bind tRNA before the RNAP reaches the decision point. From our kinetic data, this time is roughly similar to the residence time ($1/k_{off}$) of tRNA bound to stem-I of ~4 s (**Figure 4.10B**). Therefore, this suggests that the T-box riboswitch functions through the thermodynamic control mode of genetic regulation,

which means that the riboswitch achieves equilibrium with the tRNA before the RNAP reaches the decision point (50). However, kinetic control is also possible if the rate of transcription and the concentration of tRNA is higher *in vivo*. This is similar to the action of some transcriptionally acting metabolite binding riboswitches (55,56). Our work thus provides the first quantitative analysis of the kinetic pathway for T-box riboswitch function. In addition, using powerful smFRET experiments, we have studied the conformation of various T-box riboswitch variants and directly probed their heterogeneous tRNA binding kinetics that suggest a complex mode of tRNA recognition by the T-box riboswitch RNA.

4.5 Acknowledgements

I would like to thank our collaborators Prof. Edward P. Nikonowicz and Dr. Malgorzata Michnicka from Rice University for providing the materials used in this study, Paul Lund for providing the *E. coli* S100 extracts used for tRNA aminoacylation and Dr. Mario Blanco for Matlab scripts used for analyzing the binding kinetics.

4.6 References

1. Henkin, T.M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes & development*, 22, 3383-3390.
2. Winkler, W.C. and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annual review of microbiology*, 59, 487-517.
3. Breaker, R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology*, 4.
4. Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Molecular cell*, 43, 867-879.
5. Barrick, J.E. and Breaker, R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome biology*, 8, R239.
6. Garst, A.D., Edwards, A.L. and Batey, R.T. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol*, 3.
7. Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, 152, 17-24.
8. Grundy, F.J. and Henkin, T.M. (1993) tRNA as a positive regulator of transcription antitermination in B. subtilis. *Cell*, 74, 475-482.
9. Henkin, T.M. (1994) tRNA-directed transcription antitermination. *Molecular microbiology*, 13, 381-387.
10. Green, N.J., Grundy, F.J. and Henkin, T.M. (2010) The T box mechanism: tRNA as a regulatory molecule. *FEBS letters*, 584, 318-324.

11. Grundy, F.J. and Henkin, T.M. (2003) The T box and S box transcription termination control systems. *Frontiers in bioscience : a journal and virtual library*, 8, d20-31.
12. Vitreschak, A.G., Mironov, A.A., Lyubetsky, V.A. and Gelfand, M.S. (2008) Comparative genomic analysis of T-box regulatory systems in bacteria. *Rna*, 14, 717-735.
13. Henkin, T.M., Glass, B.L. and Grundy, F.J. (1992) Analysis of the Bacillus subtilis tyrS gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. *Journal of bacteriology*, 174, 1299-1306.
14. Gutierrez-Preciado, A., Henkin, T.M., Grundy, F.J., Yanofsky, C. and Merino, E. (2009) Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiology and molecular biology reviews : MMBR*, 73, 36-61.
15. Grundy, F.J., Hodil, S.E., Rollins, S.M. and Henkin, T.M. (1997) Specificity of tRNA-mRNA interactions in Bacillus subtilis tyrS antitermination. *Journal of bacteriology*, 179, 2587-2594.
16. Gerdeman, M.S., Henkin, T.M. and Hines, J.V. (2002) In vitro structure-function studies of the Bacillus subtilis tyrS mRNA antiterminator: evidence for factor-independent tRNA acceptor stem binding specificity. *Nucleic acids research*, 30, 1065-1072.
17. Grundy, F.J. and Henkin, T.M. (2004) Kinetic analysis of tRNA-directed transcription antitermination of the Bacillus subtilis *glyQS* gene in vitro. *Journal of bacteriology*, 186, 5392-5399.
18. Grundy, F.J., Yousef, M.R. and Henkin, T.M. (2005) Monitoring uncharged tRNA during transcription of the Bacillus subtilis *glyQS* gene. *Journal of molecular biology*, 346, 73-81.
19. Yousef, M.R., Grundy, F.J. and Henkin, T.M. (2005) Structural transitions induced by the interaction between tRNA(Gly) and the Bacillus subtilis *glyQS* T box leader RNA. *Journal of molecular biology*, 349, 273-287.
20. Grundy, F.J., Rollins, S.M. and Henkin, T.M. (1994) Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the Bacillus subtilis tyrS gene: a new role for the discriminator base. *Journal of bacteriology*, 176, 4518-4526.
21. Grundy, F.J., Collins, J.A., Rollins, S.M. and Henkin, T.M. (2000) tRNA determinants for transcription antitermination of the Bacillus subtilis tyrS gene. *RNA*, 6, 1131-1141.
22. Yousef, M.R., Grundy, F.J. and Henkin, T.M. (2003) tRNA requirements for *glyQS* antitermination: a new twist on tRNA. *RNA*, 9, 1148-1156.
23. Grundy, F.J., Winkler, W.C. and Henkin, T.M. (2002) tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 11121-11126.
24. Putzer, H., Condon, C., Brechemier-Baey, D., Brito, R. and Grunberg-Manago, M. (2002) Transfer RNA-mediated antitermination in vitro. *Nucleic acids research*, 30, 3026-3033.

25. Grigg, J.C., Chen, Y., Grundy, F.J., Henkin, T.M., Pollack, L. and Ke, A. (2013) T box RNA decodes both the information content and geometry of tRNA to affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 7240-7245.
26. Zhang, J. and Ferre-D'Amare, A.R. (2013) Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature*, 500, 363-366.
27. Gerdeman, M.S., Henkin, T.M. and Hines, J.V. (2003) Solution structure of the Bacillus subtilis T-box antiterminator RNA: seven nucleotide bulge characterized by stacking and flexibility. *Journal of molecular biology*, 326, 189-201.
28. Chang, A.T. and Nikonowicz, E.P. (2013) Solution NMR determination of hydrogen bonding and base pairing between the *glyQS* T box riboswitch Specifier domain and the anticodon loop of tRNA(Gly). *FEBS letters*, 587, 3495-3499.
29. Wang, J. and Nikonowicz, E.P. (2011) Solution structure of the K-turn and Specifier Loop domains from the Bacillus subtilis tyrS T-box leader RNA. *Journal of molecular biology*, 408, 99-117.
30. Lehmann, J., Jossinet, F. and Gautheret, D. (2013) A universal RNA structural motif docking the elbow of tRNA in the ribosome, RNAse P and T-box leaders. *Nucleic acids research*, 41, 5494-5502.
31. Grigg, J.C. and Ke, A. (2013) Structural determinants for geometry and information decoding of tRNA by T box leader RNA. *Structure*, 21, 2025-2032.
32. Joo, C. and Ha, T. (2012) Single-molecule FRET with total internal reflection microscopy. *Cold Spring Harbor protocols*, 2012.
33. Joo, C. and Ha, T. (2012) Prism-type total internal reflection microscopy for single-molecule FRET. *Cold Spring Harbor protocols*, 2012.
34. Collett, J.R., Cho, E.J. and Ellington, A.D. (2005) Production and processing of aptamer microarrays. *Methods*, 37, 4-15.
35. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single-molecule FRET. *Nat Methods*, 5, 507-516.
36. Suddala, K.C., Rinaldi, A.J., Feng, J., Mustoe, A.M., Eichhorn, C.D., Liberman, J.A., Wedekind, J.E., Al-Hashimi, H.M., Brooks, C.L., 3rd and Walter, N.G. (2013) Single transcriptional and translational $preQ_1$ riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure. *Nucleic Acids Res*, 41, 10462-10475.
37. Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science*, 296, 1473-1476.
38. Abelson, J., Blanco, M., Ditzler, M.A., Fuller, F., Aravamudhan, P., Wood, M., Villa, T., Ryan, D.E., Pleiss, J.A., Maeder, C. *et al.* (2010) Conformational dynamics of single pre-mRNA molecules during in vitro splicing. *Nat Struct Mol Biol*, 17, 504-512.
39. Blanco, M. and Walter, N.G. (2010) Analysis of complex single-molecule FRET time trajectories. *Methods in enzymology*, 472, 153-178.
40. Qin, F. (2007) Principles of single-channel kinetic analysis. *Methods in molecular biology*, 403, 253-286.

41. Stark, M.R., Pleiss, J.A., Deras, M., Scaringe, S.A. and Rader, S.D. (2006) An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA*, 12, 2014-2019.
42. Kohrer, C. and Rajbhandary, U.L. (2008) The many applications of acid urea polyacrylamide gel electrophoresis to studies of tRNAs and aminoacyl-tRNA synthetases. *Methods*, 44, 129-138.
43. Pereira, M.J., Behera, V. and Walter, N.G. (2010) Nondenaturing purification of co-transcriptionally folded RNA avoids common folding heterogeneity. *PLoS One*, 5, e12953.
44. Ha, T., Rasnik, I., Cheng, W., Babcock, H.P., Gauss, G.H., Lohman, T.M. and Chu, S. (2002) Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature*, 419, 638-641.
45. Daldrop, P. and Lilley, D.M. (2013) The plasticity of a structural motif in RNA: structural polymorphism of a kink turn as a function of its environment. *RNA*, 19, 357-364.
46. Goody, T.A., Melcher, S.E., Norman, D.G. and Lilley, D.M. (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA*, 10, 254-264.
47. Schroeder, K.T., McPhee, S.A., Ouellet, J. and Lilley, D.M. (2010) A structural database for k-turn motifs in RNA. *RNA*, 16, 1463-1468.
48. Grigg, J.C. and Ke, A. (2013) Sequence, structure, and stacking: Specifics of tRNA anchoring to the T box riboswitch. *RNA biology*, 10, 1761-1764.
49. Walter, N.G. (2003) Probing RNA structural dynamics and function by fluorescence resonance energy transfer (FRET). *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]*, Chapter 11, Unit 11 10.
50. Garst, A.D. and Batey, R.T. (2009) A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta*, 1789, 584-591.
51. Elenko, M.P., Szostak, J.W. and van Oijen, A.M. (2009) Single-molecule imaging of an in vitro-evolved RNA aptamer reveals homogeneous ligand binding kinetics. *J Am Chem Soc*, 131, 9866-9867.
52. Henkin, T.M. (2009) Analysis of tRNA-directed transcription antitermination in the T box system in vivo. *Methods in molecular biology*, 540, 281-290.
53. Hentzen, D., Mandel, P. and Garel, J.P. (1972) Relation between aminoacyl-tRNA stability and the fixed amino acid. *Biochim Biophys Acta*, 281, 228-232.
54. Smith, A.M., Fuchs, R.T., Grundy, F.J. and Henkin, T.M. (2010) The SAM-responsive S(MK) box is a reversible riboswitch. *Molecular microbiology*, 78, 1393-1402.
55. Wickiser, J.K., Winkler, W.C., Breaker, R.R. and Crothers, D.M. (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular cell*, 18, 49-60.
56. Haller, A., Souliere, M.F. and Micura, R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc Chem Res*, 44, 1339-1348.
57. Rhodes, G. and Chamberlin, M.J. (1974) Ribonucleic acid chain elongation by Escherichia coli ribonucleic acid polymerase. I. Isolation of ternary complexes and the kinetics of elongation. *J Biol Chem*, 249, 6675-6683.

CHAPTER 5

Conclusions and Future Directions

Non-coding RNAs (ncRNAs) are emerging as the major players of gene regulation in all forms of life (1,2). Even in simple unicellular organisms such as bacteria, a vast number of different classes of ncRNAs have been discovered that perform complex cellular functions including an RNA-based adaptive immune system and regulation of gene expression at multiple levels (2,3). Among them, a class of ncRNAs known as riboswitches regulates gene expression by directly sensing concentration changes of intracellular metabolites (4-6). Riboswitches are structured motifs in mRNAs that bind diverse classes of ligands and control expression of proteins generally involved in the metabolism or transport of the ligand. Ligand binding by the aptamer domain of a riboswitch allosterically causes structural changes in the downstream sequence (expression platform) that modulate gene expression (6). Regulation occurs largely at the level of transcription elongation or translation initiation; however, a number of other processes such as mRNA (alternative) splicing, decay, self-cleavage, recruitment of protein cofactors such as Rho and *trans* regulation of distal mRNAs have been discovered (7). Since their discovery in 2002, a large number of riboswitches have been identified that carry out vital functions mainly in bacteria. One class of riboswitches is also identified in eukaryotes and archaea, which suggests that riboswitches might have had major cellular roles during the early phase of evolution when life was predominantly RNA based (4) and later became scarcer.

In the past decade, our knowledge of riboswitch function was greatly enhanced by many biochemical and structural studies that showed how small molecule binding causes changes in gene expression (5). In addition, the crystal structures of ligand-bound riboswitches highlighted the great potential of RNA in recognizing chemically diverse small

molecules as well as complex macromolecules (5). However, due to the difficulty in studying structures of ligand-free riboswitches, our understanding of the folding process of riboswitches and how the ligand alters downstream secondary structures to affect gene expression is still far from complete (8). Nonetheless, a number of ensemble studies using NMR spectroscopy, small angle x-ray scattering (SAXS), and chemical structure probing have characterized the ligand free conformations of a few riboswitches. However, riboswitch RNAs are known to populate multiple states including functionally important conformations that are only transiently sampled in the absence of ligand (9). Therefore, ensemble biophysical methods only provide an average state that may not correspond to a true conformation. Furthermore, detecting transient, lowly populated conformations is extremely difficult using ensemble methods. In contrast, single molecule techniques are ideally suited for investigating biomolecular conformation and dynamics and are increasingly used to elucidate the ligand dependent folding processes of different riboswitches (9-11). In this dissertation, we have applied single molecule FRET using prism-based TIRF microscopy to study the conformation, dynamics and ligand binding properties of two very distinct RNAs – the $preQ_1$ and the T-box riboswitches that represent the small and big of the riboswitch world.

5.1 Conformation and folding of the $preQ_1$ riboswitch

In chapter 2, using smFRET microscopy, computational simulations and NMR spectroscopy, we have investigated the structure and dynamics of the class-I $preQ_1$ riboswitch that carries the smallest riboswitch aptamer discovered to date (12). The $preQ_1$ riboswitches from *B. subtilis* (*Bsu*) and *T. tengcongensis* (*Tte*) act through ligand mediated transcription termination and translation inhibition, respectively, and fold into similar compact H-type pseudoknot structures when bound to ligand (13,14). However, ensemble studies on the ligand-free riboswitches suggested major differences in their conformations. The *Bsu* riboswitch was proposed to exist as a largely unfolded hairpin with a non-interacting single-stranded 3'-tail, while the *Tte* riboswitch was suggested to adopt a loose pseudoknot conformation (14,15). These results were surprising, however, owing to extensive similarities

between the two riboswitches in sequence and structure, and considering the fact that the *Bsu* riboswitch is expected to be more folded given its function during transcription, which requires fast ligand mediated folding (16). Our smFRET data then clearly showed that, in the presence of Mg^{2+}, both ligand-free riboswitches exist in a major pre-folded conformational ensemble, but also sample a minor pseudoknot-like conformation. We also showed that contrary to previous studies, both riboswitches showed only subtle differences in the distribution of their FRET states and their conformational dynamics (14,15). Coarse-grained TOPRNA simulations of the riboswitches showed that the pre-folded state corresponds to conformations where the 3'-tail transiently interacts with the stem-loop P1-L1. Both riboswitches showed very different behavior of their 'pre-folded' states under increasing $preQ_1$ concentration. This result suggested that both RNAs fold through distinct ligand-mediated folding mechanisms. This observation was supported by our Gō-model folding simulations, which showed that the ligand binds to an open conformation of the *Tte* riboswitch whereas the *Bsu* riboswitch folding is concomitant with ligand binding (17). These ligand mediated folding pathways are similar to the induced-fit and conformational selection mechanisms of biomolecular recognition (18-20). Furthermore, we demonstrated that a single mutation in the A-rich 3'-tail away from the binding site renders the pre-folded state less compact and affects the ligand binding affinity, mainly for the *Bsu* riboswitch. Taken together, these results support an emerging consensus that riboswitch aptamers that belong to the same class fine-tune their ligand binding affinities through subtle sequence changes away from the binding pocket to achieve optimum gene regulation depending on the intra- and extracellular conditions of bacteria (5,21).

Our studies on the *Bsu* and *Tte* riboswitches raised a few interesting questions that can be addressed in the future. For example, the reason for slower dynamics in the ligand-free *Tte* riboswitch (dwell times of ~1–2 s in the folded-like state, as seen in smFRET traces) compared to the *Bsu* riboswitch (dwell times in the folded state close to 33-ms) is not clear. One possible explanation could be the higher stability of P2 in the *Tte* compared to the *Bsu* riboswitch. However, this does not seem to be the case as both riboswitches have a 4-bp stem with the *Tte* riboswitch consisting of two, typically weaker, non-canonical base pairs (13,14).

A more plausible reason is that nucleotide A14, that was shown to occupy the binding pocket in the *Tte* riboswitch, acts as a surrogate ligand stabilizing the folded state even in the absence of the ligand $preQ_1$ (14). One way to test this hypothesis is to mutate the A14 residue to a U and investigate its dynamics using smFRET. It is interesting to note that the *Bsu* riboswitch also has an adenine (A16) in the corresponding position, similar to the *Tte* riboswitch (13,14). However, the longer L1-loop of the *Bsu* riboswitch, that was previously shown to be dynamic, may be responsible for decreased stability of the ligand-free folded-like conformation of the *Bsu* riboswitch (22). To this end, deleting the extra bases in the *Bsu* L1-loop that are not conserved and not detected in the crystal structure may stabilize P2 in the ligand-free *Bsu* riboswitch (13). Furthermore, stem P2 in the *Bsu* riboswitch crystal structure has a terminal non-canonical C8-A34 base pair that is missing in the NMR structure. A Ca^{2+} ion is located close to this C8-A34 base pair in the crystal structure (13,15). A previous NMR study showed that Ca^{2+} is required in solution for the formation of this base pair (22). The non-canonical C8-A34 base pair does not resemble a C-A^{+} Wobble (23). However, it is possible that its stability is pH dependent and therefore can be probed by monitoring conformational dynamics of the *Bsu* riboswitch using smFRET under a constant ligand concentration. In addition, the recently developed constant pH molecular dynamics simulations can also be utilized to probe the pH dependent properties of the C8-A34 base pair to study its effect on P2 and therefore pseudoknot stability (24,25). In this respect, it is tantalizing to speculate that the *Bsu* $preQ_1$ riboswitch may act as a dual riboswitch responding to changes in both intracellular pH and ligand concentration, as opposed to riboswitches that sense either pH or metabolite concentration only (7,26). However, it is also known that many secondary and tertiary interactions in RNAs are affected by pH changes due to the protonation of nucleobases (23,27,28). Therefore, it is possible that metabolite binding riboswitch RNAs in general sense cellular pH to a certain extent. Biochemical assays probing the efficiency of gene regulation as a function of pH at a constant ligand concentration can be used to test this hypothesis. Due to its small size and simple structure the $preQ_1$ riboswitch will be a good model system to probe this hypothesis (12,13).

5.2 Ligand binding mechanism of the *Bsu* riboswitch

Despite the availability of high resolution structures of many riboswitch aptamers, there is a general lack of knowledge on their ligand-free conformations (8). Most of the relatively few available ligand-free riboswitch aptamer structures were solved using RNAs derived from thermophilic bacteria due to the ease of their crystallization at low temperature and, thus, are highly similar to their ligand-bound structures (5). Therefore, these ligand-free structures do not likely represent the true physiologically relevant conformations and do not provide an explanation for the ligand-dependent folding of the riboswitch. Characterizing the ligand-free riboswitch conformations using other ensemble methods has been difficult due to the existence of multiple states and the fact that these only provide an average conformation (8). Due to these reasons, how ligand binding drives riboswitch folding is generally not well understood (9). Ligand binding mechanisms are mainly classified into two limiting cases – conformational selection (folding first) and induced fit (binding first) mechanisms (18-20). These two mechanisms are broadly used to explain ligand binding coupled to folding of biomolecules.

High-resolution structures show that riboswitch aptamers typically form tight binding pockets that completely encapsulate the ligands (5). This begs the question of how ligands bind to such solvent inaccessible binding pockets. This problem is analogous to the case of enzymes with ‘lids’ over their active sites. In such cases, it was wrongly assumed that substrate binding should always proceed through an induced-fit like mechanism (29). However, recent studies have shown that both mechanisms are possible in specific cases and that the major mechanism can be identified by kinetic measurements or more accurately by calculating flux through both pathways (19) (30-32). As opposed to proteins, where molecular recognition coupled to folding is being actively investigated, few studies exist that investigate the molecular recognition mechanisms of riboswitches. Furthermore, the conformational selection mechanism was proposed for a few riboswitches merely from the observation of ligand-free folded-like conformations (9,33,34).

In chapter 3, we have investigated the ligand binding mechanism of the *Bsu* $preQ_1$ riboswitch by measuring the changes in kinetics of its conformational dynamics as a function of ligand concentration. The *Bsu* riboswitch encloses ~92 % of the solvent accessible surface area of its ligand $preQ_1$ (13). Our kinetic data suggest that, in the absence of Mg^{2+}, the riboswitch binds ligand majorly through an induced-fit mechanism. Surprisingly, in the presence of Mg^{2+}, the kinetics of conformational dynamics show a contrasting signature indicative of a dominant conformational selection pathway. This shift in the dominant binding mechanism can be explained by the ability of Mg^{2+} to promote pre-organized pseudoknot-like ligand-free conformations, as directly observed in our smFRET traces taken at the faster time resolution of 33ms. However, it should be noted that conformational selection and induced fit mechanisms are not mutually exclusive and a growing consensus is that both mechanisms are generally engaged in parallel during a biomolecular recognition event, just to different extents (19,35). In addition, a recent theoretical study suggests that conformational selection is characterized by a variety of kinetic signatures (31). The authors showed that while decreasing k_{obs} as a function of ligand concentration is a characteristic feature of conformational selection, an increasing k_{obs} can manifest for both mechanisms.

Given our results from smFRET kinetic measurements in chapter 3 and our Gō-model folding simulations in chapter 2, we propose that under *in vivo* conditions, due to the presence of Mg^{2+} and molecular crowding effects (36), the ligand-free *Bsu* riboswitch probably adopts even more compact conformations, resulting in conformational selection as the dominant ligand binding mechanism. This may also help in the fast ligand-binding coupled folding required for the kinetically acting *Bsu* $preQ_1$ riboswitch (37).

A direct way of testing the binding mechanism is to label the ligand with a fluorophore and observe its binding to the riboswitch labeled with a FRET pair (30). Using this method, one can clearly detect if the ligand binds to the pre-folded (0.7 FRET state, as shown in our previous study (10)) or folded (~0.9 FRET state) conformation. This requires the use of three-color FRET and was recently used to elucidate the ligand binding mechanism of the maltose-binding protein, along with smFRET kinetic measurements, similar to our work in chapter 3 (10,30). Additionally, the 3’ DY-547-labeled *Bsu* riboswitch can be used

along with a Cy5-labeled $preQ_1$ ligand to monitor the binding events using smFRET. By analyzing the FRET value/s obtained, a distance estimate and, therefore, an approximate location can be made for the initial ligand-RNA encounter, which in turn can suggest if the ligand binds to an open (with 3'-tail away from P1-L1 stem-loop) or closed RNA conformation (10). Alternatively, $preQ_1$ can be labeled with a black hole quencher dye such as BHQ-2 that quenches the fluorescence of both DY-547 and Cy5 to different extents (38). In this strategy, ligand binding will manifest as a decrease in the fluorescence intensities of both DY-547 and Cy5 and, therefore, by observing FRET before the binding event one can identify the conformation to which the ligand binds. This method can be thought of as a pseudo three-color FRET that does not require the sophisticated microscopy setup needed for performing an actual three-color FRET experiment (30). In this direction, labeling of $preQ_1$ with an n-hydroxysuccinimide (NHS) ester activated fluorophore is possible at its exocyclic amino that sticks out of the binding pocket and is solvent exposed (15). However, this conjugation may decrease the binding affinity of the fluorophore, although we expect it to still bind riboswitch. In the future, such experiments will help further characterize the binding mechanism of the *Bsu* $preQ_1$ riboswitch (10). In the future, although non-trivial, studies on riboswitches should probe the conformational dynamics and ligand-mediated folding during their transcription by RNA polymerase. Such an effort has already been made in studying the co-transcriptional ligand mediated folding of the adenine riboswitch using powerful single molecule force measurements (39). To this end, single molecule techniques will become increasingly valuable in understanding riboswitch folding during transcription and ligand mediated gene regulation.

5.3 Conformation and tRNA binding kinetics of the T-box riboswitch

In chapter 4, we have investigated the conformation and tRNA binding kinetics of the *glyQS* T-box riboswitch from *B. subtilis* using smFRET. The T-box riboswitch is one of the largest and most complex riboswitches that recognizes a macromolecular ligand, tRNA (40). Structural and biophysical studies on the T-box riboswitch have been rare until recently. Then, two crystal structures of the stem-I:tRNA complex were solved that showed details of

tRNA binding by the T-box riboswitch (41,42). However, details on the conformation of the full T-box riboswitch, the kinetics of tRNA binding and how aminoacylation destabilizes the T-box:tRNA complex were not known. In addition, the role of the newly discovered double T-loop motif on top of stem-I during tRNA binding was not clear (43). Using different T-box riboswitch designs labeled with fluorophore conjugated DNA oligonucleotides, we estimated distances between different regions of the full T-box riboswitch by smFRET. Our smFRET data on the mini T-box RNA showed that the double T-loop motif is not obligatory for tRNA binding. However, the motif stabilizes the bound tRNA and increases its lifetime in the complex by ~2-fold as demonstrated in our single molecule kinetic assays. We also showed that the conformation of the single-stranded linker region is Mg^{2+} dependent and is not affected by tRNA binding. This suggests that the full T-box is in a pre-organized conformation. In addition, we demonstrated that stem-I exists in two major FRET states, probably corresponding to extended and kinked conformations. High Mg^{2+} concentrations shift the distribution to a major (~90 %) kinked conformation. This offers an explanation for the high Mg^{2+} requirement for achieving efficient tRNA mediated transcription antitermination *in vitro* using the *glyQS* T-box riboswitch (44,45). This observed dependence on Mg^{2+} may be due to the presence of a k-turn motif and other motifs in stem-I that introduce a kink to precisely recognize tRNA geometry (13). Using single molecule kinetic assays, we have shown that stem-I by itself cannot stably bind tRNA, underscoring the importance of the anti-terminator hairpin for tRNA docking, needed for transcription anti-termination. Our kinetic assays using the mini and mutant T-box designs in comparison established that the double T-loop motif does not play a role in tRNA docking, but instead only increases the lifetime of the bound tRNA ~2-fold. It is interesting to note that certain putative translationally acting T-box riboswitches lack the sequences forming the double T-loop motif, again supporting our results that it is not functionally essential (46). Furthermore, we directly demonstrated that the presence of even a small amino acid such as glycine on the tRNA 3'-end increases its dissociation rate constant by ~6-fold relative to the uncharged tRNA. Interestingly, a ~6-fold difference was also observed in the dissociation rate constants of uncharged tRNA binding to stem-I alone and the mutant full-length T-box. These

observations show that charging of tRNA has the same effect as deleting the anti-terminator domain. This also suggests that the intervening single-stranded linker region does not contribute to tRNA binding affinity, as also shown in previous biochemical studies on the T-box riboswitch (45,47). Based on our kinetic data of tRNA binding and previous biochemical studies, we have proposed an improved quantitative kinetic model for the molecular mechanism of the T-box riboswitch (47,48).

Although the kinetics of charged tRNA binding were measured using the mutant full-length T-box, we expect that the relative difference (~6-fold) in the dissociation rate with respect to the uncharged tRNA will remain unchanged for the WT full T-box riboswitch. However, this needs to be confirmed and will be performed in the near future. In addition, it will be interesting to study the relative contributions of the specifier sequence and the double T-loop motif in tRNA binding affinity. This can be achieved by 'blocking' the double T-loop motif by using a fluorophore (Cy3) labeled DNA oligonucleotide that is complementary to the stem-I apical loop. FRET between the Cy3 and the tRNA-Cy5 can be used to monitor the tRNA binding kinetics using the stem-I only design to more sensitively dissect the function of the double T-loop motif. Furthermore, the loop residues of the anti-terminator hairpin are not conserved and therefore not involved in tRNA binding (40). A Cy5-labeled RNA oligonucleotide that forms the 3' half of the anti-terminator hairpin and a Cy3-labeled DNA oligonucleotide complementary to the stem-I apical loop can be used to doubly label the full T-box riboswitch. This design can be used to monitor the distance between the two regions and test if tRNA binding brings them closer or if they are already pre-organized, as suggested by our distance estimate between the bases of stem-I and the anti-terminator hairpin. In addition, a k-turn binding protein can be used to stabilize the kink in stem-I and test its ability to promote kinked RNA conformations under physiologically relevant Mg^{2+} concentration (13). This will also indicate whether k-turn binding proteins are required *in vivo* for proper functioning of T-box riboswitches. Furthermore, the tRNA binding kinetics in the presence of the k-turn binding protein can also be investigated.

In this dissertation, we have studied the conformation, dynamics and ligand binding of the $preQ_1$ and T-box riboswitches using smFRET. In chapter 2, we have probed the conformation and ligand-mediated folding of two structurally similar $preQ_1$ riboswitches from *B. subtilis* and *T. tencongensis* using smFRET and computational simulations. Our work suggests similar ligand-free conformational ensembles but contrasting folding pathways for the two $preQ_1$ riboswitches. In chapter 3, we further investigated the folding pathway of the *Bsu* riboswitch using kinetic analysis of $preQ_1$ dependent conformational dynamics. This also follows up on the results obtained from our Gō-model simulations and smFRET data that suggest a conformational selection-like mechanism of ligand binding by the *Bsu* riboswitch. Our single molecule kinetic analysis supports the results in chapter 2 and demonstrates a strong dependence of the folding pathway of the *Bsu* riboswitch on Mg^{2+} and ligand concentration. Furthermore, we have investigated the effect of other non-cognate ligands on the conformational dynamics of the *Bsu* riboswitch. In chapter 4, we applied the methods used in chapters 2 and 3 to studying the structure and ligand binding kinetics of the T-box riboswitch, which is a complex gene regulatory RNA element that senses the aminoacylation state of tRNA. Both the $preQ_1$ and T-box riboswitches are from *B. subtilis* and work at the level of transcription termination. However, the similarities between them end there. In terms of size, the $preQ_1$ riboswitch is the smallest natural metabolite binding RNA found to date while the T-box riboswitch is one of the largest. The T-box riboswitch recognizes a macromolecular ligand, tRNA, in contrast to the small-molecule binding $preQ_1$ riboswitch. Ligand binding leads to termination of transcription in the $preQ_1$ riboswitch whereas stable (uncharged) tRNA binding to the T-box riboswitch results in transcriptional anti-termination. In addition, our work using smFRET demonstrates a highly dynamic nature of the $preQ_1$ riboswitch while the T-box riboswitch appears to adopt a stable structure that is less dynamic with only slow transitions between its different conformations. The smFRET methods used in this dissertation can be applied to the study of ligand-free conformations and dynamics of different riboswitches and probe the effect of ligand on them. In addition, as we have shown for the *Bsu* riboswitch, the ligand dependent kinetics of conformational dynamics can be studied to elucidate the molecular recognition mechanisms of other classes

of riboswitches. This, we believe, will contribute to a deeper understanding of the metabolite binding bacterial riboswitches, with the ultimate goal to design antibiotics against them.

5.4 References

1. Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Hum Mol Genet*, **15 Spec No 1**, R17-29.
2. Repoila, F. and Darfeuille, F. (2009) Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biol Cell*, **101**, 117-131.
3. Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167-170.
4. Breaker, R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology*, **4**.
5. Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, **152**, 17-24.
6. Winkler, W.C. and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annual review of microbiology*, **59**, 487-517.
7. Peselis, A. and Serganov, A. (2014) Themes and variations in riboswitch structure and function. *Biochim Biophys Acta*.
8. Liberman, J.A. and Wedekind, J.E. (2012) Riboswitch structure in the ligand-free state. *Wiley Interdiscip Rev RNA*, **3**, 369-384.
9. Haller, A., Souliere, M.F. and Micura, R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Accounts of chemical research*, **44**, 1339-1348.
10. Suddala, K.C., Rinaldi, A.J., Feng, J., Mustoe, A.M., Eichhorn, C.D., Liberman, J.A., Wedekind, J.E., Al-Hashimi, H.M., Brooks, C.L., 3rd and Walter, N.G. (2013) Single transcriptional and translational $preQ_1$ riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure. *Nucleic acids research*, **41**, 10462-10475.
11. Roy, R., Hohng, S. and Ha, T. (2008) A practical guide to single-molecule FRET. *Nature methods*, **5**, 507-516.
12. 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. *et al.* (2007) A riboswitch selective for the queuosine precursor $preQ_1$ contains an unusually small aptamer domain. *Nat Struct Mol Biol*, **14**, 308-317.
13. Klein, D.J., Edwards, T.E. and Ferre-D'Amare, A.R. (2009) Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol*, **16**, 343-344.
14. Jenkins, J.L., Krucinska, J., McCarty, R.M., Bandarian, V. and Wedekind, J.E. (2011) Comparison of a $preQ_1$ riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *The Journal of biological chemistry*, **286**, 24626-24637.

15. Kang, M., Peterson, R. and Feigon, J. (2009) Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol Cell*, **33**, 784-790.
16. Garst, A.D. and Batey, R.T. (2009) A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta*, **1789**, 584-591.
17. Feng, J., Walter, N.G. and Brooks, C.L., III. (2011) Cooperative and directional folding of the preQ$_1$ riboswitch aptamer domain. *J Am Chem Soc*, **133**, 4196-4199.
18. Boehr, D.D., Nussinov, R. and Wright, P.E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol*, **5**, 789-796.
19. Hammes, G.G., Chang, Y.C. and Oas, T.G. (2009) Conformational selection or induced fit: a flux description of reaction mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 13737-13741.
20. Csermely, P., Palotai, R. and Nussinov, R. (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci*, **35**, 539-546.
21. Stoddard, C.D., Widmann, J., Trausch, J.J., Marcano-Velazquez, J.G., Knight, R. and Batey, R.T. (2013) Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch. *J Mol Biol*, **425**, 1596-1611.
22. Zhang, Q., Kang, M., Peterson, R.D. and Feigon, J. (2011) Comparison of solution and crystal structures of preQ1 riboswitch reveals calcium-induced changes in conformation and dynamics. *J Am Chem Soc*, **133**, 5190-5193.
23. Masquida, B. and Westhof, E. (2000) On the wobble GoU and related pairs. *Rna*, **6**, 9-15.
24. Goh, G.B., Knight, J.L. and Brooks, C.L., 3rd. (2013) pH-dependent dynamics of complex RNA macromolecules. *Journal of chemical theory and computation*, **9**, 935-943.
25. Khandogin, J. and Brooks, C.L., 3rd. (2005) Constant pH molecular dynamics with proton tautomerism. *Biophysical journal*, **89**, 141-157.
26. Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E. and Altuvia, S. (2009) A pH-responsive riboregulator. *Genes & development*, **23**, 2650-2662.
27. Houck-Loomis, B., Durney, M.A., Salguero, C., Shankar, N., Nagle, J.M., Goff, S.P. and D'Souza, V.M. (2011) An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature*, **480**, 561-564.
28. Wadkins, T.S., Shih, I., Perrotta, A.T. and Been, M.D. (2001) A pH-sensitive RNA tertiary interaction affects self-cleavage activity of the HDV ribozymes in the absence of added divalent metal ion. *J Mol Biol*, **305**, 1045-1055.
29. Sullivan, S.M. and Holyoak, T. (2008) Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13829-13834.
30. Kim, E., Lee, S., Jeon, A., Choi, J.M., Lee, H.S., Hohng, S. and Kim, H.S. (2013) A single-molecule dissection of ligand binding to a protein with intrinsic dynamics. *Nat Chem Biol*, **9**, 313-318.

31. Vogt, A.D. and Di Cera, E. (2012) Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. *Biochemistry*, **51**, 5894-5902.
32. Weikl, T.R. and von Deuster, C. (2009) Selected-fit versus induced-fit protein binding: Kinetic differences and mutational analysis. *Proteins: Structure, Function, and Bioinformatics*, **75**, 104-110.
33. Haller, A., Rieder, U., Aigner, M., Blanchard, S.C. and Micura, R. (2011) Conformational capture of the SAM-II riboswitch. *Nat Chem Biol*, **7**, 393-400.
34. Wilson, R.C., Smith, A.M., Fuchs, R.T., Kleckner, I.R., Henkin, T.M. and Foster, M.P. (2011) Tuning riboswitch regulation through conformational selection. *J Mol Biol*, **405**, 926-938.
35. Daniels, K.G., Tonthat, N.K., McClure, D.R., Chang, Y.C., Liu, X., Schumacher, M.A., Fierke, C.A., Schmidler, S.C. and Oas, T.G. (2014) Ligand concentration regulates the pathways of coupled protein folding and binding. *J Am Chem Soc*, **136**, 822-825.
36. Kilburn, D., Roh, J.H., Guo, L., Briber, R.M. and Woodson, S.A. (2010) Molecular crowding stabilizes folded RNA structure by the excluded volume effect. *J Am Chem Soc*, **132**, 8690-8696.
37. Eichhorn, C.D., Feng, J., Suddala, K.C., Walter, N.G., Brooks, C.L., III and Al-Hashimi, H.M. (2012) Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch. *Nucleic Acids Res*, **40**, 1345-1355.
38. Chen, J., Tsai, A., Petrov, A. and Puglisi, J.D. (2012) Nonfluorescent quenchers to correlate single-molecule conformational and compositional dynamics. *J Am Chem Soc*, **134**, 5734-5737.
39. Frieda, K.L. and Block, S.M. (2012) Direct observation of cotranscriptional folding in an adenine riboswitch. *Science*, **338**, 397-400.
40. Green, N.J., Grundy, F.J. and Henkin, T.M. (2010) The T box mechanism: tRNA as a regulatory molecule. *FEBS letters*, **584**, 318-324.
41. Zhang, J. and Ferre-D'Amare, A.R. (2013) Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature*, **500**, 363-366.
42. Grigg, J.C. and Ke, A. (2013) Structural determinants for geometry and information decoding of tRNA by T box leader RNA. *Structure*, **21**, 2025-2032.
43. Grigg, J.C., Chen, Y., Grundy, F.J., Henkin, T.M., Pollack, L. and Ke, A. (2013) T box RNA decodes both the information content and geometry of tRNA to affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 7240-7245.
44. Grundy, F.J., Yousef, M.R. and Henkin, T.M. (2005) Monitoring uncharged tRNA during transcription of the Bacillus subtilis *glyQS* gene. *Journal of molecular biology*, **346**, 73-81.
45. Yousef, M.R., Grundy, F.J. and Henkin, T.M. (2005) Structural transitions induced by the interaction between tRNA(Gly) and the Bacillus subtilis *glyQS* T box leader RNA. *Journal of molecular biology*, **349**, 273-287.

46. Vitreschak, A.G., Mironov, A.A., Lyubetsky, V.A. and Gelfand, M.S. (2008) Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA*, **14**, 717-735.
47. Grundy, F.J. and Henkin, T.M. (2004) Kinetic analysis of tRNA-directed transcription antitermination of the Bacillus subtilis *glyQS* gene in vitro. *Journal of bacteriology*, **186**, 5392-5399.
48. Grigg, J.C. and Ke, A. (2013) Sequence, structure, and stacking: Specifics of tRNA anchoring to the T box riboswitch. *RNA biology*, **10**, 1761-1764.