Investigating the Role of Single Stranded RNA Structure on Riboswitch Function and Activity

by

Catherine D. Eichhorn

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2012

Doctoral Committee:

       Professor Hashim M. Al-Hashimi, Chair
       Professor Charles L. Brooks III
       Professor Carol A. Fierke
       Professor Anna K. Mapp

2012

To my family and friends.

Acknowledgments

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique that reports structural and dynamic information over a wide range of biologically relevant timescales at atomic resolution. In this dissertation, a combination of NMR techniques, molecular dynamics (MD) simulations, mutagenesis, and biological assays is implemented to characterize the structural and dynamic properties of the 12 nucleotide single stranded RNA (ssRNA) tail located in the *Bacillus subtilis* prequeuosine riboswitch aptamer. ^{13}C (R_1, R_2) spin relaxation and residual dipolar couplings (RDCs) are used in combination with MD studies to gain insights into fast (picosecond to nanosecond) and slow (up to millisecond) timescale motions. We find that the ssRNA, although highly flexible, adopts a structured, A-form-like conformation within the polyadenine tract. Additionally, the MD simulation shows a similar level of order within the polyadenine tract, with a high level of dynamics at the terminal ends. A domain elongation strategy is applied to decouple internal and overall motions in order to more quantitatively assess dynamics within the ssRNA. We find that the domain elongated ssRNA has similar structural and dynamic properties to the 12 nt ssRNA, and is on average coaxially stacked with respect to the reference helix. We previously identified an A to C mutation in the middle of the polyadenine tract, which was found to destabilize the structural stability of the ssRNA. We hypothesized that ordering of the ssRNA was important for efficient riboswitch function, and that by destabilizing the ssRNA the

ability of the prequeuosine riboswitch to terminate transcription efficiently may be reduced. An *in vitro* transcription assay is developed to address the role of the mutation in proper riboswitch functioning, and finds that the mutation impacts the time-sensitive functioning of the riboswitch. Overall, this dissertation establishes that ssRNA is capable of forming structured and helical regions, and that ssRNA structure plays an important role in the kinetics of riboswitch function. Further, this dissertation lays out a general approach for assessing the structural and dynamic characteristics of other biologically relevant ssRNA systems.

CHAPTER 1

Introduction

1.1 Importance of RNA dynamics

1.1.1 Important biological roles of non-coding RNA

Ever since Friedrich Miescher's isolation and discovery of "nuclein" in 1869, scientists have sought to determine the role nucleic acids play in cellular function[1,2]. In the years that followed, significant advances were made in our understanding of the chemical composition and fundamental role of nucleic acids in the storage and transfer of genetic material. However, it was not until the structure of the DNA double helix was solved by Watson and Crick in 1953, clearly showing how its function was achievable, that nucleic acid research began to rapidly advance[3]. This pivotal discovery was one of the first to highlight the fundamental relationship between biological structure and function, paving the way for future studies. In the years that followed, a veritable explosion of biological functions for nucleic acids took place.

While the elucidation of the double helix resolved the fundamental question of how DNA can serve as the holder of genetic information, several questions regarding the relationship between RNA, DNA, and proteins arose from this finding. In 1958, Francis Crick developed the 'central dogma of biology' hypothesis, in which DNA is the carrier of inherited material and proteins carry out all catalytic and regulatory functions[4]. In this model, RNA has a passive role of serving as a messenger between DNA and proteins. One of the first lines of evidence to dispute this simplified model was the discovery that RNA can be reverse transcribed into DNA[5,6] and that RNA that does not code for protein, or non-coding RNA (ncRNA), is capable of catalyzing reactions[7,8].

Although it was known that a relatively small percentage (2-3%) of the genome accounted for protein expression, it was assumed that DNA that did not code for proteins

was merely a vestige of once protein-coding genes that, during the course of evolution, lost function through multiple mutations, rendering the DNA 'junk'. However, the discovery that up to 98% of DNA is transcribed into RNA despite only 2-3% accounting for protein coding regions hinted that ncRNA has important biological functions[9,10]. Indeed, ncRNAs have been identified that have a wide diversity of regulatory functions that play crucial roles in cellular processes. For instance, ncRNA can perform enzymatic catalysis, as seen in the ribosome and ribozymes[7,11]; regulate gene expression, as seen in riboswitches[12-14]; serve as a scaffold for protein recruitment, as seen in HOTAIR[15,16] and HIV-1 TAR[17,18]; and regulate mRNA stability, as seen in siRNA[19].

Accompanying these discoveries is the growing realization that most regulatory RNAs do not fold into a single native conformation, but rather, adopt many different conformations along a rugged free energy landscape despite consisting of only four nucleic acid building blocks[20,21]. These distinct conformations are often preferentially stabilized by cellular cues to elicit a specific biological function[21-23]. For example, riboswitches are a class of regulatory RNAs, typically located in the 5′ untranslated region of genes, that transition between different secondary structures to regulate the expression of genes in response to a wide range of cellular stimuli[24,25]. Beyond understanding function, RNA is increasing in its importance as a drug target[26] and a dynamic view of RNA structure is essential for successfully applying structure-based approaches in lead compound discovery and optimization[23,27-29].

1.1.2 Single stranded RNA serves numerous functions

RNA structure can be decomposed into three basic tiers. First is the primary sequence, which is directly transcribed from the DNA template. The second tier is secondary structure that forms as a result of two complementary nucleotides (A-U, G-C) that form hydrogen bonds to one another. Consecutive base pairs form helices, connected together by flexible unpaired residues. The helical stability results from the sequence composition as well as the sequence arrangement. RNA helices adopt an A-form conformation and are generally considered to be rigid cylindrical bodies[30]. The third tier is the three-dimensional (3D) tertiary conformation that results from the global

positioning of helices within an RNA molecule. It is the 3D conformation that is highly dynamic and adaptable[21,31-34].

A large emphasis has been placed on understanding how secondary structure transitions carry out a biological function[35-37]. However, there are many different kinds of unpaired, or single stranded, RNA (ssRNA) elements that also serve crucial roles [38,39]. Additionally, ssRNA can serve as a checkpoint to regulate RNA maturation. For example, transfer RNA (tRNA) has a 5′ ssRNA leader that must be cleaved by RNase P before it can be recognized as a fully mature tRNA[40,41]. These relatively short (1-6) single stranded regions contained within secondary structures are a large focus of current studies[42-47]. As we learn more about larger RNA systems another type of ssRNA has emerged − elements that separate domains. The secondary structure of the HIV genome solved by Weeks and coworkers reveals modular structured domains separated by ssRNA up to 50 nt long[48]. These 'interdomain' ssRNA elements are speculated to assist proper folding of each domain[48]. Perhaps the most well-studied single stranded RNA, the polyadenylated tail at the 3′ end of mRNA, has been found to have many important functions as a scaffold for regulatory proteins[49-57]. Single stranded elements also feature important roles in riboswitch function, described below, by acting as adaptable elements.

1.1.3 Riboswitches as modular regulatory elements

Among recently discovered ncRNAs, riboswitches stand out as the first example of protein-independent RNA-encoded gene regulation. Typically located in the 5′ untranslated regions (UTR) of genes, these RNA elements typically regulate gene expression at either the translational or transcriptional level by folding into different conformations in response to an array of input cellular signals, the most common of which are metabolites. Riboswitches typically function as a feedback loop by sensing a ligand involved in the function of the effected gene. Riboswitches are composed of a ligand-binding aptamer domain and an expression platform domain. These two domains carry out distinct functions: the aptamer participates in ligand recognition, whereas the expression platform directs signal transduction[58].

Prior to the discovery of riboswitches, RNA aptamers were appealing targets in the biotechnology field due to their ability to selectively bind a target of interest. Using *in*

vitro selected (SELEX) methods, RNA sequences were evolved to bind small molecule ligands with high affinity[59]. One characteristic of artificial aptamers is a high level of disorder in the unbound structure such that the RNA folds around the cognate ligand[60]. Soon after, it was discovered that RNA aptamers naturally occur as gene regulatory motifs. The first identified riboswitch was the *B. subtilis* tRNA T-box, found to be widespread in the 5′ UTR of the *tyrS* gene, encoding tyrosyl-tRNA synthetase. Henkin and coworkers found that an element within the 5′ UTR bound uncharged tRNA, resulting in the formation of an intrinsic terminator helix and thereby reducing gene expression[61]. Although these studies are the first to report an RNA-only transcription regulatory mechanism, it was unclear whether this was a unique motif or a widespread regulatory element. Several years later, Breaker and coworkers identified an RNA element that bound thiamine derivatives such as TPP and, by subsequently binding the Shine-Dalgarno region, reduced protein expression at the translation level[62]. The authors coined the term 'riboswitch', denoting the RNA's ability to turn on and off like a switch.

A large catalogue of riboswitches have since been discovered that sense cations such as Mg^{2+}[63], pH[64], temperature[65], and metabolites[66-72]. Unique scaffolds include the *glmS* riboswitch, which is a chemical rather than allosteric switch and functions as a ribozyme[66,73], the cyclic di-GMP riboswitch, which is an allosteric riboswitch that senses two ligands to regulate alternative splicing[74]; and the glycine riboswitch, which occurs as two riboswitches in tandem and binds ligand cooperatively[75,76]. Additionally, although riboswitches are commonly located at the 5′ UTR they have been found at the 3′ end to promote 3′ end stability and within introns to facilitate alternate splicing[74,77-80]. Riboswitches are found in all forms of life: algae[81], fungi[79], eukaryotes[79,82], and most commonly prokaryotes, controlling up to 3% of *B. subtilis* gene regulation[12,13,83-85]. A timeline of important riboswitch discoveries can be found in Figure 1.1.

4

artificial aptamers developed using SELEX

transcription pausing neccessary
for FMN riboswitch

allosteric cyclic di-GMP riboswitch

tRNA found to regulate transcription
termination in *tyrS* genes

first NMR study

Mg^{2+} riboswitch

translation riboswitches
are reversible

Kd << 50% activity

preQ$_1$ riboswitch

thermosensor

pH riboswitch

rho-dependent
terminator

1990 1995 2000 2005 2010

'riboswitch' term coined
TPP riboswitch

kinetic differences between
transcription and translation
riboswitches discovered

glmS riboswitch is also
a ribozyme

first free structure

alternative splicing control

TPP riboswitch discovered in eukaryotes

first bound structure

glycine riboswitch is tandem switch and binds cooperatively

Figure 1.1: Timeline of major riboswitch advances. Especially pertinent studies are highlighted in red.

Ligand recognition by the aptamer domain transduces a signal, through allosteric mechanisms, to the expression platform. The aptamer and expression domains are connected by a pliable ssRNA linker that base pairs to either domain in a mutually exclusive manner. This can be clearly demonstrated in translation-regulating riboswitches (Figure 1.2). In the absence of ligand, the linker does not base pair, leaving the Shine-Dalgarno (SD) region available for recognition by the ribosome. However, if ligand is bound, the linker base pairs to the aptamer domain, resulting in a sequestered SD region and preventing protein expression[83,86,87]. Translation-regulating riboswitches generally exist in equilibrium between an on and off state[88,89]. Upon ligand binding, the equilibrium shifts toward one state (usually off). Riboswitches that regulate translation are able to tune expression during the entire lifetime of the mRNA, and have been shown to act as thermodynamically controlled reversible switches[90].
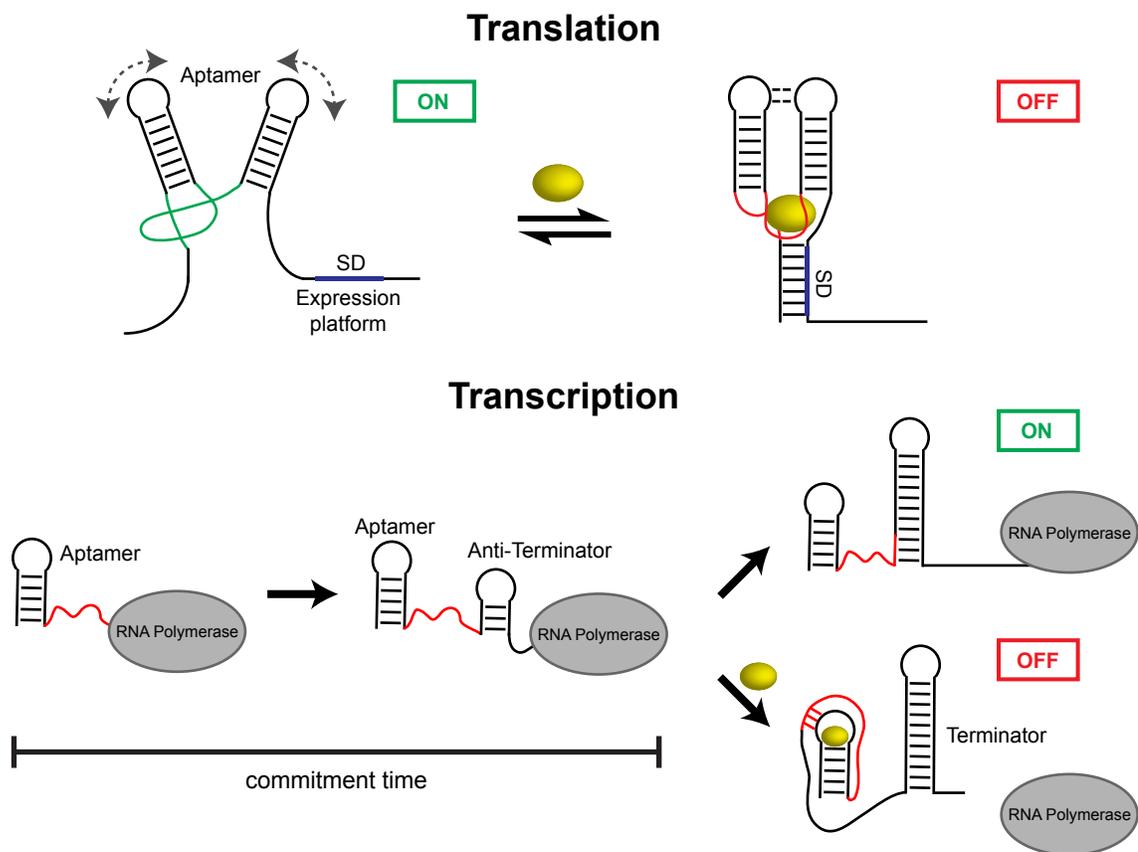
Figure 1.2: Model of translation and transcription riboswitch regulatory mechanism. *Top*: Translation riboswitches are in equilibrium between on and off states, where ligand biases the equilibrium. *Bottom*: Transcription riboswitches are kinetically controlled and depend on the timescale of RNAP transcription as well as the commitment time.

Early studies of the *in vitro* or *in vivo* activity of transcription-regulating riboswitches found that the apparent binding affinity did not correspond to the concentrations of ligand required to induce transcription termination[68,83,91]. Crothers and coworkers found that unlike translation-regulating riboswitches, riboswitches that attenuate gene expression actively during RNA transcription are not thermodynamically controlled[92]. Rather, transcription-regulating riboswitches are kinetically controlled, where the level of transcription termination efficiency is highly dependent on the rate of transcription by RNA polymerase (RNAP)[88,92]. The expression platform of transcription-regulating riboswitches is an intrinsic terminator motif[93-95], consisting of an anti-terminator/terminator hairpin system (Figure 1.2). These two hairpins overlap in the RNA sequence and are mutually exclusive, where the terminator hairpin is 3′ relative to the anti-terminator hairpin. In the absence of ligand, the anti-terminator hairpin forms,

preventing terminator hairpin formation and allowing transcription to proceed, resulting in a full mRNA product (Figure 1.2). If ligand is present the terminator hairpin will form instead, resulting in a premature RNAP dissociation and subsequent mRNA degradation by the cellular machinery. It should be noted; however, that the rates of hairpin melting are much slower than that of transcription. Therefore, once the anti-terminator has fully formed it is highly improbable that the terminator hairpin will form, even if sufficient ligand is present. As a result, transcription-regulating riboswitches have a short window of time, or a 'commitment time', by which they must commit to either the anti-terminator or terminator path (Figure 1.2) By altering the kinetics of binding or by incorporating RNAP pause sites, riboswitches are able to operate within the confines of this commitment time[88,92]. Additionally, many riboswitches have been found to resemble the bound structure in the absence of ligand, particularly when Mg^{2+} is present, likely pre-ordering the aptamer to rapidly bind ligand[96-98].

1.1.4 Prequeuosine riboswitch

Among all known riboswitches, purine riboswitches are among the smallest riboswitches with the prequeuosine (preQ$_1$) riboswitch containing the smallest naturally occurring aptamer. The *Bacillus subtilis queC* preQ$_1$-I riboswitch requires only 34 nucleotides to recognize its cognate ligand preQ$_1$, a precursor to queuosine, and to regulate genes involved in the queuosine biosynthetic pathway[71]. Queuosine is a nucleobase derived from guanine and is commonly found in the wobble position of tRNA. The minimal aptamer consists of a 4-6 base pair hairpin followed by a 12 nt adenine-rich tail. Currently two preQ$_1$-binding classes are known: the preQ$_1$-I class, which can regulate either transcription or translation[71] and the preQ$_1$-II class, which regulates translation[99]. Furthermore, preQ$_1$-I riboswitches are similar to other purine riboswitches in that a canonical Watson-Crick base pair is formed between an RNA cytosine nucleotide and its cognate ligand, whereas the preQ$_1$-II recognition motif does not appear to employ such a motif. Structural studies of the preQ$_1$-bound *queC* minimal aptamer revealed a pseudoknot structure with a host of complex tertiary interactions involved in ligand recognition, including a series of A-minor kissing interactions between tail and hairpin residues[100-105].

The small size of the *queC* riboswitch and correspondingly short commitment time (0.4-1 s) leaves very little time, in comparison to other switches, for ligand binding. For example, the *Bacillus subtilis* FMN riboswitch, which is highly dependent upon the rate of polymerase and contains sites that locally pause polymerase to lengthen the ligand-binding window, has ~70 nt between the minimal aptamer sequence and complete formation of the anti-terminator helix[92]. In comparison, the ligand-binding commitment time for the *queC* riboswitch is ~20 nt[104,105]. How rapid ligand binding is achieved is unclear given that the ssRNA tail is thought to be highly disordered, and therefore capable of sampling a wide range of competing conformations. The aim of this dissertation is to characterize the *queC* ssRNA to determine whether the ssRNA has any structural characteristics that may enable rapid binding using nucleic magnetic resonance (NMR) spectroscopy techniques, outlined in section 1.2, which report structural and dynamic information at atomic resolution and over biologically relevant timescales. In addition, an *in vitro* transcription assay is designed and implemented to determine whether the *queC* riboswitch is kinetically controlled, as observed in other transcription riboswitch systems.

1.2 Characterizing nucleic acid structure and dynamics using NMR

1.2.1 Survey of developments in nucleic acid NMR studies

The nucleic magnetic resonance (NMR) spectroscopy field has made monumental advances since the seminal discovery that spin ½ nuclei resonate at specific frequencies when placed in a magnetic field. These developments have occurred within a relatively narrow timeframe: the first one-dimensional spectrum, of water and paraffin wax, was collected by Bloch and Purcell in 1946[106,107].

The challenges of studying nucleic acids by NMR required advances in NMR methodology before it could be used to discern nucleic acid structure and dynamics. Unlike proteins, which benefit from having vastly different side chains, nucleic acids are composed of four similar building blocks: adenine, guanine, thymine/uridine, and cytosine. The chemical similarity of these residues reduces the overall chemical shift dispersion, increasing the potential for resonance overlap. In contrast, the favored nuclei

in protein studies, ^{15}N, is relatively isolated and surrounded by carbon atoms. Like spins, when close in distance via through-bond or through-space, experience couplings that can complicate spectra by the addition of artifacts or the splitting of resonances. Nucleic acid residues are aromatic bases consisting predominantly of carbon atoms, which cause a number of ^{13}C-^{13}C coupling interactions that can complicate analysis.

Early NMR studies were time consuming and plagued by severe resonance overlap, limiting the usefulness of the technique in the application of bimolecular studies. Three major advances can be credited for enabling NMR studies of biomolecules. First, the development of Fourier-transform NMR in 1966, by Ernst and Anderson, made significant advances in reducing the length of an NMR experiment while at the same time improving the signal to noise ratio[108]. Second, the development of two-dimensional (2D) NMR experiments in the 1970's greatly reduced resonance overlap by the addition of an additional nuclei dimension (i.e. ^{1}H-^{13}C)[109,110]. Multi-dimensional experiments also allow nuclei to be correlated to one another, providing a roadmap to determine biomolecules structure. Finally, the development of isotopic enrichment dramatically increased the signal to noise ratio. Previously, NMR experiments were recorded for molecules at natural abundance levels of NMR active isotopes. ^{13}C, for example, is present in approximately 1% while ^{15}N is present 0.37% of the time. The low percentage of NMR-active nuclei severely hampers NMR biomolecule studies. The commercial availability of isotopically labeled rNTPs significantly advanced the feasibility of multidimensional NMR experiments.

1.2.2 Chemical shift and RNA resonance assignments

Origin of the chemical shift

The power of NMR lies primarily in the phenomenon of the chemical shift, which is present for every spin ½ atom (i.e. ^{1}H, ^{13}C, ^{15}N, ^{31}P). Local electronic motions generate secondary or local magnetic fields. The net magnetic field experienced at a given nucleus depends both on the external magnetic field (e.g. the field strength of the NMR instrument) as well as the local fields. At a specific external magnetic field, a given nucleus type (i.e. ^{13}C) resonates at a characteristic frequency called the Larmor frequency, which varies according to the magnetic field strength. The local field

dependence, often called "nuclear shielding", either adds or subtracts to the Larmor frequency. Nuclear shielding arises from the local electronic environment. Simplistically, atoms with higher electron density will be *shielded*, and display upfield chemical shifts, whereas atoms with lower electron density will be *deshielded*, and display downfield chemical shifts.   Chemical shift is extremely sensitive to minor perturbations in electrostatics, and is therefore an excellent probe of local electronic environment and geometry.

NMR resonance assignments

        In order to characterize the structure and dynamics of a biomolecule of interest, it is imperative to first assign the chemical shifts of the biomolecule. The standard method of resonance assignment involves NMR experiments that utilize both through-space as well as through-bond correlations[111]. The most commonly used through-space solution NMR experiment relies on the nuclear Overhauser effect (NOE) that occurs between two protons (^{1}H) that are close in space (<5 Å) (Figure 1.3). An NOE is observed when nuclear spin polarization is transferred from one spin to another through dipolar cross-relaxation mechanisms[112]. The resulting 2D spectrum will have a diagonal of "autopeaks" that match the frequencies of each proton resonance; off-diagonal crosspeaks that appear are the result of magnetization transfer between two proximal protons. The sensitivity of this transfer is distance dependent ($1/r^6$), making the NOE spectroscopy, or NOESY, experiment, an extremely powerful technique for structural studies. By making basic assumptions of the biomolecule structure (for example, a canonical A-form helix), resonance assignments can be performed through what is called an "NOE walk" (Figure 1.3). For the example shown in Figure 1.3, the base H6/H8 protons are within 5 Å of both the intra-residue and inter-residue ribose H1′ proton, facilitating a straightforward walk up the contiguous helix. Other useful NOE connectivities can connect one strand to another. For instance, the uridine H3 imino proton is typically 3 Å from the adenine H2 proton of its base-pairing partner in a canonical Watson-Crick (WC) base pair, leading to strong NOE crosspeak intensities.
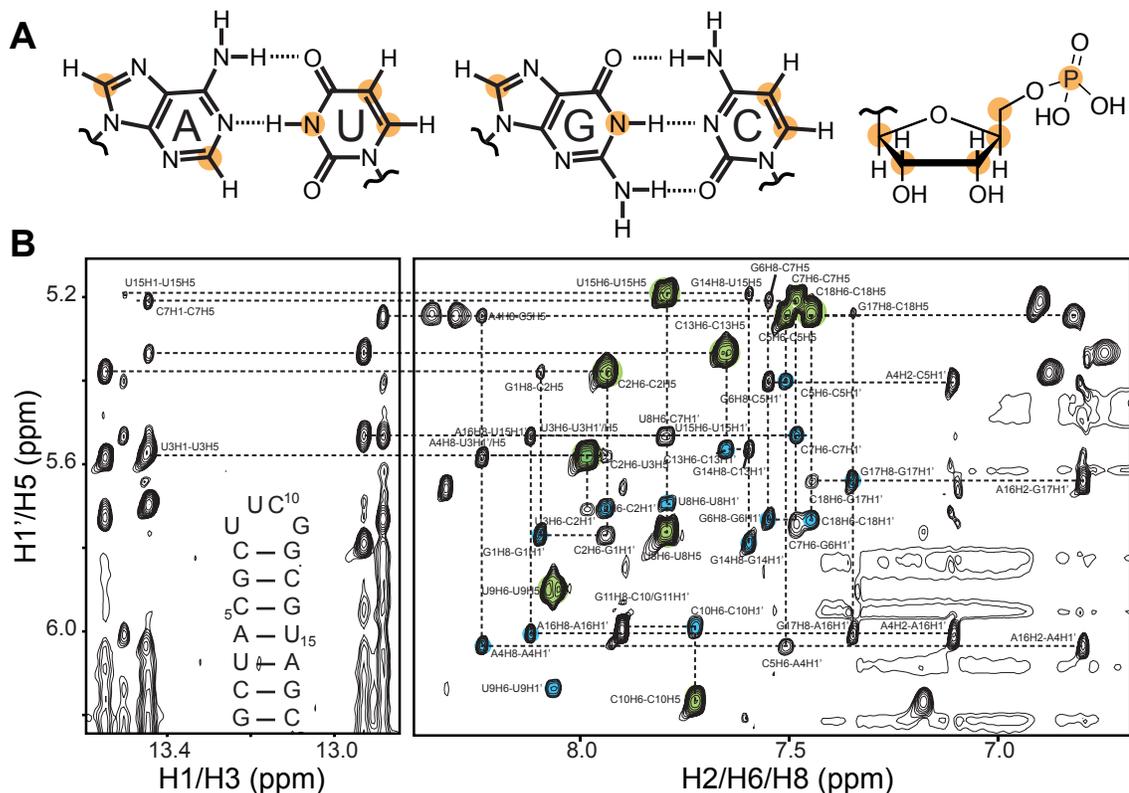
Figure 1.3: NMR resonance assignments. A) Watson-Crick base pairs with commonly used NMR probes highlighted. B) ^{1}H-^{1}H NOESY experiment allows through-space connectivities to be established. Left: imino H1/H3 connections to H5 connections shown as dashed blue lines. Right: Base to base and base to sugar connectivities. Crosspeaks highlighted in green are H5-H6 connections; highlighted in blue are H6/H8-H1′ connectivities.

As biomolecules increase in size, proton resonance overlap can become a crippling problem. NOESY experiments have been developed that filter through ^{13}C or ^{15}N, ensuring the detection of protons covalently attached to ^{13}C or ^{15}N atoms[113-117]. Although these experiments can be run as a 2D experiment, if proton overlap is particularly debilitating, they can be run as a 3D experiment, with the ^{13}C or ^{15}N frequency as the third dimension.

A host of through-bond experiments can complement NOESY data to verify or complete ambiguous assignments. The HCN experiment can be run in 2D or 3D mode and correlates the base purine C8H8 or pyrimidine C6H6 moieties to the respective N1/9 moieties through C-N scalar coupling[118-122]. The N1/9 is then correlated to the connected ribose C1′H1′ moiety. This experiment is extremely useful when NOESY H1′-H8 crosspeaks are ambiguous in assigning inter- and intra-residue crosspeaks. Additional

NMR experiments can be used to further verify assignments or to gain additional insights into structural or dynamic characteristics. If the assignment of imino resonances proves difficult, the J$_{NN}$-COSY experiment can be used to connect base pairing partners through the N-H-N hydrogen bond[123-127].

Another useful and related experiment is the constant time correlated spectroscopy (COSY) experiment, which can be used to correlate C6H6 to C5H5 spectra in addition to gaining information on the sugar pucker conformation[128,129]. Scalar couplings between two nuclei are dependent on the torsion angle of the nuclei in question. If two nuclei are connected about a rotatable bond, their scalar coupling values will be modulated by the torsion angle about this bond. According to the generalized Karplus equations[130,131], a ribose in the C3′-endo conformation will have a scalar coupling of ~0-2 Hz, resulting in a very weak crosspeak. However, a ribose in the C2′-endo conformation will correspond to a scalar coupling of ~8-14 Hz. The result is that in the COSY experiment, crosspeaks will be observed for residues with partial C2′-endo conformation. RNA in helical regions will have predominately C3′-endo sugar pucker conformations due to steric restrictions from the 2′ hydroxyl group. However, flexible residues may experience mixed sugar pucker conformations, giving rise to crosspeaks[128,132]. This experiment can be valuable in providing key insights into the sugar pucker conformation, which can assist spin relaxation and RDC measurements.

1.2.3 Methods to characterize RNA conformational dynamics by NMR

Domain elongation

The interpretation of many NMR parameters in terms of internal motions often hinges on the assumption that the internal and overall motions are not correlated to one another. This makes it possible to separate averaging contributions due to internal motions from the much larger effects arising due to overall motions[133-135]. Indeed, most formalisms developed in studies of protein dynamics invoke this so-called 'decoupling approximation' [136]. In practice, this decoupling approximation can break down in highly flexible RNA systems. Here, collective motions of A-form helical domains about flexible

junctions can lead to large changes in the overall structure of the molecule, and therefore, its overall alignment (Figure 1.4)[135,137-140].

A domain elongation strategy was developed to decouple internal and overall motions in RNA[139,141]. In this approach, a given helix in a target RNA is elongated, typically by a stretch of 22 base pairs, in order to dominate the overall shape of the molecule, and therefore, its overall alignment, in ordering media or when under the influence of the magnetic field (Figure 1.4). In this manner, internal motions occurring elsewhere in the molecule have a small effect on the overall shape and therefore alignment of the molecule. The elongated helix is not tagged onto the molecule, where tagging can give rise to complications due to mobility between the tag and target molecule. Rather, it is rigidly integrated within the natural framework of the molecule. To minimize resonance overlap the elongation can be rendered 'NMR

Figure 1.4: Domain elongation as a strategy for decoupling internal and overall motions. Right: NMR-invisible elongation approach by elongating with GC or AU helix and isotopically A/U or G/C labeling, respectively.

invisible' by using an alternating "GC/CG" elongated helix and A/U labeling or vice versa (Figure 1.4)[139,142].

The elongation has additional benefits. To a very good approximation, the elongated helix can be assumed to have an idealized A-form helical geometry. This makes it relatively straightforward to determine the overall alignment of the RNA by using RDCs measured in the elongated helix[143]. Protocols have been developed that allow accurate estimation of any uncertainty in the overall alignment tensor arising due to A-form structural noise and RDC measurement uncertainty[30]. With the overall alignment tensor in hand, the dynamic interpretation of RDCs measured in other parts of the RNA is significantly simplified. For systems composed of more than two helices, correlated

motions between helical domains can be characterized[144]. For simple hairpin structures, modulating the length of elongation can be sufficient to modulate the alignment of the RNA molecule[145].

^{13}C Spin relaxation

<u>Relaxation theory</u>

When placed in a magnetic field, a nucleus behaves in many ways like a bar magnet, having a magnetic moment that may point in any direction. Although the magnetic moment of a single nucleus is quite small, a given NMR sample contains ~10^{20} nuclei. The overall magnetic moment of these nuclei can be described by a vector model, with the net direction of magnetization called the "bulk magnetization" of the sample. Under equilibrium conditions, the net magnetization vector will be pointed along the external magnetic field, or z-axis, called longitudinal magnetization.

During an NMR experiment, bulk magnetization is tilted away from the z-axis through the use of RF pulses toward the transverse (x-y) plane. As soon as the RF pulse stops being applied, bulk magnetization begins to return to equilibrium, or relaxes. Interactions between a nucleus and the surrounding environment, commonly called the lattice, can significantly alter the relaxation mechanism. For example, consider the case of a dipole-dipole interaction between the carbon and proton nuclei in a C-H bond. The carbon nucleus will experience both the external magnetic field as well as the small magnetic field generated by the proton nucleus. As the C-H bond orientation changes, whether by internal motions or global molecular tumbling, the proton field will vary, creating an oscillating magnetic field at the carbon nucleus. A *spin-lattice* relaxation process, requiring the exchange of energy with the lattice, is called nonadiabatic and results from transverse (x-y) components of the stochastic fluctuations in the local field[146]. Alternately, *spin-spin* relaxation processes that do not require energy exchange with the lattice are called adiabatic. Spin-lattice relaxation is only caused by nonadiabatic processes and is more commonly known as longitudinal (R_1) relaxation while spin-spin relaxation, caused by both adiabatic and nonadiabatic processes, is more commonly known as transverse (R_2) relaxation[146]. R_1 and R_2 rates report on auto-relaxation

14

processes of individual nuclei whereas a third parameter, called the NOE, reports on the cross-relaxation between dipolar coupled spins[146].
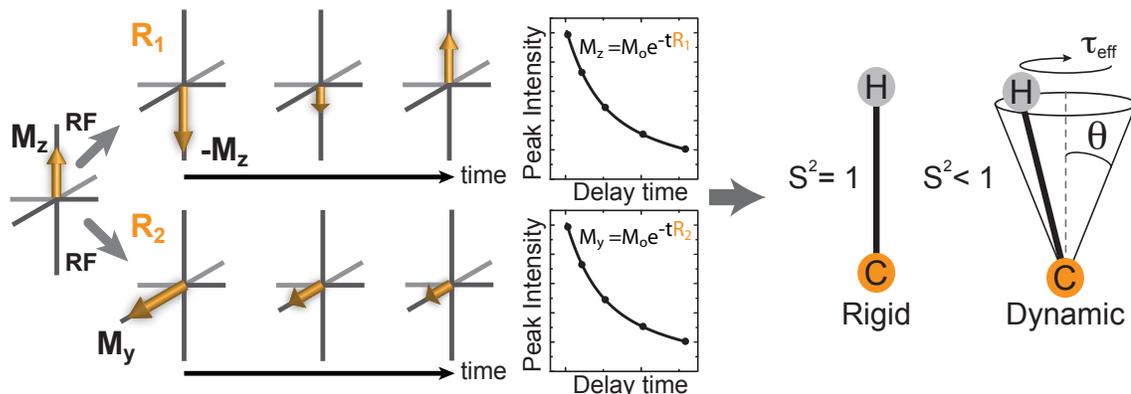
Figure 1.5: NMR spin relaxation mechanisms. An RF pulse tilts magnetization away from the z-axis, and the spin system is allowed to return to equilibrium. By plotting the peak intensity at different time points in this return to equilibrium, R_1 and R_2 values can be extracted through a monoexponential fit. Model free analysis of longitudinal (R_1) and transverse (R_2) spin relaxation data yield an order parameter (S^2) describing the amplitude of motion and a constant (τ_{eff}) describing its timescale.

Individual relaxation rate constants depend on a number of factors including correlation times, internal motions, and the orientation or geometry of the nuclei. It is important to note that internal motions occurring much slower than overall motions will have little effect on the relaxation rates since the molecule will reorient several times before the internal motions have any effect. Because of this, spin relaxation data is most sensitive to motions occurring at picosecond to nanosecond timescales. However, this sensitivity may be extended by performing the domain elongation strategy described above. If a molecule is completely rigid, bond vector reorientations will only be due to molecular tumbling causing every nucleus to have similar relaxation rates. By comparing relaxation rates for atoms within a molecule with respect to known rigid atoms, such as those involved in base pairs, a qualitative description of internal motions can be determined.

Relaxation analysis

Once measured, relaxation rates are commonly analyzed using the model-free formalism developed by Lipari and Szabo[136], which assumes internal and overall motions occur on different timescales and are uncorrelated. The model-free approach enables the

determination of the correlation function, $C(t)$, describing internal and overall motions of a bond vector:

$$J(\omega) = 2\int_0^\infty C(t)\cos(\omega t)dt \tag{1.1}$$

Decoupling the internal and overall motions divides the spectral density function in two:

$$J(\omega) = 2\int_0^\infty C_I(t)C_O(t)\cos(\omega t)dt \tag{1.2}$$

where C_I and C_O are the correlation functions describing internal and overall motions, respectively. If the further assumption is made that the overall motions are due to isotropic tumbling, C_O can be simplified to[136]:

$$C_O(t) = \frac{1}{5}e^{-6\tau_m} \tag{1.3}$$

where τ_m is the correlation time. It should be noted that this approximation is valid for molecules that can be modeled as spheres. Elongated RNA systems are more accurately modeled as rods because the long helix dominates the overall shape.

In general, the internal correlation time, C_I, is equal to the square of the generalized order parameter S^2, which varies in accordance with the range of angular motions accessed by the bond vector of interest[136]:

$$C_I(t) = S^2 + (1 - S^2)e^{\frac{-t}{\tau_e}} \tag{1.4}$$

where τ_e is the effective correlation time of the internal motion. S^2 is a measure of the spatial restriction of the internal motion of a given bond vector and ranges from 0 to 1, where 0 has maximum motions and 1 has no motions. S^2 and τ_e are sufficient to report on ps-ns motions and are model independent.

By making certain assumptions, it is possible to relate spin relaxation data to parameters describing the amplitude and timescale of internal motions for a given bond vector. However, as seen above, the assumption that internal and overall motions are uncorrelated can break down in highly flexible systems.

Residual dipolar couplings

Analogous to a pair of bar magnets, nuclear dipole-dipole interactions originate from the through-space magnetic interaction between two nuclei, where the local

magnetic field at a given nucleus is perturbed by the magnetic field of a neighboring nucleus. Consider the earlier example of the dipolar interaction between a carbon and proton nucleus in a C-H bond and how this interaction modulates the effective magnetic

field at the carbon nucleus (Figure 1.6A). The carbon nucleus experiences the sum of the static external magnetic field and the much smaller ($\sim 10^{-4}$) magnetic field generated by the proton nucleus. Because the nuclear bar magnets are always quantized parallel (or anti-parallel) to the magnetic field, the proton-induced magnetic field experienced by the carbon nucleus will vary as the C-H bond changes orientation relative to the magnetic field, either due to internal or overall motions; in some orientations the proton field adds to the external magnetic field, whereas in other orientations it subtracts or has no contribution (Figure

Figure 1.6: Physical origin and measurement of RDCs. (A) The reorientation of bond vectors leads to an oscillating local magnetic field at the nucleus of interest. (B) RDCs between spins i and j (C and H, respectively) provide long-range constraints on the average orientation (θ) of the internuclear bond vector relative to the magnetic field (B_0). (C) Measurement of RDCs as new contributions to resonance splittings (black resonances) observed upon partial alignment (green resonances).

1.6A). This angular dependence is described by $\left\langle \dfrac{3\cos^2\theta - 1}{2} \right\rangle$, where θ is the angle between the inter-nuclear vector and the magnetic field, and the angular brackets denote a time-average over all orientations sampled at a rate faster than the dipolar coupling (Figure 1.6B).

Under conditions of random molecular tumbling, the angular term averages to zero and the proton does not affect the average field at the carbon nucleus; therefore, the observed carbon frequency is unchanged. As a result, dipolar couplings are not observable under normal solution conditions. However, by imparting a small degree of order on the molecule, the angular term no longer averages to zero, and the carbon nucleus experiences a residual proton field in addition to the external magnetic field. Since half of the proton nuclei are aligned parallel and the other half anti-parallel to the

field, the proton fields add to the external field for half of the carbon nuclei and subtract for the other half[147]. Consequently, the carbon resonance frequency splits into a doublet, reflecting the addition and subtraction of the average proton field. The magnitude of this splitting is referred to as a 'residual dipolar coupling' (RDC)[148,149]. Through-space dipolar couplings (D) and through-bond scalar couplings (J) both effectively increase or decrease the average magnetic field at a given nucleus, which manifests in a splitting of resonances. This makes it possible to readily measure RDCs as new contributions to splittings when a molecule is partially aligned (Figure 1.6C).

At high magnetic fields, the dipolar interaction can be simplified to a truncated dipolar Hamiltonian[150,151], resulting in the following expression (in Hz) describing the local field contribution between nuclei i and j:

$$D_{ij} = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j h}{2\pi^2 r_{ij,eff}^3}\left\langle\frac{3\cos^2\theta - 1}{2}\right\rangle$$

(1.5)

where μ_0 is the magnetic permittivity of a vacuum, h is Planck's constant, r_{ij} is the internuclear distance between the spins, and γ is the gyromagnetic ratio. The angular bracket denotes a time average over all orientations sampled, while distance averaging of the internuclear distance is represented by the effective bond length $r_{ij,eff}$[152-154].

The utility of RDCs in studies of dynamics arises chiefly from the angular dependence of Equation 1.5[155], rendering RDCs sensitive to internal motions that reorient bond vectors at timescales faster than the inverse of the dipolar interaction. For typical levels of alignment, this encompasses a wide range of timescales spanning picoseconds to ~10 milliseconds[155]. Although RDCs do not provide information about motional timescales, they are exquisitely sensitive to the orientation distribution sampled by the bond vector and, therefore, the 3D choreography of the motion[133-135]. In addition, a wide variety of RDCs can be measured in nucleic acids (C-H, C-C, C-N, N-H, H-H, P-H, etc.), providing the basis for comprehensively mapping out nucleobase, sugar, and phosphodiester backbone dynamics.

Inducing partial alignment

The measurement of RDCs under solution conditions hinges on being able to introduce a particular level of alignment[156], either by dissolving the solute in an ordering

medium[149] or in the case of nucleic acids and paramagnetic proteins, through direct interactions with the magnetic field itself[148,157]. Alignment levels $\leq 10^{-5}$ (i.e. corresponding to 1 in 10^5 molecules being completely aligned) lead to RDCs that are too small compared to NMR line widths to allow precise measurements. Much higher degrees of alignment ($\geq 10^{-2}$) give rise to extensive dipolar couplings, compromising the spectral resolution required to analyze large biomolecules. In general, alignment levels on the order of $\sim 10^{-3}$ are optimal[156,158]. At this degree of alignment, a wide range of RDCs can be measured with a favorable magnitude-to-precision ratio while maintaining spectral resolution. A smaller subset of RDCs can be measured with alignment levels $\sim 10^{-4}$ with less than optimum magnitude-to-precision ratios.
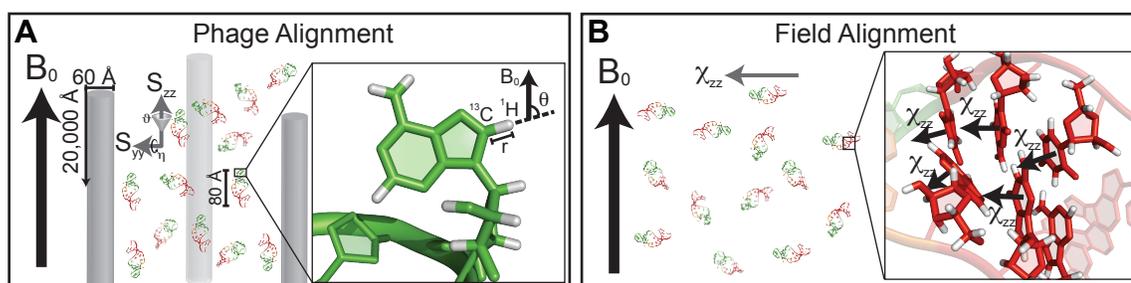
Figure 1.7: Approaches to induce partial molecular alignment using (A) ordering media such as Pf1 phage (shown in gray), which transmits order through a combination of steric and electrostatic mechanisms and (B) magnetic field alignment due to the constructive addition of anisotropic magnetic susceptibility tensors (χ) in the nucleobases.

It is now relatively straightforward to achieve alignment levels $\sim 10^{-3}$ in solution NMR by dissolving the biomolecule of interest in an inert ordering media, first demonstrated using liquid crystalline bicelles [158,147,159-161] (Figure 1.7). While this neutral bicelle medium has been used in nucleic acid studies, other media have since been introduced which have become more popular.

Since nucleic acids are highly negatively charged, the charge properties of ordering medium are an important consideration. For example, positively charged ordering media may lead to undesirable interactions with nucleic acid solutes. For nucleic acid application, the ordering medium must also be tolerant to high ionic strength conditions. The most commonly used ordering medium that satisfies the above requirements is the commercially available filamentous Pf1 bacteriophage, which induces alignment through electrostatic and steric mechanisms (Figure 1.7)[162-164].

Pf1 phage is composed of a 7.4 kb circular, single-stranded DNA genome and has a rod-like shape, estimated to be ~20,000 Å long and ~60 Å in diameter[164]. Pf1 phage is highly robust, having favorable properties largely due to its lower nematic threshold[165,166]. Its coat proteins are negatively charged, reducing the potential for adverse interactions with nucleic acids. Since polyanionic nucleic acids have a semi-uniform charge distribution[68,70], the steric and electrostatic contributions from phage are thought to have similar roles[165,167], generally aligning nucleic acids with the principal direction of order (S_{zz}) oriented along the long axis of the molecule. Positive alignment ($S_{zz} > 0$) is expected for elongated nucleic acids with S_{zz} being, on average, oriented along the magnetic field direction (Figure 1.7). Experimentally, RDCs are calculated from the difference in splittings measured in the absence (J) and presence (J+D) of Pf1 phage (Figure 1.6C). The optimum phage concentration is typically ~20 mg/mL but can vary depending on the shape of the target nucleic acid. Generally, domain elongated RNAs[139] require lower phage concentrations (5-10 mg/mL) to achieve the optimum level of alignment, whereas smaller potentially more isotropic RNAs, such as ssRNA, can require concentrations as high as 50 mg/mL[168]. The phage concentration in the NMR sample can be estimated by dividing the observed deuterium residual quadrupolar splitting by a factor of 0.886 or by measuring the UV/Vis absorbance at 270 nm using an extinction coefficient of 2.25 cm·mL/mg [164].

For proteins, the overall alignment can be modulated by changing the shape and electrostatic properties of the ordering medium used[169,170] or, by applying site-specific mutations that alter the electrostatic properties of the solute protein without affecting its functional structure[169]. This allows the measurement of multiple independent sets of RDCs from which much greater information can be obtained regarding the dynamics of bond vectors[169-173]. Attempts at using different ordering media to induce independent alignments of nucleic acids have so far been unsuccessful[174,175], likely because the uniform negative charge distribution more closely follows that of the overall molecular shape, making it difficult to independently alter shape and electrostatic contributions to alignment. The development of methods to modulate nucleic acid alignment is of key importance in enabling the extraction of the full dynamics information contained within RDCs and also in increasing the data density to allow robust cross validation of any

generated dynamic models. Two avenues can be pursued to achieve independent alignment of nucleic acids. First, magnetic field alignment has been shown to yield distinct alignments as compared to ordering media (Figure 1.7)[176,177]. Second, the systematic elongation of RNA terminal helices, which affords a change in the shape of the solute, has also been shown to modulate the overall alignment of an RNA[145].

NMR experiments to measure RDCs

Several experiments have been developed to measure a wide variety of RDCs in nucleic acids[178-189]. The choice of RDCs to be measured is generally guided by the desire to maximize the magnitude-to-precision ratio and coverage of data throughout the RNA base, sugar and backbone moieties. The most commonly and easily measured RDCs are those between directly bonded C-H, N-H, and C-C nuclei in the nucleobases and also C1′-H1′ in the sugar moieties (Figure 1.8). For small RNAs, these directly bonded C-H and N-H splittings can be measured using 2D HSQC-type experiments that employ inphase-antiphase (IPAP)[190] or spin-selective excitation methods[191-193] to encode individual components of the doublet along the ^{13}C or ^{15}N dimension. For larger RNAs (typically >40 nt), it can be advantageous to target the slowly relaxing TROSY ^{13}C or ^{15}N component of the doublet for resonances in the nucleobase that have sizeable CSAs[126,179,188,194-196]. This can be achieved either by encoding individual components of the doublets along the ^{1}H dimension[139,197,198], or through intensity-based measurements in TROSY-HSQC spectra with variable dephasing delays[188].
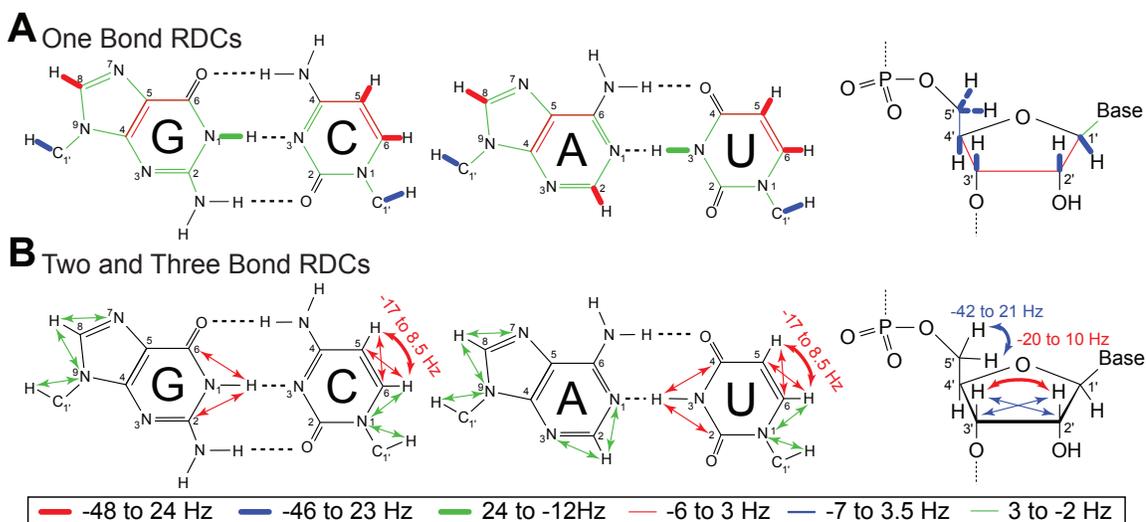
Figure 1.8: Commonly measured RDCs in nucleobase and sugar moieties. (A) One bond C-H and N-H RDCs are most often measured due to their favorable magnitude, but smaller one bond C-C and C-N as well as (B) two and three bond RDCs can be measured. Note that typically larger magnitude RDCs are shown as thick lines whereas smaller RDCs are shown as thin lines (see legend).

The alignment tensor

Central to the dynamic interpretation of RDCs is a description of the overall alignment of a molecule and specifically, the contributions to the angular term, $\left\langle \dfrac{3\cos^2\theta - 1}{2} \right\rangle$, arising due to overall reorientation. In general, overall reorientation dominates the averaging of this angular term, scaling its value down by a factor 10^{-4} compared to typically only 10^{-1} due to internal motions. The overall alignment of an internally rigid molecule relative to the magnetic field and any observed RDCs can be fully accounted for by specifying five elements of a traceless and symmetric overall order or alignment tensor[158,199]. The order tensor describes the orientation distribution of the axially symmetric magnetic field direction relative to the chiral molecular frame. The physical significance of the order tensor can be best understood using a Cartesian representation, S_{kl}[199]. Two angular terms define the orientation of a principal direction of order, S_{zz}, relative to the chiral molecular fragment. The S_{zz} axis defines the average orientation of the magnetic field relative to the fragment; it is oriented on average along and perpendicular to the magnetic field direction for $S_{zz} > 0$ and $S_{zz} < 0$, respectively. A third angular term specifies the orientation of S_{xx}/S_{yy} axes with S_{yy} pointing along the

direction of asymmetry (i.e. the direction about which the magnetic field is most likely to rotate compared to other axes perpendicular to S_{zz}). An order parameter, referred to as the generalized degree of order $(\vartheta)^{200}$ describes the degree of alignment ($\vartheta = \sqrt{\frac{2}{3}\left(S_{xx}^2 + S_{yy}^2 + S_{zz}^2\right)}$) and the extent to which the magnetic field direction is ordered relative to the molecule. Finally, an asymmetry parameter ($\eta = \frac{S_{xx} - S_{yy}}{S_{zz}}$) describes the extent of asymmetry in the distribution of the magnetic field direction relative to the chiral frame. When in the principal axis system (PAS) of the order tensor, the two order parameters are frequently expressed in terms of a magnitude, D_a and rhombicity, R^{158}:

$$D_a^{ij} = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j h}{2\pi^2 r_{ij}^3}\left(\frac{1}{2}S_{zz}\right) \quad ; \qquad R = \frac{2}{3}\eta \tag{1.6}$$

A similar order/alignment tensor type description can be used to describe internal motions within the molecule; however, in this case one describes the average orientation of an axially symmetric RDC bond vector or an axially asymmetric fragment relative to the chiral molecular frame.

For a rigid object, the time-averaged angular term in Equation 1.5 can be expressed in terms of a time-independent orientation of the internuclear vector relative to an arbitrary frame and the five order tensor elements $(S_{kl})^{199,201}$:

$$\left\langle\frac{3\cos^2\theta - 1}{2}\right\rangle = \sum_{kl=xyz}S_{kl}\cos(\alpha_k)\cos(\alpha_l) \tag{1.7}$$

where α_n is the angle between the ij^{th} internuclear vector and the nth axis of arbitrarily defined coordinates. In practice, the overall order/alignment tensor can be determined for a solute molecule provided the measurement of five or more spatially independent RDCs for bond vectors that do not undergo internal motions and whose relative orientation (but not necessarily translation) within the structure is known.

Order tensor analysis

A qualitative framework based on the order tensor analysis of RDCs was originally developed to characterize fragment orientation and dynamics in proteins[202,203] and has also been applied to characterize inter-helical motions in RNA[204]. In this

23

approach, more than five independent RDCs are used to determine five order tensor elements describing partial alignment of each helix relative to the magnetic field. Here, regions of the helices that consist of two or more non-terminal contiguous hydrogen-bonded Watson-Crick (WC) base pairs are modeled assuming a standard canonical A-form helix geometry by building sequence-specific helices using RNA structure prediction programs[30,188,204]. It is important to note that WC pairs flanked by G-U pairs or non-canonical motifs can also be used, although higher levels of structure/dynamic noise need to be considered in the analysis[30,205]. Any uncertainty arising from the assumed canonical A-form helix geometry (referred to as "structural noise"[206]) is propagated into the order tensor parameters and ultimately the relative orientation and dynamics of helices. In particular, the effects of A-form structural noise as well as RDC measurement uncertainty can be taken into account in the determination of order tensors using the program AFORM-RDC[30]. Other more general approaches for dealing with structural noise in the determination of alignment tensors have also been described[206].

The order tensor describes the average alignment of each helix relative to the applied magnetic field. For elongated RNAs, the magnetic field direction is approximately anchored along the axially symmetric axis of the elongated helix. The average orientation of fragments - one relative to the other - can be obtained by superimposing their order tensor frames[200,201,207]. The latter amounts to insisting that helical fragments share, on average, a common view of the magnetic field direction when assembled into a proper structure – similar to how countries in a properly assembled map report to a common compass bearing. The five independent parameters of the order tensor can be compared for various helices to obtain information about relative helix motions over sub-millisecond timescales[208]. While helices will report identical parameters when they are held rigid relative to one another, inter-helix motions can lead to differences. Specifically, the ϑ value for a given helix will be attenuated relative to the value observed for a helix that more strongly dominates total alignment, with the degree of attenuation generally increasing with motional amplitudes. By taking the ratio of the ϑ values for each helix ($\vartheta_{HI} / \vartheta_{HII} = \vartheta_{int}$), where ϑ_{int} ranges from 0 to 1 with 0 having maximum and 1 having minimum motions, the degree of internal motions can be determined. Although often difficult to determine reliably, the asymmetry parameter (η)

can provide insight into the directionality of inter-helix motions with spatially isotropic (directionless) motions having a smaller effect on the relative helix η values compared to anisotropic (directional) motions[200,209].

Order tensor analysis of RDCs assumes that one fragment dominates overall molecular alignment of the RNA[200,209-211]. As discussed above, this "decoupling limit" is readily satisfied in elongated RNAs or when helices are held rigidly together. Two other regimes can be identified. In the extreme coupling limit, helices have similar size and shape and contribute equally to overall alignment. Here, similar degrees of order may be observed, even in the presence of inter-helical motions, and the observation of $\vartheta_{int} = 1$ does not rule out the presence of inter-helix motions. Note that depending on the nature of inter-helical motional trajectory, different ϑ's may be observed even if the helices have equivalent size and shape. For example, twisting around the axis of a given helix will result in a reduction of its ϑ without affecting the ϑ value observed in an adjoining helix. In the intermediate coupling limit, one helix partially dominates overall alignment and the measured ϑ_{int} value will underestimate the real motional amplitudes[212]. Note that differences on the order of three base pairs can be sufficient to take an RNA system outside the extreme coupling limit and into the intermediate regime[145].

The objective of this dissertation is to use a combination of NMR techniques, molecular dynamics (MD), mutagenesis, and biological assays to characterize the structural and dynamic properties of the 12 nt prequeuosine single stranded tail. ^{13}C (R_1, R_2) and RDCs are used in combination with MD studies to gain insights on fast (ps to ns) and slow (up to ms) timescale motions. A domain elongation strategy is applied to decouple internal and overall motions in order to more quantitatively assess dynamics within the ssRNA. A mutation is found to destabilize the structural stability of the ssRNA, which is hypothesized to reduce the ability of the prequeuosine riboswitch to terminate transcription efficiently. An *in vitro* assay is developed to address the role of the mutation in proper riboswitch functioning. Overall, this dissertation establishes that ssRNA is capable of forming structured and helical regions, and that ssRNA structure plays an important role in the kinetics of riboswitch function. Further, this dissertation lays out a general approach for assessing the structural and dynamic characteristics of other biologically relevant ssRNA systems.

1.3 References

1.   Dahm, R. (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet* **122**, 565-581.
2.   Dahm, R. (2005) Friedrich Miescher and the discovery of DNA. *Dev Biol* **278**, 274-288.
3.   Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.
4.   Crick, F. H. (1958) On protein synthesis. *Symp Soc Exp Biol* **12**, 138-163.
5.   Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209-1211.
6.   Temin, H. M., and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211-1213.
7.   Cech, T. R., Zaug, A. J., and Grabowski, P. J. (1981) In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487-496.
8.   Guerrier-Takada, K., Gardiner, T., Marsh, N., Pace, N. R., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857.
9.   Eddy, S. R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**, 919-929.
10.   Huttenhofer, A., Schattner, P., and Polacek, N. (2005) Non-coding RNAs: hope or hype? *Trends Genet* **21**, 289-297.
11.   Latham, J. A., and Cech, T. R. (1989) Defining the inside and outside of a catalytic RNA molecule. *Science* **245**, 276-282.
12.   Lai, E. C. (2003) RNA sensors and riboswitches: self-regulating messages. *Curr Biol* **13**, R285-291.
13.   Mandal, M., and Breaker, R. R. (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463.
14.   Grundy, F. J., and Henkin, T. M. (2006) From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit Rev Biochem Mol Biol* **41**, 329-338.
15.   Woo, C. J., and Kingston, R. E. (2007) HOTAIR lifts noncoding RNAs to new levels. *Cell* **129**, 1257-1259.
16.   Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E., and Chang, H. Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689-693.
17.   Harrich, D., Ulich, C., Garcia-Martinez, L. F., and Gaynor, R. B. (1997) Tat is required for efficient HIV-1 reverse transcription
A critical role for the TAR element in promoting efficient human immunodeficiency virus type 1 reverse transcription. *Embo J* **16**, 1224-1235.
18.   Richter, S., Ping, Y. H., and Rana, T. M. (2002) TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. *Proc Natl Acad Sci U S A* **99**, 7928-7933.
19.   Sontheimer, E. J., and Carthew, R. W. (2005) Silence from within: endogenous siRNAs and miRNAs. *Cell* **122**, 9-12.

20. Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991) The energy landscapes and motions of proteins. *Science* **254**, 1598-1603.

21. Al-Hashimi, H. M., and Walter, N. G. (2008) RNA dynamics: it is about time. *Curr Opin Struct Biol* **18**, 321-329.

22. Leulliot, N., and Varani, G. (2001) Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **40**, 7947-7956.

23. Cruz, J. A., and Westhof, E. (2009) The dynamic landscapes of RNA architecture. *Cell* **136**, 604-609.

24. Blount, K. F., and Breaker, R. R. (2006) Riboswitches as antibacterial drug targets. *Nat Biotechnol* **24**, 1558-1564.

25. Micura, R., and Hobartner, C. (2003) On secondary structure rearrangements and equilibria of small RNAs. *Chembiochem* **4**, 984-990.

26. Cooper, T. A., Wan, L., and Dreyfuss, G. (2009) RNA and disease. *Cell* **136**, 777-793.

27. Hermann, T. (2002) Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition. *Biochimie* **84**, 869-875.

28. Fulle, S., and Gohlke, H. (2010) Molecular recognition of RNA: challenges for modelling interactions and plasticity. *J Mol Recognit* **23**, 220-231.

29. Stelzer, A. C., Frank, A. T., Kratz, J. D., Swanson, M. D., Gonzalez-Hernandez, M. J., Lee, J., Andricioaei, I., Markovitz, D. M., and Al-Hashimi, H. M. (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat Chem Biol* **7**, 553-559.

30. Musselman, C., Pitt, S. W., Gulati, K., Foster, L. L., Andricioaei, I., and Al-Hashimi, H. M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J Biomol NMR* **36**, 235-249.

31. Al-Hashimi, H. M., Gosser, Y., Gorin, A., Hu, W., Majumdar, A., and Patel, D. J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J Mol Biol* **315**, 95-102.

32. 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.

33. Rinnenthal, J., Buck, J., Ferner, J., Wacker, A., Furtig, B., and Schwalbe, H. (2011) Mapping the landscape of RNA dynamics with NMR spectroscopy. *Accounts Chem Res* **44**, 1292-1301.

34. Haller, A., Souliere, M. F., and Micura, R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Accounts Chem Res* **44**, 1339-1348.

35. Brion, P., and Westhof, E. (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* **26**, 113-137.

36. Li, P. T., Vieregg, J., and Tinoco, I., Jr. (2008) How RNA unfolds and refolds. *Annu Rev Biochem* **77**, 77-100.

37. Bailor, M. H., Mustoe, A. M., Brooks, C. L., 3rd, and Al-Hashimi, H. M. (2011) Topological constraints: using RNA secondary structure to model 3D

conformation, folding pathways, and dynamic adaptation. *Curr Opin Struct Biol* **21**, 296-305.

38. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* **264**, 968-980.

39. Costa, M., and Michel, F. (1997) Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution. *EMBO J* **16**, 3289-3302.

40. Brown, J. W. (1998) The ribonuclease P database. *Nucleic Acids Res* **26**, 351-352.

41. Spitzfaden, C., Nicholson, N., Jones, J. J., Guth, S., Lehr, R., Prescott, C. D., Hegg, L. A., and Eggleston, D. S. (2000) The structure of ribonuclease P protein from Staphylococcus aureus reveals a unique binding site for single-stranded RNA. *J Mol Biol* **295**, 105-115.

42. Shewmaker, F., Maskos, K., Simmerling, C., and Landry, S. J. (2001) The disordered mobile loop of GroES folds into a defined beta-hairpin upon binding GroEL. *J Biol Chem* **276**, 31257-31264.

43. Du, Z., Ulyanov, N. B., Yu, J., Andino, R., and James, T. L. (2004) NMR structures of loop B RNAs from the stem-loop IV domain of the enterovirus internal ribosome entry site: a single C to U substitution drastically changes the shape and flexibility of RNA. *Biochemistry* **43**, 5757-5771.

44. Eldho, N. V., and Dayie, K. T. (2007) Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: implications for catalysis. *J Mol Biol* **365**, 930-944.

45. Dethoff, E. A., Hansen, A. L., Musselman, C., Watt, E. D., Andricioaei, I., and Al-Hashimi, H. M. (2008) Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys J* **95**, 3906-3915.

46. Curuksu, J., Sponer, J., and Zacharias, M. (2009) Elbow flexibility of the kt38 RNA kink-turn motif investigated by free-energy molecular dynamics simulations. *Biophys J* **97**, 2004-2013.

47. Zhang, Q., Kim, N. K., Peterson, R. D., Wang, Z., and Feigon, J. (2010) Structurally conserved five nucleotide bulge determines the overall topology of the core domain of human telomerase RNA. *Proc Natl Acad Sci U S A* **107**, 18761-18768.

48. Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716.

49. Andreassi, C., and Riccio, A. (2009) To localize or not to localize: mRNA fate is in 3'UTR ends. *Trends Cell Biol* **19**, 465-474.

50. Regnier, P., and Hajnsdorf, E. (2009) Poly(A)-assisted RNA decay and modulators of RNA stability. *Prog Mol Biol Transl Sci* **85**, 137-185.

51. Arraiano, C. M., Andrade, J. M., Domingues, S., Guinote, I. B., Malecki, M., Matos, R. G., Moreira, R. N., Pobre, V., Reis, F. P., Saramago, M., Silva, I. J., and Viegas, S. C. (2010) The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev* **34**, 883-923.

52. Chan, S., Choi, E. A., and Shi, Y. (2011) Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev RNA* **2**, 321-335.

53. Eckmann, C. R., Rammelt, C., and Wahle, E. (2011) Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* **2**, 348-361.

54. Proudfoot, N. J. (2011) Ending the message: poly(A) signals then and now. *Genes Dev* **25**, 1770-1782.

55. Rich, A., Davies, D. R., Crick, F. H., and Watson, J. D. (1961) The molecular structure of polyadenylic acid. *J Mol Biol* **3**, 71-86.

56. Leng, M., and Felsenfeld, G. (1966) A study of polyadenylic acid at neutral pH. *J Mol Biol* **15**, 455-466.

57. Lange, H., Sement, F. M., Canaday, J., and Gagliardi, D. (2009) Polyadenylation-assisted RNA degradation processes in plants. *Trends Plant Sci* **14**, 497-504.

58. Wakeman, C. A., Winkler, W. C., and Dann, C. E., 3rd. (2007) Structural features of metabolite-sensing riboswitches. *Trends in Biochemical Sciences* **32**, 415-424.

59. Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822.

60. Meyers, L. A., Lee, J. F., Cowperthwaite, M., and Ellington, A. D. (2004) The robustness of naturally and artificially selected nucleic acid secondary structures. *Journal of molecular evolution* **58**, 681-691.

61. Grundy, F. J., and Henkin, T. M. (1993) tRNA as a positive regulator of transcription antitermination in B. subtilis. *Cell* **74**, 475-482.

62. Winkler, W., Nahvi, A., and Breaker, R. R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952-956.

63. Cromie, M. J., Shi, Y., Latifi, T., and Groisman, E. A. (2006) An RNA sensor for intracellular Mg(2+). *Cell* **125**, 71-84.

64. Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., and Altuvia, S. (2009) A pH-responsive riboregulator. *Genes Dev* **23**, 2650-2662.

65. Kortmann, J., Sczodrok, S., Rinnenthal, J., Schwalbe, H., and Narberhaus, F. (2011) Translation on demand by a simple RNA-based thermosensor. *Nucleic Acids Res* **39**, 2855-2868.

66. 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.

67. Winkler, W. C., Cohen-Chalamish, S., and Breaker, R. R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A* **99**, 15908-15913.

68. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* **10**, 701-707.

69. 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.

70. Nahvi, A., Barrick, J. E., and Breaker, R. R. (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* **32**, 143-150.

71. Roth, A., Winkler, W. C., Regulski, E. E., Lee, B. W., Lim, J., Jona, I., Barrick, J. E., Ritwik, A., Kim, J. N., Welz, R., Iwata-Reuyl, D., and Breaker, R. R. (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat Struct Mol Biol* **14**, 308-317.

72. Kim, J. N., and Breaker, R. R. (2008) Purine sensing by riboswitches. *Biol Cell* **100**, 1-11.

73. Tinsley, R. A., Furchak, J. R., and Walter, N. G. (2007) Trans-acting glmS catalytic riboswitch: locked and loaded. *RNA* **13**, 468-477.

74. 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.

75. 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.

76. Kwon, M., and Strobel, S. A. (2008) Chemical basis of glycine riboswitch cooperativity. *RNA* **14**, 25-34.

77. Weigand, J. E., and Suess, B. (2007) Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res* **35**, 4179-4185.

78. Wachter, A., Tunc-Ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K., and Breaker, R. R. (2007) Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell* **19**, 3437-3450.

79. Cheah, M. T., Wachter, A., Sudarsan, N., and Breaker, R. R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, 497-500.

80. Borsuk, P., Przykorska, A., Blachnio, K., Koper, M., Pawlowicz, J. M., Pekala, M., and Weglenski, P. (2007) L-arginine influences the structure and function of arginase mRNA in Aspergillus nidulans. *Biol Chem* **388**, 135-144.

81. Croft, M. T., Moulin, M., Webb, M. E., and Smith, A. G. (2007) Thiamine biosynthesis in algae is regulated by riboswitches. *Proc Natl Acad Sci U S A* **104**, 20770-20775.

82. Thore, S., Leibundgut, M., and Ban, N. (2006) Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **312**, 1208-1211.

83. Winkler, W. C., and Breaker, R. R. (2003) Genetic control by metabolite-binding riboswitches. *Chembiochem* **4**, 1024-1032.

84. Tucker, B. J., and Breaker, R. R. (2005) Riboswitches as versatile gene control elements. *Curr Opin Struct Biol* **15**, 342-348.

85. Pedersen, J. S., Bejerano, G., Siepel, A., Rosenbloom, K., Lindblad-Toh, K., Lander, E. S., Kent, J., Miller, W., and Haussler, D. (2006) Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol* **2**, e33.

86. 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.

87. Nudler, E. (2006) Flipping riboswitches. *Cell* **126**, 19-22.

88. Lemay, J. F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Masse, E., and Lafontaine, D. A. (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS genetics* **7**, e1001278.

89. 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.

90. 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.

91. McDaniel, B. A., Grundy, F. J., Artsimovitch, I., and Henkin, T. M. (2003) Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci U S A* **100**, 3083-3088.

92. 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. *Mol Cell* **18**, 49-60.

93. Grundy, F. J., Moir, T. R., Haldeman, M. T., and Henkin, T. M. (2002) Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: disparity between conservation and functional requirements. *Nucleic Acids Res* **30**, 1646-1655.

94. Henkin, T. M. (1996) Control of transcription termination in prokaryotes. *Annu Rev Genet* **30**, 35-57.

95. Henkin, T. M. (2000) Transcription termination control in bacteria. *Curr Opin Microbiol* **3**, 149-153.

96. Yamauchi, T., Miyoshi, D., Kubodera, T., Nishimura, A., Nakai, S., and Sugimoto, N. (2005) Roles of Mg2+ in TPP-dependent riboswitch. *FEBS Lett* **579**, 2583-2588.

97. Noeske, J., Schwalbe, H., and Wohnert, J. (2007) Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. *Nucleic Acids Res* **35**, 5262-5273.

98. Ottink, O. M., Rampersad, S. M., Tessari, M., Zaman, G. J., Heus, H. A., and Wijmenga, S. S. (2007) Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined induced fit mechanism. *RNA* **13**, 2202-2212.

99. Meyer, M. M., Roth, A., Chervin, S. M., Garcia, G. A., and Breaker, R. R. (2008) Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. *RNA* **14**, 685-695.

100. Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V., and Wedekind, J. E. (2011) Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *The Journal of biological chemistry* **286**, 24626-24637.

101. Feng, J., Walter, N. G., and Brooks, C. L., III. (2011) Cooperative and directional folding of the preQ1 riboswitch aptamer domain. *Journal of the American Chemical Society* **133**, 4196-4199.

102. 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.

103. 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.

104. 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.

105. 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.

106. Bloch, F. (1946) Nuclear Induction. *Physical Review* **70**, 460-474.

107. Purcell, E. M., Torrey, H. C., and Pound, R. V. (1946) Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review* **69**, 37-38.

108. Ernst, R. R., and Anderson, W. A. (1966) Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments* **37**, 93-102.

109. Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) 2-Dimensional Spectroscopy - Application to Nuclear Magnetic- Resonance. *J Chem Phys* **64**, 2229-2246.

110. Wüthrich, K., Wider, G., Wagner, G., and Braun, W. (1982) *J Mol Biol* **155**, 311-318.

111. Furtig, B., Richter, C., Wohnert, J., and Schwalbe, H. (2003) NMR spectroscopy of RNA. *ChemBioChem* **4**, 936-962.

112. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1-6.

113. Xia, Y. L., Man, D., and Zhu, G. (2001) 3D H-aro-NOESY-CH3NH and C-aro-NOESY-CH3NH experiments for double labeled proteins. *Journal of Biomolecular Nmr* **19**, 355-360.

114. Xia, Y. L., Sze, K. H., and Zhu, G. (2000) Transverse relaxation optimized 3D and 4D N-15/N-15 separated NOESY experiments of N-15 labeled proteins. *Journal of Biomolecular Nmr* **18**, 261-268.

115. Zhu, G., Xia, Y. L., Sze, K. H., and Yan, X. Z. (1999) 2D and 3D TROSY-enhanced NOESY of N-15 labeled proteins. *Journal of Biomolecular Nmr* **14**, 377-381.

116. Baur, M., Gemmecker, G., and Kessler, H. (1998) C-13-NOESY-HSQC with split carbon evolution for increased resolution with uniformly labeled proteins. *J. Magn. Reson.* **132**, 191-196.

117. Muhandiram, D. R., Farrow, N. A., Xu, G.-Y., Smallcombe, S. H., and Kay, L. E. (1993) A gradient ^{13}C NOESY-HSQC experiment for recording NOESY spectra of ^{13}C-labeled proteins dissolved in H_2O. *J Magn Reson* **B102**, 317-321.

118. Van Melckebeke, H., Pardi, A., Boisbouvier, J., Simorre, J. P., and Brutscher, B. (2005) Resolution-enhanced base-type-edited HCN experiment for RNA. *J Biomol NMR* **32**, 263-271.

119. Hu, W., Gosser, Y. Q., Xu, W., and Patel, D. J. (2001) Novel 2D and 3D multiple-quantum bi-directional HCNCH experiments for the correlation of ribose and base protons/carbons in 13C/15N labeled RNA. *J Biomol NMR* **20**, 167-172.

120. Brutscher, B., Simorre, J. P., Fiala, R., Czernek, J., and Sklenar, V. (2001) Transverse relaxation optimized HCN experiment for nucleic acids: combining the advantages of TROSY and MQ spin evolution
Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. *J Biomol NMR* **21**, 367-372.

121. Ramachandran, R., Sich, C., Grune, M., Soskic, V., and Brown, L. R. (1996) Sequential assignments in uniformly 13C- and 15N-labelled RNAs: the HC(N,P) and HC(N,P)-CCH-TOCSY experiments. *J Biomol NMR* **7**, 251-255.

122. Sklenar, V., Rejante, M. R., Peterson, R. D., Wang, E., and Feigon, J. (1993) Two-dimensional triple-resonance HCNCH experiment for direct correlation of ribose H1' and base H8, H6 protons in 13C,15N-labeled RNA oligonucleotides. *J. Am. Chem. Soc.* **115**, 12181-12182.

123. Dingley, A. J., and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2J_{NN}$ couplings. *J. Am. Chem. Soc.* **120**, 8293-8297.

124. Luy, B., Richter, U., DeJong, E. S., Sorensen, O. W., and Marino, J. P. (2002) Observation of H-bond mediated 3hJH2H3 coupling constants across Watson-Crick AU base pairs in RNA. *J Biomol NMR* **24**, 133-142.

125. Dingley, A. J., Masse, J. E., Feigon, J., and Grzesiek, S. (2000) Characterization of the hydrogen bond network in guanosine quartets by internucleotide 3hJ(NC)' and 2hJ(NN) scalar couplings [In Process Citation]. *J Biomol NMR* **16**, 279-289 studies of RNA.

126. Pervushin, K., Ono, A., Fernandez, C., Szyperski, T., Kainosho, M., and Wuthrich, K. (1998) NMR scaler couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation optimized spectroscopy. *Proc Natl Acad Sci U S A* **95**, 14147-14151.

127. Majumdar, A., Kettani, A., Skripkin, E., and Patel, D. J. (1999) Observation of internucleotide NH...N hydrogen bonds in the absence of directly detectable protons. *J Biomol NMR* **15**, 207-211.

128. Xia, Y., Legge, G., Jun, K. Y., Qi, Y., Lee, H., and Gao, X. (2005) IP-COSY, a totally in-phase and sensitive COSY experiment. *Magn Reson Chem* **43**, 372-379.

129. Wu, Z. R., and Bax, A. (2001) Measurement of homonuclear proton couplings based on cross-peak nulling in CT-COSY. *J Magn Reson* **151**, 242-252.

130. Karplus, M., and Grant, D. M. (1959) A Criterion for Orbital Hybridization and Charge Distribution in Chemical Bonds. *Proc Natl Acad Sci U S A* **45**, 1269-1273.

131. Karplus, M. (1963) Vicinal Proton Coupling in Nuclear Magnetic Resonance. *Journal of the American Chemical Society* **85**, 2870-2871.

132. Gonzalez, C., Stec, W., Kobylanska, A., Hogrefe, R. I., Reynolds, M., and James, T. L. (1994) Structural study of a DNA.RNA hybrid duplex with a chiral phosphorothioate moiety by NMR: extraction of distance and torsion angle constraints and imino proton exchange rates. *Biochemistry* **33**, 11062-11072.

133. Salmon, L., Bouvignies, G., Markwick, P., and Blackledge, M. (2011) Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales. *Biochemistry* **50**, 2735-2747.

134. Tolman, J. R., and Ruan, K. (2006) NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem Rev* **106**, 1720-1736.

135. Getz, M., Sun, X., Casiano-Negroni, A., Zhang, Q., and Al-Hashimi, H. M. (2007) NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers* **86**, 384-402.

136. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559.

137. Showalter, S. A., Baker, N. A., Tang, C. G., and Hall, K. (2005) Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics. *Journal Of Biomolecular Nmr* **32**, 179-193.

138. Sun, X., Zhang, Q., and Al-Hashimi, H. M. (2007) Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition. *Nucleic Acids Res* **35**, 1698-1713.

139. 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.

140. Zhang, Q., and Al-Hashimi, H. M. (2008) Extending the NMR spatial resolution limit for RNA by motional couplings. *Nat Methods* **5**, 243-245.

141. 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.

142. Zhang, Q., and Al-Hashimi, H. M. (2009) Domain-elongation NMR spectroscopy yields new insights into RNA dynamics and adaptive recognition. *RNA* **15**, 1941-1948.

143. Musselman, C., Zhang, Q., Al-Hashimi, H., and Andricioaei, I. (2010) Referencing strategy for the direct comparison of nuclear magnetic resonance and molecular dynamics motional parameters in RNA. *J Phys Chem B* **114**, 929-939.

144. Fisher, C. K., Zhang, Q., Stelzer, A., and Al-Hashimi, H. M. (2008) Ultrahigh resolution characterization of domain motions and correlations by multialignment and multireference residual dipolar coupling NMR. *J Phys Chem B* **112**, 16815-16822.

145. Dethoff, E. A., Hansen, A. L., Zhang, Q., and Al-Hashimi, H. M. (2010) Variable helix elongation as a tool to modulate RNA alignment and motional couplings. *J Magn Reson* **202**, 117-121.

146. Cavanagh, J. (2007) *Protein NMR spectroscopy : principles and practice*, 2nd ed., Academic Press, Amsterdam ; Boston.

147. Tolman, J. R., and Al-Hashimi, H. M. (2003) NMR Studies of Biomolecular Dynamics and Structural Plasticity Using Residual Dipolar Couplings. *Annual Reports on NMR Spectroscopy* **51**, 105-166.

148. Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1995) Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information

For Structure Determination in Solution. *Proc Natl Acad Sci U S A* **92**, 9279-9283.

149.   Tjandra, N., and Bax, A. (1997) Measurement of dipolar contributions to $^1J_{CH}$ splittings from magnetic-field dependence of J modulation in two- dimensional NMR spectra. *J. Magn. Reson.* **124**, 512-515.

150.   Abragam, A. (1961) The principles of nuclear magnetism. *International series of monographs on physics*, xvi, 599 p.

151.   Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *Principles of nuclear magnetic resonance in one and two dimensions*, Clarendon Press, Oxford.

152.   Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179 - 5197.

153.   Gelbin, A., Schneider, B., Clowney, L., Hsieh, S. H., Olson, W. K., and Berman, H. M. (1996) Geometric parameters in nucleic acids: Sugar and phosphate constituents. *Journal Of The American Chemical Society* **118**, 519-529.

154.   Xu, X. P., and Case, D. A. (2001) Automated prediction of 15N, 13Calpha, 13Cbeta and 13C' chemical shifts in proteins using a density functional database. *J Biomol NMR* **21**, 321-333.

155.   Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1997) NMR evidence for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.* **4**, 292-297.

156.   Tjandra, N. (1999) Establishing a degree of order: obtaining high-resolution NMR structures from molecular alignment. *Struct Fold Des* **7**, R205-R211.

157.   Bothner-By, A. A. (1995) Magnetic Field Induced Alignment of Molecules. *Encyclopedia of nuclear magnetic resonance* **5**, 2932-2238.

158.   Tjandra, N., and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-1114.

159.   Prestegard, J. H., and Kishore, A. I. (2001) Partial alignment of biomolecules: an aid to NMR characterization. *Curr. Opin. Chem. Biol.* **5**, 584-590.

160.   Ram, P., and Prestegard, J. H. (1988) Magnetic field induced ordering of bile salt/phospholipid micelles: new media for NMR structural investigations. *Biochim. Biophys. Acta.* **940**, 289-294.

161.   Sanders, C. R., Hare, B. J., Howard, K. P., and Prestegard, J. H. (1994) Magnetically-Oriented Phospholipid Micelles as a Tool for the Study of Membrane-Associated Molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **26**, 421-444.

162.   Clore, G. M., Starich, M. R., and Gronenborn, A. M. (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shaped viruses. *J. Am. Chem. Soc.* **120**, 10571-10572.

163.   Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065-1074.

164. Hansen, M. R., Hanson, P., and Pardi, A. (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol* **317**, 220-240.

165. Wu, B., Petersen, M., Girard, F., Tessari, M., and Wijmenga, S. S. (2006) Prediction of molecular alignment of nucleic acids in aligned media. *Journal of biomolecular NMR* **35**, 103-115.

166. Zweckstetter, M., and Bax, A. (2001) Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *Journal of biomolecular NMR* **20**, 365-377.

167. Zweckstetter, M., Hummer, G., and Bax, A. (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys J* **86**, 3444-3460.

168. Eichhorn, C. D., Feng, J., Suddala, K. C., Walter, N. G., Brooks, C. L., 3rd, and Al-Hashimi, H. M. (2011) Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch. *Nucleic Acids Res*.

169. Ramirez, B. E., and Bax, A. (1998) Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. *Journal of the American Chemical Society* **120**, 9106-9107.

170. Al-Hashimi, H. M., Valafar, H., Terrell, M., Zartler, E. R., Eidsness, M. K., and Prestegard, J. H. (2000) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *J Magn Reson* **143**, 402-406.

171. Tolman, J. R. (2002) A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *J Am Chem Soc* **124**, 12020-12030.

172. Peti, W., Meiler, J., Bruschweiler, R., and Griesinger, C. (2002) Model-free analysis of protein backbone motion from residual dipolar couplings. *J Am Chem Soc* **124**, 5822-5833.

173. Ruan, K., Briggman, K. B., and Tolman, J. R. (2008) De novo determination of internuclear vector orientations from residual dipolar couplings measured in three independent alignment media. *Journal of biomolecular NMR* **41**, 61-76.

174. Bondensgaard, K., Mollova, E. T., and Pardi, A. (2002) The global conformation of the hammerhead ribozyme determined using residual dipolar couplings. *Biochemistry* **41**, 11532-11542.

175. Latham, M. P., Hanson, P., Brown, D. J., and Pardi, A. (2008) Comparison of alignment tensors generated for native tRNA(Val) using magnetic fields and liquid crystalline media. *Journal of biomolecular NMR* **40**, 83-94.

176. Al-Hashimi, H. M., Majumdar, A., Gorin, A., Kettani, A., Skripkin, E., and Patel, D. J. (2001) Field- and phage-induced dipolar couplings in a homodimeric DNA quadruplex: relative orientation of G.(C-A) triad and G-tetrad motifs and direct determination of C2 symmetry axis orientation. *Journal of the American Chemical Society* **123**, 633-640.

177. Ying, J., Grishaev, A., Latham, M. P., Pardi, A., and Bax, A. (2007) Magnetic field induced residual dipolar couplings of imino groups in nucleic acids from measurements at a single magnetic field. *Journal of biomolecular NMR* **39**, 91-

96.178.   Boisbouvier, J., Bryce, D. L., O'Neil-Cabello, E., Nikonowicz, E. P., and Bax, A. (2004) Resolution-optimized NMR measurement of (1)D(CH), (1)D(CC) and (2)D(CH) residual dipolar couplings in nucleic acid bases. *Journal of biomolecular NMR* **30**, 287-301.

179.   Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D., and Simorre, J. P. (1998) Improved sensitivity and resolution in H-1-C-13 NMR experiments of RNA. *Journal of the American Chemical Society* **120**, 11845-11851.

180.   Jaroniec, C. P., Boisbouvier, J., Tworowska, I., Nikonowicz, E. P., and Bax, A. (2005) Accurate measurement of 15N-13C residual dipolar couplings in nucleic acids. *Journal of biomolecular NMR* **31**, 231-241.

181.   Miclet, E., Boisbouvier, J., and Bax, A. (2005) Measurement of eight scalar and dipolar couplings for methine-methylene pairs in proteins and nucleic acids. *Journal of biomolecular NMR* **31**, 201-216.

182.   Miclet, E., O'Neil-Cabello, E., Nikonowicz, E. P., Live, D., and Bax, A. (2003) 1H-1H dipolar couplings provide a unique probe of RNA backbone structure. *Journal of the American Chemical Society* **125**, 15740-15741.

183.   O'Neil-Cabello, E., Bryce, D. L., Nikonowicz, E. P., and Bax, A. (2004) Measurement of five dipolar couplings from a single 3D NMR multiplet applied to the study of RNA dynamics. *Journal of the American Chemical Society* **126**, 66-67.

184.   Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of Magnetic Resonance* **131**, 373-378.

185.   Schwalbe, H., Marino, J. P., King, G. C., Wechselberger, R., Bermel, W., and Griesinger, C. (1994) Determination of a complete set of coupling constants in 13C-labeled oligonucleotides. *Journal of biomolecular NMR* **4**, 631-644.

186.   Vallurupalli, P., and Moore, P. B. (2002) Measurement of H2'-C2' and H3'-C3' dipolar couplings in RNA molecules. *Journal of biomolecular NMR* **24**, 63-66.

187.   Yan, J., Corpora, T., Pradhan, P., and Bushweller, J. H. (2002) MQ-hCN-based pulse sequences for the measurement of 13C1'-1H1', 13C1'-15N, 1H1'-15N, 13C1'-13C2', 1H1'-13C2',13C6/8-1H6/8, 13C6/8-15N, 1H6/8-15N, 13C6-13C5, 1H6-13C5 dipolar couplings in 13C, 15N-labeled DNA (and RNA). *Journal of biomolecular NMR* **22**, 9-20.

188.   Ying, J., Wang, J., Grishaev, A., Yu, P., Wang, Y. X., and Bax, A. (2011) Measurement of (1)H-(1)N and (1)H-(1)(3)C residual dipolar couplings in nucleic acids from TROSY intensities. *Journal of biomolecular NMR* **51**, 89-103.

189.   Zidek, L., Wu, H., Feigon, J., and Sklenar, V. (2001) Measurement of small scalar and dipolar couplings in purine and pyrimidine bases. *Journal of biomolecular NMR* **21**, 153-160.

190.   Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* **131**, 373-378.

191.   Zidek, L., Wu, H., Feigon, J., and Sklenar, V. (2001) Measurement of small scalar and dipolar couplings in purine and pyrimidine bases. *J Biomol NMR* **21**, 153-160.

192. Permi, P. (2002) A spin-state-selective experiment for measuring heteronuclear one-bond and homonuclear two-bond couplings from an HSQC-type spectrum. *Journal Of Biomolecular Nmr* **22**, 27-35.

193. Pitt, S. W., Majumdar, A., Serganov, A., Patel, D. J., and Al-Hashimi, H. M. (2004) Argininamide binding arrests global motions in HIV-1 TAR RNA: Comparison with Mg^{2+} induced conformational stabilization. *J. Mol. Biol.* **338**, 7-16.

194. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1998) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in C-13-labeled proteins. *Journal of the American Chemical Society* **120**, 6394-6400.

195. Fiala, R., Czernek, J., and Sklenar, V. (2000) Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. *Journal of Biomolecular Nmr* **16**, 291-302.

196. Boisbouvier, J., Bryce, D. L., O'Neil-Cabello, E., Nikonowicz, E. P., and Bax, A. (2004) Resolution-optimized NMR measurement of (1)D(CH), (1)D(CC) and (2)D(CH) residual dipolar couplings in nucleic acid bases. *J Biomol NMR* **30**, 287-301.

197. Luy, B., and Marino, J. P. (2003) JE-TROSY: combined J- and TROSY-spectroscopy for the measurement of one-bond couplings in macromolecules. *J Magn Reson* **163**, 92-98.

198. Tolbert, B. S., Miyazaki, Y., Barton, S., Kinde, B., Starck, P., Singh, R., Bax, A., Case, D. A., and Summers, M. F. (2010) Major groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement. *J Biomol NMR* **47**, 205-219.

199. Saupe, A. (1968) Recent results in the field of liquid crystals. *Angew. Chem., Int. Ed. Engl.* **7**, 97-112.

200. Tolman, J. R. (2001) Dipolar couplings as a probe of molecular dynamics and structure in solution. *Curr Opin Struc Biol* **11**, 532-539.

201. Losonczi, J. A., Andrec, M., Fischer, M. W., and Prestegard, J. H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **138**, 334-342.

202. Weaver, J. L., and Prestegardgfdg, J. H. (1998) Nuclear magnetic resonance structural and ligand binding studies of BLBC, a two-domain fragment of barley lectin. *Biochemistry* **37**, 116-128.

203. Fischer, M. W., Losonczi, J. A., Weaver, J. L., and Prestegard, J. H. (1999) Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* **38**, 9013-9022.

204. Bailor, M. H., Musselman, C., Hansen, A. L., Gulati, K., Patel, D. J., and Al-Hashimi, H. M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat Protoc* **2**, 1536-1546.

205. Mustoe, A. M., Bailor, M. H., Teixeira, R. M., Brooks, C. L., 3rd, and Al-Hashimi, H. M. (2012) New insights into the fundamental role of topological

constraints as a determinant of two-way junction conformation. *Nucleic Acids Res* **40**, 892-904.

206. Zweckstetter, M., and Bax, A. (2002) Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J. Biomol. NMR* **23**, 127-137.

207. Al-Hashimi, H. M., Gorin, A., Majumdar, A., Gosser, Y., and Patel, D. J. (2002a) Towards structural genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. *J. Mol. Biol.* **318**, 637-649.

208. Tolman, J. R., Al-Hashimi, H. M., Kay, L. E., and Prestegard, J. H. (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. *Journal of the American Chemical Society* **123**, 1416-1424.

209. Al-Hashimi, H. M., and Patel, D. J. (2002c) Residual dipolar couplings: synergy between NMR and structural genomics. *J. Biomol. NMR* **22**, 1-8.

210. Briggman, K. B., and Tolman, J. R. (2003) De novo determination of bond orientations and order parameters from residual dipolar couplings with high accuracy. *J Am Chem Soc* **125**, 10164-10165.

211. Zhang, Q., Throolin, R., Pitt, S. W., Serganov, A., and Al-Hashimi, H. M. (2003) Probing motions between equivalent RNA domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. *J Am Chem Soc* **125**, 10530-10531.

212. Al-Hashimi, H. M., Gosser, Y., Gorin, A., Hu, W., Majumdar, A., and Patel, D. J. (2002b) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J. Mol. Biol.* **315**, 95-102.

CHAPTER 2

Unraveling the Structural Complexity in a Single Stranded RNA Tail:
Implications for Efficient Ligand Binding in the Prequeuosine Riboswitch

2.1 Introduction

Single stranded RNAs (ssRNAs), typically located at the ends of RNA hairpins and consisting of >3 unpaired residues, serve diverse structural and functional roles. They can fold onto neighboring RNA hairpins to form pseudoknots, essential architectural RNA elements involved in ribosomal frameshifting[1,2], hepatitis C internal ribosomal entry site (IRES) recognition[3,4], and telomerase activity[5]. Messenger RNA (mRNA) degradation is prevented or promoted by 3′ addition of a polyadenylated tail, which recruits essential protein cofactors[6]. Cleavage of the 5′ transfer RNA (tRNA) leader by RNase P is a key step in tRNA maturation[7]. In riboswitches, ssRNA links the ligand-binding aptamer domain to the expression platform, providing the basis for communication between the two[8-10].

Much of our understanding of the conformational behavior of ssRNA comes from high-resolution NMR and X-ray structures of RNA, in which ssRNA directly engages in tertiary or RNA-protein interactions. However, the atomic-level structural and dynamic behavior of these elements in the absence of these interactions remains unclear, in large part due to their high degree of flexibility. Several studies suggest that ssRNA polynucleotides adopt stacked and partially helical conformations, particularly adenine-rich sequences; however, the biological relevance of these structures is unclear[11-17]. Atomic-resolution studies of ssRNA are scarce: at present only one isosequential single stranded RNA and DNA sequence has been characterized by homonuclear NMR methods and shown to possess properties reminiscent of A-form and B-form helices, respectively[18]. Few MD studies have been performed on ssRNA, the majority of which

use the AMBER force field[19,20] to explore the impact of chemical modifications such as peptide nucleic acids (PNA) and O2′-methylation[21-24].

The class I prequeuosine riboswitch (*queC*), typically found in firmicute bacterial species, is commonly located in the 5′ untranslated region (UTR) of the queCDEF operon, which expresses proteins directly involved in the queuosine biosynthetic pathway[25]. The aptamer binds preQ$_1$, an intermediate in queuosine synthesis, with high affinity to attenuate protein expression at either the transcription or translation level[25]. This class has the smallest minimal aptamer domain (34 nucleotides, nt) discovered to date, consisting of a small hairpin followed by a 12 nt ssRNA tail (Figure 2.1A). Upon ligand recognition, the highly conserved adenine-rich tail condenses into a pseudoknot, forming a host of interactions to both the hairpin and ligand, including A-minor "kissing" interactions between the ssRNA polyadenine tract and the minor groove[26-30]. The activity of transcription-regulating riboswitches, such as the *Bacillus subtilis queC* riboswitch, has been shown to depend on the kinetics of ligand binding as well as the rate of transcription[8]. Notably, the very small size of the *queC* riboswitch leaves very little time, in comparison to other switches, for ligand binding to take place prior to formation of the anti-terminator helix which, when formed, prevents terminator helix formation, thereby allowing gene expression to continue. For example, the *Bacillus subtilis* FMN riboswitch, which is highly dependent upon the rate of polymerase and contains sites that locally pause polymerase to lengthen the ligand-binding window, has ~70 nt between the minimal aptamer sequence and complete formation of the anti-terminator helix[8]. In comparison, the ligand-binding window for the *queC* riboswitch is ~20 nt[26,27]. How efficient ligand binding is achieved is unclear given that the ssRNA tail is thought to be highly disordered, and therefore capable of sampling a wide range of competing conformations.

Here, we use NMR chemical shift, spin relaxation, and residual dipolar couplings (RDCs) in conjunction with REMD simulations using the recently updated CHARMM27 simulation nucleic acid force field [31,32] to explore the conformational properties of the 12 nt ssRNA tail from the *queC* aptamer domain and the impact of a single A-to-C mutation targeting the polyadenine tract. Our study unmasks a previously unappreciated level of

complexity in ssRNA and suggests that these structures can serve as excellent model systems for testing and developing computational force fields.

2.2 Materials and Methods

2.2.1 Sample Preparation

Uniformly ^{13}C/^{15}N labeled queC36 was prepared by *in vitro* transcription as described previously[33]. Unlabeled wild type (WT, 5′-AUAAAAAACUAA-3′) and A29C (5′-AUAACAAACUAA-3′) RNAs were purchased from Integrated DNA Technologies (IDT) and purified using a C18 column (Waters) followed by lyophilization and reconstitution in NMR buffer (15 mM sodium phosphate, pH 6.4; 25 mM sodium chloride, 0.1 mM EDTA) containing 10% D$_2$O by volume. 100% D$_2$O samples were prepared by repeatedly lyophilizing sample and replacing with 99.99% pure D$_2$O (Sigma) three times. RNA concentrations ranged from 1.5 mM to 2.8 mM. AMP, UMP, and CMP (Sigma) were directly dissolved into NMR buffer with no additional purification to 5 mM. For RDC measurements, samples were dialyzed into Millipore-purified ddH$_2$O using 1 kDa dialysis tubing (Spectrum Labs), lyophilized, and reconstituted into 52.4 mg/mL *Pf1* phage solution[34-36] in NMR buffer with 100% D$_2$O (Asla Biotech). RNA concentrations in *Pf1* phage ranged from 1.5 – 2 mM.

2.2.2 UV/Vis Melting

RNA samples (0.25-0.5 μM) were prepared in NMR buffer and the melting profiles measured between 275 K-368 K using a Varian Bio 300 UV/Vis instrument equipped with a Cary Temperature Controller. The absorbance at 260 nm was recorded every 0.5 degrees with a ramp rate of 0.5 degrees per minute. The two-state helix to coil melting transition was analyzed using $A = A_C + ((A_H - A_C)\dfrac{e^{(\frac{\Delta S}{R} - \frac{\Delta H}{RT})}}{1 + e^{(\frac{\Delta S}{R} - \frac{\Delta H}{RT})}})$, where A is the absorbance value at a given temperature T, A_H is the absorbance of the fully helical ssRNA, A_C is the absorbance of the fully random coil ssRNA, ΔS and ΔH are the entropy and enthalpy of the melting transition respectively, and R is the gas constant[37,38].

Absorbance values were fitted to the above equation using the non-linear least squares fitting function in Origin 7 to determine thermodynamic parameters. The melting temperature (T_m) was determined by dividing the enthalpy by the entropy.

2.2.3 NMR experiments

All NMR experiments were performed on a Avance Bruker 600 MHz NMR spectrometer equipped with a triple-resonance 5 mm cryogenic probe. NOESY experiments were performed at 277 K and 298 K using a mixing time of 350 ms[39]. ^{13}C spin relaxation experiments were performed at natural abundance and 298 K[40]. Relative order parameters were calculated by normalizing ($2R_2$-R_1) to either A31 (C8) or C33 (C6). Relaxation parameters were computed using HydroNMR[41,42], assuming an idealized A-form structure, to obtain diffusion tensor parameters (τ_m and D_{ratio}), and in-house written software was used to compute R_2/R_1 values as previously reported[33,40]. Motionally averaged bond lengths of 1.104 Å were used for both C8 and C6 moieties as previously described[40,43]. The following experimentally derived CSAs (σ_{xx}, σ_{yy}, σ_{zz}) were used in the analysis: (89, 15, -104); (80, 5, -85); and (98.4, 9.2, -107.5) for C2, C8, and C6 moeties[43,44].

WT		NMR				S_f^2=1		S_f^2=0.45		REMD
Residue	Carbon	R_1 (Hz)		R_2 (Hz)		R_1 (Hz)	R_2 (Hz)	R_1 (Hz)	R_2 (Hz)	S^2
A25	C8	3.276 ± 0.030		6.560 ± 0.204		2.972	20.046	3.414	11.355	0.05
A25	C2	3.369 ± 0.033		6.833 ± 0.190		3.245	22.393	3.667	12.570	*n/a*
U26	C6	3.624 ± 0.059		6.994 ± 0.397		3.625	23.916	3.993	13.431	0.09
A27	C8	3.602 ± 0.039		8.209 ± 0.266		3.043	19.767	3.449	11.232	0.26
A28	C8	3.613 ± 0.055		10.545 ± 0.419		3.074	19.695	3.465	11.201	0.34
A29	C8	3.688 ± 0.058		11.480 ± 0.454		3.076	19.659	3.463	11.182	0.43
A30	C8	3.739 ± 0.061		11.502 ± 0.513		3.077	19.668	3.463	11.186	0.41
A31	C8	3.577 ± 0.058		11.416 ± 0.503		3.077	19.661	3.464	11.183	0.34
A32	C8	3.752 ± 0.057		11.109 ± 0.433		3.081	19.682	3.469	11.197	0.29
C33	C6	4.176 ± 0.091		12.107 ± 0.597		3.627	23.882	3.992	13.413	0.32
U34	C6	3.926 ± 0.069		8.356 ± 0.451		3.584	24.014	3.968	13.468	0.08
A35	C8	3.263 ± 0.017		5.712 ± 0.115		2.981	20.071	3.423	11.372	0.04
A36	C8	3.263 ± 0.017		5.712 ± 0.115		2.979	20.089	3.422	11.380	0.02
A36	C2	3.156 ± 0.029		6.028 ± 0.165		3.245	22.393	3.667	12.570	*n/a*

Table 2.1: Table of WT NMR ^{13}C spin relaxation measurements, including HydroNMR-predicted R_1 and R_2 values assuming rigid (S^2_f=1) or semi-flexible (S^2_f=0.45) order parameters and calculated S^2 order parameters from REMD simulations.

| A29C | | NMR | | | | $S_f^2=1$ | | REMD |
Residue	Carbon	R1 (Hz)		R2 (Hz)		R1 (Hz)	R2 (Hz)	S^2
A25	C8	3.350 ± 0.040		7.059 ± 0.207		2.993	19.895	0.05
A25	C2	3.360 ± 0.041		6.660 ± 0.178		3.268	22.226	*n/a*
U26	C6	3.790 ± 0.082		8.491 ± 0.425		3.638	23.703	0.10
A27	C8	3.658 ± 0.053		8.635 ± 0.291		3.066	19.612	0.18
A28	C8	3.735 ± 0.069		11.457 ± 0.469		3.115	19.593	0.26
A28	C2	3.757 ± 0.055		10.626 ± 0.247		3.349	22.122	*n/a*
C29	C6	4.195 ± 0.292		16.601 ± 2.547		3.675	23.582	0.24
A30	C8	3.842 ± 0.086		12.054 ± 0.572		3.099	19.517	0.32
A30	C2	3.913 ± 0.064		12.612 ± 0.311		3.342	22.011	*n/a*
A31	C8	3.697 ± 0.079		13.403 ± 0.677		3.097	19.493	0.29
A32	C8	3.529 ± 0.099		14.033 ± 0.805		3.093	19.496	0.17
A32	C2	3.880 ± 0.051		10.817 ± 0.243		3.362	21.826	*n/a*
C33	C6	4.297 ± 0.119		12.509 ± 0.740		3.642	23.638	0.18
U34	C6	4.057 ± 0.087		10.792 ± 0.526		3.618	23.864	0.03
A35	C8	3.357 ± 0.020		6.801 ± 0.107		3.002	19.923	0.03
A35	C2	3.512 ± 0.044		7.026 ± 0.185		3.301	22.026	*n/a*
A36	C8	3.357 ± 0.020		6.801 ± 0.107		3.002	19.946	0.01
A36	C2	3.279 ± 0.036		5.785 ± 0.147		3.272	22.229	*n/a*

Table 2.2: Table of A29C ^{13}C spin relaxation measurements with HydroNMR-predicted R_1 and R_2 values assuming rigid structure and REMD-calculated order parameters.

IP-COSY experiments were performed at 277 K and 298 K to observe relative $^3J_{H1'-H2'}$ scalar coupling crosspeak intensities[45]. Base and sugar ^{1}H-^{13}C splittings were measured from the difference between the upfield and downfield components of the ^{1}H-^{13}C doublet along the ^{1}H component using the narrow transverse relaxation-optimized spectroscopy (TROSY) component in the ^{13}C dimension as implemented in 2D ^{1}H-^{13}C S^{3}CT-heteronuclear single quantum correlation (HSQC) experiments[46]. ^{2}H splittings were 71 Hz and 69 Hz for WT and A29C, respectively. Idealized A-form structures were constructed using Insight II (Molecular Simulations, Inc.) correcting the propeller twist angles from +15° to -15° using an in-house program, as previously described[47]. The complementary strand was removed and the resulting ssRNA used in NMR data analysis. B-form helices were constructed using W3DNA[48].

WT Residue	Bond	NMR RDC			MD RDC	A29C Residue	Bond	NMR RDC			MD RDC
A25	*C1'H1'*	*6.02*	*±*	*1.35*	*-4.66*	*A25*	*C1'H1'*	*8.07*	*±*	*1.04*	*-18.90*
A25	*C2H2*	*18.00*	*±*	*0.93*	*24.67*	*A25*	*C2H2*	*18.70*	*±*	*1.04*	*28.17*
A25	*C8H8*	*17.58*	*±*	*0.69*	*25.36*	*A25*	*C8H8*	*16.87*	*±*	*0.57*	*27.71*
U26	*C1'H1'*	*-2.77*	*±*	*1.35*	*-15.42*	*U26*	*C1'H1'*	*-3.58*	*±*	*1.04*	*-29.47*
U26	*C5H5*	*0.99*	*±*	*1.11*	*28.44*	*U26*	*C5H5*	*0.47*	*±*	*0.85*	*29.64*
U26	*C6H6*	*13.73*	*±*	*0.62*	*9.57*	*U26*	*C6H6*	*19.46*	*±*	*0.76*	*27.12*
A27	C8H8	25.51	±	0.69	32.62	A27	C8H8	20.40	±	0.57	39.57
A28	C8H8	39.16	±	0.69	42.53	A28	C8H8	36.32	±	0.57	48.74
A29	C8H8	51.47	±	0.69	48.63	A28	C2H2	23.35	±	1.04	49.55
A30	C8H8	46.89	±	0.69	49.30	A28	C1'H1'	-14.83	±	1.04	-16.50
A31	C8H8	46.89	±	0.69	40.23	A30	C8H8	46.57	±	0.57	36.23
A32	C1'H1'	17.93	±	1.35	-29.80	A30	C2H2	46.85	±	1.04	45.39
C33	C6H6	37.33	±	0.62	33.52	A30	C1'H1'	9.57	±	1.04	-12.66
U34	C6H6	15.68	±	0.62	25.09	A31	C8H8	38.11	±	0.57	29.63
A35	*C1'H1'*	*-0.90*	*±*	*1.35*	*9.45*	A31	C1'H1'	19.19	±	1.04	-22.96
A35	*C8H8*	*0.67*	*±*	*0.69*	*8.64*	A32	C2H2	33.62	±	1.04	31.14
A36	*C1'H1'*	*-1.67*	*±*	*1.35*	*15.31*	A32	C1'H1'	15.47	±	1.04	-33.50
A36	*C2H2*	*2.70*	*±*	*0.93*	*16.98*	U34	C6H6	18.12	±	0.76	13.83
A36	*C8H8*	*0.67*	*±*	*0.69*	*10.12*	U34	C5H5	10.22	±	0.85	33.47
						A35	*C1'H1'*	*-2.21*	*±*	*1.04*	*8.04*
						A35	*C2H2*	*5.14*	*±*	*1.04*	*13.17*
						A35	*C8H8*	*0.16*	*±*	*0.57*	*-0.95*
						A36	*C1'H1'*	*-1.74*	*±*	*1.04*	*10.40*
						A36	*C2H2*	*1.38*	*±*	*1.04*	*10.29*
						A36	*C8H8*	*0.16*	*±*	*0.57*	*0.74*

Table 2.3: Table of measured and computed RDC values, where italicized terminal end residues are not included in order tensor analysis. WT RDC values on left and A29C RDC values on right.

2.2.4 MD Simulation

REMD simulations were performed with the CHARMM simulation package[49] using the recently updated CHARMM27 nucleic acid force field[31,32] and the MMTSB[50] tool set. Each REMD comprised 40 replicas exponentially distributed over a temperature range from 278 K to 330 K, resulting in an average exchange acceptance ratio of 30%. Each replica was first equilibrated for 0.5 ns, restraining nucleotide heavy atoms, and subsequently run without any restraints for 10 ns, with exchange moves attempted every 0.5 ps.

Both WT and A29C RNAs were initially built in an ideal A-form helical configuration and served as the starting conformation in every simulation of REMD. The RNA was

solvated in an 80 Å cubic box of pre-equilibrated TIP3P water (approximately 50,000 atoms). Twelve pairs of sodium chloride with an additional 11 sodium ions were added to the box, corresponding to the experimental ionic concentration of 40 mM.

2.2.5 MD Simulation Analysis

We utilized the last 5 ns of the REMD trajectory at 298 K for all the following analysis. Base stacking energies were defined as the electrostatic and van der Waals interaction energies between the adjacent bases. The molecular orientation was expressed by the order parameters S^2 of the C-H bond vectors employing the model-free approach of Lipari and Szabo[51]. After a translational and rotational fit of each RNA snapshot to the ideal A-form helical structure, the order parameters were taken from the plateau phase of the correlation function, given by $C(t) = \langle P_2(\bar{\mu}(0) \cdot \bar{\mu}(t)) \rangle$, where P_2 is the second order Legendre polynomial and $\bar{\mu}$ is the unit vector along the C-H dipole. Additionally, from the atomic coordinates we constructed the RDC values by first orienting an idealized A-form ssRNA helix into the principal axis system determined from the order tensor analysis of the experimental RDCs. Each frame of the trajectory was superimposed with this ideal helix followed by calculating the average of $\left\langle \dfrac{3\cos^2\theta - 1}{2} \right\rangle$, where θ is the angle between a given bond vector (e.g. C1′H1′) and the z-axis. The RDC values were then scaled by $-82/r^3$, in which r is the C-H bond length and a factor of -82 is applied to shift the computed RDCs to the same scale as the NMR values. The average structure of the ssRNA was calculated as the structure with the minimal root-mean-square deviations from all RNA conformations in the 5 ns REMD trajectory.

2.3 Results And Discussion

2.3.1 NMR chemical shift and NOE-based analysis of the ssRNA tail conformation

Previous studies have shown that in the absence of ligand, the *queC* aptamer domain folds into a non-native hairpin, in which the 5′ strand frame-shifts to allow the first two guanine residues to base pair, with the 12 nt ssRNA tail lacking any tertiary

interactions[26]. The 2D C-H NMR spectra of the 36 nt *queC* minimal aptamer domain (Figure 2.1A), in the absence of ligand, show severe resonance overlap and large variations in resonance intensities indicating a highly disordered conformation (Figure 2.1B). Excess imino proton resonances as well as ^{1}H-^{15}N NOE data indicate that the unbound *queC* aptamer domain is in equilibrium between native and non-native hairpin conformations (data not shown), consistent with previous NMR studies[26].
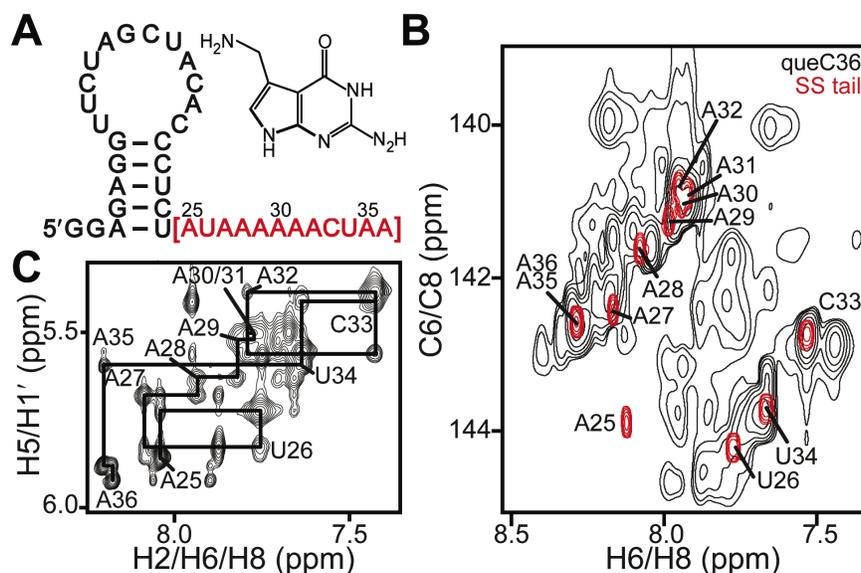
Figure 2.1 NMR resonance assignment of the prequeuosine riboswitch aptamer ssRNA tail. A) Sequence and secondary structure of the *Bacillus subtilis queC* riboswitch minimal aptamer with PreQ$_1$ inset, B) 2D C-H NMR chemical shifts show near-identical agreement between SS and unbound minimal aptamer, C) NOE crosspeaks at 277 K enable resonance assignment and indicate base stacking within single strand.

The NMR spectra suggest that the unbound 36 nt *queC* minimal aptamer domain is highly disordered and that the ssRNA tail is not involved significantly in any tertiary interactions. To test this hypothesis further, we compared NMR spectra of the isolated 12 nt ssRNA tail with the corresponding spectra of the unbound *queC* aptamer. Remarkably, NMR spectra of the isolated 12 nt ssRNA tail overlay almost perfectly with the *queC* aptamer domain and specifically onto the highly intense resonances corresponding to highly disordered residues (Figure 2.1B). The only significant deviations are observed for A25 and U26, which are located at the junction site between the hairpin and the tail (Figure 2.1B). This indicates that in the absence of ligand, the ssRNA tail is not involved

in any significant tertiary interactions under the NMR conditions (1 mM RNA, 25 mM sodium chloride, 15 mM sodium phosphate, pH 6.4, 0.1 mM EDTA, 298 K).

Similarly to the *Fusobacterium nucleatum queC* riboswitch, the *B. subtilis queC* aptamer forms kissing dimers, as observed in non-denaturing polyacrylamide gels (Figure 2.2)[52]. To ensure that the dimer does not obstruct hairpin-tail interactions, we compared a
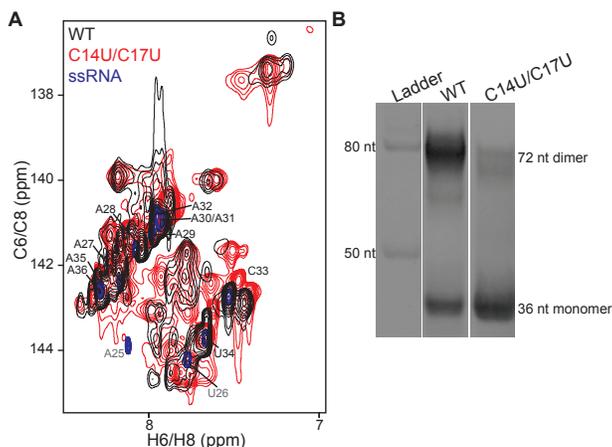
Figure 2.2: Double C14U/C17U mutant ablates kissing dimer formation. A) Comparison of nucleobase chemical shifts among wild-type (WT), C14U/C17U mutant, and 12 nt ssRNA constructs. B) 12% native PAGE shows significant dimer species in the WT construct with little to no dimer present in the C14U/C17U mutant construct.

mutant C14U/C17U construct characterized previously by Kang and coworkers to generate a ligand-bound solution NMR structure[26] to the wild-type aptamer. MFold predicts the C14U/C17U mutations will reduce the dimer stability from -6.1 kcal/mol to -0.9 kcal/mol[53]. While we observe removal of the kissing dimer, chemical shifts overall overlay extremely well between the wild-type *queC* aptamer and the C14U/C17U mutant (Figure 2.2A). Specifically,

tail chemical shifts correspond extremely well to the 12 nt ssRNA, further suggesting that the tail does not participate in tertiary interactions in the absence of ligand under our NMR conditions.

Strikingly, the spectra of the 12 nt ssRNA are well resolved, indicating that it does not adopt a completely random conformation (Figure 2.1B and Figure 2.3). This stands in stark contrast to corresponding spectra of a 12 nt polyuridine (polyU) ssRNA, well established to have a random coil conformation[16], which exhibits severe spectral overlap indicative of a highly disordered conformation (Figure 2.3). This structural order is observed in the ssRNA despite the lack of any observable imino protons and therefore any base pairing or secondary structure (Figure A1.1A).
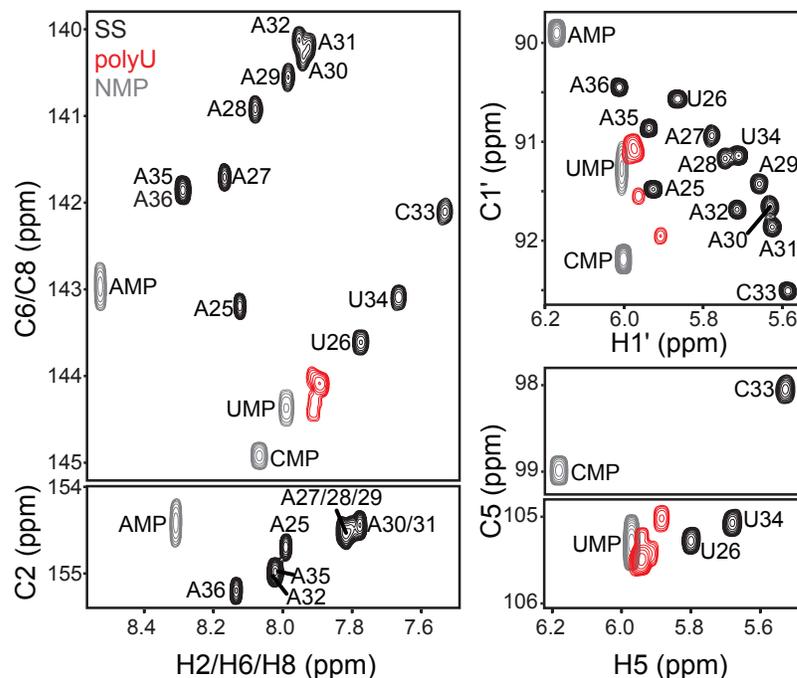
Figure 2.3: Comparison of chemical shifts of the 12 nt *queC* ssRNA (black) with free nucleotide monophosphates (gray) and a 12 nt polyuridine (red) ssRNA.

The 2D ^{1}H-^{1}H NOESY spectrum of the ssRNA shows abundant nuclear Overhauser effect (NOE) connectivities expected for a helical conformation, allowing the near complete assignment of base and sugar (H1′) protons at 298 K (Figure A1.1B). Particularly noteworthy are inter-base NOEs observed between adenine H8 protons within the polyadenine tract and between C33-U34 H6 protons, indicating significant base stacking within the polyadenine core at 277 K and decreased at 298 K (Figure A1.1)[54]. Sequential NOEs are only observed for A25, U26, A35, and A36 upon decreasing the temperature from 298 K to 277 K, indicating a higher level of disorder at the terminal ends (Figure 2.1C and Figure A1.1). Furthermore, homonuclear three bond scalar couplings (^{3}J$_{H1′-H2′}$) indicate that residues within the polyadenine core adopt a C3′-endo sugar pucker conformation, consistent with an A-form-like geometry, with the tendency to adopt alternative sugar pucker conformations increasing towards terminal residues (Figure A1.2A).

NMR chemical shifts are extremely sensitive probes of the local electronic environment for a given bond vector and can provide useful structural information[55-58].

Highly disordered residues are expected to have chemical shifts similar to nucleotide monophosphates (NMPs). While the chemical shifts of terminal residues are similar to their NMP analogs, increasing differences are observed when approaching the polyadenine core with the greatest differences observed for A30-32 (Figure 2.3 and A1.2B). The directionality of the chemical shifts is consistent with increased formation of stacking interactions towards the center of the tail[57]. This is further supported by chemical shift perturbations in a trajectory toward the NMPs with increasing temperature (data not shown). Alternatively, addition of magnesium up to 4 mM results in slight chemical shift perturbations farther from NMPs, consistent with previous studies suggesting that increases in ionic strength stabilize ssRNA stacking interactions[59] (data not shown). In contrast polyU has near-identical ($\leq$0.1 ppm) chemical shifts to UMP (Figure 2.3 and A1.2B). Thus, consistent with NOE data, the chemical shift data suggest a comparatively stacked core with a growing level of disorder towards the terminal ends. Normalized resonance intensities[33] further support these observations, which gradually increase towards the terminal ends, consistent with a higher level of pico- to nanosecond motions (Figure A1.2C).

2.3.2 Thermal stability by experiment and REMD computation

The abundance of NOEs indicates significant base stacking interactions, which likely contribute to ordering of the tail. To probe the thermodynamic stability of the tail, we performed UV/Vis melting experiments to determine the melting temperature of the helix to coil transition. Consistent with previous studies of single stranded nucleic acids, the melting profile of the ssRNA is extremely broad, characteristic of a non-cooperative transition (Figure 2.4A)[37]. Previous studies of a 7 nt polyadenine ssRNA in similar buffer conditions yield analogous melting temperatures to those observed (~35 °C compared to 31.7 ± 1.90 °C)[37].

We then used our REMD simulations to explore the temperature dependence of base stacking compared to the described UV/Vis melting curves. Base stacking energies from the REMD simulation between temperatures 278-330 K show a similar gradual decrease with increasing temperature and a similar, although reduced, T_m value (experimental 31.7 ± 1.90 °C compared to computed 20-25 °C, as estimating the T_{50}

value from melting curve, Figure 2.4A). However, the calculated base stacking energy plateaus around 320 K while the experimental slope begins to plateau around 330 K, indicating that stacking energies may be under-estimated in the REMD simulation or that additional unaccounted-for factors contribute to the ssRNA stability. Nevertheless, our data suggest that base stacking is the guiding force behind ssRNA stability, consistent with previous studies.
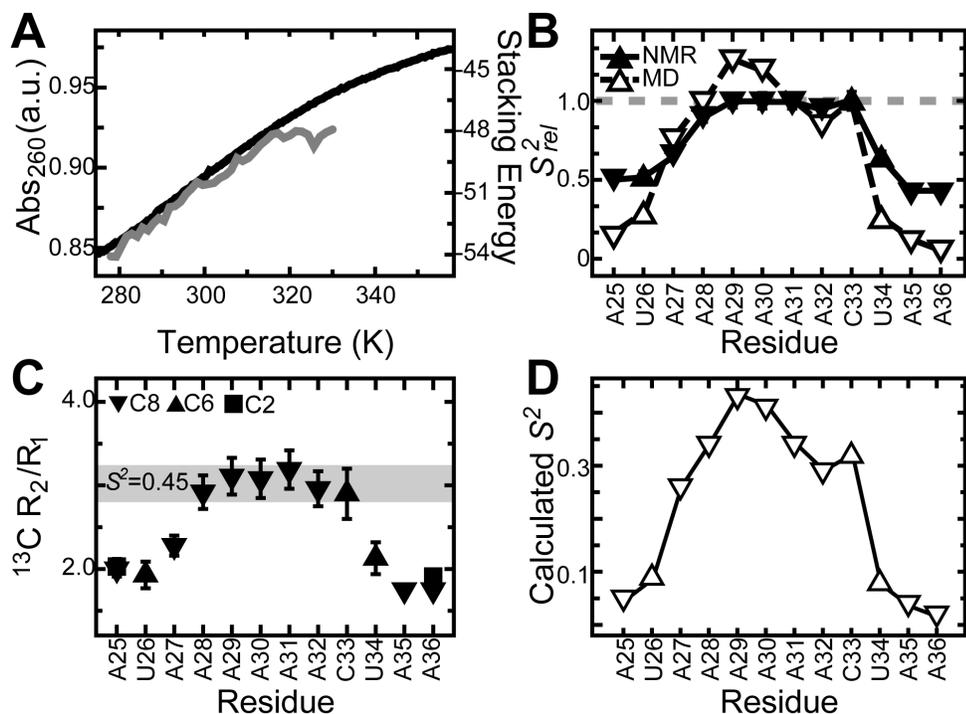
Figure 2.4: Comparison of experimental and computational thermal stabilities and spin relaxation measurements. A) UV melting profile (black) compared to base stacking energy calculated from REMD simulations (gray), B) NMR (closed) and REMD (open) relative order parameters suggest central polyadenine residues are more ordered with flexible terminal ends, C) NMR ^{13}C spin relaxation R_2/R_1 values, with HYDRONMR-predicted values assuming a global order parameter of 0.45 shown as a gray bar, D) REMD-calculated order parameters.

2.3.3 Picosecond to nanosecond dynamics by NMR spin relaxation and comparison with REMD simulations

To gain further insights into the dynamic properties of the ssRNA at pico- to nanosecond timescales, we measured longitudinal (R_1) and transverse (R_2) carbon relaxation data for the nucleobases (C2 C6 C8) using 2D ^{13}C relaxation R_1 and $R_1\rho$ NMR experiments[40], where R_1 and R_2 values are determined using in-house software (Table

2.1). These measurements represent the first nucleobase ^{13}C relaxation measurements performed on a ssRNA. The measured R_1 and R_2 values were used to compute order parameters[51] using $S^2 = (2R_2-R_1)$[60], and normalized to yield a relative order parameter (S^2_{rel}) describing the relative degree of order within a molecule ranging from 0 to 1, where 0 and 1 represent minimum and maximum order, respectively. The S^2_{rel} values were normalized against central residues A31 (C8) and C33 (C6). Resonance overlap prevented the normalization of C2 spins. Again, we observe a gradual reduction in S^2_{rel} indicating higher levels of disorder moving from central polyadenine residues (A28-C33) towards the terminal ends (Figure 2.4B).

We also computed the S^2_{rel} values based on the REMD simulation described above. The REMD simulations reproduce the general trends observed in the experiments; however, the simulations show significantly increased dynamics at the terminal ends compared to experimental values, with S^2_{rel} values approaching the dynamic limit (Figure 2.4B). Additionally, while experimental values have similar relative order parameters from A28-C33, large variations are observed in the REMD simulation, with A29-A30 being more ordered and A32 less ordered than experimentally observed (Figure 2.4B). These differences may reflect shortcomings in the force field and/or mismatch in the experimental/computational timescales since the REMD simulations likely probe fluctuations that extend beyond the picosecond timescales sensed by spin relaxation data.

The high level of disorder and motional coupling in the ssRNA prevents quantitative analysis of relaxation data using the model-free formalism, which assumes that internal and overall motions are decoupled from one another[51]. This makes it difficult if not impossible to assess the absolute level of disorder in the ssRNA; one can only make qualitative assessments about the relative disorder across different residues. However, it is noteworthy that even the comparatively high R_2/R_1 values measured in the rigid core (~2.9, Figure 2.4C) remain significantly lower than values predicted for a perfectly rigid helical ssRNA (~6.4, Table 2.1) as estimated using the program HYDRONMR[41,42]. If we assume an overall diffusion tensor predicted by HYDRONMR, we find that central polyadenine residues are highly flexible with an estimated average NMR spin relaxation order parameter S^2 of ~0.45 (Figure 2C and Table 2.1). Interestingly, similar though slightly smaller absolute S^2 values are calculated from the REMD simulations (on

average $S^2 \sim 0.36$ for core residues, Figure 2.4D). These data indicate that despite measurable stacking interactions and a helical-like average conformation, the polyadenine core is highly disordered with residues experiencing fluctuations on the order of a ±40º cone angle[61] at pico- to nanosecond timescales.

2.3.4 Overall conformation and sub-millisecond dynamics by NMR residual dipolar couplings and comparison with REMD simulations

To further probe the conformation of the ssRNA and extend the NMR timescale sensitivity to milliseconds, we measured RDCs[62,63] using 52.4 mg/mL *Pf1* phage as an ordering medium. While most RNAs align optimally in ~25 mg/mL of phage, a much higher concentration of phage was used for the ssRNA to ensure optimal alignment. To our knowledge, these are the first RDC measurements reported on a single stranded nucleic acid. The RDCs measured between two nuclei depend on $\left\langle \dfrac{3\cos^2 \theta - 1}{2} \right\rangle$, where θ is the angle between the inter-nuclear vector and the magnetic field and the angular bracket denotes a time-average over all orientations sampled at sub-millisecond timescales[62,63]. RDCs were measured for base C5H5, C6H6, C8H8, C2H2 and sugar C1′H1′ moieties[47].

In general, isotropic motions tend to reduce the observed RDC value, approaching zero at the limit of spatially unrestricted isotropic motions[61,64,65]. In the ssRNA, large base C-H RDCs are measured in the polyadenine tract residues that gradually decrease at the termini (Figure 2.5A). Although small RDC values can also arise from static placement of the bond vector near the magic angle relative to the principal direction of order, the overall trends observed are consistent with NMR chemical shift and S^2_{rel} data suggesting that the RDCs indicate increased dynamic averaging at the termini (Figure 2.5A). Interestingly, the near-zero RDCs measured at terminal residues (Figure 2.5A and Table 2.3) agree more closely to the REMD simulations compared to the S^2_{rel} values, indicating that the discrepancy between the measured and computed S^2_{rel} values may be due to truncation of the S^2 sensitivity to motions faster than nanoseconds. These results add to a growing number of NMR studies on different types of RNA showing that RDC

data are capable of probing motions that are incompletely sensed by spin relaxation due to truncation of the time-sensitivity by overall correlation time of the molecule[64,66,67]. Unfortunately, severe spectral overlap, particularly pronounced in the *Pf1* phage sample, prevented measurement of several C1′H1′ RDCs for the polyadenine core.
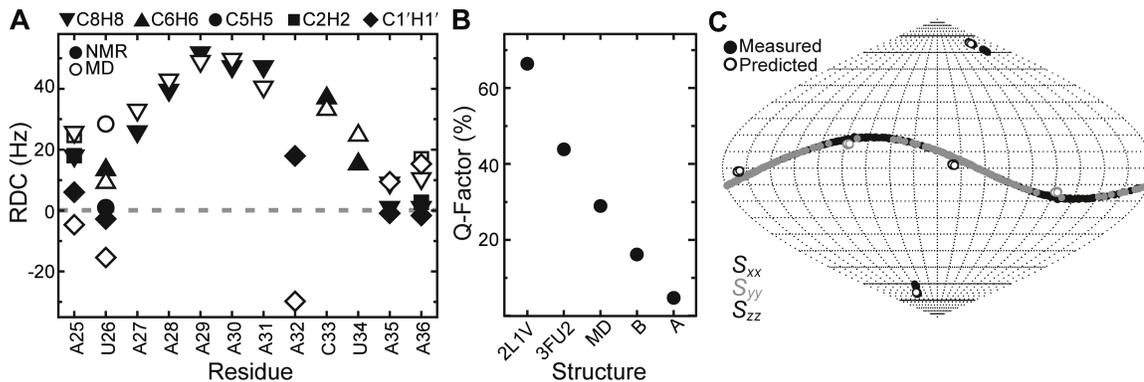
Figure 2.5: RDCs and order tensor analysis of the 12 nt *queC* aptamer tail A) Measured (closed) and computed (open) RDCs show reduced values at the terminal ends, indicating increased dynamics, B) Q-factor comparison indicates SS adops A-form conformation C) Sauson-Flamsteed map shows good agreement between predicted (open) and experimental (closed) order tensors.

We subjected the RDCs (excluding RDCs for the two flexible residues from the terminal ends) to an order tensor analysis[47,68,69] assuming different input structures including single strands derived from idealized A-form and B-form helices, the REMD-averaged structure, and available ligand-bound X-ray and NMR structures[26,27]. Despite the relatively small number of RDCs used in this analysis, we clearly observe a better fit with an A-form geometry (Q-factor 4.77%) as compared to all other conformations (Q-factor $\geq$ 16%) (Figure 2.5B). This is consistent with independently observed $^3J_{H1′-H2′}$ scalar coupling crosspeaks, which indicate a C3′-endo sugar conformation for core residues in the tail, suggesting an A-form (and not B-form) helical geometry. The RDCs are in strong disagreement with preQ$_1$-bound X-ray and NMR structures (PDBID: 3FU2 and 2L1V) indicating that the tail must undergo a transition from an A-form helical geometry towards the distinct helical conformation observed in the X-ray and NMR structures in which the A-form geometry is perturbed at the hairpin-tail junction, likely due to torsional strain from the ssRNA folding back upon the hairpin. The REMD-

averaged structure has a Q-factor of 30%, indicating a better fit than ligand-bound structures, but is still outside the range considered to represent a good fit. Together, these data suggest that, on average, the ssRNA tail adopts an A-form like conformation. The good RDC fit to the A-form structure also suggests that averaging of the RDCs due to internal motions is largely isotropic in nature, causing a semi-uniform attenuation of the RDCs relative to values expected for an A-form structure. The dynamics could involve exchange between a stacked ordered conformation and unstacked highly disordered conformation, or local isotropic motions about the average A-form conformation.

As a further check on the accuracy of the A-form structure, we compared the principal direction of alignment (S_{zz}) determined experimentally using RDCs assuming a ssRNA A-form structure with the orientation predicted by PALES[70] using a ssRNA A-form structure. Surprisingly, we find that the experimentally determined S_{zz} deviates from the helix axis by ~19.8° (Figure 2.5C). Interestingly, PALES predicts a principal direction of order that deviates from the helix axis by 14.4°; the S_{zz} orientation predicted using PALES is in good agreement from that measured experimentally (deviation ~6°). The deviation from the helix axis can be attributed to the absence of the complementary strand, resulting in an overall shape with a long axis that is not coincident with the helical axis, as reported previously for a quadruplex DNA topology[71].

To further test the conformational distribution from the REMD simulations, we used a number of simplifying assumptions to compute RDCs from the REMD trajectory. Snapshots from the REMD simulations were superimposed onto an idealized A-form helix oriented in the principal axis system determined using the experimental RDCs and the order tensor fit. RDCs were then arbitrarily scaled by $-82/r^3$, in which r is the C-H bond length and accounts for bond length variations during the dynamics. We find excellent agreement between experimental and computed nucleobase RDCs; however, computed C1′H1′ RDCs fail to reproduce observed RDCs, particularly for A32: while the magnitude is similar (18 Hz compared to -30 Hz) the sign differs, suggesting the orientation of the C1′H1′ bond vector differs between experiment and simulation (Table 2.3 and Figure A1.3A). C1′H1′ RDCs are generally opposite in sign to base RDCs in a double stranded A-form helix. However, back-calculated C1′H1′ RDCs from the order tensor analysis assuming a ssRNA A-form helix are positive in sign (Figure A1.3B),

suggesting the C1'H1' orientation in the REMD simulations deviates from an A-form structure.

2.3.5 Impact of A-to-C mutation within polyadenine core

Taken together, the data show that the polyadenine tract is relatively ordered at 298 K, with a gradual reduction in order approaching the termini and that base stacking interactions are the guiding force behind this order. To determine whether disrupting the polyadenine tract will destabilize the global structure, we substituted A29 within the polyadenine tract with a cytosine residue (referred to as A29C). Other types of mutations involving placements of uridine were not explored as these were expected to yield partially base paired conformations. As with the wild-type construct (WT), we observed no imino protons, indicating the absence of any detectable base pairing and secondary structure (Figure A2.1A).
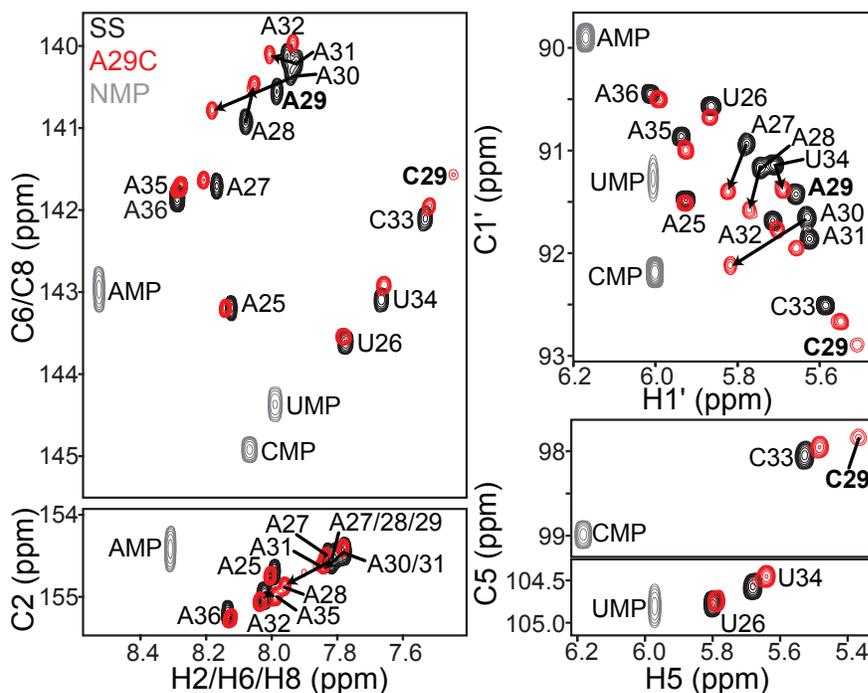
Figure 2.6: A29C chemical shift assignments and comparison to WT and NMPs.

The 2D C-H spectra for the A29C mutant remain highly disperse, and the chemical shift perturbations relative to WT are clustered around the site of mutation (A28 and A30) (Figure 2.6 and Figure 2.8A). However, small but significant chemical shift

perturbations relative to WT are also observed at more distant residues, including A27, A31, C33, and U34. These perturbations diminish when moving away from the center of the ssRNA and are basically absent in the highly flexible terminal residues (Figure 2.6 and Figure 2.8A). Such longer-range perturbations suggest that the mutation may have a long-range effect possibly by influencing the stacking interactions of several nucleobases. A perturbation to stacking interactions is also supported by distinct NOE connectivities in A29C,

Figure 2.7: UV/Vis melting curve for WT (black) and A29C (red) with base stacking energies from the REMD simulation have good agreement between a reduction in stacking energy and a reduced melting temperature.

which show weakened cross peaks to C29, and new crosspeaks between A28 (H2) and A30 (H1′) that indicate C29 partially loops out to allow A28 to stack onto A30 (Figure A2.1). The melting temperature of the mutant is reduced by ~5 °C, and the base stacking energies are computed to be ~2 kcal/mol lower compared to WT, indicating that the mutation likely destabilizes the stacking interactions (Figure 2.7).

Interestingly, many of the residues that experience chemical shift perturbations following the A29 to C29 mutation also exhibit a greater degree of dynamics as assessed by normalized resonance peak intensities in 2D C-H HSQC spectra (Figure A2.2A) and carbon relaxation data (R_1 and R_2) (Figure 2.8B and Table 2.2). In particular, severe line broadening consistent with a slow exchange process occurring at micro- to millisecond timescales manifesting as reduced resonance intensities in 2D spectra and higher R_2 values is observed for C29 in the A29C mutant (Table 2.2 and Figure 2.6). This is not observed for A29 in WT. Smaller but significant line broadening is also observed for residues A31, A32, and U34 (Table 2.2). This line broadening across several residues may reflect exchange between stacked and unstacked conformations. Higher intensities as well as reduced S^2_{rel} values are observed for residues A27 and A28, indicating a greater degree of fast pico- to nanosecond dynamics (Figure A2.2A). Note that the high R_2 and weak signal intensity leads to a higher error in the R_2/R_1 measurements, particularly for C29.

57

Although the A29C RDCs are generally in good agreement with the WT RDCs, variations are observed for a number of residues (U26, A27, A31) that indicate differences in conformation and/or dynamic behavior (Figure 2.8C). Though an order tensor analysis of 13 RDCs shows best agreement with an A-form structure, the quality of the fit is not as good as that observed for WT (Q-factor = 8.77%, Figure 2.8C). The S_{zz} direction measured for A29C when assuming an A-form structure deviates substantially from that predicted using PALES (~11°, Figure A2.3). These data suggest that A29C deviates from an idealized A-form structure as compared to WT. These deviations may reflect static and/or dynamic bending about the C29 pivot point, possibly arising from looping out of this residue from the helical stack. Such a conformation is observed in the REMD simulations of A29C ~1% but not in WT (data not shown).

Figure 2.8: Comparison of WT and A29C constructs A) Chemical shift perturbations between WT and A29C are largely localized about mutation site, B) NMR spin relaxation parameters between WT and A29C are similar, with deviations occurring several residues from mutation site, C) *Left*: Measured RDCs between WT and A29C values show good agreement, *Right*: Q-factor indicates A29C adopts A-form conformation. Gray circles indicate quality of fit upon removal of A28 C8H8, A30 C1′H1′, and A30 C2H2 measured NMR values.

In general, the REMD simulations predict the NMR data measured for A29C with reduced quality to that noted for WT. Interestingly, the computed absolute S^2 values indicate a global reduction in order for A29C, particularly for residues A27-C33, (Figure A2.2) whereas NMR relaxation parameters between WT and A29C are more similar, suggesting comparable global order parameters. The REMD simulations reveal enhanced dynamics at C29 consistent with the NMR chemical exchange data. The REMD simulation also suggests increased dynamics at A32, which is not observed

experimentally: although slightly reduced, the S^2_{rel} is within error of A29-A31 values (S^2_{rel} of 1) (Figure A2.2). Computed RDCs agree reasonably with measured RDCs, although the C1′H1′ RDCs are opposite in sign as observed in the comparison between WT NMR and REMD-calculated RDCs. The Q-factor comparing the average REMD structure to measured RDCs is 70%; however, removal of A28 C8H8, A30 C2H2, and A30 C1′H1′ RDCs improves the Q-factor significantly. This improvement is observed only for the REMD structure (Figure 2.8C), indicating that these residues, localized about the mutation site, adopt non-A-form conformations and likely experience perturbations from the increased dynamics at C29. The difference in timescales between the REMD simulations and NMR is another factor that may lead to the observed discrepancies. Nevertheless, MD and NMR data both indicate significant dynamics at the mutation site with perturbations extending toward the 3′ end of the ssRNA.

2.3.6 ssRNA tail conformation and dynamics optimized for ligand docking in *queC* aptamer

One of the main questions we set out to explore during the course of our studies was how the *queC* aptamer manages to efficiently bind its cognate ligand despite the small commitment time available in the kinetic switch and the large conformational space that may be available to a highly disordered ssRNA, which would have to search many competing conformations before arriving at the ligand bound pseudoknot conformation. Our study reveals that the ssRNA is not entirely disordered, but rather, has the character of a stacked A-form-like helical conformation, which may effectively reduce the conformational search of the ssRNA, promoting efficient docking onto the hairpin to form the pseudoknot. Moreover, our study uncovers a greater degree of flexibility towards the terminal ends, particularly the 5′ end which forms the pivot point for docking the ssRNA tail onto the hairpin loop.

The NMR data clearly show the absence of any pre-existing tertiary interactions involving the ssRNA tail in the unbound *queC* aptamer domain. This together with our findings regarding the conformational behavior of the unbound ssRNA tail suggests the following model for ligand binding (Figure 2.9). In the absence of ligand, the ssRNA tail

is disordered but on average forms an A-form helix-like conformation, which can efficiently explore conformational space about a highly flexible junction. The ligand may transiently form encounter complexes when the tail is close in space to the P1 hairpin, and possibly with the help of divalent ions such as calcium[27,72], triggering the necessary conformational changes required to form the pseudoknot and binding pocket. This finding is
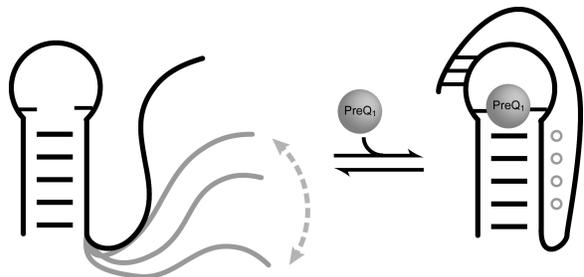
Figure 2.9: A tentative model for *queC* riboswitch ligand recognition. In the absence of ligand the ssRNA tail rotates freely about the helix-tail pivot point in a stacked, A-form helical-like conformation. Upon ligand recognition the pseudoknot is stabilized. A-minor tertiary interactions are shown as open gray circles.

consistent with computational modeling of the ligand binding mechanism in which A-minor tertiary interactions form first, followed by pseudoknot formation[30] and may explain the fast ligand binding rate observed in the related *F. nucleatum queC* riboswitch[52]. Our results, including the observation of greater dynamics in the mutant, provide a framework for more rigorous testing of this proposed model with future *in vitro* and *in vivo* studies.

2.4 Conclusion

Our study shows that ssRNA can exhibit complex conformational behavior, including variable levels of stacking and propensities to form an A-form helical conformation across the polynucleotide chain, and also, the ability to interrupt stacked residues by introducing sequence-specific kinks and/or distortions. While it has been known for some time that polyadenine stretches tend to stack and form helical conformations[13,14,16,18,37], the details of this helical geometry were difficult to decipher based solely on NOE-based NMR data. Our RDC measurements on the ssRNA, together with scalar coupling constant measurements, strongly suggest that the polyadenine tract forms an A-form-like conformation in the WT ssRNA. Our results also unveil dynamic complexity in ssRNA, including a gradual increase in disorder occurring towards the terminal ends that is reminiscent of unfolded polypeptide chains[73], and also, slower sequence-specific dynamics occurring at micro- to millisecond timescales that may

involve transient stacking/unstacking motions that may result in kinking of the ssRNA. All together, our studies show that 'structured' ssRNA exhibits exquisite quality spectra and can be studied quantitatively using NMR based structure and dynamics measurements.

The REMD simulations recapitulate many of the key features and trends observed based on the melting and NMR data, including the existence of stacking interactions that are weakened by the A29C mutation, the formation of helical geometry that may be kinked in A29C at the mutation site, and an increase in dynamic disorder towards the terminal ends and localized about C29 in the mutant. Having said that, the REMD simulations showed weaker agreement with sugar RDCs or sugar conformation, particularly in the A29C mutant, and had increased dynamics compared to the NMR data. Prior studies on HIV-1 TAR RNA noted higher levels of dynamics in CHARMM simulations compared to NMR measurements[74]. Our studies indicate that suboptimal base stacking energies may be a source of these excess dynamics. However, a quantitative assessment of the simulations requires the application of domain-elongation[33,74-76] methods to rigorously decouple internal and overall motions, and make it possible to quantitatively predict NMR measurements. In addition, MD simulations that retain aspects of time are required to compare the rates of dynamics observed by relaxation and exchange broadening type measurements. The simplicity of ssRNA offers a much needed model system for such studies directed at rigorously examining currently used nucleic acid force fields.

Finally, our results suggest that the conformational properties of the ssRNA tail are optimized to allow the *queC* riboswitch to efficiently bind ligands within the short commitment time available to this kinetic switch. In particular, the pre-stacked ssRNA tail can efficiently rotate about a flexible hinge against the hairpin loop, and explore conformational space efficiently for rapid ligand binding. This pre-stacking about dynamic hinges may be a general feature of many ssRNAs that can play different architectural roles in a variety of RNA contexts.

2.5 References

1. Brierley, I., Digard, P., and Inglis, S. C. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**, 537-547.

2. Brierley, I., Rolley, N. J., Jenner, A. J., and Inglis, S. C. (1991) Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* **220**, 889-902.

3. Rijnbrand, R., van der Straaten, T., van Rijn, P. A., Spaan, W. J., and Bredenbeek, P. J. (1997) Internal entry of ribosomes is directed by the 5' noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. *J Virol* **71**, 451-457.

4. Wang, C., Le, S. Y., Ali, N., and Siddiqui, A. (1995) An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5' noncoding region. *RNA* **1**, 526-537.

5. Gilley, D., and Blackburn, E. H. (1999) The telomerase RNA pseudoknot is critical for the stable assembly of a catalytically active ribonucleoprotein. *Proc Natl Acad Sci U S A* **96**, 6621-6625.

6. Lange, H., Sement, F. M., Canaday, J., and Gagliardi, D. (2009) Polyadenylation-assisted RNA degradation processes in plants. *Trends Plant Sci* **14**, 497-504.

7. Kurz, J. C., and Fierke, C. A. (2000) Ribonuclease P: a ribonucleoprotein enzyme. *Curr. Opin. Chem. Biol.* **4**, 553-558.

8. 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. *Mol Cell* **18**, 49-60.

9. Montange, R. K., and Batey, R. T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys* **37**, 117-133.

10. Breaker, R. R. (2010) Riboswitches and the RNA World. *Cold Spring Harb Perspect Biol*.

11. Fasman, G. D., Lindblow, C., and Grossman, L. (1964) The Helical Conformations of Polycytidylic Acid: Studies on the Forces Involved. *Biochemistry* **3**, 1015-1021.

12. McDonald, C. C., Phillips, W. D., and Penman, S. (1964) Nucleic Acids: A Nuclear Magnetic Resonance Study. *Science* **144**, 1234-1237.

13. Brahms, J., Michelson, A. M., and Van Holde, K. E. (1966) Adenylate oligomers in single- and double-strand conformation. *J Mol Biol* **15**, 467-488.

14. Leng, M., and Felsenfeld, G. (1966) A study of polyadenylic acid at neutral pH. *J Mol Biol* **15**, 455-466.

15. BouchemalChibani, N., duPenhoat, C. H., Abdelkafi, M., Ghomi, M., and Turpin, P. Y. (1996) Characterizaten of the dynamic behavior of r(ACC) and r(AAC) with NMR relaxation data and both Metropolis Monte Carlo and molecular dynamics simulations. *Biopolymers* **39**, 549-571.

16. Seol, Y., Skinner, G. M., and Visscher, K. (2004) Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys Rev Lett* **93**, 118102.

17. Seol, Y., Skinner, G. M., Visscher, K., Buhot, A., and Halperin, A. (2007) Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys Rev Lett* **98**, 158103.

18.   Isaksson, J., Acharya, S., Barman, J., Cheruku, P., and Chattopadhyaya, J. (2004) Single-stranded adenine-rich DNA and RNA retain structural characteristics of their respective double-stranded conformations and show directional differences in stacking pattern. *Biochemistry* **43**, 15996-16010.

19.   Kollman, P. A., Wang, J. M., and Cieplak, P. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* **21**, 1049-1074.

20.   Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* **92**, 3817-3829.

21.   Kaukinen, U., Venalainen, T., Lonnberg, H., and Perakyla, M. (2003) The base sequence dependent flexibility of linear single-stranded oligoribonucleotides correlates with the reactivity of the phosphodiester bond. *Org Biomol Chem* **1**, 2439-2447.

22.   Yeh, I. C., and Hummer, G. (2004) Diffusion and electrophoretic mobility of single-stranded RNA from molecular dynamics simulations. *Biophys J* **86**, 681-689.

23.   Biyani, M., and Nishigaki, K. (2005) Single-strand conformation polymorphism (SSCP) of oligodeoxyribonucleotides: an insight into solution structural dynamics of DNAs provided by gel electrophoresis and molecular dynamics simulations. *J Biochem* **138**, 363-373.

24.   Panecka, J., Mura, C., and Trylska, J. (2011) Molecular dynamics of potential rRNA binders: single-stranded nucleic acids and some analogues. *J Phys Chem B* **115**, 532-546.

25.   Roth, A., Winkler, W. C., Regulski, E. E., Lee, B. W., Lim, J., Jona, I., Barrick, J. E., Ritwik, A., Kim, J. N., Welz, R., Iwata-Reuyl, D., and Breaker, R. R. (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat Struct Mol Biol* **14**, 308-317.

26.   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.

27.   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.

28.   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.

29.   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.

30.   Feng, J., Walter, N. G., and Brooks, C. L., III. (2011) Cooperative and directional folding of the preQ1 riboswitch aptamer domain. *Journal of the American Chemical Society* **133**, 4196-4199.

31.   Denning, E. J., Priyakumar, U. D., Nilsson, L., and Mackerell, A. D., Jr. (2011) Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update

of the CHARMM all-atom additive force field for RNA. *J Comput Chem* **32**, 1929-1943.

32.   Foloppe, N., and MacKerell, A. D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem* **21**, 86-104.

33.   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.

34.   Clore, G. M., Starich, M. R., and Gronenborn, A. M. (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shaped viruses. *J. Am. Chem. Soc.* **120**, 10571-10572.

35.   Hansen, M. R., Hanson, P., and Pardi, A. (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol* **317**, 220-240.

36.   Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065-1074.

37.   Dewey, T. G., and Turner, D. H. (1979) Laser temperature-jump study of stacking in adenylic acid polymers. *Biochemistry* **18**, 5757-5762.

38.   Freier, S. M., Hill, K. O., Dewey, T. G., Marky, L. A., Breslauer, K. J., and Turner, D. H. (1981) Solvent effects on the kinetics and thermodynamics of stacking in poly(cytidylic acid). *Biochemistry* **20**, 1419-1426.

39.   Hwang, T. L., and Shaka, A. J. (1995) Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson. Ser. A* **112**, 275-279.

40.   Hansen, A. L., and Al-Hashimi, H. M. (2007) Dynamics of large elongated RNA by NMR carbon relaxation. *Journal of the American Chemical Society* **129**, 16072-16082.

41.   Garcia de la Torre, J., Huertas, M. L., and Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138-146.

42.   Bernado, P., Garcia de la Torre, J., and Pons, M. (2002) Interpretation of 15N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J Biomol NMR* **23**, 139-150.

43.   Ying, J., Grishaev, A., Bryce, D. L., and Bax, A. (2006) Chemical shift tensors of protonated base carbons in helical RNA and DNA from NMR relaxation and liquid crystal measurements. *J Am Chem Soc* **128**, 11443-11454.

44.   Stueber, D., and Grant, D. M. (2002) 13C and (15)N chemical shift tensors in adenosine, guanosine dihydrate, 2'-deoxythymidine, and cytidine. *Journal of the American Chemical Society* **124**, 10539-10551.

45.   Xia, Y., Legge, G., Jun, K. Y., Qi, Y., Lee, H., and Gao, X. (2005) IP-COSY, a totally in-phase and sensitive COSY experiment. *Magn Reson Chem* **43**, 372-379.

46.   Meissner, A., and Sorensen, O. W. (1999) The role of coherence transfer efficiency in design of TROSY-type multidimensional NMR experiments. *J. Magn. Reson.* **139**, 439-442.

47.   Bailor, M. H., Musselman, C., Hansen, A. L., Gulati, K., Patel, D. J., and Al-Hashimi, H. M. (2007) Characterizing the relative orientation and dynamics of

RNA A-form helices using NMR residual dipolar couplings. *Nat Protoc* **2**, 1536-1546.

48.   Zheng, G., Lu, X. J., and Olson, W. K. (2009) Web 3DNA--a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res* **37**, W240-246.

49.   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., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614.

50.   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.

51.   Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *Journal of the American Chemical Society* **104**, 4559-4570.

52.   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.

53.   Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415.

54.   Adamiak, R. W., Blazewicz, J., Formanowicz, P., Gdaniec, Z., Kasprzak, M., Popenda, M., and Szachniuk, M. (2004) An algorithm for an automatic NOE pathways analysis of 2D NMR spectra of RNA duplexes. *J Comput Biol* **11**, 163-179.

55.   Case, D. A. (1995) Calibration of ring-current effects in proteins and nucleic acids. *J Biomol NMR* **6**, 341-346.

56.   Cromsigt, J., Hilbers, C. W., and Wijmenga, S. S. (2001) Prediction of proton chemical shifts in RNA - Their use in structure refinement and validation. *Journal of Biomolecular NMR* **21**, 11-29.

57.   Fares, C., Amata, I., and Carlomagno, T. (2007) 13C-detection in RNA bases: revealing structure-chemical shift relationships. *Journal of the American Chemical Society* **129**, 15814-15823.

58.   Zhao, Y., Alipanahi, B., Li, S. C., and Li, M. (2010) Protein secondary structure prediction using NMR chemical shift data. *J Bioinform Comput Biol* **8**, 867-884.

59.   Kankia, B. I. (2003) Binding of Mg2+ to single-stranded polynucleotides: hydration and optical studies. *Biophys Chem* **104**, 643-654.

60.   Dethoff, E. A., Hansen, A. L., Musselman, C., Watt, E. D., Andricioaei, I., and Al-Hashimi, H. M. (2008) Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys J* **95**, 3906-3915.

61.   Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1997) NMR evidence for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.* **4**, 292-297.

62. Tjandra, N., and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-1114.

63. Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A* **92**, 9279-9283.

64. Getz, M., Sun, X., Casiano-Negroni, A., Zhang, Q., and Al-Hashimi, H. M. (2007) NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers* **86**, 384-402.

65. Tolman, J. R., and Ruan, K. (2006) NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem Rev* **106**, 1720-1736.

66. Sun, X., Zhang, Q., and Al-Hashimi, H. M. (2007) Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition. *Nucleic Acids Res* **35**, 1698-1713.

67. 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.

68. Losonczi, J. A., Andrec, M., Fischer, M. W. F., and Prestegard, J. H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* **138**, 334-342.

69. Musselman, C., Pitt, S. W., Gulati, K., Foster, L. L., Andricioaei, I., and Al-Hashimi, H. M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J Biomol NMR* **36**, 235-249.

70. Zweckstetter, M. (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc* **3**, 679-690.

71. Al-Hashimi, H. M., Majumdar, A., Gorin, A., Kettani, A., Skripkin, E., and Patel, D. J. (2001) Field- and phage-induced dipolar couplings in a homodimeric DNA quadruplex: relative orientation of G.(C-A) triad and G-tetrad motifs and direct determination of C2 symmetry axis orientation. *Journal of the American Chemical Society* **123**, 633-640.

72. 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. *Journal of the American Chemical Society* **133**, 5190-5193.

73. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein. *Science* **295**, 1719-1722.

74. Musselman, C., Zhang, Q., Al-Hashimi, H., and Andricioaei, I. (2010) Referencing strategy for the direct comparison of nuclear magnetic resonance and molecular dynamics motional parameters in RNA. *J Phys Chem B* **114**, 929-939.

75. Dethoff, E. A., Hansen, A. L., Zhang, Q., and Al-Hashimi, H. M. (2010) Variable helix elongation as a tool to modulate RNA alignment and motional couplings. *J Magn Reson* **202**, 117-121.

76. Zhang, Q., and Al-Hashimi, H. M. (2009) Domain-elongation NMR spectroscopy yields new insights into RNA dynamics and adaptive recognition. *RNA* **15**, 1941-1948.

CHAPTER 3

Domain Elongation of a Single Stranded RNA

3.1 Introduction

The last several years have brought to light the importance of structure to biomolecular function, and the role sequence plays in modulating the structure/function relationship. Throughout the course of numerous studies on this relationship, it is now being increasingly realized that dynamics are critical to function[1-5]. Dynamics are particularly important in the functioning of noncoding RNA, which, unlike proteins, have a high degree of structural plasticity. For example, riboswitches drastically alter their secondary structure upon binding cognate ligands[6-8].

However, RNA structure and dynamics studies have largely focused on helical regions, with particular emphasis on inter-helical dynamics. Single stranded RNA (ssRNA) serves many diverse functions in cellular processes, detailed in section 1.1.3. Low-resolution circular dichroism and UV/melting studies have hinted that ssRNA can form helical conformations, and that the degree of helicity was more prevalent in polyadenine oligomers[9-15]. However, the general consensus in the field has been that ssRNA is highly disordered, with little to no structural characteristics.

Our NMR studies on the 12 nucleotide (nt) ssRNA derived from the prequeuosine riboswitch showed that the adenine-rich ssRNA contains a high level of stacking interactions and an overall conformation resembling an A-form helix[16]. Concomitantly, the ssRNA has a high level of motions, particularly at the 3′ and 5′ terminal ends, with an estimated order parameter (S^2) of ~0.4. While these studies were among the first to show that biologically relevant ssRNA sequences have unique structural characteristics, the small size and subsequently short correlation time, combined with the high level of dynamics, resulted in motional coupling of internal and overall motions, preventing

complete structural and dynamic characterization. Moreover, studies of the isolated ssRNA neglect the potential effects of conjoined helices and there is great interest in understanding the structure and dynamic behavior of such helix-ssRNA junctions.

Motional coupling is a significant problem for many flexible RNA systems[17-19] and can be overcome through the use of domain elongation[20,21]. Domain elongation typically elongates one helix; however, since no initial helix is present in ssRNA, rather, a helix is appended to the 5′ end of the ssRNA. Since many functional ssRNAs occur as 3′ overhangs, the domain-elongated construct allows us to determine the role of adjacent helices on ssRNA structure and dynamics. Many biophysical studies have been carried out on systems containing 3′ overhangs[22,23]; however, little to no attention has focused on the ssRNA but rather how the ssRNA overhang impacts helical stability. The elongation, in addition to decoupling internal and overall motions, also provides a helix-ssRNA junction providing an opportunity to study the behavior of such junctions.

3.2 Materials and Methods

3.2.1 Sample preparation

The $^{13}C/^{15}N$ A/U-labeled, elongated single strand (E-SS) was prepared by *in vitro* transcription using T7 RNA polymerase as described previously[24]. The RNA was repeatedly exchanged into NMR buffer (25 mM NaCl, 15 mM Sodium Phosphate pH 6.4, 0.1 mM EDTA) using an Ultra-4 amicon (Millipore Corp.). The final RNA concentration was ~0.4 mM. The helical fragment (CUAC) was purchased from Integrated DNA Technologies as a lyophilized powder at natural abundance. To prepare the sample, the RNA was dissolved in ddH$_2$O and annealed at 95 °C for 5 minutes followed by multiple steps of buffer exchange into NMR buffer. The final RNA concentration was ~2 mM.

3.2.1 Resonance assignments

All NMR experiments were performed at 298 K, unless specified otherwise, on an Avance Bruker 600 MHz spectrometer equipped with a triple-resonance cryogenic (5 mm) probe. NMR spectra were analyzed using NMR Draw[25] and Sparky 3[26]. The ^{1}H, ^{13}C,

and ^{15}N resonances in E-SS were assigned using standard homonuclear and heteronuclear 2D experiments as well as a 'divide and conquer' strategy to assign helical resonances. The CUAC construct was assigned with the [1]H-^{1}H NOESY experiment using a mixing time of 250 ms (section 1.2.2). The NOESY data was supplemented with 2D HCN experiments to correlate nucleobase H6/H8 to ribose C1′H1′ through the shared N1/N9 atom.

3.2.3 Carbon spin relaxation

Longitudinal (R_1) and transverse (R_2) carbon relaxation data for the nucleobases (C2, C6, and C8) were obtained using 2D ^{13}C relaxation R_1 and $R_{1\rho}$ NMR experiments [27], where R_1 and $R_{1\rho}$ rates are determined by fitting peak intensities to a mono-exponential decay using in-house software. $R_{1\rho}$ rates were converted to R_2 using the following equation:

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta \tag{3.1}$$

where $\theta = \arctan(\dfrac{\omega_{SL}}{\Omega})$ and is the effective tilt angle of the spin lock field from the static field, ω_{SL} is the amplitude of the spin lock field in Hz, and Ω is the resonance offset from the spin lock carrier frequency.

Due to significant differences in the relaxation properties between the elongated helix and the ssRNA, relaxation delays were optimized to each domain. The relaxation parameters used are listed in Table 3.1. Relaxation delays were performed in an interleaved manner with alternating short and long relaxation delays. R_1 and R_2 rates are listed in Table 3.2.

	C2/C6/C8 delays (ms)	
	Helix	Single Strand
R_1	20, 160, 320, 400(x2), 480	20, 160, 320, 400(x2), 480
$R_{1\rho}$	4, 16, 40(x2)	4, 16, 40(x2), 60(x3), 80(x2)

Table 3.1: Relaxation delays used in R_1 and $R_{1\rho}$ experiments. Duplicate and triplicate measurements used for error estimation are indicated by (x2) or (x3), respectively. Relaxation delays were run in an interleaved manner.

Residue	moiety	R1 (Hz)			R2 (Hz)		
U-41	C6	1.566	±	0.101	57.81	±	9.61
A-40	C8	1.123	±	0.035	71.68	±	4.14
A-40	C2	1.106	±	0.039	109.80	±	18.91
U-26	C6	1.970	±	0.050	73.20	±	7.93
U-25	C6	1.743	±	0.018	58.63	±	1.63
A-8/A-30	C8	1.161	±	0.023	73.17	±	2.68
A-30	C2	1.246	±	0.034	99.48	±	19.41
U-19	C6	1.852	±	0.059	75.20	±	6.85
U-9	C6	1.696	±	0.072	63.27	±	11.85
A-8	C2	1.084	±	0.041	72.92	±	13.44
U26t	C6	2.373	±	0.023	67.64	±	1.81
A27	*C8*	*3.216*	*±*	*0.007*	*6.31*	*±*	*0.05*
A28	*C8*	*2.961*	*±*	*0.004*	*12.72*	*±*	*0.06*
A29	C8	2.840	±	0.005	22.87	±	0.12
A30	C8	2.847	±	0.004	25.03	±	0.13
A31	C8	2.952	±	0.003	22.78	±	0.11
A32	C8	3.019	±	0.005	19.37	±	0.10
C33	C6	3.802	±	0.108	23.63	±	1.70
U34	C6	3.873	±	0.005	11.84	±	0.04
A35	C8	3.487	±	0.007	7.77	±	0.04
A36	C8	3.581	±	0.009	7.42	±	0.05
A36	C2	3.387	±	0.005	7.80	±	0.03

Table 3.2: Nucleobase ^{13}C spin relaxation measurements of E-SS construct. Resonances with ambiguous assignments are italicized.

The measured R_1 and R_2 values were used to compute relative order parameters[28] using $S^2 = (2R_2-R_1)$[29], and normalized to yield a relative order parameter (S^2_{rel}) describing the relative degree of order within a molecule ranging from 0 to 1, where 0 and 1 represent minimum and maximum order, respectively. The S^2_{rel} values were normalized against the helical residues with the highest values: A-30 (C8), A-40 (C2) and U-19 (C6).

3.2.4 Measurement and order tensor analysis of RDCs

Base and sugar ^1H-^{13}C splittings were measured from the difference between the upfield and downfield components of the ^1H-^{13}C doublet along the ^1H component using the narrow transverse relaxation-optimized spectroscopy (TROSY) component in the ^{13}C dimension as implemented in 2D ^1H-^{13}C S^{3}CT-heteronuclear single quantum correlation

(HSQC) experiments[30]. The measured ^{2}H splitting was ~8 Hz in the presence of ~8 mg/mL Pf1 phage (Asla Biotech). Idealized A-form structures were constructed using Insight II (Molecular Simulations, Inc.) correcting the propeller twist angles from +15° to -15° using an in-house program, as previously described[31]. The measured RDCs are listed in Table 3.3. The experimental error was estimated to be ~3 Hz, determined by repeated measurements of the weaker helical resonances.

Residue	Bond	RDC (Hz)	Residue	Bond	RDC (Hz)
A-40	C8H8	22.2	U26	C1'H1'	-13.0
A-40	C2H2	36.4	U26	C5H5	26.0
A-30	C8H8	33.6	U26	C6H6	19.5
A-30	C2H2	42.1	*A27*	*C8H8*	*0.5*
U-26	*C5H5*	*4.9*	A28	C8H8	4.4
U-26	*C6H6*	*17.0*	A29	C8H8	7.8
U-26	C1'H1'	-36.9	A29	C1'H1'	-2.0
U-25	*C5H5*	*25.8*	A30	C8H8	24.6
U-25	*C6H6*	*23.0*	A30	C1'H1'	-4.2
U-25	*C1'H1'*	*30.9*	A31	C8H8	15.6
U-19	C6H6	27.0	A31	C1'H1'	-5.7
U-19	C1'H1'	-30.5	A32	C8H8	10.4
U-9	C5H5	29.5	A32	C1'H1'	-6.1
U-9	C6H6	24.1	U34	C6H6	3.7
U-9	C1'H1'	-30.5	U34	C5H5	2.9
A-8	C8H8	26.9	*A35*	*C8H8*	*2.6*
			A36	*C1'H1'*	*-0.5*
			A36	*C2H2*	*2.3*
			A36	*C8H8*	*2.2*

Table 3.3: RDCs measured in E-SS. Flexible residues, shown in italics, were excluded from order tensor analysis.

RDCs from the elongated helix and 3' single strand (Table 3.3) were independently subjected to order tensor analysis using idealized A-form helices as described in Chapter 1.2.3[31]. Previous NMR studies of the 12 nt ssRNA showed that the ssRNA can be modeled as an idealized helix for RDC order tensor analysis[16]. Briefly, the measured RDCs and idealized A-form helices were used to determine the best-fit order tensors for both helical and single stranded domains using singular value decomposition, implemented by the in-house written program RAMAH. Another in-house program,

Aform-RDC, was used to determine the order tensor errors due to inherent structural noise as well as RDC uncertainty[32]. The final RNA structure was assembled by rotating each domain into the principal axis system (PAS) of each best-fit order tensor and assembling the two helices. RDCs from tetraloop residue U-34 as well as the terminal end residues A35-A35 were excluded from analysis due to a high level of dynamics as observed in ^{13}C spin relaxation measurements. The inter-helical angles were calculated using an in-house program as previously described[31].

3.3 Results and Discussion

3.3.1 Elongated construct design and resonance assignments

Domain elongation provides the basis for decoupling internal and overall motions and the quantitative analysis of the dynamics of highly flexible nucleic acid systems using NMR spectroscopy[20,21,24,33]. The domain elongation strategy typically extends an existing helix in an RNA molecule such that the overall shape – and therefore overall motions- are dominated by the resulting cylindrical rod shape[20,21]. By elongating helices with either G-C or A-U base pairs, and using AU and GC isotopically labeled rNTPs, respectively, it is possible to elongate an RNA while rendering elongation residues "NMR invisible".

In the case of ssRNAs, one must append a helix to the single strand, requiring a somewhat different construct design. For our studies, we used the construct shown in Figure 3.1A, featuring a 22 base pair hairpin capped with a stable cUUCGg tetraloop. The construct is prepared with the use of only A and U isotopically labeled nucleotides. The proper choice of sequence within the helix allows for reporter residues while minimizing the potential for resonance overlap. Analysis of NMR spectra confirmed that the construct folds into the predicted secondary structure. In particular, 1D proton spectra of the imino region (Figure 3.1) revealed the expected number of ^{15}N-labeled and unlabeled Watson-Crick base pairs and that residues U26 and U34 remain single stranded and do not participate in base pairing interactions.

Previous studies showed that elongating terminal helices minimally affects the structure, dynamics and functional characteristics of the target RNA[20,24,33]. This issue is a

greater concern for flexible ssRNAs given the potential for greater perturbations. Comparison of 2D ^{1}H-^{13}C HSQC spectra of the 12 nt ssRNA derived from the prequeuosine riboswitch aptamer with (E-SS) and without (SS) the elongation helix revealed very good agreement. The largest discrepancies, A25, U26, A27, and A28, are observed for residues near the site of elongation. Otherwise, residues A29-A36 in the E-SS construct could be assigned based on spectral overlap and transferring assignments.

Perturbed resonances belonging to the helix-ssRNA junction (A25-A28) proved difficult to assign. We were not able to obtain reliable assignments for A25 while U26 nucleobase C6H6 and C5H5 resonances could only be assigned by a process of elimination as the sole unassigned uridine resonance. Base C6H6/C8H8 and sugar C1′H1′ moieties were assigned using 2D HCN NMR experiments. A25, A27 and A28 C8H8 chemical shifts were not significantly perturbed from the SS construct; however, several additional peaks appear in the vicinity of these peaks (marked by asterisks in Figure 3.1B). It is possible that the junction residues may slowly exchange between a stacked and unstacked state, leading to multiple resonances at these sites. These residues are included in Tables 3.2 and 3.3 in section 3.2 for visual comparison to proceeding residues, however it should be noted that these assignments remain ambiguous and were not used in the RDC order tensor analysis.
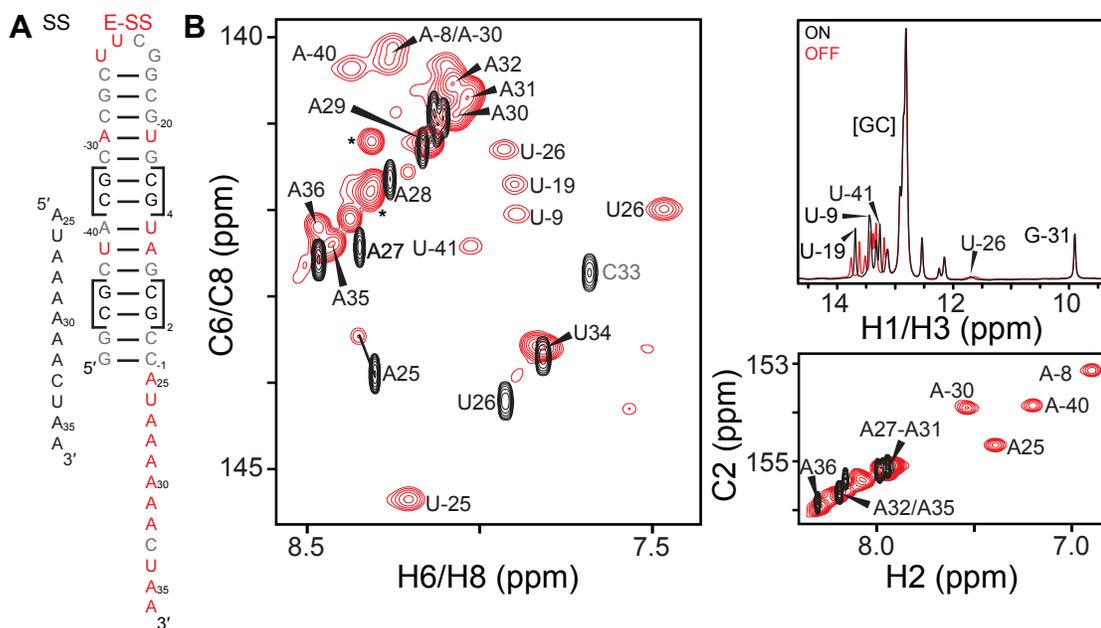
Figure 3.1: E-SS construct and chemical shift comparison with 12 nt SS construct. A) 12 nt SS and 60 nt E-SS construct. B)^{1}H-^{13}C HSQC of aromatic C6H6, C8H8, and C2H2 moieties show good agreement between SS and E-SS constructs. Top right: ^{15}N-decoupling on (black) or off (red) shows three base paired ^{15}N-labeled uridines and several base paired guanines, confirming proper secondary structure.

Assignment of resonances in the helix also proved difficult in part because of the large size of the elongated RNA but also because lack of uniform labeling rendered it difficult to establish NOE connectivity. We employed a 'divide and conquer' approach to obtain helix resonance assignments. Here, trinucleotide steps from the helix are compared to assigned spectra in other RNAs (Appendix 3). On this basis, we were able to assign A-8, U-9, U-19, A-30, and A-40. Few NMR studies have been performed on samples containing CUAC steps, resulting in a lack of available chemical shift assignments for comparison. A "CUAC" fragment, surrounded on both 5′ and 3′ end with two C-G base pairs and capped with the highly stable cUUCGg tetraloop was purchased and NMR experiments recorded at natural abundance. Resonance assignments were performed using the conventional [1]H-^{1}H NOESY experiment. When overlaid with the E-SS construct, excellent agreement was found between unassigned helical peaks and the CUAC fragment (Appendix 3).

As a further confirmation of the assignments; resonances belonging to the 22 base pair helix have weakened intensities, characteristic of a slowly-tumbling sample. In

contrast, single stranded residues are highly intense (Figure 3.1). The remaining unassigned resonances, which are also highly intense, are assumed to belong to single stranded residues. This assumption was tested by recording spectra at reduced temperatures. As the temperature decreased, these unassigned residues shifted upfield in a manner consistent with increased base stacking (Figure 3.2). In contrast, helical residues were unaffected by temperature changes. These extra peaks may be due to slow exchange processes occurring at the junction site, resulting in multiple peaks. Additionally, *in vitro* RNA transcription often yields N+1 products[34-37]: a small population of RNA molecules may have additional residues after A36, accounting for the additional intense peaks.

3.3.2 Stacking differences

Comparison of E-SS with the 12 nt SS sample showed that the 3′ terminal residues (U34-A36) overlaid well (within 0.05 ppm) suggesting the same highly flexible chemical environment. These data fits well with the previous comparison of terminal residues to NMPs, showing that the terminal ends are extremely disordered[16]. Although chemical shifts of the polyadenine tract have a similar overall pattern, A29-A32 in particular are shifted upfield relative to the SS construct, indicating increased base stacking. This trend is also observed in the C14C17U sample in the same NMR buffer conditions (Figure 2.2), suggesting increased base stacking of the polyadenine tract may occur by appending a helix. Interestingly, A30-32 chemical shifts agree with the 12 nt construct in the presence of 2 mM $MgCl_2$ and also follows the trajectory in a temperature titration (Figure 3.2). The presence of divalent cations such as Mg^{2+} has been shown to arrest dynamics and assist stacking interactions[38-42]. These observations suggest that the chemical shift differences are due to small changes in stacking of the polyadenine tract, possibly due to further stacking on the double helix.
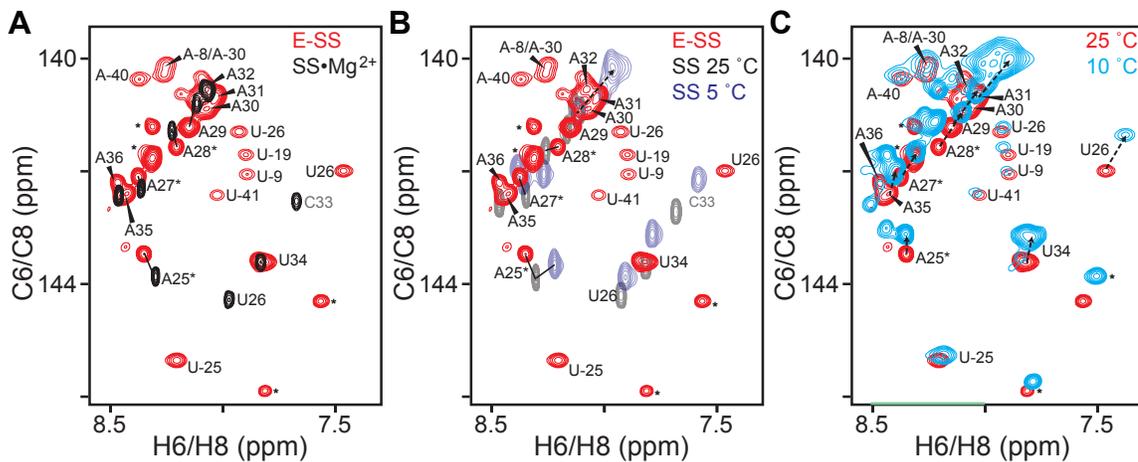
Figure 3.2: Influence of single stranded stacking energy differences on chemical shift perturbations. A) Comparison of E-SS chemical shifts in NMR buffer and SS in NMR buffer with an additional 2 mM Mg^{2+} show good agreement, suggesting similar base stacking energies; B) E-SS chemical shifts of the polyadenine tract fall along the trajectory of a temperature titration of SS, showing chemical shift perturbations are due to base stacking stability; C) E-SS at reduced temperatures shows increased single strand stabilization, evidenced by upfield chemical shift perturbations.

To further assess base stacking in the single strand, ^1H-^{13}C HSQC spectra were recorded as a function of temperature. As expected, helical chemical shifts, including the UUCG tetraloop, do not experience significant chemical shift perturbations signifying they are stable and are not further stabilized by a reduction in temperature. However, the single stranded residues experience upfield shifts up to ~0.15 ppm in both carbon and proton dimensions, suggesting a further increase in base stacking (Figure 3.2). At 10 °C the chemical shifts agree extremely well with the 12 nt construct at 5 °C, suggesting the helix may impart a small degree of stabilization to the single strand. It is important to note that the single strand likely undergoes a rapid transition between an ordered conformation and an highly disordered state, observed in previous RDC studies[16]. A system in fast exchange between two states will have an observed chemical shift frequency that is a population-weighted average of the frequencies of the two states[43]. Therefore the slight chemical shift changes observed as a function of temperature, ionic strength, or by appending a helix are likely due to stabilization of stacked/ordered state and changes in the relative percentages of ordered and disordered conformations.

3.3.3 ^{13}C Spin relaxation to determine ps-ns motions

To characterize the pico- to nanosecond dynamic behavior of the ssRNA, we measured R_1 and R_2 ^{13}C spin relaxation rates targeting base carbon spins on four constructs: 1) the wild-type (WT) prequeuosine aptamer, which predominantly forms a 72 nt dimer; 2) the double mutant C14U/C17U prequeuosine aptamer, which prevents the aptamer's ability to form a dimer, resulting in a 36 nt construct; 3) the 60 nt E-SS construct; and 4) the 12 nt ssRNA construct. The magnitude of R_2/R_1 values depends on the size of the biomolecule as well as internal motions. As shown in Figure 3.3A, the residue nearest to the appended helix has the largest R_2/R_1 value for all constructs except E-SS. The magnitude of each R_2/R_1 value matches the size of each construct (WT > E-SS > C14U/C17U > SS), as expected. However, each construct experiences a gradual decrease in R_2/R_1 values indicating increased dynamics as residues move away from the helix-ssRNA junction. The R_2/R_1 values converge for all samples at U34-A36 indicating that after ~10 residues, single strands attached to a helix have the same properties as an isolated single strand. The number of residues required to achieve this state are likely sequence dependent and coupled to the intrinsic persistence length of the ssRNA.

Interestingly, the E-SS construct has reduced R_2/R_1 values for A28 and A29, with A30 having the maximum R_2/R_1 value. This observation may reflect extensive motions occurring at the junction site. Differences in the apparent stability of the junction site between WT, C14U/C17U, and E-SS samples may be due to the helix stability. Both WT and C14U/C17U constructs interconvert between a five base paired and a seven base paired secondary structure. In sharp contrast, E-SS consists of a highly stable GC-rich 22 base pair helix. The high level of disorder in WT and C14U/C17U constructs may also contribute to motional coupling, convoluting relaxation analysis. In contrast, the domain elongation in the E-SS construct, by removing motional coupling, is able to reveal internal motions at the junction site.
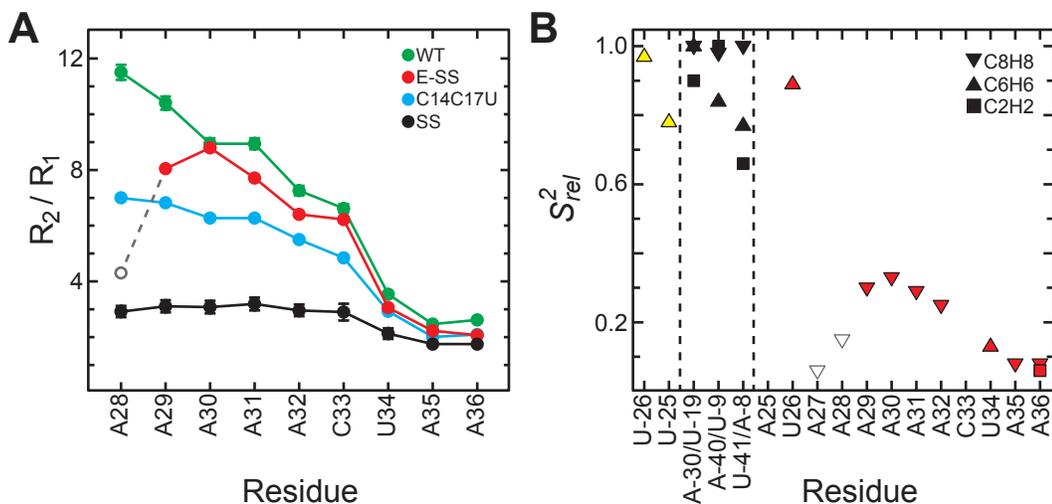
Figure 3.3: [13]C spin relaxation of E-SS. A) R_2/R_1 values show similar ssRNA behavior with different helices. B) Relative S^2 (S^2_{rel}) parameters for E-SS construct.

E-SS R_1 and R_2 relaxation rates were used to calculate $2R_2-R_1$, which is a good approximation of the order parameter S^2, describing motions occurring faster than the overall molecular tumbling[27,44]. Relative S^2 (S^2_{rel}) values were calculated by normalizing values to that of the largest value measured in the stable A-form helix (Figure 3.3B)[27]. Each bond vector type was individually normalized. S^2_{rel} reports on the amplitude of internal motions on the ps-ns timescale and ranges from 0 to 1, with maximum and minimum motions, respectively. While U26 has similar order parameters to the elongated helix, a sharp drop is observed for the polyadenine tract residues from ~0.9 to ~0.3 for residues A29-A32, which were shown in previous studies to be the most ordered[16]. As residues approach the 3′ terminus S^2_{rel} values gradually reduce, matching the previous observation for the SS construct. Interestingly, the tentatively assigned A27 and A28 have dramatically reduced S^2_{rel} values, suggesting extensive dynamics on the scale of the 3′ terminus. However, S^2_{rel} interpretation in elongated samples can be complicated by anisotropic tumbling caused by the elongation. Spin relaxation data for anisotropic rotational diffusion is strongly dependent on the orientation of the principal axis of overall rotational diffusion (D_{zz}) relative to the dipolar and CSA tensors[27]. It is for this reason that a small S^2_{rel} does not necessarily correspond to internal motions since a unique static structure can also contribute to attenuation.

3.3.4 Residual dipolar coupling measurements

       [13]C spin relaxation parameters report on ps-ns motions whereas residual dipolar couplings (RDCs) access all sub-ms motions. RDCs report on the time-averaged orientation of a bond vector with respect to the magnetic external field, described in detail in section 1.2.3[17,45-47]. Furthermore, RDCs provide additional dynamics and orientation information that can resolve the apparent discrepancy in the low S^2_{rel} values of A27 and A28. In an elongated system, the principal axis of alignment (S_{zz}) can be assumed to be identical to D_{zz} because alignments in both cases are determined by the global shape of the RNA, which is dominated by the long helix. It is assumed that the 22 base pair helix is aligned parallel to the external magnetic field (Chapter 1.2.3).

       RDCs were measured for base C5H5, C6H6, C8H8, C2H2, and ribose C1′H1′ moieties (Figure 3.4A). Since the elongated helix will dominate alignment with its long helix aligned parallel to the field, nucleobases will be aligned perpendicular with respect to the magnetic field and as such represent, to a good approximation, minimum values for a rigid bond vector. Residues with nucleobase RDCs smaller than those of helical RDCs can be interpreted as having local dynamics, with the level of attenuation proportional to the level of dynamics present. Similar to the trend observed with the ^{13}C spin relaxation values, the RDC magnitude of U26 approaches those of helical residues. Interestingly, A29 also has a relatively large RDC value compared to other single stranded residues. Additionally, when the long axis of a helix is parallel to the magnetic field, ribose C1′H1′ bond vectors will generally be negative in sign. This is observed in the helical residues as well as the single stranded residues. The same trend observed in spin relaxation data is observed where a gradual reduction in RDC magnitude is observed moving toward terminal end residues, with terminal residues having 0-2 Hz RDC values indicating extensive motions. This observation is in good agreement with previous studies described in Chapter 2 on the 12 nt ssRNA, showing that terminal end residues are extremely flexible.
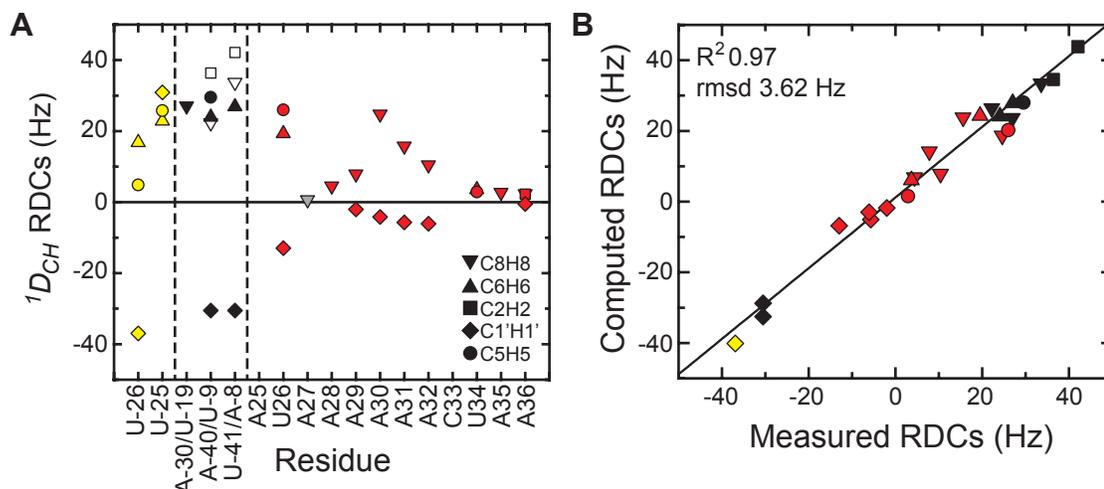
Figure 3.4: E-SS RDCs and order tensor analysis. A) $^1D_{CH}$ RDCs as a function of secondary structure. 5′ strand residues depicted as open symbols. B) Back-calculated RDC values computed from cross-validation analysis compared to measured RDCs shows excellent agreement.

To perform the order tensor analysis to determine the orientation of the single strand with respect to the reference helix, singular value decomposition analysis was performed independently for each domain (helix or single stranded) using an idealized A-form helix for each domain[31,48]. Each domain was rotated into its respective experimentally determined order tensor. The order tensor was validated with a 'leave one out' analysis and computing the RDC for the absent bond vector, which showed excellent agreement (Figure 3.4B).

The rotated domains were then assembled to bring A25 to the backbone of the terminal helical C-1 residue. The order tensor has inherent 4^{n-1} degeneracy, where n is equal to the number of domains. The degeneracy, conserving the chirality of the molecule, is 180° about each axis ($S_{xx}+180°$, $S_{yy}+180°$, $S_{zz}+180°$), resulting in four potential orientations of the single stranded domain with respect to the reference helix (Figure 3.5). $S_{yy}+180°$ and $S_{xx}+180°$ were ruled out due to steric clashes with the reference helix (Figure 3.5). However, degeneracy about the S_{zz} axis remains undefined, which is an existing problem in RDC analysis for rotations about the RNA long helical axis[31]. Conformation A in Figure 3.5 is the more likely conformation due to retention of coaxial stacking on the reference helix, whereas conformation B has no stacking interactions between the single strand and reference helix and involves a sharp bend

about the backbone at the intersection between the two domains. Further order tensor analysis was performed on conformation A.
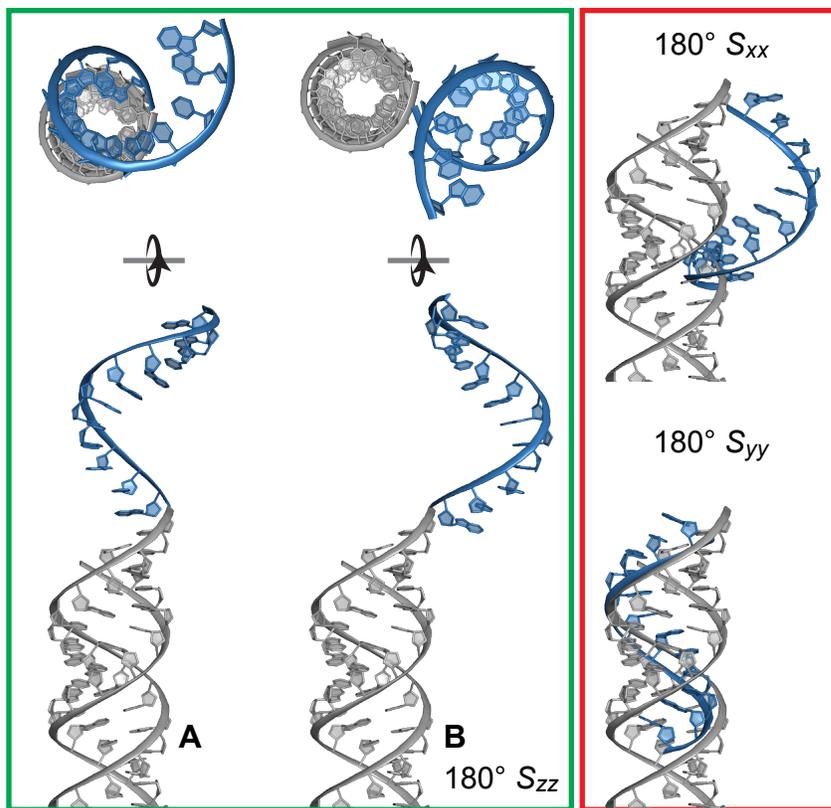
Figure 3.5: Potential structures resulting from RDC order tensor analysis. 180° rotations about S_{xx} and S_{yy} are excluded due to steric clashes with helix. Either S_{zz} orientation (conformation A or B) are possible, although conformation B is less likely due to loss of stacking interactions and a sharp kink about the backbone.

To further assess the accuracy of the conformation resulting from the order tensor analysis, we compared the alignment tensor determined experimentally against that predicted by PALES[49,50]. We found very good agreement for conformation A (S_{zz} deviation ~1.7°) with slightly poorer agreement for conformation B (~ 5.6° deviation). RDCs computed from the PALES-defined alignment tensor fit well with experimentally measured RDCs (R^2=0.83, Figure 3.6A).
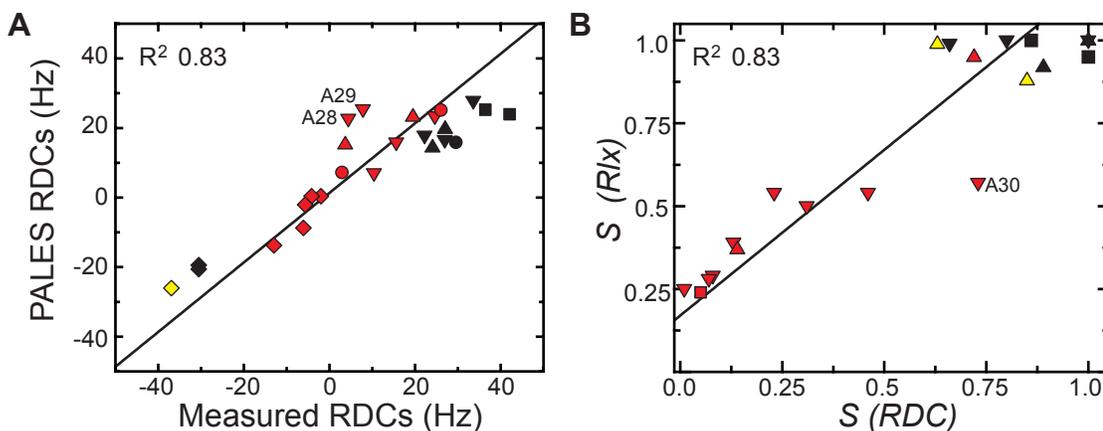
Figure 3.6: E-SS RDCs and order tensor analysis. A) Computed RDCs from PALES-defined order tensor compared to measured RDCs. B) Comparison of normalized ^{13}C spin relaxation order parameters and RDC order parameters show good agreement. Both S RDC and S Rlx were normalized to the same helical residues.

Analogous to the order parameters determined from spin relaxation measurements, RDCs also report on the degree of order. By normalizing RDC values to the same helical residues used in the spin relaxation analysis, an "*S RDC*" was determined and compared to the square square root of the S^2_{rel} determined from spin relaxation measurements to yield an "*S Rlx*". We find good agreement in relative order parameters between RDCs and spin relaxation measurements (R^2=0.83), suggesting that the motions reported on by both methods belong to the same process (Figure 3.6B). However, relaxation order parameters are consistently higher in magnitude than RDC order parameters, suggesting that the RDCs capture additional motions that are inaccessible by spin relaxation. Since spin relaxation tends to report local motions whereas RDCs report global as well as local motions, the differences observed may be due to large amplitude motions of the ssRNA.

The order tensor consists of five parameters (Chapter 1.2.3)[17,45]. The relative degree of order (ϑ_{int}) between the two domains is 0.51, which is similar to the three-nt bulge HIV-1 TAR system[17,20,24]. ϑ_{int} is also proportional to the order parameter S (corresponding to a global S^2 of 0.26) and agrees well to the observed S^2_{rel} of ~0.3 for the single stranded domain. The observed order parameters for the single strand are similar to earlier estimates of ~0.4 for the degree of order, determined from spin relaxation and MD simulations[16].

Domain	N	CN	RMSD (Hz)	R^2	η	$\vartheta \times 10^{-3}$	ϑ_{int}	θ	ξ
Helix	11	4.37	1.5	0.99	0.15 ± 0.04	1.66 ± 0.07	0.51	15.8	-0.81
SS	14	2.46	2.9	0.95	0.67 ± 0.13	0.85 ± 0.12	± 0.07	± 3.1	

Table 3.4: Summary of order tensor parameters

The interhelical bend (θ) and twist (ξ) angles were computed for conformation A using an in-house program. The interhelical bend was found to be ~16°, which is quite small considering the degree of allowed space that can be sampled by the single strand. The relatively small bend angle suggests significant coaxial stacking with the reference helix, explaining the relatively high order parameter and RDC value for junction residue U26. The interhelical twist is ~1 degree, which is also quite small. The small bend and twist angles but low ϑ_{int} provide a picture of the types of motions present in the single strand that can be combined with our previous knowledge of the system. On average the single strand is ordered and coaxially stacked upon the helix. However, high levels of motions are present, particularly at the helix-ssRNA junction, suggesting that the ssRNA alternates between an ordered A-form like conformation and a highly disordered state.

3.4 Conclusion

Domain elongation removed the motional coupling present in the 12 nt ssRNA construct, enabling a more quantitative description of the internal motions of the single strand. We find that the helix appears to impart a small degree of stacking stability to the single strand, particularly for the polyadenine tract. However, the single strand remains highly flexible with an average relative order parameter of ~0.3. RDCs show that, on average, the single strand coaxially stacks upon the reference helix with an interhelical bend of ~16°. These findings agree well with those from our previous NMR study, discussed in Chapter 2.

This single stranded sequence is located in the prequeuosine riboswitch as a 3′ overhang. On binding ligand, the single strand forms a sharp kink at the helix-junction site and forms a pseudoknot, where the polyadenine tract forms A-minor interactions to the helix and 3′ residues base pair to the apical loop[51-55]. A degree of flexibility at the

junction site may be necessary in order to efficiently bind ligand: if the single strand rigidly stacked upon the helix it would be unable to fold into the pseudoknotted structure. Concomitantly, if the single strand were infinitely flexible in a random-coil type conformation, the large amount of accessible conformational space may inhibit efficient ligand binding. A combination of a flexible junction with an ordered central region likely allows for competent binding in a timely manner, while maintaining structural plasticity to rapidly adopt the bound conformation in the presence of ligand.

The idea was conceived by Al-Hashimi, H.M. and Eichhorn, C.D. All sample preparation, NMR experiments, and data analysis were performed by Eichhorn, C.D.

3.5 References

1.   Rinnenthal, J., Buck, J., Ferner, J., Wacker, A., Furtig, B., and Schwalbe, H. (2011) Mapping the landscape of RNA dynamics with NMR spectroscopy. *Accounts Chem Res* **44**, 1292-1301.
2.   Stelzer, A. C., Kratz, J. D., Zhang, Q., and Al-Hashimi, H. M. (2010) RNA dynamics by design: biasing ensembles towards the ligand-bound state. *Angew Chem Int Ed Engl* **49**, 5731-5733.
3.   Bailor, M. H., Sun, X., and Al-Hashimi, H. M. (2009) Decoding RNA Dynamics and Adaptation. **In Preparation**.
4.   Al-Hashimi, H. M., and Walter, N. G. (2008) RNA dynamics: it is about time. *Curr Opin Struct Biol* **18**, 321-329.
5.   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.
6.   Henkin, T. M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* **22**, 3383-3390.
7.   Lai, E. C. (2003) RNA sensors and riboswitches: self-regulating messages. *Curr Biol* **13**, R285-291.
8.   Mandal, M., and Breaker, R. R. (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463.
9.   Fasman, G. D., Lindblow, C., and Grossman, L. (1964) The Helical Conformations of Polycytidylic Acid: Studies on the Forces Involved. *Biochemistry* **3**, 1015-1021.
10.   McDonald, C. C., Phillips, W. D., and Penman, S. (1964) Nucleic Acids: A Nuclear Magnetic Resonance Study. *Science* **144**, 1234-1237.
11.   Brahms, J., Michelson, A. M., and Van Holde, K. E. (1966) Adenylate oligomers in single- and double-strand conformation. *J Mol Biol* **15**, 467-488.
12.   Leng, M., and Felsenfeld, G. (1966) A study of polyadenylic acid at neutral pH. *J Mol Biol* **15**, 455-466.

13. BouchemalChibani, N., duPenhoat, C. H., Abdelkafi, M., Ghomi, M., and Turpin, P. Y. (1996) Characterization of the dynamic behavior of r(ACC) and r(AAC) with NMR relaxation data and both Metropolis Monte Carlo and molecular dynamics simulations. *Biopolymers* **39**, 549-571.

14. Seol, Y., Skinner, G. M., and Visscher, K. (2004) Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys Rev Lett* **93**, 118102.

15. Seol, Y., Skinner, G. M., Visscher, K., Buhot, A., and Halperin, A. (2007) Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys Rev Lett* **98**, 158103.

16. Eichhorn, C. D., Feng, J., Suddala, K. C., Walter, N. G., Brooks, C. L., 3rd, and Al-Hashimi, H. M. (2011) Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch. *Nucleic Acids Res*.

17. Getz, M., Sun, X., Casiano-Negroni, A., Zhang, Q., and Al-Hashimi, H. M. (2007) NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers* **86**, 384-402.

18. Getz, M. M., Andrews, A. J., Fierke, C. A., and Al-Hashimi, H. M. (2006) Structural plasticity and Mg2+ binding properties of RNase P P4 from combined analysis of NMR residual dipolar couplings and motionally decoupled spin relaxation. *RNA*.

19. Sun, X., Zhang, Q., and Al-Hashimi, H. M. (2007) Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition. *Nucleic Acids Res* **35**, 1698-1713.

20. 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.

21. Zhang, Q., and Al-Hashimi, H. M. (2009) Domain-elongation NMR spectroscopy yields new insights into RNA dynamics and adaptive recognition. *RNA* **15**, 1941-1948.

22. O'Toole, A. S., Miller, S., and Serra, M. J. (2005) Stability of 3' double nucleotide overhangs that model the 3' ends of siRNA. *RNA* **11**, 512-516.

23. O'Toole, A. S., Miller, S., Haines, N., Zink, M. C., and Serra, M. J. (2006) Comprehensive thermodynamic analysis of 3' double-nucleotide overhangs neighboring Watson-Crick terminal base pairs. *Nucleic Acids Res* **34**, 3338-3344.

24. 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.

25. 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. *Journal Of Biomolecular Nmr* **6**, 277-293.

26. Goddard, T. D., and Kneller, D. G. (2004) *SPARKY 3*, University of California, San Franscisco.

27. Hansen, A. L., and Al-Hashimi, H. M. (2007) Dynamics of large elongated RNA by NMR carbon relaxation. *Journal of the American Chemical Society* **129**, 16072-16082.

28. Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *Journal of the American Chemical Society* **104**, 4559-4570.

29. Dethoff, E. A., Hansen, A. L., Musselman, C., Watt, E. D., Andricioaei, I., and Al-Hashimi, H. M. (2008) Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys J* **95**, 3906-3915.

30. Meissner, A., and Sorensen, O. W. (1999) The role of coherence transfer efficiency in design of TROSY-type multidimensional NMR experiments. *J. Magn. Reson.* **139**, 439-442.

31. Bailor, M. H., Musselman, C., Hansen, A. L., Gulati, K., Patel, D. J., and Al-Hashimi, H. M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat Protoc* **2**, 1536-1546.

32. Musselman, C., Pitt, S. W., Gulati, K., Foster, L. L., Andricioaei, I., and Al-Hashimi, H. M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J Biomol NMR* **36**, 235-249.

33. Dethoff, E. A., Hansen, A. L., Zhang, Q., and Al-Hashimi, H. M. (2010) Variable helix elongation as a tool to modulate RNA alignment and motional couplings. *J Magn Reson* **202**, 117-121.

34. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**, 8783-8798.

35. Cazenave, C., and Uhlenbeck, O. C. (1994) RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc Natl Acad Sci U S A* **91**, 6972-6976.

36. Pleiss, J. A., Derrick, M. L., and Uhlenbeck, O. C. (1998) T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA* **4**, 1313-1317.

37. Helm, M., Brule, H., Giege, R., and Florentz, C. (1999) More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA* **5**, 618-621.

38. Casiano-Negroni, A., Sun, X., and Al-Hashimi, H. M. (2007) Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. *Biochemistry* **46**, 6525-6535.

39. Tan, Z. J., and Chen, S. J. (2006) Electrostatic free energy landscapes for nucleic acid helix assembly. *Nucleic Acids Res* **34**, 6629-6639.

40. Draper, D. E., Grilley, D., and Soto, A. M. (2005) Ions and RNA folding. *Annu Rev Biophys Biomol Struct* **34**, 221-243.

41. Kankia, B. I. (2003) Binding of Mg2+ to single-stranded polynucleotides: hydration and optical studies. *Biophys Chem* **104**, 643-654.

42. 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.

43. Cavanagh, J. (2007) *Protein NMR spectroscopy : principles and practice*, 2nd ed., Academic Press, Amsterdam ; Boston.

44. Fushman, D., Tjandra, N., and Cowburn, D. (1999) An approach to direct determination of protein dynamics from N-15 NMR relaxation at multiple fields, independent of variable N-15 chemical shift anisotropy and chemical exchange contributions. *Journal of the American Chemical Society* **121**, 8577-8582.

45. Tolman, J. R., and Al-Hashimi, H. M. (2003) NMR Studies of Biomolecular Dynamics and Structural Plasticity Using Residual Dipolar Couplings. *Annual Reports on NMR Spectroscopy* **51**, 105-166.

46. Lipsitz, R. S., and Tjandra, N. (2004) Residual dipolar couplings in NMR structure analysis. *Annu Rev Bioph Biom* **33**, 387-413.

47. Prestegard, J. H., Bougault, C. M., and Kishore, A. I. (2004) Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* **104**, 3519-3540.

48. Losonczi, J. A., Andrec, M., Fischer, M. W., and Prestegard, J. H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **138**, 334-342.

49. Zweckstetter, M., Hummer, G., and Bax, A. (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys J* **86**, 3444-3460.

50. Zweckstetter, M. (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc* **3**, 679-690.

51. 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.

52. 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.

53. 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.

54. 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.

55. Feng, J., Walter, N. G., and Brooks, C. L., III. (2011) Cooperative and directional folding of the preQ1 riboswitch aptamer domain. *Journal of the American Chemical Society* **133**, 4196-4199.

CHAPTER 4

Sequence Conservation and Role of the Single Stranded RNA in Prequeuosine Riboswitch Activity

4.1 Introduction

In the last decade, riboswitches have emerged as noncoding RNA elements that control gene expression[1-4]. Riboswitches are found in all three branches of life and are abundant in bacteria, controlling the expression of over 3% of the *B. subtilis* genome[3,5,6]. Riboswitches sense a diverse set of cellular cues ranging from pH, metabolites, temperature, and divalent cations, discussed in section 1.1.3. The majority of riboswitches control gene expression at either the transcription or translation levels. Riboswitches are modular in nature, consisting of a ligand-binding aptamer domain and a downstream expression platform that carries out the required function, discussed in section 1.1.3. These two domains are connected by a pliable RNA element that transduces ligand recognition to the expression platform through its ability to form base pairing interactions to either aptamer or expression platform.

Recent studies have highlighted important differences between transcription and translation-regulating riboswitches[7,8]. Riboswitches that regulate translation are located in fully formed messenger RNA (mRNA) transcripts. Ligand binding sequesters the Shine-Dalgarno region, preventing ribosome recognition and subsequent protein expression (section 1.1.3). Translation-regulating riboswitches have been shown to operate under thermodynamic equilibrium[7], provided that ligand binding occurs faster than the mRNA lifetime, which is typically on the minute timescale[9].

In contrast, riboswitches that regulate transcription must actively function during RNA transcription, harnessing co-transcriptional folding processes to modulate the secondary structure of the expression platform using a canonical intrinsic terminator

system. In the absence of ligand, the expression platform will form an "anti-terminator" hairpin adjacent to the aptamer domain (Figure 1.2), allowing production of the full-length mRNA transcript. However, if ligand is available to bind before the anti-terminator has fully formed, ligand binding will cause residues that would otherwise form the anti-terminator hairpin to be sequestered. Since the anti-terminator hairpin can no longer form, residues that would form part of the anti-terminator are free to form alternate secondary structures as the nascent RNA chain is elongated. The next available hairpin that can form is a canonical terminator hairpin, defined as a highly stable hairpin followed by a series of 3-8 uridines[10]. The combination of a helix at the RNA exit channel coupled with a weak DNA-RNA hybrid within the RNA polymerase (RNAP) results in premature dissociation from the RNAP and an incomplete transcript, which is then recycled by the cellular machinery.

Unlike translation-regulating riboswitches, which operate under equilibrium conditions, transcription-regulating riboswitches must function within the time constraints of transcription and, rather than minutes, have on the order of seconds to bind their ligand targets. Recent evidence has shown that transcription-regulating riboswitches operate under kinetic rather than thermodynamic equilibrium[11]. It is important to note that binding must occur prior to anti-terminator formation since the timescale of helix melting is much slower than ligand binding. Even if sufficient ligand is available, it is unlikely that the anti-terminator helix will melt and allow the terminator helix to form. This 'commitment time' is likely to be system-specific and can be attenuated through the evolution of polymerase pause sites to allow more time for ligand binding to occur[11]. Comparisons of the binding kinetics of translation and transcription regulating riboswitches have found that transcription riboswitches tend to bind their cognate ligand much faster than their translation counterpart[8,12].

The prequeuosine riboswitch contains the smallest aptamer discovered to date, with a minimal aptamer sequence of 34 nucleotides (nt), consisting of a short 4-6 base pair helix followed by a 11-13 nt single stranded RNA (ssRNA) element[13]. In *Bacillus subtilis*, the transcription-regulating *ykvJ queC* riboswitch has ~20 nt between the minimal aptamer sequence and full formation of the anti-terminator helix. The estimated rate of RNAP transcription is 25-50 nt/s[14-16], giving the *queC* riboswitch ~0.4-1 second to

bind ligand. This short commitment time, coupled with a presumably disordered single stranded element capable of accessing a large conformational space, poses a fundamental question as to how this riboswitch is able to operate under kinetic control.

Previous NMR studies, detailed in Chapter 2, showed that rather than intrinsically disordered, the adenine-rich single stranded element has residual structure resembling an A-form helix[17]. Mutating an adenine within the six-adenine tract to a cytosine reduced the stability of the single strand. A hypothesis was raised from these conclusions that pre-ordering of the single strand was necessary for efficient ligand binding within the time constraints of transcription. In order to test this hypothesis the sequence conservation of the single stranded element between transcription and translation mechanisms was compared to identify if operating under thermodynamic or kinetic control influences the single stranded sequence. An *in vitro* assay was designed to determine whether the *queC* riboswitch acts under kinetic control as well as to test whether the A29C mutation affects riboswitch activity.

4.2 Materials and Methods

4.2.1 Sequence conservation algorithm

The starting dataset was obtained from the Rfam database[18-20] and contained a total of 415 putative prequeuosine riboswitches from 285 species. This dataset only contains the minimal aptamer sequence. To obtain the entire riboswitch sequence, including the protein start site, a program (pullfullsequence.pl) was written to remotely access the NCBI nucleotide database. The sequence 50 nucleotides before the aptamer start position and 400 nucleotides after the aptamer end position was retrieved using the accession number and aptamer position in the genome, provided by the Rfam dataset. The NCBI protein database was then searched for proteins near putative aptamers: sequences with no protein start site identified within 400 nt of the aptamer were labeled 'no protein' and excluded from further analysis.

Sequences adjacent to protein coding regions were classified as Type I or Type II using previously established definitions[13] using the in-house program analyze_sequence.pl. Sequences were then classified as transcription or translation

regulating using the following criteria. Aptamer sequences located within 20 nt of the protein start codon were classified as regulating translation. The remaining sequences were assessed for the presence of terminator and anti-terminator hairpins. Intrinsic terminator systems are defined as the presence of two mutually exclusive hairpins, where the terminator hairpin is located downstream of the anti-terminator hairpin and is more thermodynamically stable with a 3′ polyuridine overhang of three or more uridines. The following criteria were used to identify the intrinsic terminator motif. First, secondary structures of the minimal aptamer were determined using mFold[21]. MFold was then performed on the 40 nt region following the aptamer hairpin, starting at A28 of the ssRNA, to obtain free energies for the putative anti-terminator hairpin. Ten nt were then added to this sequence and mFold was performed again to obtain a putative terminator hairpin. Secondary structures containing a 3′ uridine overhang of at least three nt were selected. The free energies of each terminator/anti-terminator system were then assessed: if the terminator hairpin had a lower free energy than the anti-terminator hairpin and contained >3 uridines as a single stranded 3′ overhang, the sequence was classified as regulating transcription. A total of 76 Type I transcription riboswitches and 108 Type I translation-regulating riboswitches were identified (Appendix 5). A total of three Type II translation and seven Type II transcription regulating riboswitches were identified. Eight transcription regulating riboswitches were identified with no Type; two riboswitches were identified with no type or apparent regulatory mechanism; and nine Type I riboswitches were identified with no apparent regulatory mechanism. All programs can be found in Appendix 4.

4.2.2 *E. coli in vitro* transcription assay

Lyophilized PAGE-purified DNA template (Integrated DNA Technologies) was suspended in ddH$_2$O to 25 μM with 3 mM MgCl$_2$. The DNA template was annealed at 95 °C for 5 minutes, followed by incubation on ice for one hour. PreQ$_1$ was donated by Prof. George Garcia (University of Michigan) and suspended as a 10 mM stock in 10 mM NH$_4$OH followed by serial dilution to desired concentrations using ddH$_2$O. 7-deazaguanine (Sigma) was suspended identically to preQ$_1$. *In vitro* transcription reactions

were performed using the following buffer conditions: 40 mM Tris (pH 7.5-8), 10 mM MgCl$_2$, 50 mM NaCl, 0.01% Triton-X, 2 mM DTT, 100 or 600 µM rNTPs (as specified), 0.625 µM DNA template, varying PreQ$_1$ concentrations, and 1.5 U *E. coli* RNA polymerase holoenzme (Epicentre Biotechnologies) in a 20 µL reaction volume. The transcription reaction was allowed to proceed for five hours at either 37 °C or 25 °C, as specified. The reaction was quenched with denaturing formamide loading dye supplemented with 10 ng/µL 29 nt RNA as a loading standard. The quenched reaction was heat-denatured at 95 °C for 2-3 minutes and immediately loaded onto a 10% denaturing polyacrylamide (PAGE) gel. The gel was run at 16 W for 1.5 hours using 1X TBE as running buffer. The gel was then stained with Sybr Gold (Invitrogen) for 45 minutes, per the manufacturer-recommended protocol, and visualized using a UV transilluminator attached to a camera setup.

4.2.3 *In vitro* transcription assay data analysis

Gel images were analyzed using Adobe Photoshop CS5. To perform baseline subtraction for each lane, normalized band intensities (N. intensity) were determined using the following equation: $Intensity = Band - (\dfrac{U+L}{2})$, where U and L are the background intensities above and below the band, respectively. Normalized band intensities were fit to a sigmoid curve using Origin 7.0 using the following equation: $y = \dfrac{A1 - A2}{1 - e^{(x-x_0)/dx}}$, where y is the normalized band intensity, A1 and A2 are the lower and upper baselines, respectively, x is the log[Ligand], and x_0 is the 50% transcript termination (T$_{50}$) value.

4.3 Results and Discussion

4.3.1 Sequence conservation differences between translation and transcription prequeuosine riboswitches

It has been discovered in other riboswitch systems that small sequence differences far from the ligand-binding pocket can affect ligand binding affinity[8,22]. Although the *B.*

93

subtilis queC riboswitch has an adenine-rich single stranded tail, this sequence does not necessarily represent the consensus sequence. Additionally, the consensus sequence may differ between translation and transcription regulating riboswitches since they are thought to be controlled by different mechanisms. However, the literature reveals a lack of sequence conservation information available for the prequeuosine riboswitch: the seminal paper by Breaker and coworkers lists only 36 prequeuosine riboswitches[13], only one of which has a known mechanism. A further layer of complexity is that prequeuosine riboswitches fall under either Type I or Type II classifications, which have slightly different sequence conservation, occurring predominantly in the apical loop[13].

To gain a larger sample size to better assess the sequence conservation of transcription and translation riboswitches, we constructed a program to classify a given sequence as Type I or Type II and also to identify as transcription or translation regulating, using a key set of criteria, detailed in section 4.2.1. The dataset was obtained from the Rfam database[19], containing a total of 415 putative prequeuosine riboswitch aptamers. It is important to note that the Rfam-derived dataset was generated through sequence homology, with only one sequence confirmed to function *in vivo*. It is therefore possible that the Rfam database could have a biased dataset and may miss prequeuosine riboswitches with unique sequences. A set of stringent criteria was

Figure 4.1: Sequence conservation differences between transcription and translation regulating Type I prequeuosine riboswitches.

developed, described in section 4.2.1, to select likely riboswitch candidates. Since riboswitches control gene expression, only aptamer sequences adjacent to protein-coding regions were selected. The riboswitch type was assigned using the respective Type I and Type II consensus sequences defined by Breaker and coworkers[13]. To discern between a transcription or translation regulatory mechanism, sequences were assessed for the intrinsic terminator motif, described in detail in section 4.2.1.

These studies revealed that the Rfam database is highly skewed in favor of Type I riboswitches (91% Type I vs. 5% Type II). Due to the lack of Type II riboswitches we continued our analysis of the single stranded tail on Type I prequeuosine riboswitches to remain specific to a given type. It was observed that transcription-regulating riboswitches have a high level of sequence conservation in general (Figure 4.1 and Appendix 5). Full conservation is observed for a six-adenine repeat, followed by CU, which forms the pseudoknot to the apical loop in the ligand-bound conformation. The helix-ssRNA junction site always contains at least one pyrimidine, typically a uridine, within 2 nucleotides of the helix. The single stranded element length is typically 12 nt, although occasionally an additional adenine is present at the junction site (Figure 4.1).

In contrast, translation riboswitches have marked differences in the consensus sequence, particularly at the junction site. Full conservation is observed for a four-adenine repeat followed by CY, where Y is a pyrimidine. Six-adenine tracts are commonly found; however, several cases have pyrimidine-rich sequences (Figure 4.1 and Appendix 5) permeating up to 4 residues into the single-stranded repeat. While the length of transcription-regulating riboswitches is highly conserved at 12 nt, the length of translating riboswitches tends to be 11 nt long but can be up to 13 nt. In some species a small helix is found between the single strand and the P1 helix, previously identified by Breaker and coworkers[13].

While sequence conservation differences exist between transcription and translation, the two retain an adenine-rich sequence, which is required for A-minor 'kissing' interactions between the P1 helix and ssRNA[23-26]. However, the differences in the length and helix-ssRNA junction sequence might directly relate to the helical pre-ordering of polyadenine tracts in ssRNA. It should be noted that since the Rfam database has skewed Type I/Type II populations, comprehensive sequence conservation information may be incomplete. However, the sequence conservation patterns agree well with our previous observations as well as the existing literature, leading us to propose that performing the A29C mutation, previously shown to destabilize the single strand, should reduce the function of the transcription regulating riboswitch.

4.3.2 *In vitro* transcription assay

E. coli polymerase assay design

We designed an *in vitro* assay to monitor the transcription efficiency of the prequeuosine riboswitch. Although the *queC* prequeuosine riboswitch is located in *B. subtilis*, *B. subtilis* RNAP availability is limited. *E coli* RNA polymerase is commercially available and has been shown to have similar transcription rates to *B. subtilis* RNAP[11,14]. This assay exploits the transcription termination mechanism, yielding a difference in RNA product lengths as a function of ligand, which can then be analyzed on a denaturing PAGE gel. During transcription, in the absence of ligand, the full-length 106 nt RNA transcript will be generated (Figure 4.2). However, if preQ$_1$ is present, tertiary interactions between the single strand and aptamer apical loop prevent anti-terminator formation and the resultant terminator hairpin generates a truncated 75 nt transcript (Figure 4.2).
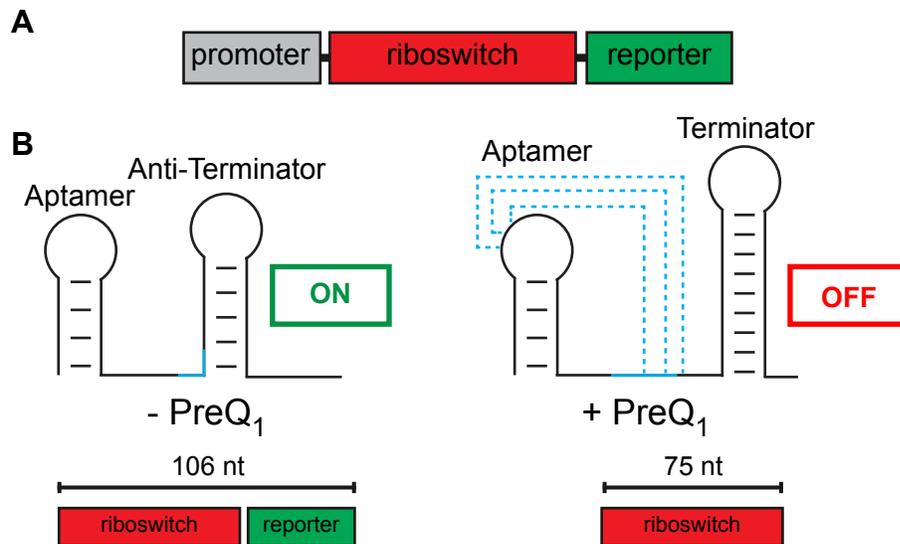
Figure 4.2: Schematic describing the prequeuosine riboswitch mechanism. A) DNA template consists of the recA promoter (gray), the riboswitch sequence (red), and a reporter sequence (green). B) *Left*: Absence of ligand will result in downstream anti-terminator hairpin formation, allowing gene expression to proceed, resulting in a 106 nt RNA product; *Right*: presence of ligand results in tertiary interactions to single strand (depicted as blue dashed lines), preventing anti-terminator hairpin formation. The downstream terminator hairpin will form, causing polymerase dissociation and a truncated 75 nt RNA product.

Upon performing transcription reactions with increasing ligand concentration, we observed increasing amounts of the truncated transcript, denoted as T, on a denaturing PAGE gel, consistent with increasing transcription termination as a function of ligand

concentration. The ligand concentration dependence of T exhibited the expected titration curve (Figure 4.3). The data was fit to obtain a T_{50}, defined as the ligand concentration needed for 50% termination (Figure 4.3). At 37 °C the wild-type (WT) construct appears to be saturated at ~0.5 mM, with an apparent T_{50} of 35.6 μM. This value is three orders of magnitude weaker than the apparent K_d value of 20 nM for the aptamer[13]. If the prequeuosine riboswitch acts under thermodynamic control the T_{50} value should match the K_d value. The T_{50} value is in agreement with previous *in vitro* transcription assays performed on other riboswitch systems, although slightly reduced,[11,27] and supports the hypothesis that the prequeuosine riboswitch acts as a kinetic rather than thermodynamic switch. It should be noted that the presence of ammonium hydroxide, necessary to solubilize the preQ1 ligand, may perturb the activity of inherent RNAP function. Control reactions performed in the presence of the maximum concentration (2.5 mM) of ammonium hydroxide resulted in band intensity values close to the zero ligand point, indicating that the changes in band intensity observed are specific to preQ1 concentration.

Figure 4.3: Example PAGE gel and titration curve.

4.3.3 Ligand specificity further suggests binding kinetics are important for transcription efficiency

The specificity of the prequeuosine riboswitch aptamer to preQ1 over a variety of near-cognate small molecules was previously determined by Breaker and coworkers using in-line probing experiments[13]. However, the ligand specificity during transcription is unknown. To test the ligand-specific transcription efficiency the termination efficiency of the near-cognate 7-deazaguanine ligand was tested. 7-deazaguanine lacks the aminomethyl substituent at the N7 position, and has been shown to have a ~100 fold

reduction in the binding affinity. We observed a 10-fold reduction in termination efficiency compared to preQ$_1$, which is smaller than might be expected given the 100-fold difference in K$_d$ (Figure 4.3).
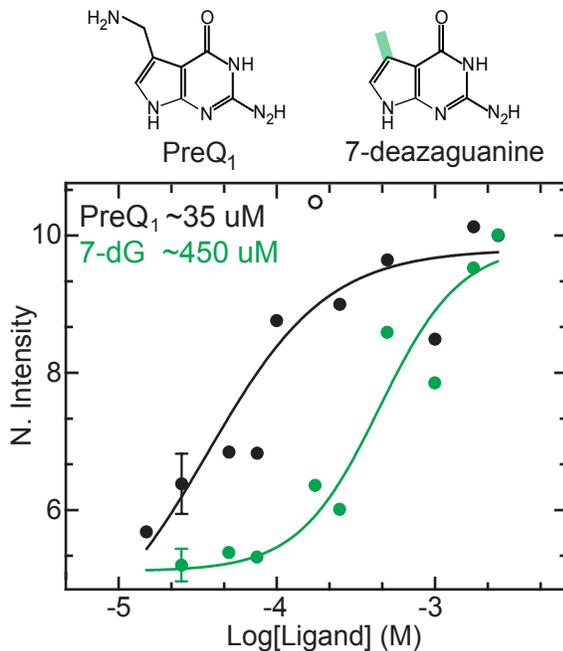
Figure 4.4: WT prequeuosine riboswitch termination efficiency as a function of ligand identity. PreQ$_1$ titration curve shown in black, 7-deazaguanine (7-dG) titration curve shown in green.

One potential explanation is that the k$_{on}$ rate constants for both ligands may be similar whereas k$_{off}$ rate constants differ significantly, leading to the apparent reduction in binding affinity. In such a situation, since the riboswitch commitment time is short compared to the time for ligand dissociation (estimated at ~6 seconds for preQ$_1$[28]), as soon as the ligand-binding event occurs the terminator helix forms, terminating transcription. As a result, the k$_{off}$ rate constant is not important so long as the k$_{on}$ rate constant is sufficiently faster than the commitment time and the k$_{off}$ rate. Alternately, the k$_{on}$ rate constant could be 10-fold slower for 7-deazaguanine compared to preQ$_1$ combined with a 10-fold faster observed k$_{off}$ rate constant. These differences could amount to a 100-fold K$_d$ difference, yet only the k$_{on}$ rate constant difference will be reflected in the transcription efficiency.

4.3.4 A29C mutation effect and temperature dependence

Our previous NMR studies in Chapter 2 showed that the A29C mutation in the aptamer single stranded region destabilizes the ordering of the single strand, resulting in a ~5 °C reduction in melting temperature. Single molecule FRET experiments estimate that the A29C mutation reduces the apparent K_d by 100-fold to ~5 µM[29]. To test the impact of this mutation on the transcription efficiency of the prequeuosine riboswitch, the *in vitro* transcription reaction was performed at 37 °C on a DNA template containing the A29C mutation. In stark contrast to the clear increase of band intensity as a function of ligand, the A29C mutant exhibited little to no termination products, suggesting the near complete loss of responsiveness (Figure 4.5).
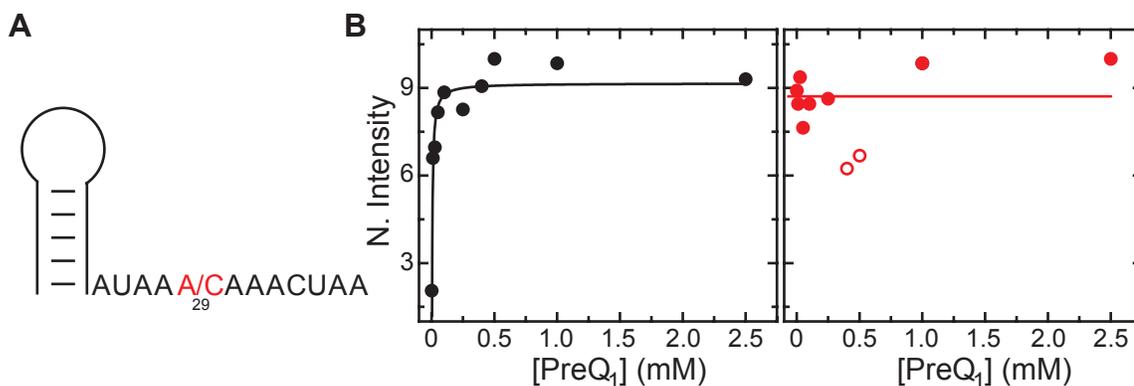
Figure 4.5: A29C mutation reduces prequeuosine riboswitch responsiveness. A) *queC* minimal aptamer single stranded sequence. A29C mutation is highlighted in red. B) *In vitro* assay termination product intensities as a function of ligand show that A29C mutation (right, red) significantly reduces termination efficiency.

If the T_{50} value is 100-fold higher than the WT value, corresponding to ~3.5 mM, a change in termination products should be observable at ligand concentrations used, although ligand saturation would not be reached (maximum preQ$_1$ concentration was 2.5 mM). It is possible that the A29C construct is able to weakly respond to ligand but is outside the sensitivity range of this assay. These data suggest that the A29C mutation reduces the k_{on} rate constant of ligand binding such that it is significantly longer in comparison to the commitment time. Indeed, if we assume that the 100-fold drop in K_d will reflect in a 100-fold reduction in k_{on} rate constants, we might expect a k_{on} of ~6×10^3 $M^{-1}s^{-1}$ [28], minimally requiring 0.17 mM preQ$_1$ to operate within the ~1 s commitment time. Although the concentration of free preQ$_1$ in the cell is unknown and is likely low

given preQ$_1$ is an intermediate in a biosynthetic pathway, other purine concentrations are on the order of 30 μM[30], suggesting that the A29C mutation would be nonfunctional *in vivo*. By destabilizing the ordering of the single stranded RNA tail, the aptamer has many more accessible conformational states that are nonproductive to binding, thereby increasing the entropic cost of binding. This can conceivably reduce the k_{on} rate constant, resulting in a weaker apparent K_d. This reduction in the transcription efficiency is likely due to the inability of the aptamer to bind preQ$_1$ rapidly compared to the short commitment time.

Additionally, the difference in binding affinity alone cannot explain the discrepancy in transcription termination behavior. The apparent binding affinities of the A29C mutation to PreQ$_1$ and the WT construct to 7-dG are similar (5 μM[29] and 2 μM[13], respectively). If the riboswitch operated under thermodynamic equilibrium, it may be expected that similar T_{50} values would be observed for the two systems. While ligand responsiveness, albeit 10-fold weaker than WT, is observed for 7-dG, no function is observed for the A29C construct (Figure 4.5B). This observation further supports the hypothesis that a) the prequeuosine riboswitch operates under kinetic control and b) the destabilization of the ssRNA tail by the A29C mutation reduces the activity of the prequeuosine riboswitch. These observations support the hypothesis that the A29C mutation reduces the rate of the formation of the bound conformation.
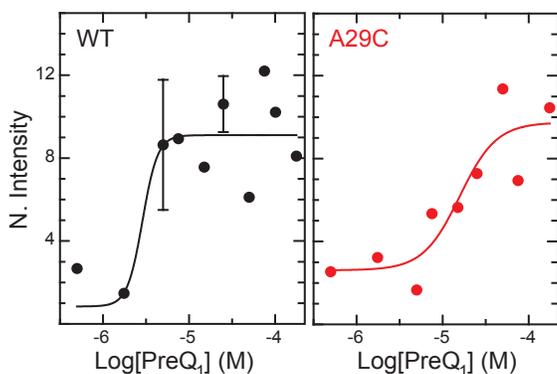
Figure 4.6: Termination efficiency of WT (left) and A29C (right) constructs at 25 °C.

The loss of function of the A29C mutant construct hints at the importance of single stranded RNA structure in the prequeuosine riboswitch efficiency. However, if the system functions under kinetic control, by altering the kinetics of the system the termination efficiency should also be affected. Previous *in vitro* transcription studies have shown that transcription termination efficiency is improved by reducing the rate of RNAP transcription[11]. The RNAP transcription rate is commonly attenuated by changing rNTP concentration or by changing the temperature of the reaction. First, the temperature of the transcription

reaction was reduced from 37 °C to 25 °C. Altering the temperature alone did not appear to significantly improve A29C function in preliminary studies and did not have a significant effect on WT T_{50} values (Table 4.1). Reducing the rNTP concentration results in a longer nucleotide addition step, reducing the overall rate of transcription. Reducing the temperature and available rNTP concentration improves the WT T_{50} by one order of magnitude, supporting the hypothesis that slowing the rate of transcription improves the termination efficiency and that the prequeuosine riboswitch is a kinetic rather than thermodynamic switch.

By reducing both the temperature and the rNTP concentration, the A29C construct became functional (Figure 4.6). The A29C mutant has a four-fold reduction in T_{50} values over WT at 25 °C, suggesting that although the rate of transcription is slow enough to allow riboswitch activity in both constructs, A29C remains a suboptimal construct compared to WT.

4.4 Conclusions

The sequence conservation data agrees with our current understanding of the prequeuosine riboswitch. Using the domain elongation approach in Chapter 3 it was shown that the helix-ssRNA junction site is highly flexible, which is likely due to weakened stacking of the uridine at the junction. Translation-regulating riboswitches tend to have more variable sequences at the helix-ssRNA junction site, suggesting that differences in the junction dynamics may contribute to binding rate differences. The difference in the lengths of conserved adenine tracts between transcription and translation riboswitches suggests that the additional adenines in the transcription riboswitch may confer additional stacking stability, as seen in Chapter 2 when comparing the structural characteristics of the A29C mutant, which has a three-adenine tract, to the WT construct, which has a six-adenine tract. Stabilizing the ssRNA residues that are important for ligand binding while retaining a flexible helix-ssRNA junction may increase the k_{on} rate constant.

Table 4.1 highlights all results and conditions studied. In all cases the 50% termination activity is much weaker than the apparent binding affinity of the aptamer for ligand, suggesting a kinetic mechanism. However, reducing the rate of transcription has a

pronounced effect on the T_{50} value, and the transcription assay identifies ligand selectivity. The observed low μM T_{50} values are in agreement with other riboswitch values, although slightly reduced[11,27]. It should be pointed out that experimentally determined K_d values were obtained in different buffer conditions and at 20-25 °C and may not be equivalent to K_d values at transcription temperatures of 37 °C.

Surprisingly, we find that the magnitude in binding affinity differences do not translate to the magnitude in transcription efficiency differences, suggesting that differences in the k_{on} rate constant may have a much greater importance than the k_{off} rate constant for transcription termination efficiency. Kinetic measurements by Wedekind and coworkers on the related TTE translation switch, containing a 4 nt adenine stretch with an inserted C, reveal ~8-fold slower k_{on} rate constants compared to a transcription-regulating prequeuosine riboswitch. This observation agrees well with our observation that a disruption of the six-nucleotide polyadenine tract impacts the kinetics of termination efficiency. Interestingly, the k_{off} rate constant for the transcription riboswitch was 1000-fold faster than the translation regulating riboswitch, further suggesting that a fast k_{on} rate constant is critical in transcription riboswitch activity compared to the k_{off} rate constant[12].

Construct	T (°C)	[rNTP] (μM)	Ligand	T_{50} (μM)			K_d (μM)
WT	37	600	PreQ$_1$	35.6	±	2.7	0.02
WT	25	600	PreQ$_1$	43.6	±	5.1	0.02
WT	25	100	PreQ$_1$	3.84	±	0.4	0.02
A29C	37	600	PreQ$_1$	*na*	±	*na*	5
A29C	25	100	PreQ$_1$	15.8	±	0.7	5
WT	37	600	7-dG	451.4	±	24.6	2

Table 4.1: Summary of *in vitro* transcription results.

The sequence-structure relationship has been a source of avid study in double stranded systems for many years. However, the same relationship is poorly understood for ssRNA. This study is one of the first to show a biological function for a structured, yet single-stranded RNA, and how this ordering fundamentally affects the activity of a kinetically controlled system.

Additional experiments may be performed to validate and expand upon these observations. In order to gain a larger sampling pool for the sequence conservation studies, rather than use the limited Rfam dataset the existing program can be modified to *a priori* identify putative prequeuosine riboswitch sequences. Staining the gel with Sybr Gold, although more sensitive than ethidium bromide, is relatively insensitive compared to radiometric assays. Labeling RNA transcripts with $5'$-$^{32}PO_4$ or using ^{32}P-labeled rNTPs would dramatically improve sensitivity. Rather than use *E. coli* RNAP, more physiologically relevant results can be obtained by using B. subtilis RNAP. Kinetics measurements determining the k_{on} and k_{off} rate constants would quantitatively establish the source of the binding affinity differences between WT and A29C to preQ$_1$.

4.5 References

1. Winkler, W. C., and Breaker, R. R. (2003) Genetic control by metabolite-binding riboswitches. *Chembiochem* **4**, 1024-1032.
2. Nudler, E. (2006) Flipping riboswitches. *Cell* **126**, 19-22.
3. Henkin, T. M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* **22**, 3383-3390.
4. Montange, R. K., and Batey, R. T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys* **37**, 117-133.
5. Blount, K. F., and Breaker, R. R. (2006) Riboswitches as antibacterial drug targets. *Nat Biotechnol* **24**, 1558-1564.
6. Breaker, R. R. (2010) Riboswitches and the RNA World. *Cold Spring Harb Perspect Biol*.
7. 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.
8. Lemay, J. F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Masse, E., and Lafontaine, D. A. (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS genetics* **7**, e1001278.
9. Narsai, R., Howell, K. A., Millar, A. H., O'Toole, N., Small, I., and Whelan, J. (2007) Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *Plant Cell* **19**, 3418-3436.
10. Grundy, F. J., Moir, T. R., Haldeman, M. T., and Henkin, T. M. (2002) Sequence requirements for terminators and antiterminators in the T box transcription

antitermination system: disparity between conservation and functional requirements. *Nucleic Acids Res* **30**, 1646-1655.

11. 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. *Mol Cell* **18**, 49-60.

12. Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V., and Wedekind, J. E. (2011) Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *The Journal of biological chemistry* **286**, 24626-24637.

13. Roth, A., Winkler, W. C., Regulski, E. E., Lee, B. W., Lim, J., Jona, I., Barrick, J. E., Ritwik, A., Kim, J. N., Welz, R., Iwata-Reuyl, D., and Breaker, R. R. (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat Struct Mol Biol* **14**, 308-317.

14. Artsimovitch, I., Svetlov, V., Anthony, L., Burgess, R. R., and Landick, R. (2000) RNA polymerases from Bacillus subtilis and Escherichia coli differ in recognition of regulatory signals in vitro. *J Bacteriol* **182**, 6027-6035.

15. Tolic-Norrelykke, S. F., Engh, A. M., Landick, R., and Gelles, J. (2004) Diversity in the rates of transcript elongation by single RNA polymerase molecules. *The Journal of biological chemistry* **279**, 3292-3299.

16. Greive, S. J., Weitzel, S. E., Goodarzi, J. P., Main, L. J., Pasman, Z., and von Hippel, P. H. (2008) Monitoring RNA transcription in real time by using surface plasmon resonance. *Proc Natl Acad Sci U S A* **105**, 3315-3320.

17. Eichhorn, C. D., Feng, J., Suddala, K. C., Walter, N. G., Brooks, C. L., 3rd, and Al-Hashimi, H. M. (2011) Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch. *Nucleic Acids Res*.

18. Griffiths-Jones, S. (2005) Annotating non-coding RNAs with Rfam. *Curr Protoc Bioinformatics* **Chapter 12**, Unit 12 15.

19. Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., and Bateman, A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**, D121-124.

20. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003) Rfam: an RNA family database. *Nucleic Acids Res* **31**, 439-441.

21. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415.

22. Lu, C., Smith, A. M., Ding, F., Chowdhury, A., Henkin, T. M., and Ke, A. (2011) Variable sequences outside the SAM-binding core critically influence the conformational dynamics of the SAM-III/SMK box riboswitch. *J Mol Biol* **409**, 786-799.

23. 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.

24. 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.

25.   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.

26.   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.

27.   McDaniel, B. A., Grundy, F. J., Artsimovitch, I., and Henkin, T. M. (2003) Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci U S A* **100**, 3083-3088.

28.   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.

29.   Suddala, K. C., Rinaldi, A. J., Feng, J., Mustoe, A. M., Eichhorn, C. D., Al-Hashimi, H. M., Brooks III, C. L., and Walter, N. G. (submitted) Differential conformational selection and induced fit in structurally similar transcriptional and translational riboswitch molecules.

30.   Nygaard, P., and Saxild, H. H. (2005) The purine efflux pump PbuE in Bacillus subtilis modulates expression of the PurR and G-box (XptR) regulons by adjusting the purine base pool size. *J Bacteriol* **187**, 791-794.

CHAPTER 5

Conclusions and Future Directions

RNA dynamics are increasingly being realized to play critical cellular roles[1-3]. However, a great deal of focus has been placed on studying the inter-helical motions of relatively structured RNA elements[4-8]. As high-throughput genomic screening techniques improve and RNA transcriptomes and corresponding secondary structures are revealed[9,10], an array of complex structures is being discovered. One of the first major discoveries of large-scale structural elements was the secondary structure of the 9 kb HIV-1 RNA genome solved by Weeks and coworkers[11]. Upon visualizing the structure, it was immediately apparent that in between large structured domains were long single stranded RNAs (ssRNAs). As the field has moved toward larger systems more and more ssRNA regions are being uncovered, although the biological relevance is unclear at present. However, ssRNAs are already well known to have important and diverse roles as scaffolds for proteins and RNA-RNA interactions[12-17]. Despite the biological significance of these RNA elements, ssRNA structural studies are severely lacking, largely due to an assumption that they are intrinsically disordered and highly flexible.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique that reports at atomic resolution on structural and dynamic information over a wide range of biologically relevant timescales[2,18]. In this dissertation, by combining NMR, UV spectroscopy, molecular dynamics (MD) simulations, and biological assays a ssRNA element located in the *Bacillus subtilis queC* riboswitch has been characterized and found to play an important structural role in carrying out riboswitch function.

5.1 Structural Characteristics of Single Stranded RNA

The majority of ssRNA structural studies have been performed using low-resolution techniques such as UV/Vis and circular dichroism spectroscopy[19-27]. While

providing insights into pH effects on ssRNA and general conformational behavior, sequence dependence studies and quantitative structural characteristics were lacking. In Chapter 2, we developed an approach to quantitatively characterize structural and dynamic aspects of a ssRNA at atomic resolution using a combination of NMR ^{1}H-^{1}H NOESY, H1′-H2′ scalar couplings, ^{13}C spin relaxation, residual dipolar couplings (RDCs), and MD simulations. Our results showed that rather than access a large conformational space, as might be expected for an unpaired oligomer, the 12 nucleotide (nt) ssRNA had an unexpectedly high degree of order at the six-nt adenine tract, with motions gradually increasing toward the terminal ends. Replica exchange MD simulations agreed with the trends observed by NMR, suggesting that the MD simulation accurately captured structural and dynamic characteristics within the ssRNA. While RDC data indicated that the ssRNA adopted, on average, an A-form-like conformation rather than a variety of potential conformations, ^{13}C spin relaxation parameters suggested the ssRNA remained highly flexible. The low level of cooperativity in the helix to coil melting behavior indicated that the ssRNA may be in equilibrium between an ordered, A-form-like conformation and a highly disordered state. However, the lack of a proper reference frame in addition to complications in data analysis arising from motional couplings as a result of the short correlation time and high level of dynamics required the application of a domain elongation strategy.

In Chapter 3 we expanded our studies of the ssRNA by performing domain elongation to gain more quantitative insights into the behavior of the ssRNA. Measuring ^{13}C spin relaxation rates and RDCs we found that, again, the single strand has a high degree of stacking at the central polyadenine tract. Surprisingly, the helix-ssRNA junction appeared to have a high level of motions, suggesting that 3′ overhang nucleotides do not simply stack upon their conjoined helix but have interesting dynamic characteristics. We found that the 12 nt ssRNA appended to the helix on average coaxially stacks upon the helix, with an average interhelical bend angle of ~16°. Similar to our previous study outlined in Chapter 2, ^{13}C spin relaxation parameters showed a high level of flexibility, with an average S^2_{rel} of ~0.3, and the degree of order gradually decreasing toward terminal ends. The ϑ_{int} determined from RDC order tensor analysis

was 0.51, agreeing well with spin relaxation data indicating the ssRNA is highly flexible, despite the high degree of stacking and average helical conformation.

Chapters 2 and 3 outline a strategy to study other biologically relevant ssRNA systems. The level of ordering within a ssRNA is likely to be highly sequence-dependent: adenines are well-characterized as being the most favorable at stacking while polyuridine oligomers are known to have minimal structural characteristics. Future studies to gain insights into the length and sequence dependency on ssRNA order can provide a framework toward understanding structural characteristics of biologically interesting ssRNA systems.

5.2 ssRNA has Sequence-Specific Dynamics

A number of low-resolution studies previously showed that polyadenine tracts have a high propensity to stack[19,26,28]. In Chapter 2, we disrupted the six-nt adenine tract by mutating an adenine to a cytosine to determine the sequence-dependent effects on ssRNA structural characteristics. We found that, as expected, the A29C mutation disrupted stacking about the mutation site. However, the presence of NOE crosspeaks between adenines surrounding C29 suggested the cytosine residue was looping out to allow the two adenine residues to stack. Although the melting temperature, determined using UV/Vis measurements, was reduced ~5 °C, the A29C mutation does not appear to significantly change the ssRNA structural or dynamic characteristics. Rather, ^{13}C spin relaxation measurements suggest that, in general, the A29C ssRNA has similar characteristics to the wild-type ssRNA, with the exception of residues about the mutation site as well as several residues 3′ with respect to the mutation site, which appear to have chemical exchange. This chemical exchange is likely stacking/unstacking motions occurring at us-ms timescales. Furthermore, RDC measurements suggest an A-form-like conformation, as observed in the wild-type sequence. However, unlike the wild-type ssRNA, the alignment tensor could not be reproduced by alignment prediction software, indicating that the molecule may exhibit bending motions.

These studies indicate that ssRNA is capable of complex dynamics approaching those of RNA helices. By placing a poor-stacking cytosine residue between two adenine tracts, the ssRNA appears to exhibit bending motions about the mutation site. This can be

compared to a helix with a one-nucleotide bulge that has a localized pivot point about the bulge. This finding has many exciting implications for ssRNA playing important roles in directing the 3D architecture of large RNA systems, rather than playing a passive role as a highly flexible element. Future studies employing a domain elongation strategy will be useful to gain insights into the directionality and extent of these bending motions.

5.3 ssRNA Dynamics Serve Functional Roles

The ssRNA studied in Chapters 2 and 3 is located in the transcription-regulating prequeuosine riboswitch aptamer. As discussed in Chapter 1, riboswitches that regulate transcription are increasingly being realized to operate under kinetic rather than thermodynamic control[29-33]. The kinetics of ligand recognition are therefore extremely important to biological function. Our studies suggested that the structural characteristics of the ssRNA were important in pre-ordering the ssRNA to allow the aptamer to more efficiently bind ligand. The A29C mutation indicated a reduction in stacking stability as well as dynamic bending motions, which were hypothesized to inhibit riboswitch function.

In Chapter 4, an *in vitro* transcription assay was developed to first determine whether the prequeuosine riboswitch functioned under kinetic control. We found that the transcription termination efficiency (T_{50}) values were 1000-fold weaker than the binding affinity, suggesting that this riboswitch is kinetically rather than thermodynamically controlled. We found that the A29C mutation dramatically reduced riboswitch function to undetectable levels at 37 °C, indicating that the mutation reduces the ligand binding rate, leading to the reduction in activity. By slowing the rate of transcription the function of the A29C mutant construct was rescued and found to have similar T_{50} values to the wild-type construct. Additionally, by reducing the rate of transcription the termination efficiency of the wild-type construct improves 10-fold, further supporting the hypothesis that the prequeuosine riboswitch is a kinetic switch.

Together, these studies show that ssRNA is capable of complex structural and dynamic characteristics and that ssRNA structure can serve important functional roles. We are hopeful that our findings will prompt future studies that will shed additional insights into the sequence dependency of these characteristics, the types of complex

motions taking place with ssRNA, and discover a multitude of important ssRNA structural roles.

5.4 References

1.   Haller, A., Souliere, M. F., and Micura, R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Accounts Chem Res* **44**, 1339-1348.
2.   Rinnenthal, J., Buck, J., Ferner, J., Wacker, A., Furtig, B., and Schwalbe, H. (2011) Mapping the landscape of RNA dynamics with NMR spectroscopy. *Accounts Chem Res* **44**, 1292-1301.
3.   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.
4.   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.
5.   Casiano-Negroni, A., Sun, X., and Al-Hashimi, H. M. (2007) Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. *Biochemistry* **46**, 6525-6535.
6.   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.
7.   Bailor, M. H., Sun, X., and Al-Hashimi, H. M. (2010) Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**, 202-206.
8.   Bailor, M. H., Mustoe, A. M., Brooks, C. L., 3rd, and Al-Hashimi, H. M. (2011) 3D maps of RNA interhelical junctions. *Nat Protoc* **6**, 1536-1545.
9.   Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aidinis, V., Allen, J. E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L., Bansal, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crowe, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S. P., Kruger, A., Kummerfeld, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V.,

Pang, K. C., Pavan, W. J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J. F., Ring, B. Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S. L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C. A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S. L., Tang, S., Taylor, M. S., Tegner, J., Teichmann, S. A., Ueda, H. R., van Nimwegen, E., Verardo, R., Wei, C. L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S. M., Teasdale, R. D., Liu, E. T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J. S., Hume, D. A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., and Hayashizaki, Y. (2005) The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-1563.

10. Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., Giddings, M. C., and Weeks, K. M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**, e96.

11. Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716.

12. Andreassi, C., and Riccio, A. (2009) To localize or not to localize: mRNA fate is in 3'UTR ends. *Trends Cell Biol* **19**, 465-474.

13. Regnier, P., and Hajnsdorf, E. (2009) Poly(A)-assisted RNA decay and modulators of RNA stability. *Prog Mol Biol Transl Sci* **85**, 137-185.

14. Arraiano, C. M., Andrade, J. M., Domingues, S., Guinote, I. B., Malecki, M., Matos, R. G., Moreira, R. N., Pobre, V., Reis, F. P., Saramago, M., Silva, I. J., and Viegas, S. C. (2010) The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev* **34**, 883-923.

15. Chan, S., Choi, E. A., and Shi, Y. (2011) Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev RNA* **2**, 321-335.

16. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* **264**, 968-980.

17. Zheng, M. X., Wu, M., and Tinoci, I. (2001) Formation of a GNRA tetraloop in P5abc can disrupt an interdomain interaction in the Tetrahymena group I ribozyme. *Proc Natl Acad Sci U S A* **98**, 3695-3700.

18. Bothe, J. R., Nikolova, E. N., Eichhorn, C. D., Chugh, J., Hansen, A. L., and Al-Hashimi, H. M. (2011) Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods* **8**, 919-931.

19. Seol, Y., Skinner, G. M., Visscher, K., Buhot, A., and Halperin, A. (2007) Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys Rev Lett* **98**, 158103.

20.	Seol, Y., Skinner, G. M., and Visscher, K. (2004) Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys Rev Lett* **93**, 118102.

21.	Kaukinen, U., Lyytikainen, S., Mikkola, S., and Lonnberg, H. (2002) The reactivity of phosphodiester bonds within linear single-stranded oligoribonucleotides is strongly dependent on the base sequence. *Nucleic Acids Res* **30**, 468-474.

22.	Hingerty, B. E., Broyde, S. B., and Olson, W. K. (1982) The poly(rU) coil: a minimum-energy model that matches experimental observations. *Biopolymers* **21**, 1167-1188.

23.	Freier, S. M., Hill, K. O., Dewey, T. G., Marky, L. A., Breslauer, K. J., and Turner, D. H. (1981) Solvent effects on the kinetics and thermodynamics of stacking in poly(cytidylic acid). *Biochemistry* **20**, 1419-1426.

24.	Dewey, T. G., and Turner, D. H. (1979) Laser temperature-jump study of stacking in adenylic acid polymers. *Biochemistry* **18**, 5757-5762.

25.	Michelson, A. M., Ulbricht, T. L., Emerson, T. R., and Swan, R. J. (1966) Optical rotatory dispersion of oligoadenylic acids and a consideration of the factors stabilizing helical polynucleotides. *Nature* **209**, 873-874.

26.	Leng, M., and Felsenfeld, G. (1966) A study of polyadenylic acid at neutral pH. *J Mol Biol* **15**, 455-466.

27.	Fasman, G. D., Lindblow, C., and Grossman, L. (1964) The Helical Conformations of Polycytidylic Acid: Studies on the Forces Involved. *Biochemistry* **3**, 1015-1021.

28.	Rich, A., Davies, D. R., Crick, F. H., and Watson, J. D. (1961) The molecular structure of polyadenylic acid. *J Mol Biol* **3**, 71-86.

29.	McDaniel, B. A., Grundy, F. J., Artsimovitch, I., and Henkin, T. M. (2003) Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci U S A* **100**, 3083-3088.

30.	Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G. M., and Breaker, R. R. (2005) Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem Biol* **12**, 1325-1335.

31.	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. *Mol Cell* **18**, 49-60.

32.	Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V., and Wedekind, J. E. (2011) Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *The Journal of biological chemistry* **286**, 24626-24637.

33.	Lemay, J. F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Masse, E., and Lafontaine, D. A. (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS genetics* **7**, e1001278.

Summary of WT queC ssRNA NMR Assignments, [13]C Spin Relaxation, and RDC Order Tensor Analysis
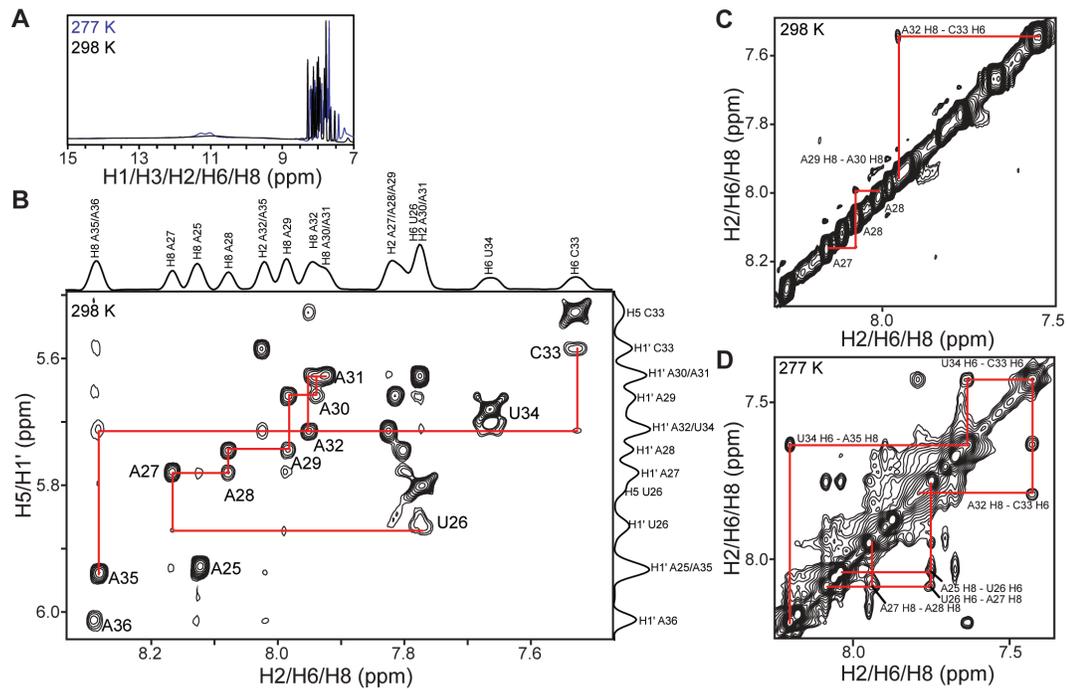
Figure A1.1: NMR resonance assignments of the 12 nt queC ssRNA construct. A) 1D imino proton spectrum shows no base pairing, even at 277 K, in which broad peak at ~11 ppm is characteristic of single stranded nucleic acids and peaks at 7-9 ppm are nucleobase H2/H6/H8 resonances for comparison, B) base to sugar NOE connectivities at 298 K enable a complete walk from U26-A35, C) base to base NOE connectivities in polyadenine tract at 298 K and D) at 277 K.
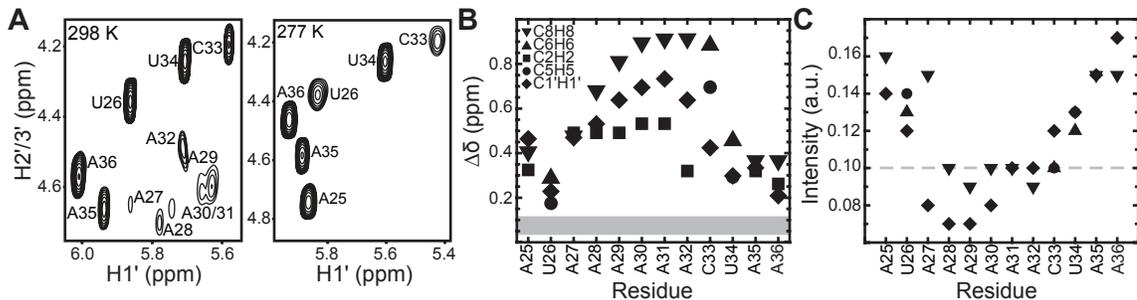
Figure A1.2: Summary of ssRNA sugar conformation behavior, comparison to free NMPs, and normalized peak intensities. A) homonuclear three bond scalar couplings ($^3J_{H1'-H2'}$) at 277K and 298K show C3'-endo sugar conformations for the central residues with significant C'-endo/C'-endo heterogeneity at the terminal ends, B) weighted-average chemical shift differences between NMPs and SS tail show terminal ends are more similar to NMPs than central polyadenine residues, with gray bar representing analogous chemical shift differences for polyU, C) Normalized peak intensities show terminal ends are more disordered than central residues (normalized to A31 and C33). Intensities were obtained from nonconstant time ^{1}H-^{13}C HSQC experiments at 298 K, where intensities for a given C-H bond vector are normalized relative to a value of 0.1 indicated by the dashed gray line.
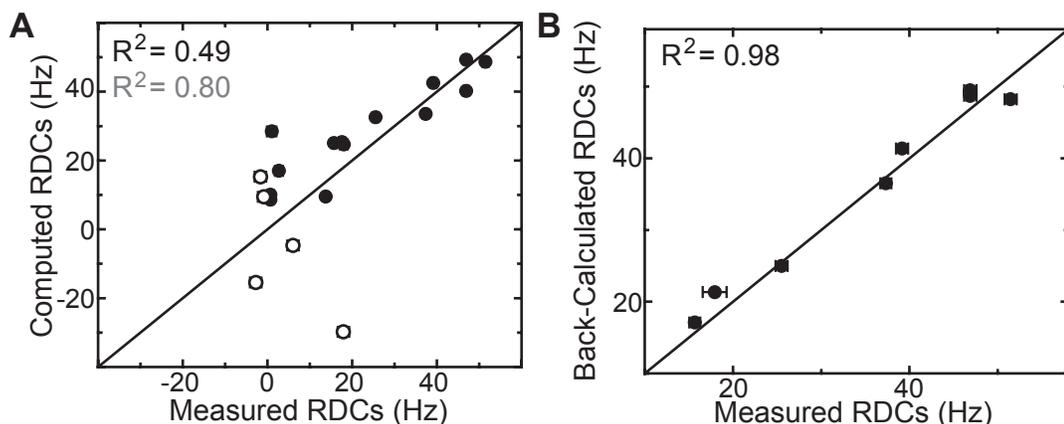
Figure A1.3: WT ssRNA order tensor validation. B) Comparison of computed and measured RDCs for all RDCs (R^2=0.49) and nucleobase-only RDCs (R^2=0.80), with C1'H1' as open circles, C) comparison of back-calculated RDCs from order tensor analysis with measured RDCs show excellent agreement.

Summary of A29C ssRNA NMR Assignments, ^{13}C Spin Relaxation, and RDC Order Tensor Analysis
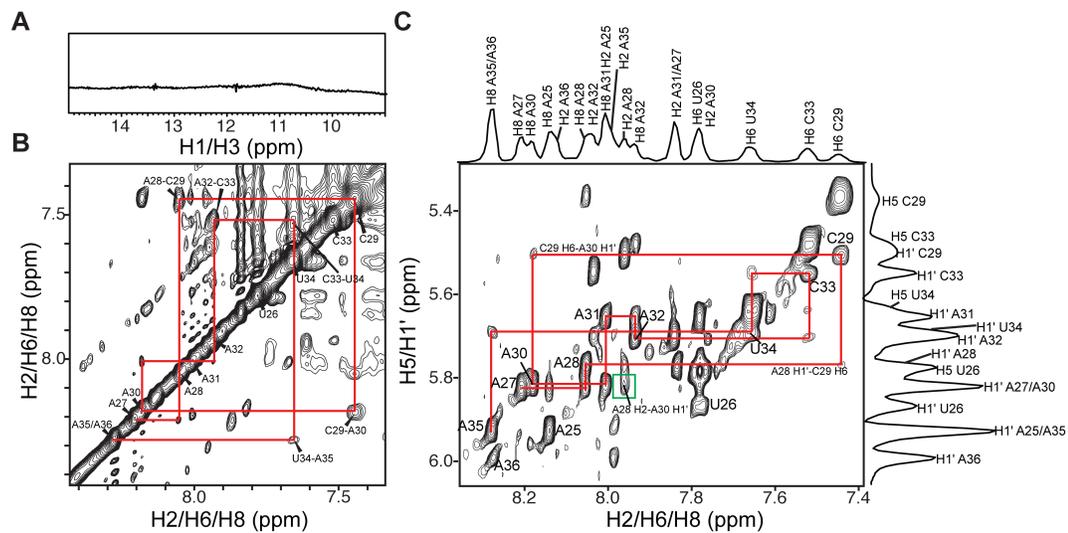
**Figure A2.1:** A29C NMR resonance assignments. A) 1D imino proton resonances show no base pairing, B) base to base NOE connectivities show base stacking in A29C, C) sequential NOE walk for A29C mutant, showing weak crosspeaks between A28-C29-A30 and a crosspeak between A28 H2 – A30 H1′ (boxed in green).
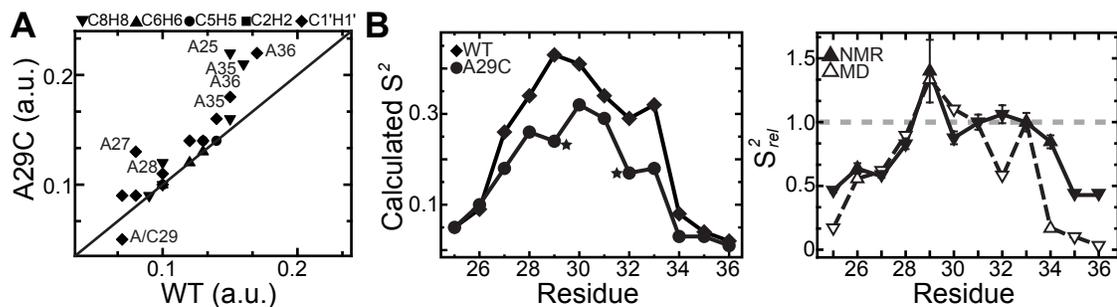
Figure A2.2: Comparison of WT and A29C peak intensities and order parameters. A) Comparison of normalized peak intensities for A29C and WT show similarities, with increased intensities at A27 and A28 indicating increased ps-ns motions and decreased intensity at the sugar

of the 29 position indicating chemical exchange, B) *Left*: Calculated WT S^2 values are increased relative to A29C in the polyadenine core, indicating reduced global ordering. Starred residues (C29 and A32) are significantly reduced, indicating increased dynamics, *Right*: S^2_{rel} values show good agreement between NMR (closed) and MD simulations (open). Although seemingly in disagreement, measured A32 S^2_{rel} value of >1 indicates chemical exchange whereas calculated A32 S^2_{rel} value of <1 indicates increased dynamics.
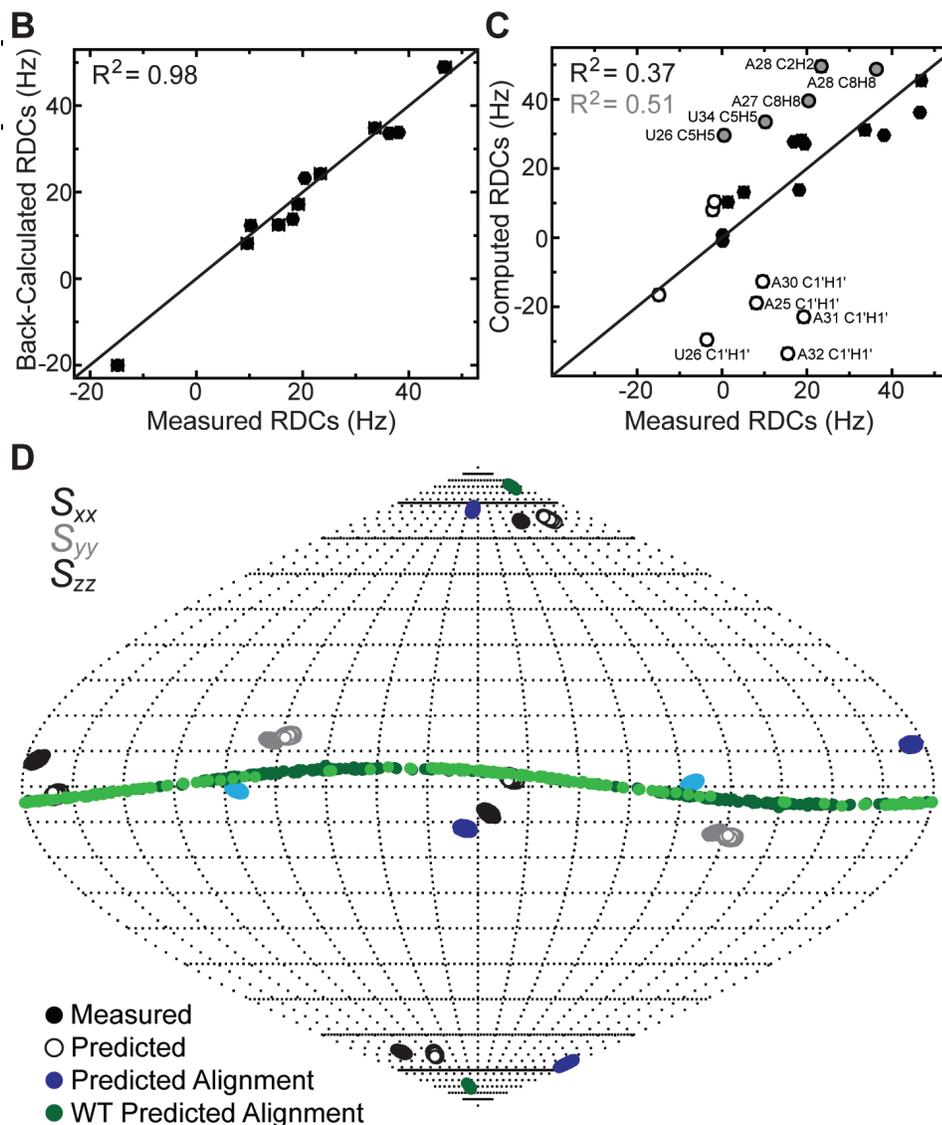
Figure A2.3: A29C order tensor analysis. B) *Left*: comparison of back-calculated RDCs from order tensor analysis with measured RDCs show excellent agreement, *Right*: computed RDCs vs measured RDCs show reasonable agreement (R^2=0.37 for all RDCs, R^2=0.51 excluding C1′H1′), with C1′H1′ as open circles and nucleobase outliers in gray, D) Sauson-Flamsteed map showing deviation of measured (black) and PALES-predicted (open) S_{xx}, S_{yy}, and S_{zz} tensors from the helical axis. PALES predicts the alignment frame for WT (green, deviation ~6°) more accurately than A29C (blue, deviation ~11°), suggesting A29C may be kinked.

Comparison of CUAC with E-SS
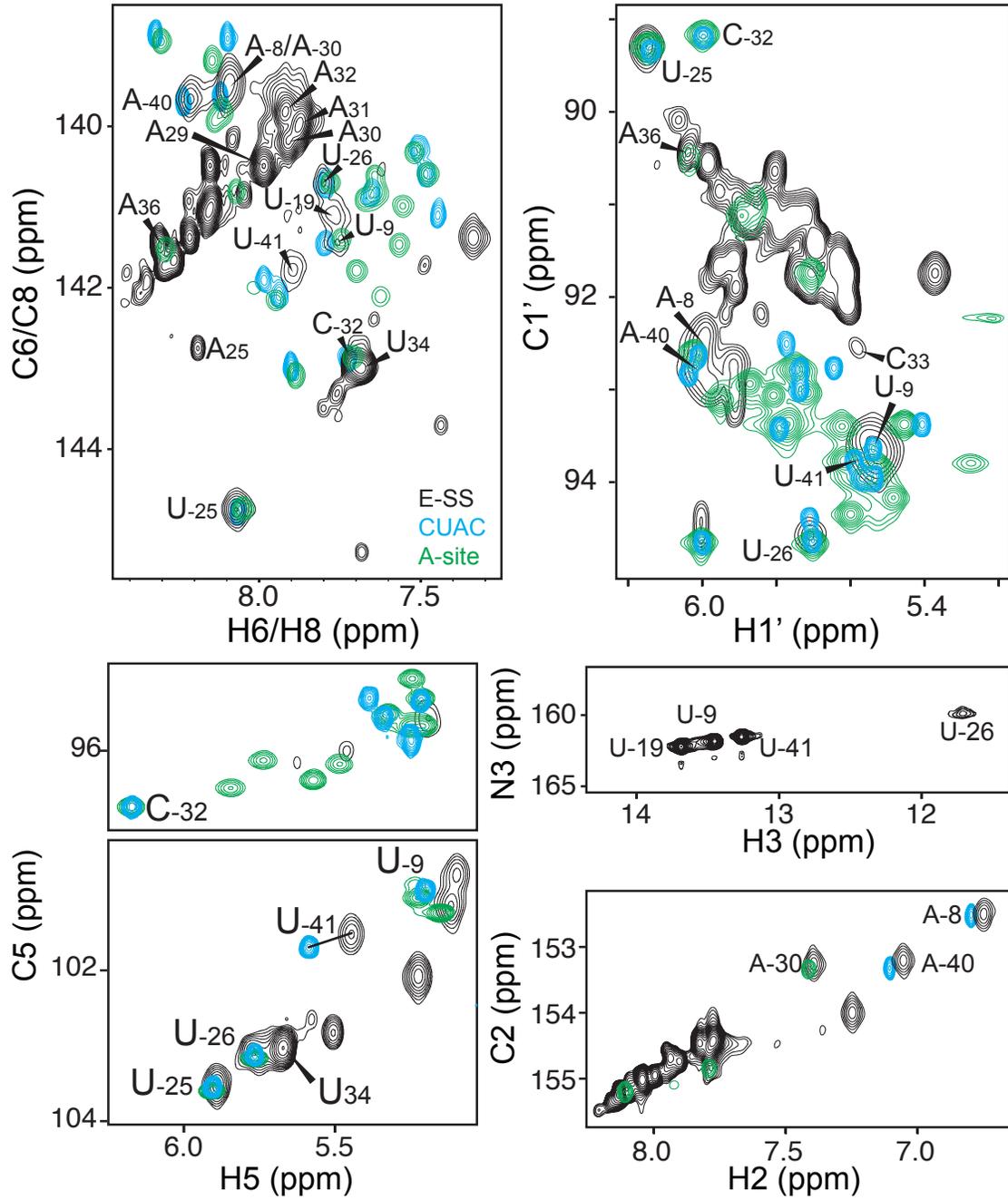
Appendix 4

Sequence conservation Program Code

"pullfullsequence.pl"

```perl
#!/usr/bin/perl

use warnings;
use strict;
use LWP::Simple;
use List::Util 'max';
use Benchmark;

my $nbase="http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nucleotide"; #base
for fetching info from nucleotide
my $pbase="http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=protein"; #base
for fetching info from protein
my $opt1="&rettype=gb&retmode=text"; #options database output
my $opt2="&rettype=fasta&retmode=text"; #options for database output

open(NAME,$ARGV[0]);
my @quecseqlist=<NAME>; #input the queC sequence list into an array
close(NAME);

open(FILE, ">NoProteinFound.dat");

open(FILE2, ">ReversedStartandEnd.dat");

print "SpeciesName Accession Start End JustAptmer AptmerToProteinStart EntireSearch
ProteinStart Definition\n";

foreach my $Line (@quecseqlist) {
```

```perl
next if $Line !~ /\s+AC\s+/ || $Line !~ /\/\//;

#print "made it!\n";

my @spl = split /\s+/, $Line;

my ($id, $start, $end) = ($spl[3] =~ /(\S+)\/(\d+)-(\d+)/);

if($start > $end) {

        print FILE2 $spl[1] . " " . $id . " " . $start . " " . $end . " " . "\n";
          next;

}

#next if $id ne "CP000922.1";

my ($orgstart, $orgend) = ($start,$end);

$start -= 50;
$end += 400;

my $url1=$nbase."&id=".$id."&seq_start=".$start."&seq_stop=".$end.$opt2;

my $output=get($url1); #fetch fasta sequence

next if ! defined $output;

my $Sequence = "";

foreach my $line (split /\s+/, $output) {
```

119

```perl
    $Sequence .= $1 if $line =~ /([AGCT]{20,})\n?/;

}

my $url2=$nbase."&id=".$id."&seq_start=".$start."&seq_stop=".$end.$opt1;

my $output2=get($url2);

next if ! defined $output2;

my @split2 = split /ORIGIN/, $output2;

my @prot_sequences = ($split2[1] =~ /\s*([agct]+)\s*\n?/g);

my $prot_sequence = uc(join("", @prot_sequences));

my $result = rindex($Sequence, substr($prot_sequence, 0, 30));

#print $result . "\n";

my ($definition) = ($split2[0] =~ /DEFINITION([\s\S]+)\nACCESSION/);

my ($position) = ($split2[0] =~ /gene\s+(\d+)/);

#print $output2 . "\n";

my $justaptmer = substr($prot_sequence, 50, $orgend - $orgstart + 1);

if(! defined $position) {
```

```perl
        print FILE $spl[1] . " " . $id . " " . $orgstart . " " . $orgend . " " . $justaptmer . " "
. $Sequence . "\n";


        next;


  }


  my $toproteinstart = substr($prot_sequence, 50, $position + 2 - 50);


  #print $prot_sequence . "\n";


  print $spl[1] . " " . $id . " " . $orgstart . " " . $orgend . " " . $justaptmer . " " .
$toproteinstart . " " . $Sequence . " " . ($position + $orgstart - 51) . " " . $definition . "\n";


}


close(FILE);


close(FILE2);
```



```perl
"AnalyzeSequence.pl"
#!/usr/bin/perl


use strict;
use warnings;


open(FILE, "align_data.txt");
```

```perl
my @FileContents = <FILE>;

close(FILE);

my @Sequences;

my $Count = 0;

my @Names = qw(SpeciesName Accession Start End JustAptmer AptmerToProteinStart
EntireSearch ProteinStart Definition);

foreach my $Line (@FileContents) {

        next if $Line !~ /\d+\s+\d+/;

        my @split = split /\s+/, $Line;

        my $Hash;

        foreach my $i (0 .. @Names-2) {

                $Hash->{$Names[$i]} = $split[$i];

        }

        $Hash->{$Names[-1]} = join(" ", @split[@Names-1, @split-1]);

        push @Sequences, $Hash;

}
```

```perl
#Type I vs Type II

my $R = qr/[AG]/;
my $Y = qr/[TU]/;

my $TypeI1 = qr/[AG]GTTC[AG]/;
my $TypeI2 = qr/TCC[TC]/;
my $TypeI3 = qr/AAAAC[TC]A\wG/;

my @distances = (10, 31);

my $TypeII1 = qr/GT/;
my $TypeII2 = qr/[TC]TA[AG]/;
my $TypeII3 = qr/ACC\w\w\wT/;
my $TypeII4 = qr/A\wAAAA\wT\w[AG][AG]/;

my @distances_II = (12,15, 40);

foreach my $Seq (@Sequences) {

  my $Aptmer = $Seq->{JustAptmer};

  my @rindexes = (0,0,0);

  my $Count = 0;

  foreach my $Pattern ($TypeI1, $TypeI2, $TypeI3) {

    my ($Result) = ($Aptmer =~ /($Pattern)/);
```

```perl
  last if ! defined $Result;

  $Count++;

  $rindexes[$Count] = rindex($Aptmer, $Result);

}

my $Fail = 0;

$Fail = 1 if $Count < 2;


my @actual_distances = ($rindexes[1] - $rindexes[0], $rindexes[2] - $rindexes[1]);

foreach my $i (0 .. @actual_distances-1) {

  if($actual_distances[$i] > $distances[$i]) {

    $Fail = 1; last;

  }

  elsif($actual_distances[$i] < 0) {

        $Fail = 1; last;

  }

}
```

```perl
  if(!$Fail) {

    $Seq->{Type} = 1;

    next;

  }

  @rindexes = (0,0,0,0);

  $Count = 0;

  foreach my $Pattern ($TypeII1, $TypeII2, $TypeII3, $TypeII4) {

    my ($Result) = ($Aptmer =~ /($Pattern)/);

    last if ! defined $Result;

    $Count++;

    $rindexes[$Count] = rindex($Aptmer, $Result);

  }

  $Fail = 0;

  $Fail = 1 if $Count < 3;

  @actual_distances = ($rindexes[1] - $rindexes[0], $rindexes[2] - $rindexes[1],
$rindexes[3] - $rindexes[2]);
```

```perl
foreach my $i (0 .. @actual_distances-1) {

  if($actual_distances[$i] > $distances_II[$i]) {

    $Fail = 1; last;

  }

  if($actual_distances[$i] < 0) { $Fail = 1; last}

}

if(!$Fail) {

  $Seq->{Type} = 2;

}

else {

  $Seq->{Type} = 0;

}

}

#Is Translation

foreach my $Seq (@Sequences) {
```

```perl
my $AptmerToProtein = $Seq->{AptmerToProteinStart};

my $ToProtein = -length($Seq->{JustAptmer}) + length($AptmerToProtein);

next if $ToProtein > 20;

$Count++;

$Seq->{RiboType} = "Translation";

}

foreach my $Seq (@Sequences) {

  next if defined $Seq->{RiboType};

  #Find last bp in just Aptmer

  my $Aptmer = $Seq->{JustAptmer};

  my $Accession = $Seq->{Accession};

  open(FILE, ">$Accession.seq");

  print FILE substr($Aptmer,4) . "\n";

  close(FILE);

  system("mfold SEQ=$Accession.seq");

  open(FILE, "$Accession\_1.ct");
```

```perl
my @FileContents = <FILE>;

close(FILE);

my $lastbp = 0;

shift @FileContents;

foreach my $Line (reverse @FileContents) {

  my @spl = split /\s+/, $Line;

  if($spl[4] != 0) {

    $lastbp = $spl[0]+1; last;

  }

}

my $FromSSToProtein = substr($Seq->{AptmerToProteinStart},$lastbp+4);

open(FILE, ">$Accession\_SS40.seq");

print FILE substr($FromSSToProtein,0,40) . "\n";

close(FILE);

system("mfold SEQ=$Accession\_SS40.seq");
```

```perl
open(FILE, "$Accession\_SS40_1.ct");

@FileContents = <FILE>;

close(FILE);

my ($DeltaGA) = (shift(@FileContents) =~ /\s+dG\s+=\s+(-?\d+\.\d+)/);

my $BPCount = LengthOfHelix(@FileContents);

print $DeltaGA . " " . $BPCount . "\n";

open(FILE, ">$Accession\_SS50.seq");

print FILE substr($FromSSToProtein,0,50) . "\n";

close(FILE);

system("mfold SEQ=$Accession\_SS50.seq");

open(FILE, "$Accession\_SS50.ct");

@FileContents = <FILE>;

close(FILE);

$lastbp = 0;

my ($DeltaGB) = (shift(@FileContents) =~ /\s+dG\s+=\s+(-?\d+\.\d+)/);

next if $DeltaGB > $DeltaGA;
```

```perl
foreach my $Line (reverse @FileContents) {

  my @spl = split /\s+/, $Line;

  if($spl[4] != 0) {

    $lastbp = $spl[0]; last;

  }

}

print $lastbp . "\n";

my $BestTCount = 0;
my $TCount = 0;

foreach my $i ($lastbp .. @FileContents-1) {

  my @spl = split /\s+/, $FileContents[$i];

  if($spl[1] eq "T") { $TCount++ }

  else {

    $BestTCount = $TCount if $TCount > $BestTCount;

    $TCount = 0;

  }
```

```perl
    }


    if($BestTCount < 3) {

      $TCount = 0;

      foreach my $i (reverse(0 .. $lastbp)) {

        my @spl = split /\s+/, $FileContents[$i];

        if($spl[1] eq "T") { $TCount++ }

        else {

          $BestTCount = $TCount if $TCount > $BestTCount;

          $TCount = 0;

        }

      }

    }

    $Seq->{RiboType} = "Transcription" if $BestTCount >= 3;

}

open(DATAFILE, ">summary.txt");
```

```perl
foreach my $Seq (@Sequences) {

 print DATAFILE $Seq->{SpeciesName} . " ". $Seq->{Accession} . " " . $Seq-
>{JustAptmer} . " " . $Seq->{Type} . " ";

 if(! defined $Seq->{RiboType}) {

  print DATAFILE "NoType\n";

 }

 else {

  print DATAFILE $Seq->{RiboType} . "\n";

 }


}

close(FILE);

sub LengthOfHelix {

 my @FileContents = @_;

 my $LongestBpCount = 0;

 my $BPcount = 0;
```

```perl
foreach my $Line (@FileContents) {

  my @split = split /\s+/, $Line;

  if($split[4] != 0) {

    $BPcount++;

  }

  else {

    $LongestBpCount = $BPcount if($LongestBpCount < $BPcount);

    $BPcount = 0;
  }

}

return $LongestBpCount;

}
```

Appendix 5

Sequence Conservation Results For Transcription And Translation Prequeuosine Riboswitches

Translation

Species	Accession No.	ssRNA Sequence	Length	Species	Accession No.	ssRNA Sequence	Length
Actinobacillus_minor.1	ACFT01000021.1	CAAAAAAACTAA	12	Bacillus_mycoides_Ro.2	ACMV01000119.1	AAAAAAACTAA	11
Actinobacillus_pleur.1	CP001091.1	TCAAAGACACTCTTTGAAATCAAAACTAA	12	Bacillus_pseudomycoi.2	ACMX01000029.1	AAAAAAACTAA	11
Actinobacillus_pleur.3	CP000569.1	CCAAAGATTCTCTTTGAAATCAAAACTAA	13	Bacillus_subtilis_su.3	AL009126.3	ATAAAAAACTAA	12
Actinobacillus_pleur.4	CP000687.1	CCAAAGATTCTCTTTGAAATCAAAACTAA	13	Bacillus_thuringiens.11	ACNK01000039.1	AAAAAAACTAA	11
Actinobacillus_succi.1	CP000746.1	CAAAAAAACTAA	12	Bacillus_thuringiens.17	ACNH01000032.1	AAAAAAACTAA	11
B.anthracis.1	AE017334.2	AAAAAAACTAA	11	Bacillus_thuringiens.19	ACNI01000037.1	AAAAAAACTAA	11
B.anthracis.10	AE016879.1	AAAAAAACTAA	11	Bacillus_thuringiens.20	ACNA01000041.1	AAAAAAACTAA	11
B.anthracis.2	AE017225.1	AAAAAAACTAA	11	Bacillus_thuringiens.22	AE017355.1	AAAAAAACTAA	11
B.thuringiensis.2	CP000485.1	AAAAAAACTAA	11	Bacillus_thuringiens.23	ACNG01000049.1	AAAAAAACTAA	11
Bacillus_amyloliquef.1	CP000560.1	ATAAAAAACTAA	12	Bacillus_thuringiens.24	ACMY01000044.1	AAAAAAACTAA	11
Bacillus_anthracis_s.10	CP001598.1	AAAAAAACTAA	11	Bacillus_thuringiens.28	ACNB01000063.1	AAAAAAACTAA	11
Bacillus_anthracis_s.7	ABLH01000004.1	AAAAAAACTAA	11	Bacillus_thuringiens.29	ACND01000042.1	AAAAAAACTAA	11
Bacillus_anthracis_s.9	ABJC01000002.1	AAAAAAACTAA	11	Bacillus_thuringiens.3	ACNF01000046.1	AAAAAAACTAA	11
Bacillus_cereus_03BB.4	CP001407.1	AAAAAAACTAA	11	Bacillus_thuringiens.32	ACNL01000055.1	AAAAAAACTAA	11
Bacillus_cereus_1725.1	ACLV01000037.1	AAAAAAACTAA	11	Bacillus_thuringiens.5	ACMZ01000038.1	AAAAAAACTAA	11
Bacillus_cereus_9582.2	ACMF01000034.1	AAAAAAACTAA	11	Bacillus_thuringiens.6	ACNC01000061.1	AAAAAAACTAA	11
Bacillus_cereus_AH12.2	ACMR01000074.1	AAAAAAACTAA	11	Bacillus_thuringiens.8	ACNJ01000035.1	AAAAAAACTAA	11
Bacillus_cereus_AH12.3	ACMT01000087.1	AAAAAAACTAA	11	Bacillus_thuringiens.9	ACNE01000034.1	AAAAAAACTAA	11
Bacillus_cereus_AH12.5	ACMS01000078.1	AAAAAAACTAA	11	Bacillus_weihensteph.2	CP000903.1	AAAAAAACTAA	11
Bacillus_cereus_AH18.2	CP001177.1	AAAAAAACTAA	11	Bacillus.7	AAXV01000007.1	AAAAAAAACTAA	12
Bacillus_cereus_AH60.1	ACMP01000053.1	AAAAAAACTAA	11	Dorea_longicatena_DS.1	AAXB02000001.1	ATAAAAAACTAG	12
Bacillus_cereus_AH62.2	ACLX01000046.1	AAAAAAACTAA	11	Enterococcus_faecali.14	ACGM01000006.1	TAAAAAAACTCG	12
Bacillus_cereus_AH67.1	ACMQ01000057.1	AAAAAAACTAA	11	Enterococcus_faecali.9	ACHK01000087.1	TAAAAAAACTCG	12
Bacillus_cereus_AH82.1	CP001283.1	AAAAAAACTAA	11	H.influenzae.1	AAZI01000005.1	AATAAAAAACTAG	13
Bacillus_cereus_ATCC.2	AE017194.1	AAAAAAACTAA	11	H.influenzae.2	CP000671.1	AATAAAAAACTAG	13
Bacillus_cereus_ATCC.3	ACLZ01000022.1	AAAAAAACTAA	11	H.influenzae.4	AAZH01000008.1	AATAAAAAACTAG	13
Bacillus_cereus_ATCC.4	ACLT01000039.1	AAAAAAACTAA	11	Haemophilus_influenz.1	AAZF01000004.1	AATAAAAAACTAG	13
Bacillus_cereus_ATCC.5	AE017194.1	ATAAAAAACTAT	12	Haemophilus_influenz.4	AAZD01000001.1	AATAAAAAACTAG	13
Bacillus_cereus_ATCC.7	AE016877.1	AAAAAAACTAA	11	Haemophilus_influenz.6	ABWW01000004.1	AATAAAAAACTAG	13
Bacillus_cereus_B426.1	CP001176.1	AAAAAAACTAA	11	Haemophilus_parasuis.2	ABKM01000055.1	CAAAAAAACTAA	12
Bacillus_cereus_BDRD.2	ACME01000036.1	AAAAAAACTAA	11	L.innocua.1	AL596166.1	AAAAAAACTAG	11
Bacillus_cereus_BDRD.3	ACMB01000039.1	AAAAAAACTAA	11	L.monocytogenes.1	AL591977.1	AAAAAAACTAG	11
Bacillus_cereus_BDRD.5	ACMC01000037.1	AAAAAAACTAA	11	Lactobacillus_planta.1	ACGZ01000094.1	TAAAAAAACTAG	12
Bacillus_cereus_BDRD.7	ACMD01000065.1	AAAAAAACTAA	11	Lactobacillus_planta.2	AL935253.1	TAAAAAAACTAG	12
Bacillus_cereus_BGSC.2	ACLU01000039.1	AAAAAAACTAA	11	Lactobacillus_sakei_.2	ACLD01000058.1	AAAAAAACTAG	11
Bacillus_cereus_E33L.1	CP000001.1	AAAAAAACTAA	11	Lactobacillus_saliva.4	CP000233.1	ATAAAAAACCAA	12
Bacillus_cereus_F651.2	ACMO01000029.1	AAAAAAACTAA	11	Leuconostoc_citreum_.2	DQ489736.1	TATCAAAACTAG	12
Bacillus_cereus_G984.2	CP001186.1	AAAAAAACTAA	11	Listeria_grayi_DSM_2.1	ACCR01000022.1	CAAAAAAACTAA	12
Bacillus_cereus_H308.2	ABDL02000004.1	AAAAAAACTAA	11	Listeria_monocytogen.10	AE017262.2	AAAAAAACTAG	11
Bacillus_cereus_m129.1	ACLS01000052.1	AAAAAAACTAA	11	Listeria_monocytogen.14	AARU02000001.1	AAAAAAACTAG	11
Bacillus_cereus_m155.1	ACMA01000031.1	AAAAAAACTAA	11	Listeria_monocytogen.16	AARR02000003.1	AAAAAAACTAG	11
Bacillus_cereus_MM3.1	ACLW01000046.1	AAAAAAACTAA	11	Listeria_monocytogen.2	AARX02000001.1	AAAAAAACTAG	11
Bacillus_cereus_Q1.2	CP000227.1	AAAAAAACTAA	11	Listeria_monocytogen.20	FM242711.1	AAAAAAACTAG	11
Bacillus_cereus_R309.1	ACLY01000033.1	AAAAAAACTAA	11	Listeria_monocytogen.3	AATL02000002.1	AAAAAAACTAG	11
Bacillus_cereus_Rock.1	ACMJ01000036.1	AAAAAAACTAA	11	Listeria_monocytogen.7	AADR01000005.1	AAAAAAACTAG	11
Bacillus_cereus_Rock.10	ACMG01000028.1	-AAAAAAACTAA	10	Listeria_monocytogen.9	AARQ02000004.1	AAAAAAACTAG	11
Bacillus_cereus_Rock.13	ACMI01000112.1	AAAAAAACTAA	11	Listeria_welshimeri_.1	AM263198.1	AAAAAAACTAG	11
Bacillus_cereus_Rock.5	ACML01000096.1	AAAAAAACTAA	11	Neisseria_cinerea_AT.1	ACDY01000012.1	ATTAAAAAACTAA	13
Bacillus_cereus_Rock.7	ACMM01000135.1	AAAAAAACTAA	11	Neisseria_flavescens.1	ACEN01000025.1	ATTAAAAAACTAA	13
Bacillus_cereus_Rock.8	ACMK01000049.1	AAAAAAACTAA	11	Neisseria_gonorrhoea.13	AE004969.1	ACTAAAAAACTAA	13
Bacillus_cereus_Rock.9	ACMH01000040.1	AAAAAAACTAA	11	Neisseria_gonorrhoea.4	CP001050.1	ACTAAAAAACTAA	13
Bacillus_cereus_subs.2	CP000764.1	AAAAAAACTAA	11	Neisseria_lactamica_.1	ACEQ01000070.1	ACTAAAAAACTAA	13
Bacillus_coagulans_3.1	AAWV02000002.1	TAAAAAAACTAA	12	Neisseria_meningitid.3	AM421808.1	ACTAAAAAACTAA	13
Bacillus_mycoides_DS.2	ACMU01000036.1	AAAAAAACTAA	11	Neisseria_meningitid.4	AL157959.1	ACTAAAAAACTAA	13
Bacillus_mycoides_Ro.1	ACMW01000039.1	AAAAAAACTAA	11	Neisseria_mucosa_ATC.1	ACDX01000133.1	ATTAAAAAACTAA	13

Transcription

Species	Accession No.	ssRNA Sequence	Length	Species	Accession No.	ssRNA Sequence	Length
B.anthracis.6	AE017334.2	ATAAAAAACTAT	12	Bacillus_cereus_R309.2	ACLY01000029.1	ATAAAAAACTAT	12
B.anthracis.8	AE017225.1	ATAAAAAACTAT	12	Bacillus_cereus_Rock.11	ACMI01000108.1	ATAAAAAACTAT	12
B.anthracis.9	AE016879.1	ATAAAAAACTAT	12	Bacillus_cereus_Rock.12	ACMK01000041.1	ATAAAAAACTAT	12
B.thuringiensis.1	CP000485.1	ATAAAAAACTAT	12	Bacillus_cereus_Rock.16	ACMG01000025.1	ATAAAAAACTAT	12
Bacillus_anthracis_s.22	CP001598.1	ATAAAAAACTAT	12	Bacillus_cereus_Rock.3	ACML01000083.1	ATAAAAAACTAT	12
Bacillus_anthracis_s.3	ABJC01000006.1	ATAAAAAACTAT	12	Bacillus_cereus_Rock.4	ACMH01000035.1	ATAAAAAACTAT	12
Bacillus_anthracis_s.8	ABKG01000007.1	ATAAAAAACTAT	12	Bacillus_cereus_Rock.6	ACMM01000130.1	ATAAAAAACTAT	12
Bacillus_anthracis_T.1	ABDN02000012.1	ATAAAAAACTAT	12	Bacillus_cereus_subs.1	CP000764.1	ATAAAAAACTAT	12
Bacillus_cereus_03BB.1	ABDM02000015.1	ATAAAAAACTAT	12	Bacillus_mycoides_DS.1	ACMU01000033.1	ATAAAAAACTAT	12
Bacillus_cereus_03BB.3	CP001407.1	ATAAAAAACTAT	12	Bacillus_mycoides_Ro.3	ACMV01000059.1	ATAAAAAACTAT	12
Bacillus_cereus_1725.2	ACLV01000035.1	ATAAAAAACTAT	12	Bacillus_mycoides_Ro.4	ACMW01000034.1	ATAAAAAACTAT	12
Bacillus_cereus_9582.1	ACMF01000023.1	ATAAAAAACTAT	12	Bacillus_pseudomycoi.1	ACMX01000029.1	ATAAAAAACTAT	12
Bacillus_cereus_AH11.2	ABDA02000001.1	ATAAAAAACTAT	12	Bacillus_thuringiens.1	ACNI01000033.1	ATAAAAAACTAT	12
Bacillus_cereus_AH12.1	ACMS01000069.1	ATAAAAAACTAT	12	Bacillus_thuringiens.10	ACNK01000033.1	ATAAAAAACTAT	12
Bacillus_cereus_AH12.4	ACMR01000061.1	ATAAAAAACTAT	12	Bacillus_thuringiens.12	AAJM01000029.1	ATAAAAAACTAT	12
Bacillus_cereus_AH12.6	ACMT01000074.1	ATAAAAAACTAT	12	Bacillus_thuringiens.13	ACMY01000039.1	ATAAAAAACTAT	12
Bacillus_cereus_AH18.1	CP001177.1	ATAAAAAACTAT	12	Bacillus_thuringiens.14	ACND01000039.1	ATAAAAAACTAT	12
Bacillus_cereus_AH60.2	ACMP01000046.1	ATAAAAAACTAT	12	Bacillus_thuringiens.16	ACNF01000038.1	ATAAAAAACTAT	12
Bacillus_cereus_AH62.1	ACLX01000040.1	ATAAAAAACTAT	12	Bacillus_thuringiens.18	ACNA01000036.1	ATAAAAAACTAT	12
Bacillus_cereus_AH67.2	ACMQ01000044.1	ATAAAAAACTAT	12	Bacillus_thuringiens.2	ACNL01000049.1	ATAAAAAACTAT	12
Bacillus_cereus_AH82.2	CP001283.1	ATAAAAAACTAT	12	Bacillus_thuringiens.21	ACNJ01000028.1	ATAAAAAACTAT	12
Bacillus_cereus_ATCC.1	AE016877.1	ATAAAAAACTAT	12	Bacillus_thuringiens.25	ACMZ01000036.1	ATAAAAAACTAT	12
Bacillus_cereus_ATCC.6	ACLZ01000019.1	ATAAAAAACTAT	12	Bacillus_thuringiens.26	ACNC01000054.1	ATAAAAAACTAT	12
Bacillus_cereus_ATCC.8	ACLT01000036.1	ATAAAAAACTAT	12	Bacillus_thuringiens.27	AE017355.1	ATAAAAAACTAT	12
Bacillus_cereus_B426.2	CP001176.1	ATAAAAAACTAT	12	Bacillus_thuringiens.30	ACNB01000056.1	ATAAAAAACTAT	12
Bacillus_cereus_BDRD.1	ACME01000023.1	ATAAAAAACTAT	12	Bacillus_thuringiens.31	ACNH01000028.1	ATAAAAAACTAT	12
Bacillus_cereus_BDRD.4	ACMB01000029.1	ATAAAAAACTAT	12	Bacillus_thuringiens.4	ACNE01000028.1	ATAAAAAACTAT	12
Bacillus_cereus_BDRD.6	ACMC01000030.1	ATAAAAAACTAT	12	Bacillus_thuringiens.7	ACNG01000040.1	ATAAAAAACTAT	12
Bacillus_cereus_BDRD.8	ACMD01000059.1	ATAAAAAACTAT	12	Bacillus_weihensteph.1	CP000903.1	ATAAAAAACTAT	12
Bacillus_cereus_BGSC.1	ACLU01000032.1	ATAAAAAACTAT	12	Brevibacillus_brevis.1	AP008955.1	ATAAAAAACTAA	12
Bacillus_cereus_E33L.2	CP000001.1	ATAAAAAACTAT	12	Dichelobacter_nodosu.1	CP000513.1	CTAAAAAACTAG	12
Bacillus_cereus_F651.1	ACMO01000026.1	ATAAAAAACTAT	12	Enterococcus_faecali.11	ACIX01000103.1	AATAAAAAACTAA	13
Bacillus_cereus_G984.1	CP001186.1	ATAAAAAACTAT	12	Enterococcus_faecali.19	ACGL01000175.1	TAAAAAAACTCG	12
Bacillus_cereus_H308.1	ABDL02000025.1	ATAAAAAACTAT	12	Enterococcus_faecali.3	AE016830.1	AATAAAAAACTAA	13
Bacillus_cereus_m129.2	ACLS01000040.1	ATAAAAAACTAT	12	Lactobacillus_saliva.2	ACGT01000007.1	TAAAAAAACTAG	12
Bacillus_cereus_m155.2	ACMA01000028.1	ATAAAAAACTAT	12	Streptococcus_agalac.1	AAJQ01000020.1	AATAAAAAACTAA	13
Bacillus_cereus_MM3.2	ACLW01000041.1	ATAAAAAACTAT	12	Streptococcus_agalac.4	AAJO01000118.1	AATAAAAAACTAA	13
Bacillus_cereus_Q1.1	CP000227.1	ATAAAAAACTAT	12	Streptococcus_infant.1	ABJK02000022.1	AATAAAAAACTAG	13