PROBING EXCITED CONFORMATIONAL STATES INVOLVED IN MICRORNA BIOGENESIS

Jared T. Baisden

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics in the School of Medicine.

Chapel Hill
2020

Approved by:

Qi Zhang

Andrew Lee

William Marzluff

Xian Chen

Scott Hammond

ABSTRACT

Jared T. Baisden: Probing excited conformational states involved in microRNA biogenesis
(Under the direction of Qi Zhang)

MicroRNAs are evolutionarily conserved small, non-coding RNAs, ~20-22 nucleotides in length, which serve as critical gene regulators in many aspects of biological processes, including development, cellular proliferation, differentiation, and apoptosis. Due to this central regulatory role of microRNAs, their biogenesis is tightly regulated through the binding of multiple protein activators/suppressors and sequence-specific recognition motifs that are utilized by the biogenesis machineries. While the importance of these microRNA sequences in their self-regulation of biogenesis has just begun to be unveil, most of these RNA elements are highly dynamic and have evaded high-resolution structural studies by conventional structural biology techniques. In my thesis work, I have developed and applied state-of-the-art solution-state nuclear magnetic resonance (NMR) techniques, which enabled me to overcome technical challenges in high-resolution structural and dynamic characterization of these flexible RNAs. Together with mutagenesis and functional assays, I have first characterized structure, dynamics, and function of the precursor of microRNA-21 (pre-miR-21), one of the first identified human microRNAs, which also functions as an oncogene involved in tumorigenesis, progression, and metastasis. My results unveiled a pH-dependent conformational ensemble in pre-miR-21 that spontaneously reshuffles the secondary structure of the entire apical stem-loop region, where the alternative excited conformation transiently sequesters the bulged adenine into a non-canonical protonated A+–G mismatch and confers a two-fold enhancement in Dicer processing. Next, I have studied the

mismatch GHG (mGHG) motif in primary microRNAs, an essential RNA recognition site for the microprocessor (Drosha/DGCR8). Remarkably, my results revealed that the mGHG motif also exists as a pH-dependent conformational ensemble that involves a sparsely populated and short-lived protonated state. Furthermore, I showed that these inherent excited-state-mediated dynamics correlate with microRNA recognition and/or productive cleavage by the microprocessor complex. Together, these studies suggest that microRNA can encode a complex conformational landscape to direct functional outcomes, adopting multiple conformations that result in differential enzymatic fitness for each cleavage step of the canonical microRNA biogenesis pathway and exemplifying a novel RNA-centric regulation of microRNA biogenesis.

To my mother, Elaine A. Baisden. Your constant encouragement and support has always given me the motivation and inspiration to continue learning more each and every day.

ACKNOWLEDGEMENTS

To my thesis committee, in particular Prof. Andrew Lee and Prof. William Marzluff, thank you for pressing my PhD forward. Thank you to Prof. Xian Chen for continued optimism with my work, despite the change in focus away from our collaborative project. And thank you to Prof. Scott Hammond, for graciously providing enzymatic expertise and samples, which were essential to connecting my biophysical data with biology.

To the Department of Biochemistry and Biophysics, the Biological and Biomedical Sciences Program and the Initiative for Maximizing Student Diversity, I thank you for supporting my research and engaging me intellectually throughout my journey to complete this degree.

Thank you to members of the Zhang Lab. I appreciate all of the helpful comradery, and the willingness to discuss any portion of research or research problems at any time. To Dr. Joshua Boyer, I appreciate all of the technical expertise that you were so willing to share, regardless of how many projects you already had on your plate. I would also like to thank Dr. Bo Zhao, who mentored me in NMR RD experiments, and was an insightful co-author on my publications.

Thank you to my family and to my friends, who were always supportive and loving, and who have been patient in my adventure of graduate school. To my mother, who stood with me through *all* of my decisions. Lastly, thank you to two of my earliest mentors, Melanie Vinion, who inspired my love for science with her own excitement, and coach Doug Bennett, who taught me that hard work and perseverance can push us through the hardest of problems.

PREFACE

Parts of the work described in this dissertation has been previously published as:

Chapter 2 is reproduced with permission from Jared T. Baisden, Joshua A. Boyer, Bo Zhao, and Qi Zhang. (2020) Visualizing a transient RNA protonation state that modulates oncogenic microRNA-21 maturation. doi: https://doi.org/10.1101/852822 (bioRxiv), *Nature Chemical Biology* (under revision).

Chapter 5 is reproduced with permission from Jared T. Baisden*, Bo Zhao* and Qi Zhang. (2020) Probing excited conformational states of nucleic acids by nitrogen CEST NMR spectroscopy. *Journal of Magnetic Resonance* (310): e106642 p1-7.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

CEST	Chemical Exchange Saturation Transfer
CPMG	Carr-Purcell-Meiboom-Gill
Cryo-EM	Cryogenic electron microscopy
DGCR8	DiGeorge Critical Region 8
dsRBD	double-stranded RNA binding domain
ES	Excited State
GS	Ground State
HSQC	Heteronuclear single quantum correlation
J	Scalar Coupling
K	Kelvin
k_{ex}	Exchange rate
KSRP (KHSRP)	KH-Type Splicing Regulatory Protein
MCSF	Mutate and Chemical Shift Fingerprint
mGHG	mismatch GHG
miR-21	mature microRNA-21
miRISC	microRNA-Induced Silencing Complex
ncRNA	non-coding RNA
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
PDCD4	Programmed Cell Death 4
p	Population
PAZ	Piwi Argonaut Zille

ppm	Parts Per Million
pre-E-miR-21	Precursor element loop construct of miR-21
pre-miRNA	precursor microRNA
pre-miR-21	precursor microRNA-21
pri-miRNA	primary microRNA
pri-miR-21	primary microRNA-21
$R_{1\rho}$	R_1 in the rotating frame
R_1	Longitudinal NMR relaxation
R_2	Transverse NMR relaxation
RD	Relaxation Dispersion
RMSD	Root Mean Square Deviation
RRM	RNA Recognition Motif
s.d.	Standard Deviation
TRBP	Transactivation-Responsive RNA Binding Protein

CHAPTER 1 - INTRODUCTION

MicroRNA Biogenesis

MicroRNAs are highly conserved, small noncoding RNAs that regulate more than 60% of protein coding genes at the post-transcriptional level[1–5]. Most microRNAs are initially transcribed by RNA polymerase II as introns of protein-coding genes or from independent coding genes into long primary transcripts (pri-miRNAs) that feature 5' 7-methylguanosine caps and 3' poly-A tails[6,7]. In the canonical biogenesis pathway, pri-miRNAs are subsequently processed into ~70 nucleotide precursor hairpins (pre-miRNAs) by the Microprocessor complex, consisting of one RNase III family enzyme, Drosha, and two DiGeorge Critical Region 8 proteins (DGCR8)[8,9]. Pre-miRNAs are then exported from the nucleus to the cytoplasm by Exportin-5 (Ref. [10]) and further processed into ~20 base-pair miRNA/miRNA* duplexes by another RNase III family enzyme, Dicer, in complex with transactivation-responsive RNA binding protein (TRBP)[11,12]. The resulting single-stranded mature miRNA is incorporated into the miRNA-inducing silencing complex (miRISC), which regulates protein expression by repressing translation, promoting deadenylation, and/or cleaving target mRNA[13].

Due to their essential regulatory roles, miRNA biogenesis is tightly regulated to ensure proper gene expression[3–5], and abnormal miRNA regulation has often been associated with cancer, neurological disorders, cardiovascular diseases and others[14,15]. Although they share the same set of enzymes in the canonical biogenesis pathway, individual microRNAs exhibit cell-type and cell-state specific expressions. Remarkably, even those clustered on the same primary transcript can be differentially processed in a tissue-specific manner[3–5]. Over the past decade, it has been shown that

specific sequences and structures of pri-/pre-miRNAs can be recognized by processing machineries and protein factors for regulation[16–25]. For example, pri-miRNAs that possess a UGU motif in the apical loops are preferentially processed by the Microprocessor[18,19], pre-miRNAs encoding a two-nucleotide distance between the cleavage sites and the apical bulge/loop structures are more accurately processed by Dicer[20], and miRNAs that feature stable basal stems in pri-miRNAs and flexible apical loops in pri-/pre-miRNAs are more efficiently processed by biogenesis machineries[21]. In addition, altering secondary structures and even primary sequences of pri-/pre-miRNAs via protein binding[22], enzymatic-driven nucleotide modification[23,24] and disease-linked mutation[25] can further influence the outcome of microRNA biogenesis[3]. During these regulatory processes, it is often perceived that protein factors act on the largely passive pri-/pre-miRNAs to direct their maturation outcome. However, despite many non-coding RNAs having been shown to actively explore their conformational dynamics for function[26], it remains elusive whether pri-/pre-miRNAs can play an active role in modulating miRNA biogenesis in the absence of protein factors. This is largely due to our limited high-resolution structural and dynamic knowledge of most pri-/pre-miRNAs, where key regulatory elements of pri-/pre-miRNAs, such as the apical stem-loop region, also known as the pre-element region, are often too flexible to be studied by conventional structural biology approaches.

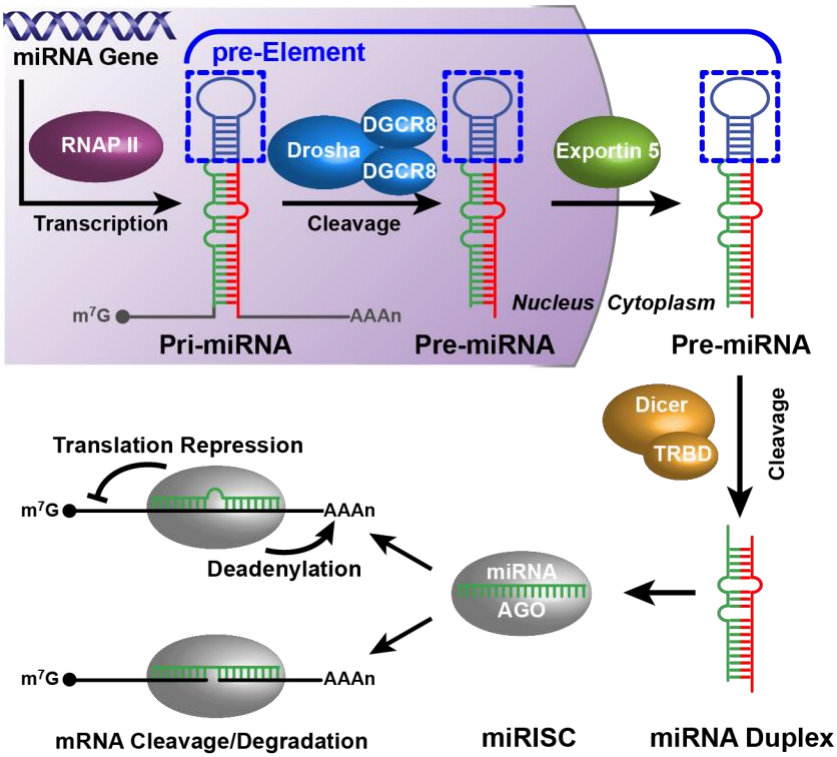

Figure 1.1 - Canonical biogenesis pathway of microRNAs
Most microRNAs are produced through the canonical biogenesis pathway, where they are recognized for their similar hairpin structure and apical loop. After cleavage by Drosha and Dicer, one strand of the miRNA duplex is loaded into AGO2 to form the miRISC, where it functions to repress tune translation through repression and mRNA degradation.

MicroRNA-21

MicroRNA-21, one of the first identified human miRNAs, functions as an oncogene involved in tumorigenesis, progression, metastasis, and cell survival, where its biogenesis is regulated at both transcriptional and post-transcriptional levels. Originally described by Thomas Tuschl in 2001[27], further sequence requirements for function in human cells was determined by Bryan Cullen in 2003. Primary microRNA-21 is transcribed from a coding region of chromosome 17[28]. Despite being located in an intronic region, it has its own promoter, resulting in a 3434 nucleotide-long transcript. miR-21 is processed through the canonical microRNA biogenesis pathway; changes in the levels of mature sequence has been linked to a large number of human cancers[28–30]. Many miR-21 tumors classify as "oncogene addicted," where addition of anti-miR-21 is sufficient to inhibit tumor cell growth, increasing apoptosis and decreasing cell proliferation[31]. The increased production of this microRNA has been correlated with more aggressive tumors, with mature levels being correlated with higher rates of metastasis and cell growth[30]. These essential roles in cancer survival and progression have guided research towards a comprehensive, mechanistic understanding of miR-21 regulation and upregulation in cancer (Fig 1.2). Initial work on mRNA targets revealed that the seed sequence for this particular microRNA interacts Programmed Cell Death 4 (PDCD4)[32]. miR-21 has also been shown to directly target Tropomyosin (TPM1), Sprouty RTK Signaling Antagonist 2 (SPRY2), and mapsin, as well as influencing other important pathways[29,30]. A number of transcription factors, including Nuclear Factor 1 B (NF1B), Signal Transducer and Activator of Transcription 3 (STAT3), Activator Protein 1 (AP-1), and Transforming Growth Factor Beta (TGFß) have been shown to regulate miR-21 transcription. Aside from increasing transcription, the binding of KH-type Splicing Regulatory Proteins (KSRP) has been shown to promote enzymatic processing throughout the biogenesis

pathway. Sequence focused studies[33] have shown that the pre-element region of pri-/pre-miR-21 serves as an important element for regulating miR-21 biogenesis, as mutations to the apical loop result in decreased biogenesis. The sequence for the precursor of miR-21 is highly conserved across vertebrates, spanning fish, avian, and mammalian species[28], likely due to its essential role in regulating cell survival.

Due to the prevalent upregulation of miR-21 in cancer, its potential as a therapeutic target has been widely recognized. Numerous studies have focused on the inhibition of a particular step in the biogenesis of miR-21, including small molecule inhibitors, modified peptides, and RNA binding proteins to varying levels of success. Designed peptide like inhibitors for the cropping step - Peptoids[34] - and dicing - macrocyclic peptides[35] -have reached K_D's on the lower micromolar range, but these studies are largely proof-of-principle, and still require significant progress in design and optimization before they would be clinically relevant. A peptide developed through phage display[36], which interacts with the apical loop of pri-/pre-miR-21, was shown to permeate cells and therapeutically inhibit miR-21 production at concentrations of ~1μM. RNA binding proteins that include an optimized RNA recognition motif (RRM) have been de novo designed to target the terminal loop of miR-21, inhibiting both major endonuclease steps[37]. Chimeras of this designed RRM attached to Giardia Dicer were shown to specifically degrade precursor miR-21. Approaches for high-throughput testing of microRNA inhibitors[38] , termed "AbsorbArray," have directly verified that approved anti-cancer drugs currently target miR-21 production. Given all of these developments, there is still no treatment in the clinic for tumors overexpressing miR-21, and targeting miR-21 remains an area of extensive research.

A major roadblock to the design and improvement of miR-21 therapeutics has been the lack of a high-resolution structure of the apical loop of pri-/pre-miR-21. The lack of structural

information on such a critical region for RNA biogenesis is due to the intrinsically dynamic nature of the apical loop in microRNAs. Dicer preferentially cleaves substrates with unfolded apical loops[39], which explains the conservation of unfolded apical loops across various microRNAs. miR-21 has previously been shown to be capped by an unfolded apical loop. One group created an oligonucleotide model of the miR-21 apical loop in an attempt to solve the structure. This NMR construct, which used native sequence, resulted in imino proton exchange rates that were too rapid for classical ^{1}H-^{1}H NMR characterization, likely due to two stacked G$_{o}$U wobble base pairs that tentatively formed in the apical stem. In order to create a refined structural model, this group stabilized these potential G$_{o}$U base pairs to G-C base pairs[40], locking down the pre-element region and enabling structure modeling by restrained molecular dynamics to generate an NMR ensemble of modified miR-21 to an average Root Mean Square Deviation (RMSD) of 2.53Å. Another NMR group[35] expanded on this work using exchangeable ^{1}H-^{1}H NMR experiments performed on an oligonucleotide model of native pri-/pre-miR-21 in concert with distance constraints from non-exchangeable ^{1}H-^{1}H NOEs. This resulted in a largely unfolded apical loop, which was attributed to changes in secondary structure, or intrinsic unfolding among the precursor element. Further structure refinement using chemical shift values with FARFAR[41] yielded a more uniform, yet unfolded apical loop that was subject to perturbation by small molecules. Aside from these two major studies, only one other microRNA precursor element structure is deposited in the Protein Data Bank - miR-20 (Ref [42]), which has a more rigid apical loop that opens up upon binding to protein cofactors. These studies highlight the dynamic nature of pri-/pre-miRNAs, implying the requirement of a dynamic ensemble and a diverse folding landscape for proper microRNA biogenesis and regulation. A computational study[25] on structural dynamics in microRNA precursor regions highlighted miR-125a. In this work done by François Major's group, a single nucleotide

polymorphism (SNP) in miR-125a was shown to preserve the lowest energy solution structure but alter the higher energy folding landscape - causing a larger population of the apical loop to occupy higher energy structures. This computationally proposed change in the ensemble of free energy folding structures correlated with a change in biogenesis, revealing that changes in the conformational space of microRNA precursors had the potential to alter biological function. All together, these previous works illustrate the dynamic nature of the apical loop of pri-/pre-miRNA and bring into question the contribution of the RNA substrate, and its dynamic ensemble, to regulation in microRNA biogenesis.

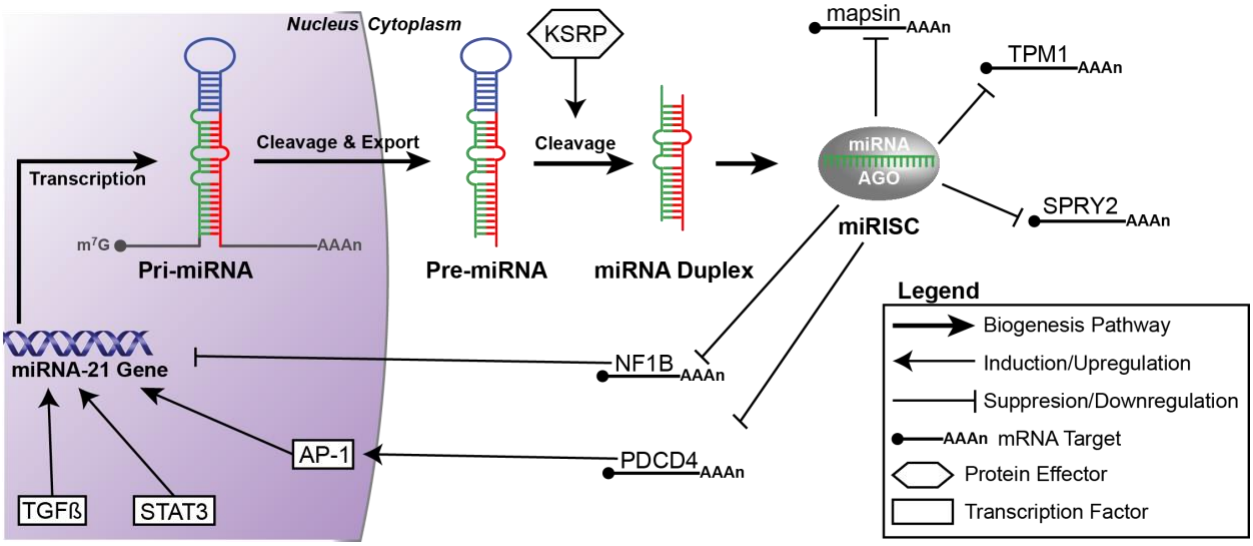

Figure 1.2 - miR-21 activity in cancer cells.
miR-21 is independently transcribed, with its own transcription factors. It interacts with a number of cellular mRNA targets when upregulated in cancer cells. Figure adapted from Secuklu, S.D. et al, 2009[28] and Krischecky A.M. and Gabriely G, 2009[30].

Mismatch GHG Motif

The microprocessor complex, which consists of Drosha and DGCR8, is the first endonuclease step in the biogenesis pathway. This cropping step, which cleaves primary microRNA into their corresponding precursors, was discovered co-currently by three research groups[9,43,44], which described the functional complex between Drosha and DGCR8. Since primary microRNAs are transcribed by Pol II and have the same characteristics as messenger RNA, their secondary structure must contain unique elements that enable the microprocessor to distinguish them from other RNA sequences (Fig 1.2). Overall, primary microRNAs adopt very similar structures, and usually contain a double stranded hairpin of approximately 3 helix turns capped with an unfolded terminal loop[45]. A model structure, including the RNA motifs, secondary structure, and Drosha/DGCR8 binding sites is shown in Figure 1.3. The 60-80 nucleotide primary microRNA hairpin is canonically flanked by long sequences of single stranded RNA at its base. The regions between the helix and terminal loop have been named the 'apical junction' and 'basal junction' respectively (Figure 1.3). Aside from the hairpin structure, a number of essential RNA motifs[18,33,45,46] have been described as determinants of primary microRNA hairpin processing. These conserved motifs influence the cleavage site of the microprocessor complex. 5' of the basal junction, a UG motif, a UGUG motif in the terminal loop after the apical junction, and a CNNC motif 3' of the basal junction, which reside at -11, +22, and -14 nucleotides from the endonuclease cleavage sites, respectively (Figure 1.3). Interestingly, most primary microRNAs do not contain all three of these motifs, meaning the microprocessor uses a combination of motifs to enable accurate recognition and cleavage. More recently, *de novo* design of primary microRNAs[47] illuminated another essential motif that regulates recognition and licensing of cropping. A mismatch GHG motif residing ~7 base pairs away from the Drosha cleavage site has been shown

to enable single nucleotide precision for the cropping step. The microprocessor complex has been shown to tolerate multiple mismatches and wobbles throughout the hairpin structure, which made this specific sequence motif indiscernible from the intrinsic variability of native hairpins until it was bioinformatically determined through an exhaustive search using sequentially modified microRNAs[48]. As of yet, little work has been done to determine the exact solution structure of these primary microRNA motifs, and if a particular structure in the RNA structural ensemble is essential for recognition and cleavage by Drosha. As the processing machinery has been shown to tolerate multiple bulges and mismatches in the hairpin, and a simple mismatch has been shown to preferentially interact with Drosha, we hypothesis that RNA dynamics may play an important role in the regulation of microRNA biogenesis at this essential step.

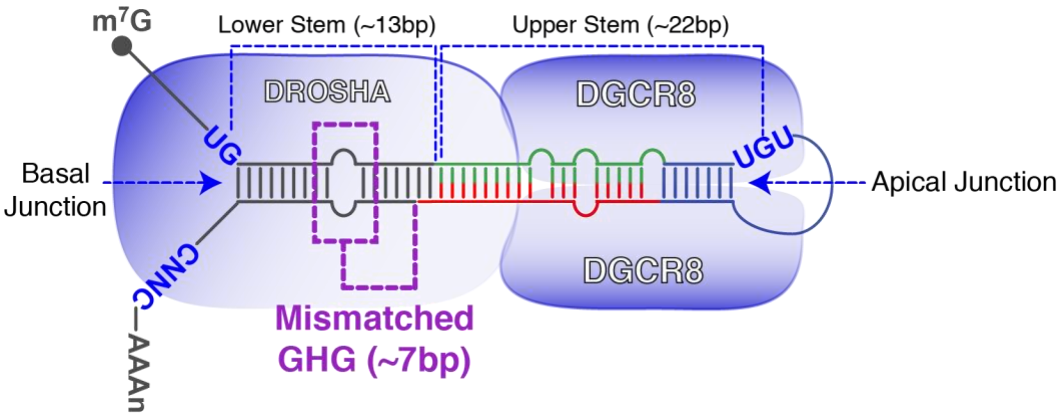

Figure 1.3 - Recognition sites for the microprocessor
Including the overall hairpin structure, 4 sequence motifs enable precise cleavage by the microprocessor. UG and CNNC motifs reside at the basal junction, UGU resides at the apical junction, and the mismatch GHG motif is located in the lower stem of the hairpin. Figure adopted from Kwon et. al, 2019[48].

CHAPTER 2 - VISUALIZING A PROTONATED RNA STATE THAT MODULATES MICRORNA-21 MATURATION

Introduction

Here, using NMR RD spectroscopy and Dicer processing assays, we set out to characterize the structural dynamics of microRNA-21 precursor (pre-miR-21) and examine how the intrinsic RNA conformational plasticity may contribute to miRNA maturation. MicroRNA-21, one of the first identified human miRNAs[49], functions as an oncogene involved in tumorigenesis, progression, metastasis, and cell survival[50], where its biogenesis is regulated at both transcriptional and post-transcriptional levels[51]. Previous studies have shown that the pre-element region of pri-/pre-miR-21 serves as an important element for regulating miR-21 biogenesis. Mutations that stabilize the pre-element inhibit Microprocessor processing of pri-miR-21 (Ref. [33]), whereas binding of KSRP at this location promotes enzymatic processing of pri-/pre-miR-21s[52]. By carrying out NMR RD measurements, we discovered that the pre-element region of pri-/pre-miR-21 exists as a pH-dependent ensemble, which undergoes a two-state structural transition and dynamically accesses a low-populated (~ 1 – 15%) transient, yet kinetically stable (lifetime ~ 0.8 ms) state referred to as an excited state (ES) across physiologically relevant ranges of pH (pH ~ 6.5 – 8.0). With ^{15}N CEST NMR spectroscopy, we were able to directly measure, for the first time, an adenine N1 protonation event, which occurs at the Dicer cleavage site and underlies this unique, pH-dependent structural transition. This adenine protonation corresponds to a concerted secondary structural reshuffling of the entire pre-element region, transitioning the adenine from a bulged residue in the ground-state (GS) conformation to being sequestered into a non-canonical A$^+$(*anti*)–

G(*syn*) base pair in the ES. We further demonstrated that these distinct structures are processed differently by Dicer, where the ES-mimicking substrate is processed to mature miR-21 with a two-fold enhancement in efficiency over its GS counterpart. Hence, despite adopting an apparently simple secondary structure, pre-miR-21 encodes a dynamic ensemble at its pre-element region that encapsulates environmentally sensitive states with distinct fitness for processing. With the emerging view of RNA ES as a 'hidden' layer for regulation[26], our results further suggest that miRNA processing intermediates may employ ES-encoded dynamic ensembles as potential means to regulate microRNA biogenesis in response to environmental and cellular stimuli.

Results

The pre-element region of pre-miR-21 samples distinct conformational states

The miR-21 precursor consists of the pre-element region and miR-21/miR-21* helix, and is predicted to fold into a hairpin structure with four double-stranded helices, three bulges, and one apical loop (Fig. 2.1a). To focus on the pre-element region, we designed a shorter RNA construct, preE-miR-21, which contains the entire pre-element and the adjacent helix from the miR-21/miR-21* stem (Fig. 2.1b). NMR ^{1}H-^{1}H Nuclear Overhauser Effect Spectroscopy (NOESY) experiment on the imino region provides an excellent characterization of RNA secondary structure, as one imino resonance is expected for a canonical Watson-Crick base pair and two imino resonances are expected for a G-U wobble pair. Except for the formation of Watson-Crick base pairs at the lower stem, the pre-element of miR-21 does not adopt a stable conformation, as only weak imino resonances of G-U wobble pairs can be observed in the NMR ^{1}H-^{1}H NOESY spectrum (Fig. 2.1c). This observation is consistent with previous NMR studies on miR-21 precursor[35,40], where mutations of the pre-element were made to quench the structural flexibility into a single conformational state[40]. When we carried out an NMR ^{13}C-^{1}H heteronuclear single quantum

correlation (HSQC) experiment that probes non-solvent-exchangeable signals, surprisingly only 20 out of a total of 29 expected NMR resonances from preE-miR-21 were observed at room temperature (Fig. 2.1c). This spectroscopic behavior resembles the typical NMR phenomenon of exchange broadening, where interconversion between two or more states can lead to the disappearance of NMR signals. Indeed, by raising the temperature from 25$_{o}$C to a more physiologically relevant 35$_{o}$C, most of the missing resonances reappeared in the NMR ^{13}C-^{1}H HSQC spectrum (Fig. 2.1c), confirming the presence of conformational exchange.

Recent developments of NMR R_1 in the rotating frame ($R_{1\rho}$) RD spectroscopy have opened new avenues to quantify microsecond-to-millisecond conformational changes and made it possible to study RNA ESs that are too low-populated and short-lived to be detected by conventional techniques[53–57]. Here, we carried out both on-resonance and off-resonance low spin-lock field $R_{1\rho}$ RD experiments to quantify the exchange process in preE-miR-21. For residues from the miR-21/miR-21* stem region, we observed flat RD profiles for base (C2, C5, C6, and C8) and sugar (C1') carbons (Fig. 2.1d and 2.2), which are consistent with one stable helical conformation of this region. In contrast, residues within the pre-element region, ranging from the bulge residue A22 to the stem residue A35 (Fig. 2.1d and 2.3), display power and offset dependent RD profiles. Indeed, these RD profiles can be global-fitted to a single two-state (GS ↔ ES) exchange process. These results reveal that the pre-element region is not only conformationally flexible, but also dynamically interconverts between at least two structurally and kinetically distinct states, where the ES has a low population (p_{ES}) of 15.2 ± 0.3% and a short lifetime ($\tau_{ES} = 1/k_{EG}$) of 816 ± 15 μs (Fig. 2.1d).

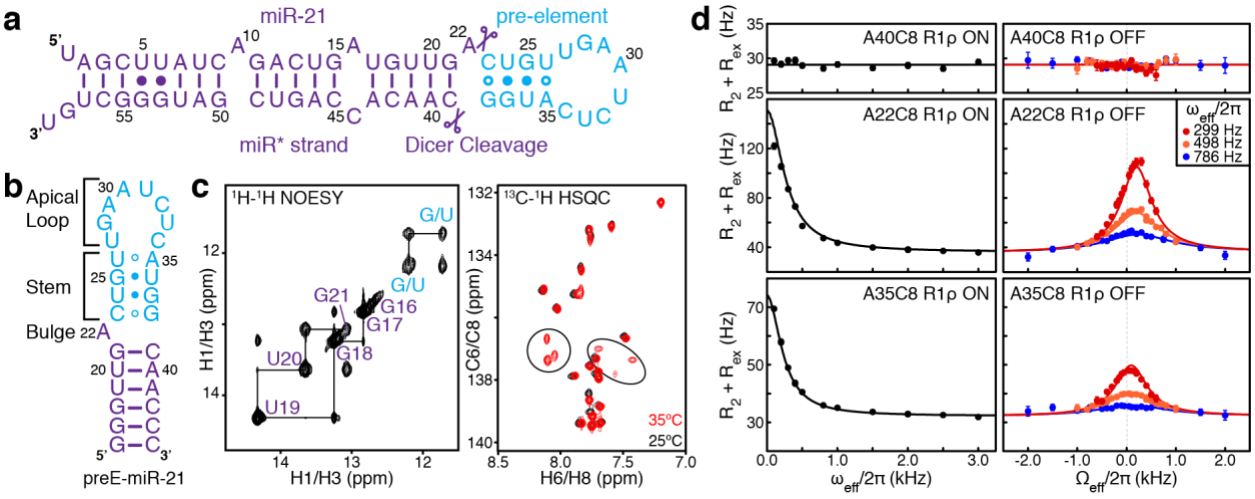

Figure 2.1 - NMR characterization of preE-miR-21.
(**a**) Secondary structure of pre-miR-21 with dicer cleavage sites highlighted as scissors. (**b**) Secondary structure of preE-miR-21 construct derived from NMR data, where Watson-Crick base pairs, GU wobbles, and potential Watson-Crick base pairs are highlighted with lines, filled circles, and open circles, respectively. (**c**) NMR $_{1}$H-$_{1}$H NOESY spectrum of the imino proton region of preE-miR-21 at 10°C and $_{13}$C-$_{1}$H HSQC spectra of base carbon (C6 and C8) region of preE-miR-21 at 25°C and 35°C. (**d**) Representative $_{13}$C on-resonance and off-resonance relaxation dispersion (RD) profiles at 35°C showing dependence of $R_2 + R_{ex}$ on spin-lock power ($\omega_{eff}/2\pi$) and offset ($\Omega/2\pi$), respectively, where Ω is the difference between the spin-lock carrier frequency and the observed resonance frequency. RD profiles of A40 are fit to a single-state model and RD profiles of A22 and A35 are fit to a global two-state model using the Bloch-McConnell equation. Error bars are experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

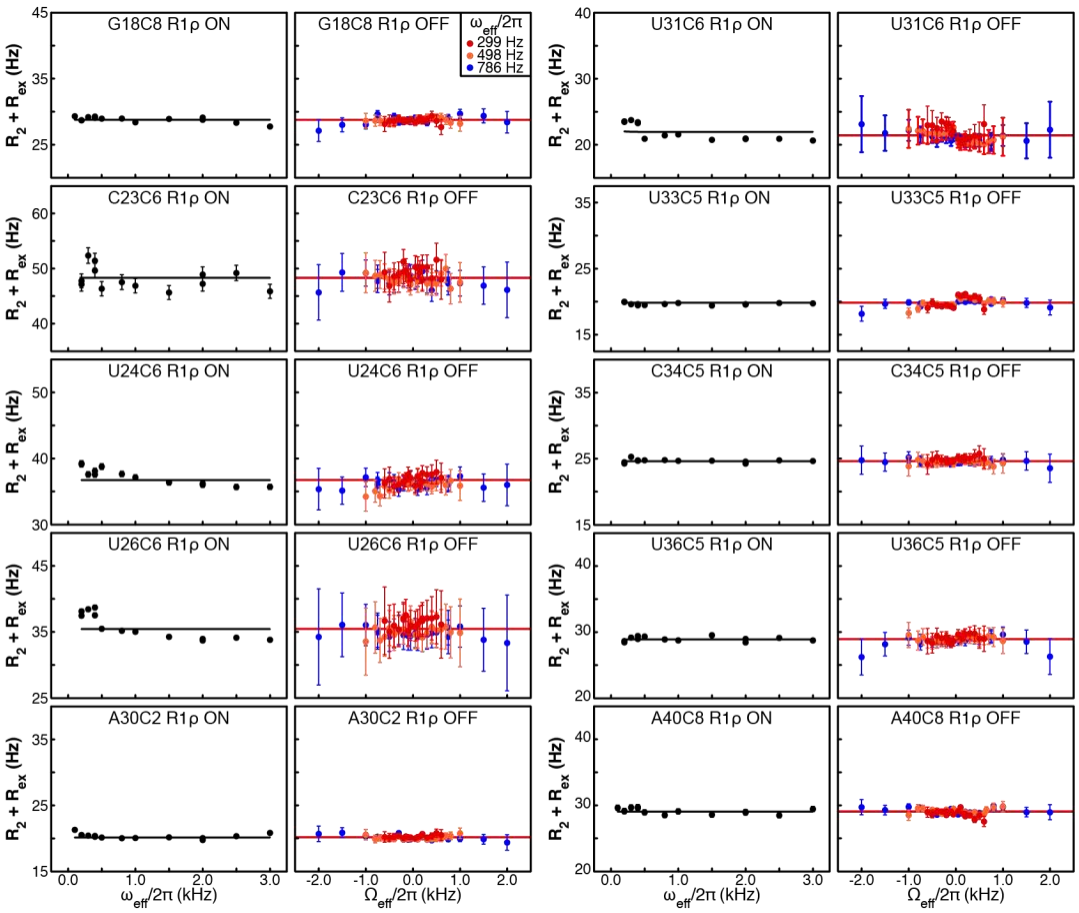

Figure 2.2 - ^{13}C R1ρ RD profiles of preE-miR-21 residues without apparent chemical exchange.

On- and off-resonance ^{13}C RD profiles depicting spin-lock power (ωeff/2π) and offset (Ω/2π) dependence of R2 + Rex measured at pH 6.45. Solid lines represent the best fits to a single-state model using the Bloch-McConnell equation. Error bars, experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

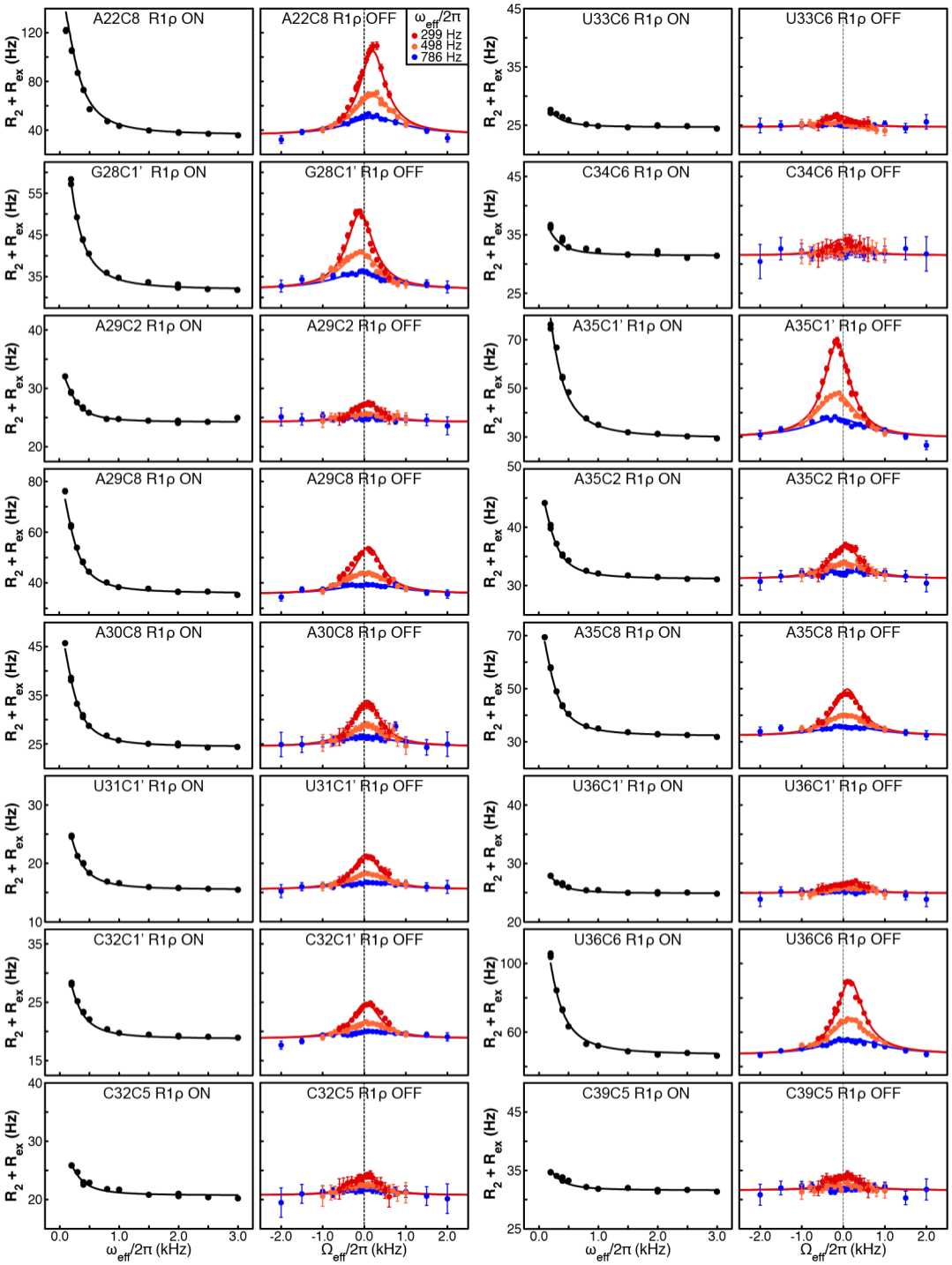

Figure 2.3 - ^{13}C $R_{1\rho}$ RD profiles of preE-miR-21 residues that undergo chemical exchange at pH 6.45.

Solid lines represent the best fits to a global two-state exchange model (k_{ex} = 1445 ± 17 s^{-1} and p_E = 15.2 ± 0.3%) using the Bloch-McConnell equation. Error bars, experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

The ES involves transient protonation at the Dicer cleavage site

To gain structural insights into the ES, we utilized NMR chemical shifts, which are one of the most sensitive measurements for probing local chemical environments. We examined ES carbon chemical shifts ($\varpi_{ES} = \varpi_{GS} + \Delta\varpi$), where $\Delta\varpi$ is the chemical shift difference between ES (ϖ_{ES}) and GS (ϖ_{GS}) extracted from the two-state analysis of an $R_{1\rho}$ RD profile (Fig. 2.4). Among all the extracted ES chemical shifts, the base carbon C8 of bulge A22, which resides at the Dicer cleavage site, displays the largest deviation ($\Delta\varpi$) with a ~1.49 ppm down-field shift from its GS position (Fig. 2.1d and 2.4). Notably, we were not able to obtain the RD profile for base carbon C2 of A22, as the C2H2 resonance remains severely broadened beyond detection in the NMR ^{13}C-^{1}H HSQC (Fig. 2.5a), suggesting even larger perturbations in carbon C2 and/or proton H2 chemical shifts between ES and GS. The dramatically different behavior of C8H8 and C2H2 resonances from the same base is reminiscent of recent NMR studies on transiently N1-protonated adenines[58,59].

To examine whether the exchange process could be due to possible protonation events, we increased the pH of the sample from 6.45 to 8.04, aiming to shift the equilibrium towards non-protonated states. Indeed, the NMR ^{13}C-^{1}H HSQC spectrum recorded at pH 8.04 exhibits much higher quality, where exchange broadening of most resonances is substantially reduced, such that the A22-C2H2 resonance can be readily observed (Fig. 2.5a). Low spin-lock field NMR $R_{1\rho}$ RD measurements provide further quantitative support that the observed exchange involves a protonated ES of the pre-element region of miR-21 (Fig. 2.5b and 2.6). Global fit of RD profiles showed that the GS ↔ ES equilibrium is significantly shifted towards GS at pH 8.04, where the ES population (p_{ES}) is reduced to a mere 1.1 ± 0.1%. In addition, a two-state analysis of the RD profile of base carbon A22-C2 further revealed a remarkable 7.9 ppm difference between its GS

and ES chemical shifts, resulting in an ES chemical shift of 144.5 ppm (Fig. 2.5b and 2.6). This significantly up-field shifted C2 chemical shift is consistent with C2 chemical shifts reported for stably N1-protonated adenines, strongly suggesting that A22 is protonated at the N1 site in the ES.

Recently, we have developed nucleic-acid-specific ^{15}N CEST NMR spectroscopy to study RNA conformational exchanges using non-proton-bonded nitrogens as probes[60]. This technique also enables direct evaluation of the protonation status of adenines in low-populated and short-lived states, which have remained elusive to date. By measuring ^{15}N CEST profiles at pH 8.04, we were able to unambiguously identify that A22 is transiently protonated at N1 (Fig. 2.5c). Unlike N1 of A30, which is not protonated and displays an apparent single-dip CEST profile, the nitrogen CEST profile of A22-N1 exhibits two distinct intensity dips that correspond to two alternative conformations. A two-state analysis of the CEST profile validates that A22-N1 probes the same two-state exchange process, where extracted ES population ($p_{ES\text{-}CEST}$ ~ 1.1 ± 0.1%) and lifetime ($\tau_{ES\text{-}CEST}$ ~ 645 ± 114 μs) from ^{15}N CEST are very similar to those obtained from ^{13}C $R_{1\rho}$ RD at pH 8.04 ($p_{ES\text{-}R1\rho}$ ~ 1.1 ± 0.1%, $\tau_{ES\text{-}R1\rho}$ ~ 823 ± 109 μs). The ES chemical shift of A22-N1 directly supports a protonated N1 with an unprecedented up-field shift of 67.3 ± 0.2 ppm to 157.7 ppm, residing well among resonances of proton-bonded imino nitrogens in RNA (~134-152 ppm in Gs and ~154-165 ppm in Us).

To obtain more insights into the A22 (GS) ↔ A22$^+$ (ES) transition, we further carried out carbon $R_{1\rho}$ RD and nitrogen CEST measurements on base (C2, C8, N1) and sugar (C1') moieties of A22 at pH 6.96 and 7.47 (Fig. 2.5d and 2.7). Consistent with being a protonation-dependent process, the population of A22$^+$ gradually increases from ~ 1% at pH 8.04 to ~ 15% at pH 6.45. Surprisingly, the lifetime of the ES A22$^+$ remains largely unperturbed between pH 6.45 and pH 8.04, where the average lifetime is τ_{ES} ~ 847 ± 49 μs (Fig. 2.5e). In contrast, the apparent lifetime

of the GS, which is derived from the extracted rate of exchange ($\tau_{GS} = 1/k_{GE}$), reduces substantially from τ_{GS} ~ 74 ms at pH 8.04 to τ_{GS} ~ 5 ms at pH 6.45 (Fig. 2.5e). The high population of $A22^{+}$ at pH 6.45, which otherwise would be close to zero based on the intrinsic pK_a (~ 3.5) of free adenine N1 site[61], further suggests that A22 has a distinct protonation propensity when compared with other unstructured adenines. Consistent with this observation, pH-dependent chemical shift analyses showed a pK_a value of 5.84 ± 0.08 for A22, which is substantially shifted towards neutral pH from adenines in the apical-loop that have an average pK_a value of 4.17 ± 0.06 (Fig. 2.5f and 2.8). Taken together, these results unambiguously revealed that preE-miR-21 undergoes a pH-dependent conformational transition, where A22 at the Dicer cleavage site is specifically protonated in the ES.

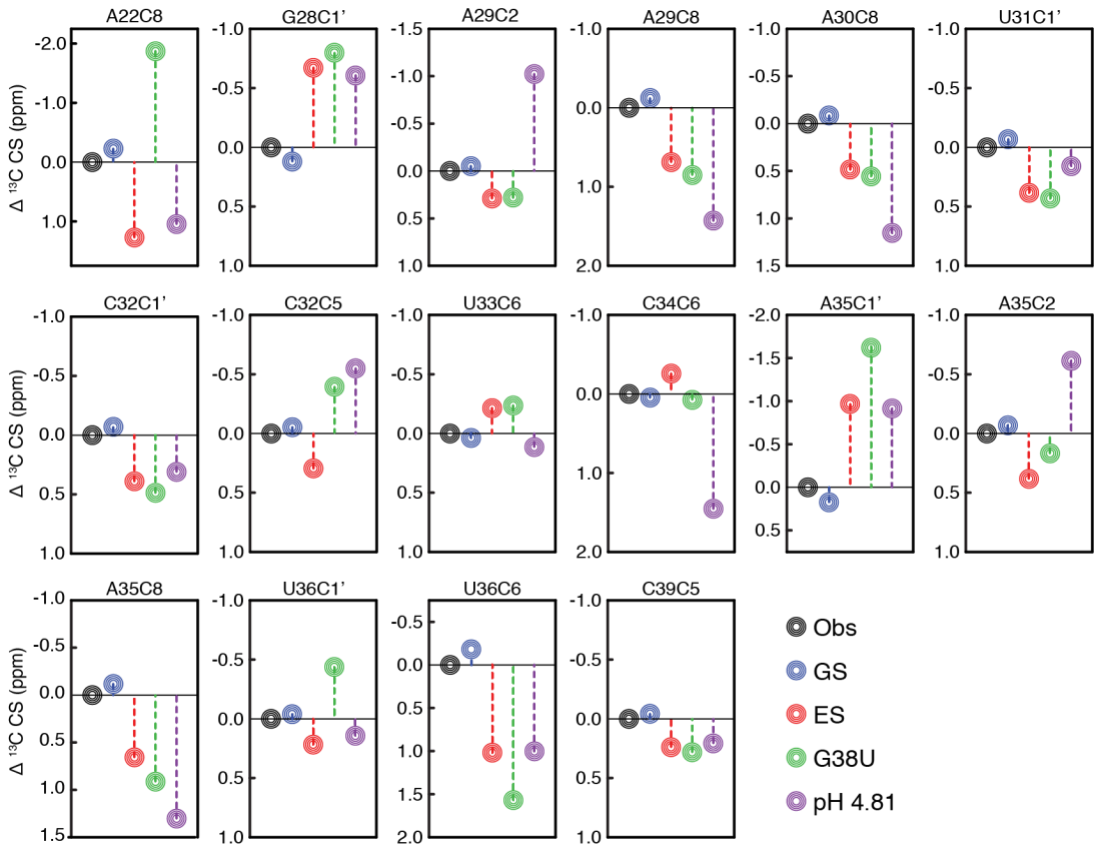

Figure 2.4 - MCSF plots for preE-miR-21.

Comparison of GS, ES, mutant carbon chemical shifts. Shown are differences between the observed chemical shifts (black) and the chemical shifts of GS (blue) and ES (red) extracted from 13C R1ρ RD profiles at pH 6.45, G38U mutant at pH 6.45, and preE-miR-21 at pH 4.81. The apparent discrepancies between ES base carbon chemical shifts of A-C2/C8s and C-C5/C6s and their corresponding chemical shifts at pH 4.81 are likely due to intrinsic protonation at A-N1s and C-N3s at pH 4.81.

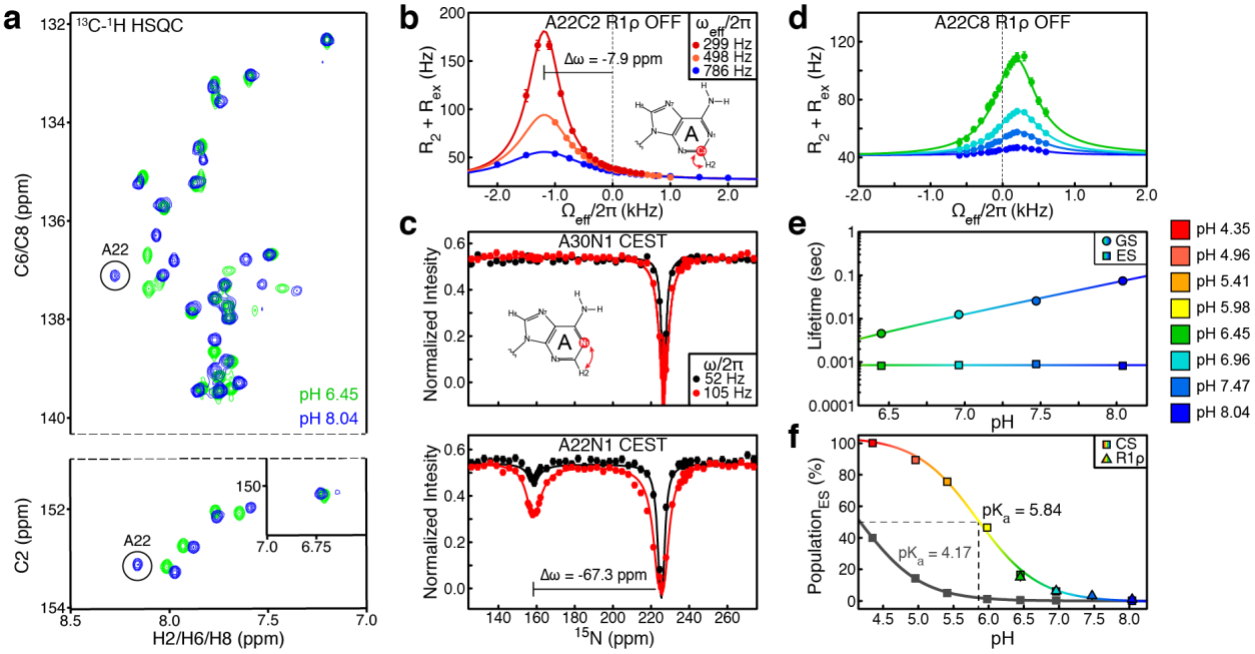

Figure 2.5 - PreE-miR-21 populates a protonated excited state with a neutral shifted pKa
(**a**) NMR ^{13}C-^{1}H HSQC spectra of base carbon (C2, C6, and C8) region of preE-miR-21 at pH 6.45 and pH 8.04. (**b**) ^{13}C off-resonance RD profiles of A22-C2 at pH 8.04. (**c**) ^{15}N CEST profiles of A30-N1 and A22-N1 at pH 8.04, which are fit to a single-state and a two-state model, respectively, using the Bloch-McConnell equation. (**d**) The pH-dependent ^{13}C off-resonance RD profiles of A35-C8 at spin-lock power of $\omega_{eff}/2\pi$ =299 Hz. (**e**) The pH-dependent apparent lifetimes of GS and ES from RD analysis. (**f**) The pH-dependent population of the excited state based on $R_{1\rho}$ RD data and A22-C8 chemical shift (CS) for extracting an apparent pK_a of A22. Representative pK_a derived from unpaired A29, A30 and A35 is shown in black. Error bars are experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

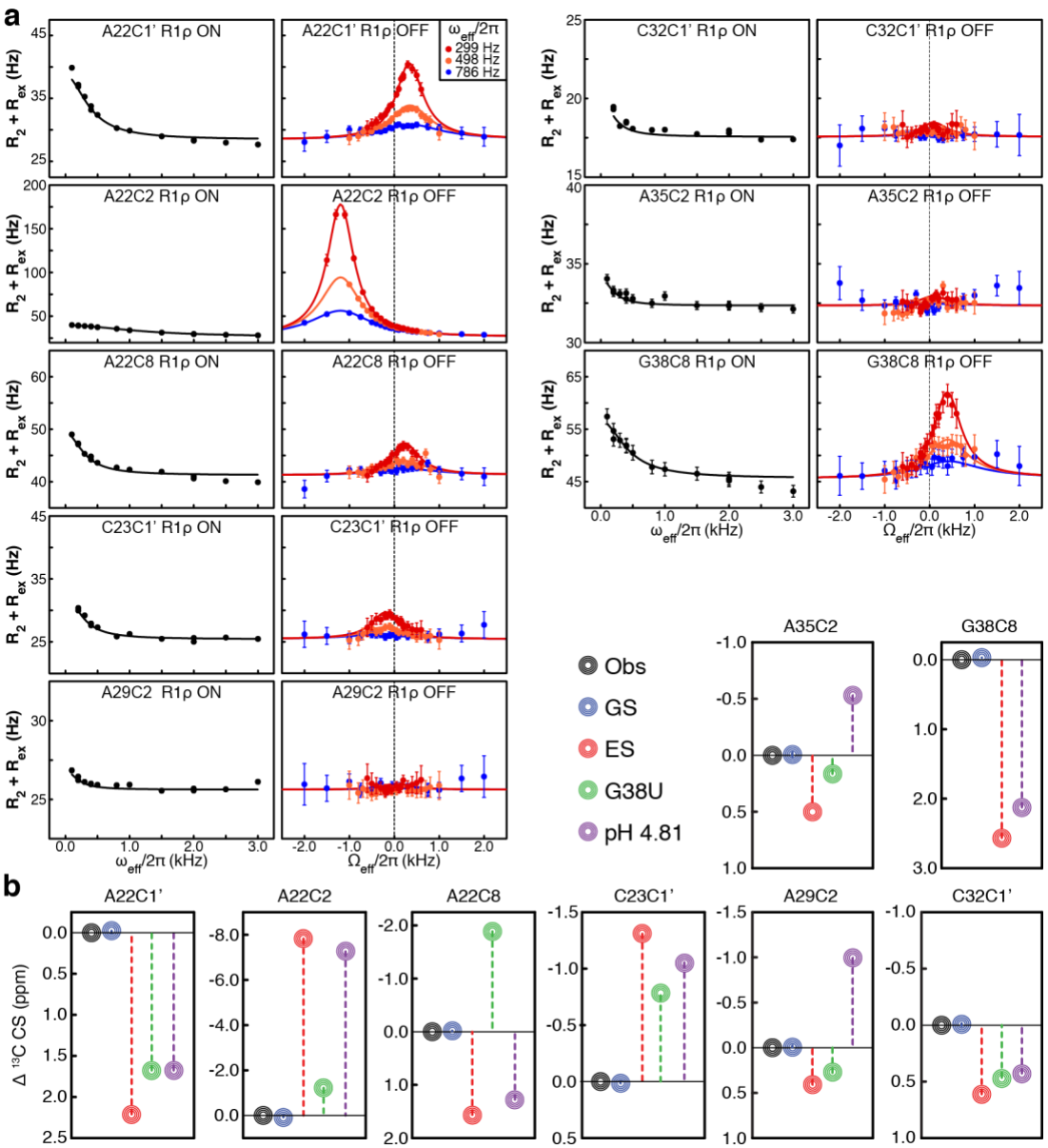

Figure 2.6 - ^{13}C $R_{1\rho}$ RD characterization of preE-miR-21 at pH 8.04.
(**a**) On- and off-resonance ^{13}C RD profiles of residues that are exchange-broadened at lower pH points, including A22-C2, G38-C8, and G38-C1'. Solid lines represent the best fits to a global two-state exchange (k_{ex} = 1228 ± 51 s^{-1} and p_E = 1.1 ± 0.1%) using the Bloch-McConnell equation. (**b**) Comparison of GS, ES, mutant carbon chemical shifts. Shown are differences between observed chemical shifts (black) and chemical shifts of GS (blue) and ES (red) extracted from ^{13}C $R_{1\rho}$ RD profiles at pH 8.04, G38U mutant at pH 6.45, and preE-miR-21 at pH 4.81. Error bars, experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

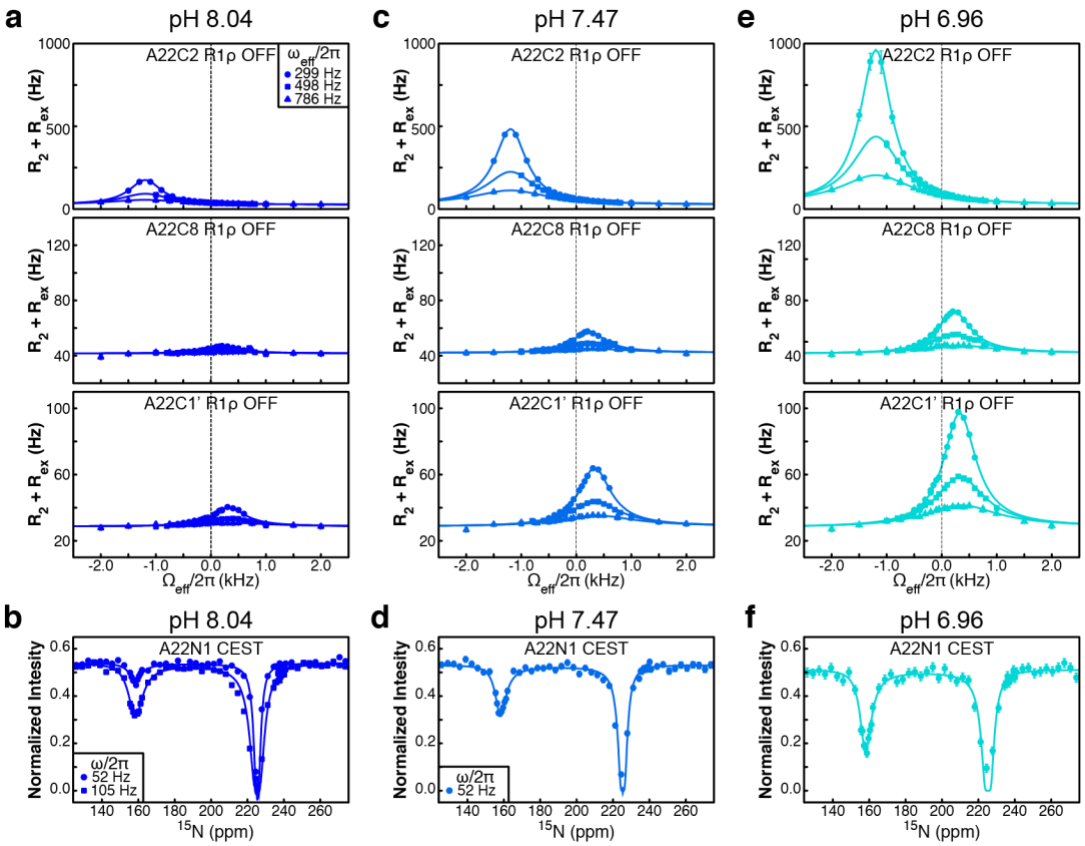

Figure 2.7 - NMR RD characterization of pH-dependent chemical exchange of residue A22. Shown are ^{13}C $R_{1\rho}$ RD profiles of A22-C2/C8/C1' and ^{15}N CEST profile of A22-N1 at (**a-b**) pH 8.04, (**c-d**) pH 7.47, and (**e-f**) pH 6.96. Solid lines represent the best fits to a global two-state exchange model at individual pH condition using the Bloch-McConnell equation, resulting in p_E = 1.1 ± 0.1% at pH 8.04, p_E = 3.4 ± 0.1% at pH 7.47, and p_E = 6.4 ± 0.1% at pH 6.96. Error bars, experimental uncertainties (s.d.) estimated from n = 3 independently measured peak intensities for CEST profiles and mono-exponential fitting of n = 3 independently measured peak intensities for $R_{1\rho}$ RD profiles.

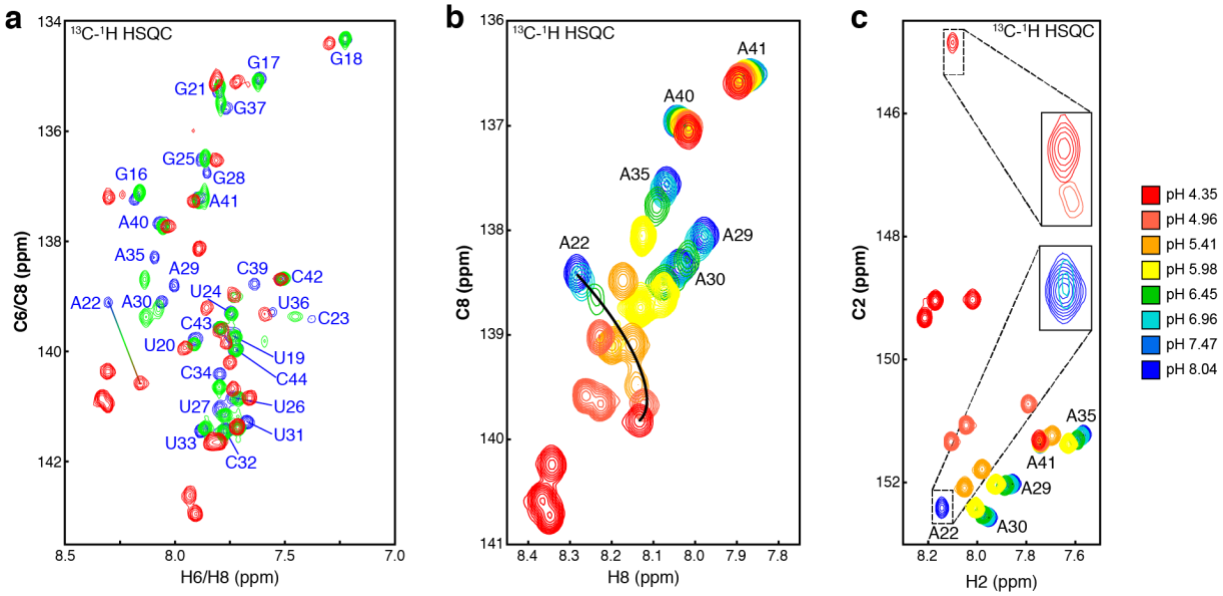

Figure 2.8 - NMR characterization of pH-dependent changes of preE-miR-21.
(a) ^{13}C-^{1}H HSQC spectra of base carbons (C6 and C8) of uniformly ^{13}C/^{15}N-labeled preE-miR-21 at pHs 4.35, 6.45 and 8.04. (**b**) ^{13}C-^{1}H HSQC spectra of base carbons (C8) of adenine-specifically ^{13}C/^{15}N-labeled preE-miR-21 with pH ranging from 4.35 to 8.04. (**c**) ^{13}C-^{1}H HSQC spectra of base carbons (C2) of adenine-specifically ^{13}C/^{15}N-labeled preE-miR-21 with pH ranging from 4.35 to 8.04.

The transient protonation couples global secondary structural reshuffling

How does the excited state stabilize a site-specific protonation? To address this, we first evaluated the role of each structural motif of the pre-element region – the bulge, the stem, and the apical loop – in the observed conformational transition (Fig. 2.9a-c and 2.10). Bulge A22 is the site of protonation. Without A22, we could not detect any conformational exchange within the rest of the pre-element region, as evidenced with flat RD profiles for the A22-deletion mutant (Fig. 2.9a). Not only is a protonated A22 the result of the structural transition, this protonation may likely be the chemical basis that triggers the larger transition across the entire pre-element region. In addition to the indispensable bulge A22, we found that both a weak stem and a flexible apical loop are needed to achieve the structural transition. Stabilizing the two G-U wobble pairs with G-C Watson-Crick pairs completely quenches the exchange (Fig. 2.9b); replacing the apical loop with a highly structured UUCG tetraloop also eliminates the transition (Fig. 2.9c). These results suggest that the pre-element region serves as a unified structural entity to enable a concerted transition towards stabilizing the protonated excited state. The rate of exchange ($k_{ex} = k_{GE} + k_{EG} \sim 1445$ s^{-1}) is an order of magnitude slower than rates observed from local structural changes involving transient adenine protonation[62], but similar to the secondary-structure-based long-range communication observed HIV-1 TAR RNA[63], further supporting a global secondary structural reshuffling of the pre-element region.

To further delineate the secondary structure of the excited state, we employed a mutate-and-chemical-shift-fingerprinting strategy (MCSF)[62]. In this approach, mutations are introduced to stabilize conformational features unique to a proposed/predicted ES secondary structure, which are then validated by comparing chemical shift differences between the mutant and wild-type ($\Delta\varpi_{mut}$) to those extracted from NMR RD profiles ($\Delta\varpi_{RD}$). Here, we used MC-fold[64] to predict

possible alternative low-energy secondary structures of preE-miR-21. Strikingly, most of the predicted structures share a common feature of an A22-G38 base pair, whereas the remaining pre-element adopts various secondary structures that are distinct from the ground-state conformation (Fig. 2.11). Being protonated at N1, A22 could potentially be base paired with G38 in the ES, albeit in the A^+(*anti*)–G(*syn*) form rather than the conventional A(*anti*)–G(*anti*) pair, where the *syn* conformation of G38 is indicated with the down-field chemical shift of base carbon C8 ($\Delta\varpi_{ES}$ = 2.6 ppm) at pH 8.04 (Fig. 2.6). An interesting structural feature of the A^+(*anti*)–G(*syn*) pair is that it largely retains an overall A-form-like geometry with an inter-sugar distance of 10.4 Å[65], whereas the A(*anti*)–G(*anti*) pair substantially widens this distance to 12.9 Å[66] and subsequently distorts the helical geometry of neighboring base pairs (Fig. 2.9d).

To test this proposed ES structural feature, we mutated G38 to a uridine, which not only sequesters A22 into a base pair, but also maintains an A-form geometry at the site of mutation. The G38U mutant converges into a single state as evidenced by flat RD profiles (Fig. 2.9e and 2.12), and largely represents the ES of preE-miR-21 based on chemical shifts. Good agreement was observed between $\Delta\varpi_{G38U}$ and $\Delta\varpi_{RD}$ for 15 out of 19 base and sugar carbon resonances from the pre-element residues with detectable RD profiles, including A22(C1'), C23(C1'), G28(C1'), A29(C2/C8), A30(C8), U31(C1'), C32(C1'), U33(C6), C34(C6), A35(C1'/C2/C8), U36(C6), and C39(C5) (Fig. 2.9f, 2.6 and 2.4). The agreement of sugar carbon C1's of A22 and C23 further supports the ES adopting an A-form-like backbone geometry at the site of protonation. The deviations between $\Delta\varpi_{G38U}$ and $\Delta\varpi_{RD}$ for base carbons C2 and C8 of A22 can be attributed to protonation-induced major chemical shift perturbations in the wild-type, which cannot be recapitulated with this mutation. However, the deviations for C32-C5 and U36-C1' could be due to their relatively small chemical shift differences (<0.5 ppm) and/or local conformational

perturbations in the wild-type from the mutation, which is subject to future investigation. As the G38U mutant closely mimics the ES, it also provides some insights into the two residues (U24 and U26) in the pre-element region that showed no detectable RD. For U24-C6, its flat RD profile can be explained with essentially identical chemical shifts between the ground and excited states as indicated with $\Delta\varpi_{G38U} \sim 0$, whereas the lack of detectable RD for U26-C6 may be due to additional local conformational perturbations that are subject to further studies (Fig. 2.12). To provide independent validation of the chemical shift fingerprints of the ES, we compared $\Delta\varpi_{G38U}$ to chemical shift differences of the wild type between pH 8.04 and pH 4.81 ($\Delta\varpi_{pH}$), where the low pH value was chosen to shift the population towards the protonated ES without inducing global protonation of adenines and cytosines. Good agreement was observed between $\Delta\varpi_{RD}$ and $\Delta\varpi_{pH}$, except for some major deviations from unpaired adenines and cytosines that are likely due to rapid protonation at pH 4.81 given their intrinsic pK_as (~ 3.5 – 4.2) when unpaired (Fig. 2.9f, 2.4 and 2.6). In particular, excellent agreement between $\Delta\varpi_{RD}$ and $\Delta\varpi_{pH}$ of A22-C2 and A22-C8 complements the G38U mutant. Taken together, low pH and G38U are each able to recapitulate a portion of the ES structure, with low pH chemical shifts matching changes in A22 residues, and G38U chemical shifts matching throughout the rest of the structure. These results strongly suggest that pre-miR-21 undergoes a global structural reshuffling at the pre-element region, where A22 is transiently protonated and forms a distinct A^+(*anti*)–G(*syn*) base pair in the ES.

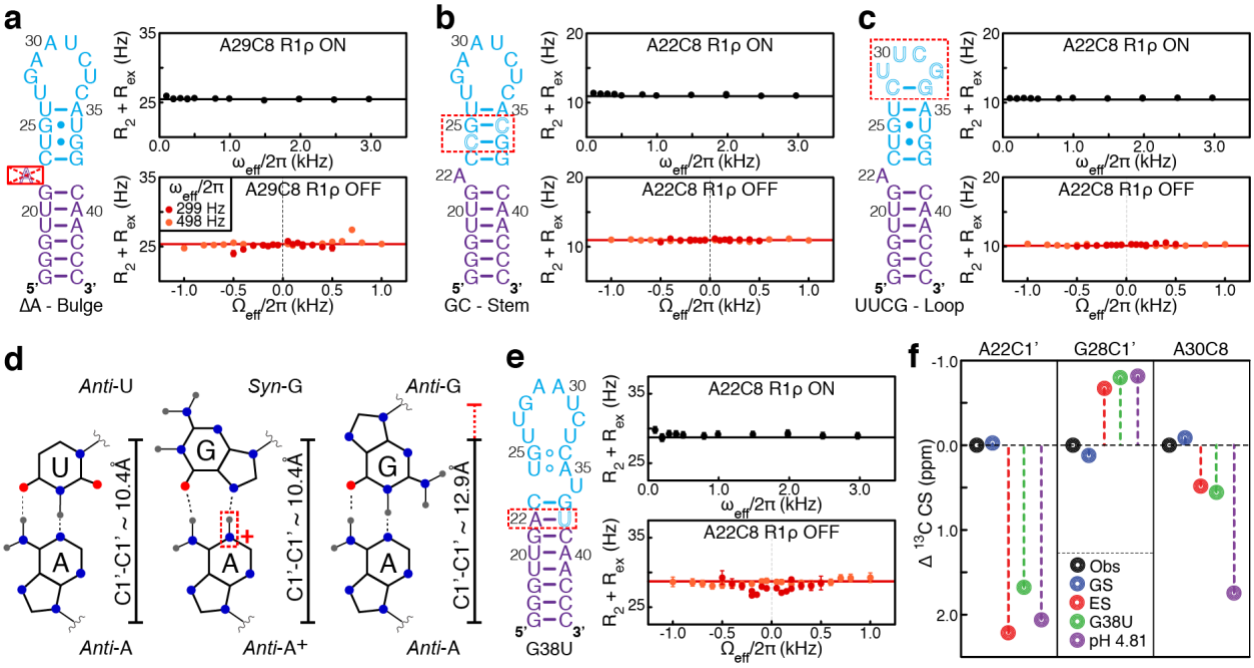

Figure 2.9 - Excited-state structure of preE-miR-21

(**a-c**) Secondary structures and representative $_{13}$C on-resonance and off-resonance RD profiles of bulge, stem, and loop mutants. (**d**) Sugar-sugar (C1'-C1') distances of A-U Watson-Crick base pair, A-G mismatch, and A$_+$-G mismatch. (**e**) Secondary structure and representative $_{13}$C on-resonance and off-resonance RD profiles of ES-mimic mutant. (**f**) Comparison of carbon chemical shifts for the GS, ES, the mutant mimics, and wild-type construct at pH 4.81. Error bars are experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

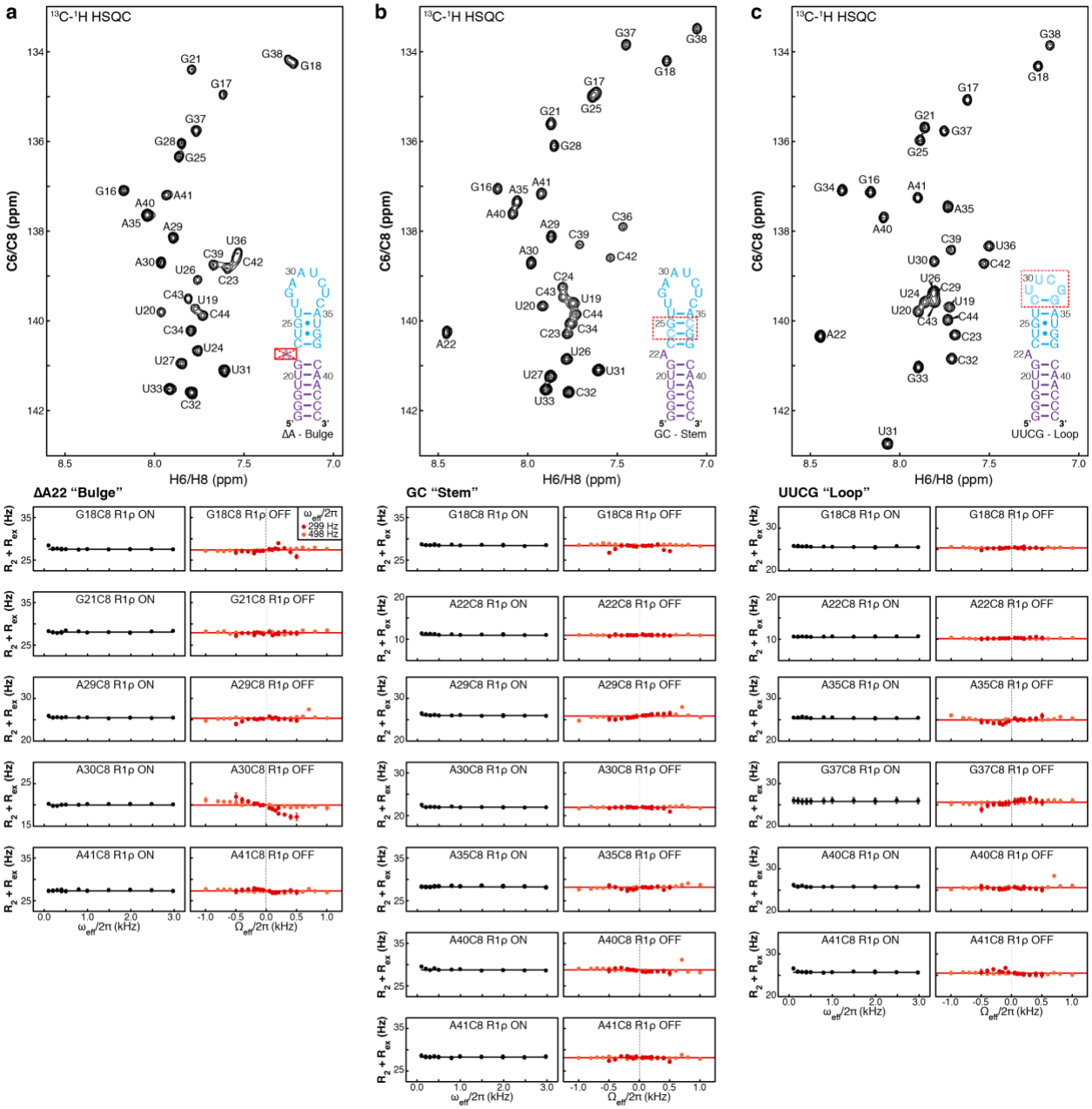

Figure 2.10 - NMR characterization of preE-miR-21 mutants.
Shown are 13C-1H HSQC spectra of base carbons (C6 and C8) and 13C R1ρ RD profiles of uniformly 13C/15N-labeled (a) ΔA22 "bulge", (b) GC "Stem", and (c) UUCG "Loop" mutants at pH 6.45. Solid lines represent the best fits to a single-state model using the Bloch-McConnell equation. Error bars, experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

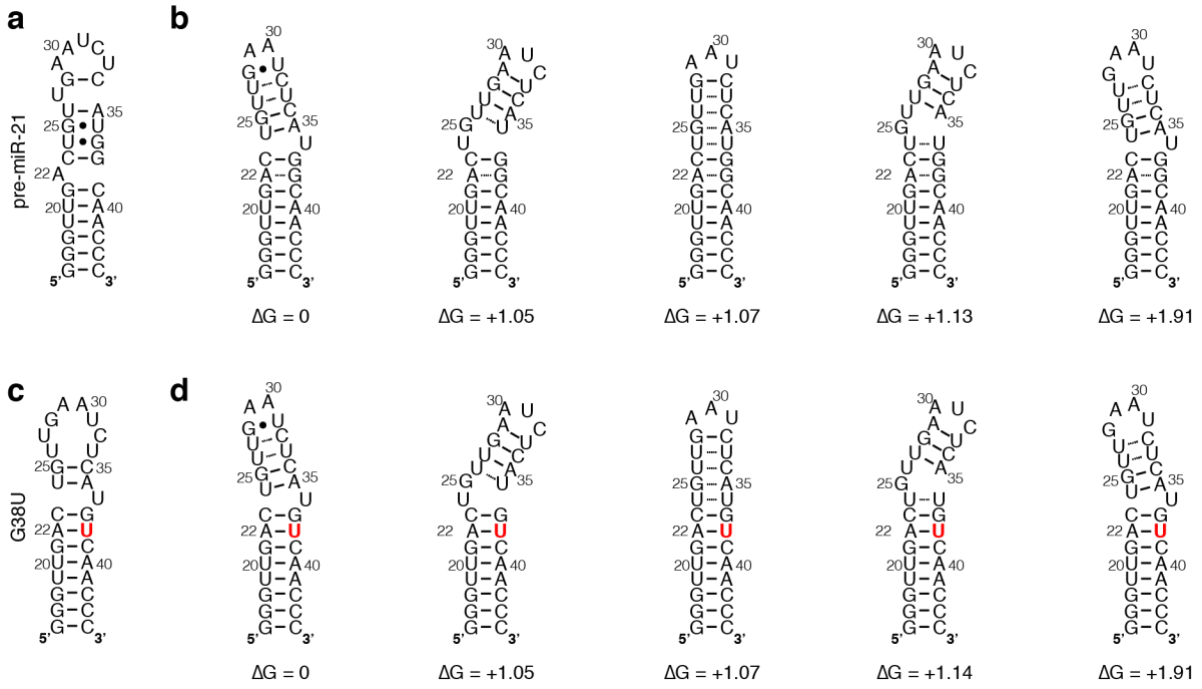

Figure 2.11 - Secondary structure prediction of preE-miR-21 and mutant by Mfold and MC-Fold.

(**a**) M-fold predicted secondary structure of preE-miR-21, which contains G-U base pairs as observed in NMR data of the ground state. (**b**) MC-Fold predicted secondary structures of preE-miR-21. Shown are the top 5 lowest-energy structures, ranked with ΔG relative to the lowest predicted structure. The A22-G38 base pair is predicted in all MC-Fold structures. (**c**) M-fold predicted secondary structure of preE-miR-21 G38U mutant. (**d**) MC-Fold predicted secondary structures of preE-miR-21 G38U mutant. Shown are the top 5 lowest-energy structures, ranked with ΔG relative to the lowest predicted structure. Both M-fold and MC-Fold predict the same lower stem structure of preE-miR-21 G38U mutant, which is also consistent with NMR data.

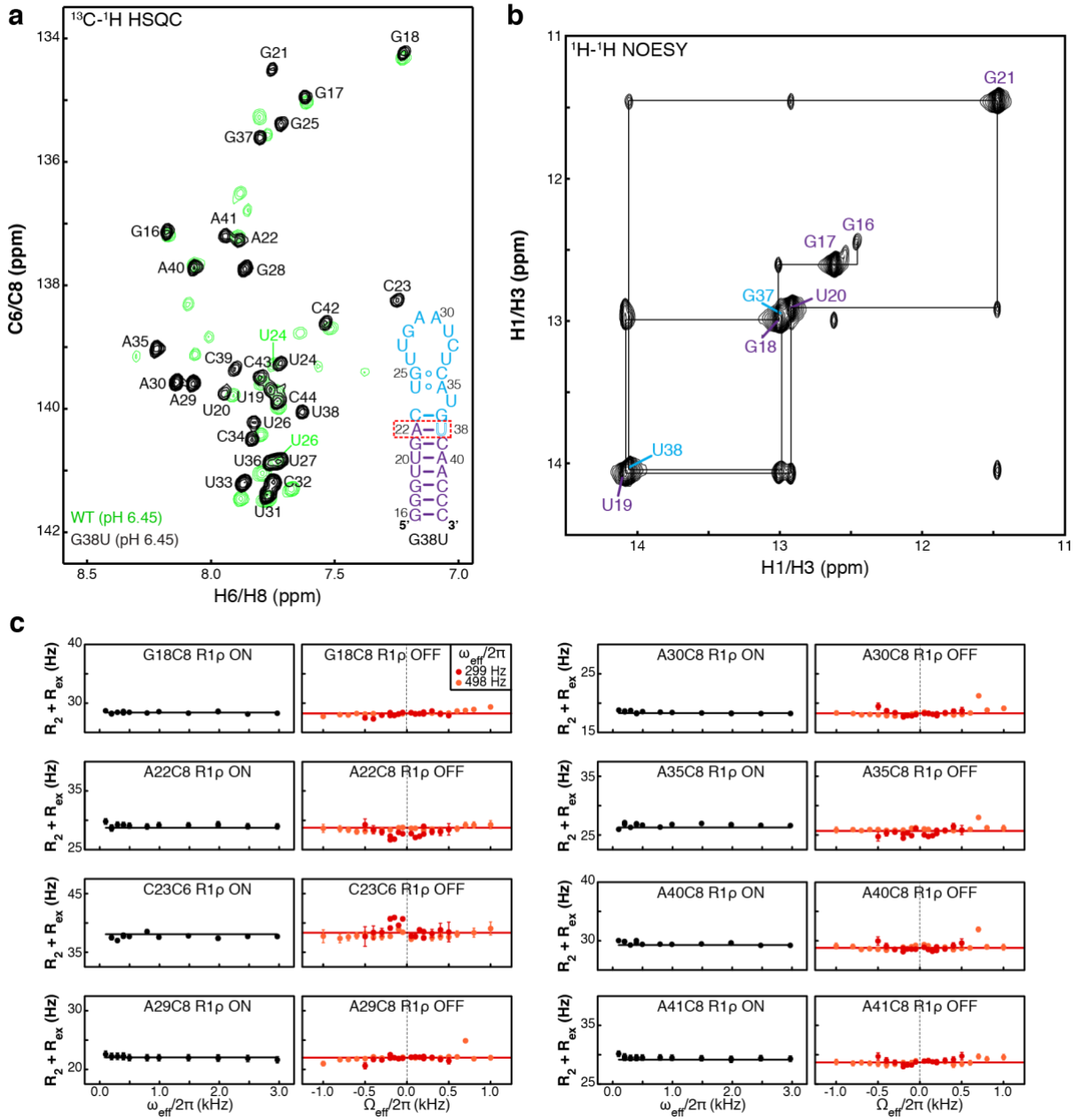

Figure 2.12 - NMR characterization of preE-miR-21 excited-state mimic.
Shown are (a) 13C-1H HSQC spectrum of base carbons (C6 and C8) (black) that overlays with 13C-1H HSQC spectrum of WT preE-miR-21 at pH 6.45 (green) (b) 1H-1H imino NOESY spectrum, and (c) 13C R1ρ RD profiles of uniformly 13C/15N-labeled G38U mutant at pH 6.45. Solid lines represent the best fits to a single-state model using the Bloch-McConnell equation. Error bars, experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

The miR-21 precursor encodes states with differential Dicer processivities

The conformational transition of pre-miR-21 is the protonation-driven base pairing of the bulged A22 that resides specifically at the location of Dicer cleavage. Hence, it is of interest to see how these structural changes may affect Dicer cleavage of the miR-21 precursor. To examine this, we designed GS- and ES-mimicking substrates and performed processing assays using commercially available recombinant human Dicer (Fig. 2.13a). For the GS-mimicking substrate (pre-miR-21$_{GS}$), we mutated the two G–U wobbles with two G–C pairs, which was shown to stabilize the flexible stem that forms in the ground state. For the ES-mimicking structure (pre-miR-21$_{ES}$), we incorporated a G38U mutation to the full-length pre-miR-21. We would like to note that the G38U mutant, which recapitulates key ES structural features, cannot perfectly mimic the electrostatic property of the protonated ES. However, since Dicer cleaves the phosphate backbone, we anticipate structures of the backbone, rather than the base pairing identity of the ES, may influence Dicer activity. In order to generate a native-like precursor with 5'-terminal phosphate group and sequence, we fused a hammerhead ribozyme to the 5'-end of full-length miR-21 precursor. During *in vitro* transcription, the hammerhead ribozyme self-cleaves, and the resulting miR-21 precursor is 5'-end phosphorylated and labelled with [γ-^{32}P] ATP. Remarkably, GS- and ES-mimicking substrates exhibited substantially different Dicer processivities (Fig. 2.13b). As can be seen, ~ 23 ± 6% of pre-miR-21$_{GS}$ was cleaved by Dicer to generate mature miR-21, which is essentially identical to a processivity of ~ 26 ± 9% for the wild-type substrate (pre-miR-21$_{WT}$). This is consistent with pre-miR-21$_{WT}$ occupying ~99% GS under the assay condition of pH 8.04. In contrast, Dicer processed pre-miR-21$_{ES}$ much more efficiently than its GS counterpart, where double the amount of substrate (~ 48 ± 6%) was converted to mature miR-21. Together, these results not only unveil differential fitness of the GS and ES of the pre-element region of the pre-

miR-21 for miR-21 maturation, but further exemplify the importance of RNA structures in directing the overall outcome of miRNA biogenesis.

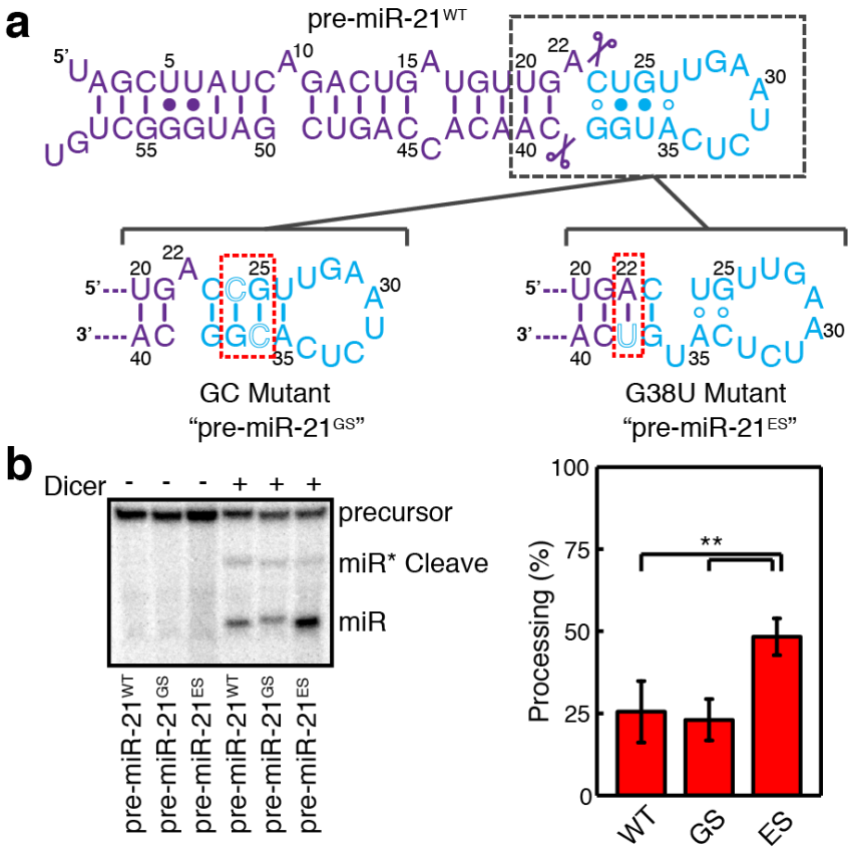

Figure 2.13 - Excited state of preE-miR-21 enhances dicer processing
(**a**) Secondary structures of the wild-type, ground-state mimic, and excited-state mimic of pre-miR-21. (**b**) Dicer processing assays of wild-type and mutant pre-miR-21s, quantification using ImageQuant shown on right (** denotes $p \leqslant 0.01$ via student T-Test). Shown are means and standard deviations (s.d.) from $n = 4$ independent assays.

Discussion

Here, by integrating structural, dynamic, and functional analyses on miR-21 precursor, we showed that the intrinsic conformational plasticity of miRNA processing intermediates can serve as a new layer of regulation for miRNA biogenesis. Often, structural changes in pri-/pre-miRNAs can be induced upon binding to protein regulators[22], nucleotide modifications such as ADAR1-mediated adenine-to-inosine editing[23] and METTL1-mediated methylation[24], and disease-linked mutations[25]. RNA structural motifs are also important factors in recognition by processing machineries for miRNA biogenesis, where altering primary and/or precursor structures of a target miRNA can further lead to altered biogenesis, inducing a change in physiological outcomes[3]. In contrast to these adaptive changes, we found that pre-miR-21 encodes a dynamic ensemble in its apical stem-loop region that undergoes spontaneous structural transitions between two kinetically and functionally distinct states (Fig. 2.14). In the ground state, the two Dicer cleavage sites reside within a largely unstructured region; in contrast, both locations become structured in the excited state. Relative to the ground state, the excited state conformation more closely resembles the optimal structure for Dicer, where both cleavage sites are base-paired and positioned two-nucleotides away from a flexible apical loop[20], hence, providing a better topology for Dicer cleavage.

A hallmark of the ES of pre-miR-21 is protonation of the adenine residue at the Dicer cleavage site. Protonation is a fundamental chemical property and one of the smallest chemical modifications on nucleic acids[61]. The intrinsic pK_as for protonation of adenines and cytosines are acidic and reside far from the physiological pH ranges. However, by adopting sophisticated structures, RNA can shift acidic pK_as toward neutral pHs, such that specific ionization can be achieved under physiological conditions for function[61]. For example, the universally conserved

adenine residue at the active site of the ribosome has a shifted pK_a to serve as a general acid-base catalyst for peptide formation[67], whereas the murine leukemia virus (MLV) recoding signal employs a protonated adenine as a structural factor to stabilize a compact pseudoknot in order to allow the ribosome to bypass the Gag stop codon[58]. Here, protonation provides the crucial chemical basis for pre-miR-21 to form the A^+(*anti*)–G(*syn*) base pair, which ensures the cleavage site adopting an overall A-form-like topology in the ES. Without being protonated, the adenine residue may still be able to pair with the upper stem guanine residue, albeit in the form of A(*anti*)–G(*anti*) mismatch, which is likely functionally indistinguishable from the GS. Another feature of the protonation event in pre-miR-21 is that the underlying structural transition occurs at the millisecond timescale, which is substantially faster than those involved in major structural changes, such as the adenine protonation in MLV[58]. This fast GS ↔ ES interconversion could enable pre-miR-21 to rapidly reach new equilibrium upon a transient high acid load due to disease-induced metabolic shifts, modulating maturation of miR-21 in response to environmental stimuli. Despite displaying distinct *in vitro* outcomes, a functional understanding of the role of pre-miR-21 protonation in regulating biogenesis will require future investigations that evaluate the response of pre-miR-21 under various cellular conditions such as hypoxia and acidosis.

Interestingly, the spontaneous conformational transition in pre-miR-21, which involves secondary structural reshuffling of the pre-element region, is reminiscent of that observed in Lin28-dependent regulation of the biogenesis of let-7 family of miRNAs[68]. In general, RNA secondary structural changes encounter large kinetic barriers, hence, need to be catalyzed by external factors such as RNA-binding proteins. Both domains of Lin28 work cooperatively to bind two independent RNA elements to induce the regulatory structural changes of pri-/pre-let-7s. In contrast, pre-miR-21 accomplishes such structural changes without protein factors by accessing

an excited state, where the bulged adenine base pairs with the upper stem guanine residue, propagating a global change in strand register. Here, all three structural elements of the apical stem-loop – the bulge, metastable stem, and flexible loop – are essential to achieve this concerted movement, and eliminating any of them abolishes the spontaneous transitions in pre-miR-21. While our observation represents the first example of an excited state in pre-miRNAs, an ES-based mechanism for remodeling distant motifs has also been recently reported in HIV-1 TAR RNA[63]. Since the apical stem-loop is a common structural feature among all pri-/pre-miRNAs, we speculate that many miRNA processing intermediates may encode similar ES-based conformational plasticity for long-range communication across their regulatory pre-element regions, which is further supported by a recent computational modeling of secondary structural ensembles of miRNAs[25].

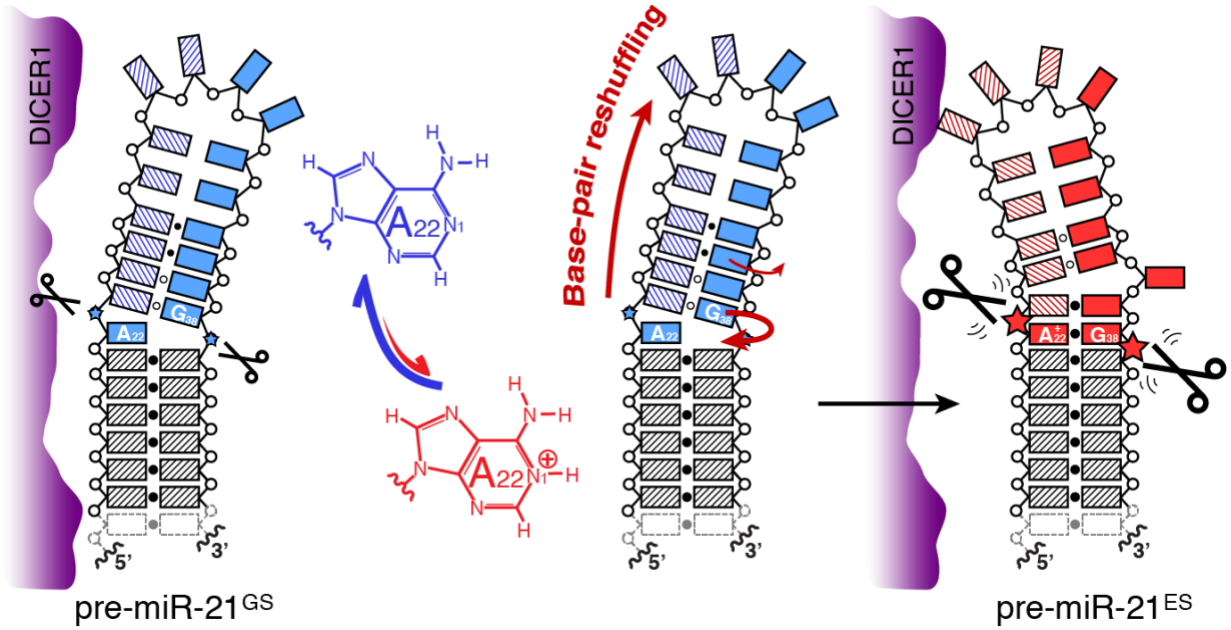

Figure 2.14 - Modulation of miR-21 maturation with a protonation-dependent structural ensemble

Protonation at N1 site of A22 sequesters the bulged adenine into a non-canonical A+–G mismatch and is associated with a long-range conformational reshuffling of the pre-element region. This structural rearrangement results in a conformation that is better suited for Dicer processing to generate mature miR-21.

Methods

Sample preparation

Unlabeled, uniformly $^{13}C/^{15}N$-labeled, and adenine-specifically $^{13}C/^{15}N$-labeled preE-miR-21 samples and mutants were prepared as previously described[69]. Briefly, after *in vitro* transcription, samples were ethanol precipitated, gel purified using 15% denaturing polyacrylamide gel, electro-eluted with the Elutrap system (Whatman), purified with an 5 mL Hi-Trap Q anion-exchange column (GE Healthcare), and desalted by exchanging into H_2O using an Amicon filtration unit with 3K Da MW cut-off membrane (Millipore). Samples were then concentrated and exchanged into NMR buffers with 50 mM KCl and 50 μM EDTA, where 10 mM acetate buffers were used for pHs at 4.35, 4.96, and 5.41, and 10 mM sodium phosphate buffers were used for pHs at 5.98, 6.45, 6.96, 7.47, and 8.04. For H_2O samples, 5% D_2O was added. For D_2O samples in sodium phosphate buffer, H_2O samples were lyophilized and redissolved in the same volume of 99.996% D_2O (Sigma). For D_2O samples in sodium acetate buffer, samples were exchanged into 10 mM acetate buffer with 50 mM KCl and 50 μM EDTA in D_2O.

NMR spectroscopy

All NMR experiments were carried out on a Bruker Avance III 600 MHz spectrometer equipped with a 5-mm triple-resonance (TCI) cryogenic probe. Exchangeable proton spectra were recorded using H_2O samples at 283 K, and nonexchangeable proton spectra were recorded at 298 K and 308 K using H_2O and D_2O samples. Spectra were processed and analyzed with TOPSPIN 3.5 (Bruker), NMRPipe[70], NMRView[71], and Sparky 3.110 (University of California, San Francisco, CA). The assignments for preE-miR-21 were obtained with samples at pHs 6.45 and 8.04 and the assignments for its mutants were obtained with samples at pH 6.45 using 2D NOESY, 2D TOCSY, $^{1}H-^{15}N$ HSQC, $^{1}H-^{13}C$ HSQC, 2D HCCH-COSY, and HCN experiments on

unlabeled, uniformly labeled and adenine-specifically $^{13}C/^{15}N$ labeled samples with standard protocols[72]. The apparent pKa values of adenines were obtained by fitting pH-dependent excited-state population and chemical shift to the Henderson-Hasselbalch equation as described previously[73].

^{13}C $R_{1\rho}$ relaxation dispersion measurements and data analysis

The on- and off-resonance relaxation dispersion profiles were measured using the 1D selective $R_{1\rho}$ experiment developed by Al-Hashimi and co-workers[55] and a constant-time approach described by Kay and co-workers[54], where $R_{1\rho}$ values were obtained from a single delay period (T_{EX}). For on-resonance experiments, the relaxation delay was set to T_{EX} = 32 ms; for off-resonance experiments, the relaxation delay was set to T_{EX} = 40 ms, except T_{EX} = 16ms for A22-C8 and A35-C1' and T_{EX} = 24 ms for C23-C6, U36-C6, and G28-C1'. Relaxation rates were determined by fitting peak intensity to a single exponential decay as $R_{1\rho} = -\ln(I_T/I_0)/T_{EX}$, where I_T is the decayed peak intensity and I_0 is the reference peak intensity. Relaxation rate errors were estimated by intensity deviations between three duplicates at T_{EX} = 0 and the signal-to-noise ratios in 1D spectra. The largest of the two errors was reported. For on-resonance experiments, eleven ^{13}C spin-lock fields ($\omega/2\pi$) of 100, 199 (x2), 299, 399, 498 (x2), 786, 982, 1474, 1965 (x2), 2456, and 2947 Hz were employed, (x2) indicates performed in duplicates. Due to large C-C couplings, the lowest ^{13}C spin-lock field ($\omega/2\pi$) of 100 Hz was not used in measuring on-resonance C5/C6/C1' RD profiles. For off-resonance experiments, three ^{13}C spin-lock fields ($\omega/2\pi$) of 299, 498, and 786 Hz were used. For $\omega/2\pi$ = 299 Hz, the ^{13}C offsets ranged between -600 and 600 Hz with a spacing of 100 Hz and a smaller spacing of 50 Hz between -200 and 200 Hz; for $\omega/2\pi$ = 498 Hz, the ^{13}C offsets ranged between -1000 and 1000 Hz with a spacing of 200 Hz from -1000 to -800 Hz and from 800 to 1000 Hz, a spacing of 100 Hz from -800 to -100 Hz and from 100 to 800 Hz, and single points

at -50 and 50 Hz; for $\omega/2\pi$ = 786 Hz, the ^{13}C offsets ranged between -2000 and 2000 Hz with a spacing of 500 Hz from -2000 to -1000 Hz and 1000 to 2000 Hz, a spacing of 250 Hz from -1000 to -500 Hz and 500 to 1000 Hz, and a spacing of 100 Hz from -500 to -100 Hz and from 100 to 500 Hz, and single points at -50 and 50 Hz. ^{13}C spin-lock powers were calibrated according to the 1D approach by Guenneugues *et al*[74]. as previously described[69,75].

Relaxation dispersion profiles were analyzed as described previously[69]. Briefly, on- and off-resonance relaxation dispersion profiles were obtained by measuring the rate of decay of magnetization over the spin-lock period as a function of spin-lock power ($\omega_{eff}/2\pi$) and spin-lock offset (Ω), respectively, where $\Omega = \omega_{rf} - \Omega_{obs}$ is the frequency difference between the spin-lock carrier frequency (ω_{rf}) and the observed resonance frequency (Ω_{obs}). The $R_{1\rho}$ profiles of residues displaying conformational exchange were fit to a two-state exchange model between the ground (G) and excited (E) states based on the Bloch-McConnell equation[76],

$$\frac{d}{dt}\begin{pmatrix} I_x^G \\ I_y^G \\ I_z^G \\ I_x^E \\ I_y^E \\ I_z^E \end{pmatrix} = \begin{pmatrix} -R_2^G - k_{GE} & -\omega_G & 0 & k_{EG} & 0 & 0 \\ \omega_G & -R_2^G - k_{GE} & -\omega_1 & 0 & k_{EG} & 0 \\ 0 & \omega_1 & -R_1^G - k_{GE} & 0 & 0 & k_{EG} \\ k_{GE} & 0 & 0 & -R_2^E - k_{EG} & -\omega_E & 0 \\ 0 & k_{GE} & 0 & \omega_E & -R_2^E - k_{EG} & -\omega_1 \\ 0 & 0 & k_{GE} & 0 & \omega_1 & -R_1^E - k_{EG} \end{pmatrix} \begin{pmatrix} I_x^G \\ I_y^G \\ I_z^G \\ I_x^E \\ I_y^E \\ I_z^E \end{pmatrix}$$

where $R_{1G/E}$ is the longitudinal relaxation rate of the ground/excited state, $R_{2G/E}$ is the transverse relaxation rate of the ground/excited state, $\omega_{G/E}$ is the offset of the applied ^{13}C spin lock with a strength of ω_1 from the chemical shift ($\Omega_{G/E}$) of the ground/excited state, and k_{GE} and k_{EG} are forward and backward exchange rates as defined by $k_{GE} = p_E k_{ex}$ and $k_{EG} = p_G k_{ex}$. Here, $k_{ex} = k_{GE} + k_{EG}$ is the rate of exchange, p_G and p_E are populations of ground and excited states, respectively, and $\Omega_{obs} = p_G \Omega_G + p_E \Omega_E$ and $\Omega_E = \Omega_G + \Delta\omega$, where $\Delta\omega$ is the chemical shift difference between the ground and excited states. Ground state and excited state magnetizations at the beginning of

the T_{EX} period are along the effective spin-lock field as, $I_{xG/E} = p_{G/E} \sin(\theta)$, $I_{yG/E} = 0$, $I_{zG/E} = p_{G/E} \cos(\theta)$, where $\theta = \arctan(\omega_1/\Omega)$ is the effective tilt angle. Fitting parameters are $R_1 = R_{1G/E}$, $R_2 = R_{2G/E}$, $\Delta\omega$, k_{ex}, and p_E, where we assume $R_{1G} = R_{1E}$ and $R_{2G} = R_{2E}$. Since the applied ^{13}C spin-lock powers are strong enough to decouple C-C couplings, the relaxation dispersion profiles of C1', C5, and C6 were analyzed the same as C2 and C8 profiles. For global fitting of dispersion profiles at individual pH condition, spin specific R_1, R_2, and $\Delta\omega$ were used, whereas k_{ex} and p_E were fit globally. For residues without conformational exchange, the two-state model was simplified to a one-state model by fixing all exchange parameters (rate of exchange k_{ex} and population of excited state p_E) to 0. All profiles were fitted using an in-house OriginLab® program with a Levenberg-Marquardt algorithm.

^{15}N CEST measurements and data analysis

^{15}N CEST profiles were measured using a recently developed $^2J_{NH}$-based 2D ^1H-^{15}N HSQC CEST experiment pulse sequence that monitors longitudinal two-spin order ($2N_zH_z$) as described previously[60]. The ^{15}N carrier was set to 224.9 ppm with a spectral width of 6 ppm, and the ^{15}N offsets ranged between -6000 Hz and 3000 Hz with a spacing of 200 Hz, except for a of spacing of 50 Hz from -4200 to -3800 Hz and from 600 to 1000 Hz. With a relaxation period of T_{EX} = 0.1 s, two ^{15}N B_1 fields ($\omega/2\pi$) of 52.1 Hz and 104.7 Hz were used at pH 8.04 and one ^{15}N B_1 field ($\omega/2\pi$) of 52.1 Hz was used at pH 7.47 and 6.96. ^{15}N spin-lock powers were calibrated according to the 1D approach by Guenneugues *et al*[74]. as previously described[69,75]. For all measurements, three spectra with T_{EX} = 0 s were recorded for reference in data fitting and error estimation.

CEST profiles were analyzed as described previously[60]. Briefly, CEST profiles were obtained by normalizing peak intensity as a function of spin lock offset Ω to the peak intensity recorded at $T_{EX} = 0$, where $\Omega = \omega_{rf} - \Omega_{obs}$ is the difference between the spin-lock carrier (ω_{rf}) and

the observed peak (Ω_{obs}) frequencies. Measurement errors were estimated based on both triplicates performed at T_{EX} = 0 s and the baseline of CEST profiles. Two-spin order ^{15}N CEST profiles for A22-N1 at various pHs displaying conformational exchange were fit to a two-state exchange model between the ground (G) and excited (E) states based on the Bloch-McConnell equation[76] that describes magnetization evolution in a coupled two-spin ^{15}N-^{1}H system[77,78],

$$\frac{d}{dt}\mathbf{v}^{G/E} = -\mathbf{R}^{G/E}\mathbf{v}^{G/E} = \begin{pmatrix} R_2^{G/E} & \omega_N^{G/E} & 0 & \eta_{xy}^{G/E} & \pi J_{NH}^{G/E} & 0 \\ -\omega_N^{G/E} & R_2^{G/E} & \omega_1 & -\pi J_{NH}^{G/E} & \eta_{xy}^{G/E} & 0 \\ 0 & -\omega_1 & R_1^{G/E} & 0 & 0 & \eta_z^{G/E} \\ \eta_{xy}^{G/E} & \pi J_{NH}^{G/E} & 0 & R_{2HN}^{G/E} & \omega_N^{G/E} & 0 \\ -\pi J_{NH}^{G/E} & \eta_{xy}^{G/E} & 0 & -\omega_N^{G/E} & R_{2HN}^{G/E} & \omega_1 \\ 0 & 0 & \eta_z^{G/E} & 0 & -\omega_1 & R_{1HN}^{G/E} \end{pmatrix} \begin{pmatrix} N_x^{G/E} \\ N_y^{G/E} \\ N_z^{G/E} \\ 2H_zN_x^{G/E} \\ 2H_zN_y^{G/E} \\ 2H_zN_z^{G/E} \end{pmatrix}$$

$$\frac{d}{dt}\boldsymbol{\sigma}(\mathrm{t}) = -\mathbf{L}\cdot\begin{bmatrix}\mathbf{v}^G \\ \mathbf{v}^E\end{bmatrix} = \left(\begin{bmatrix}\mathbf{R}^G & \mathbf{0}_6 \\ \mathbf{0}_6 & \mathbf{R}^E\end{bmatrix} + \begin{bmatrix}-k_{GE} & k_{EG} \\ k_{GE} & -k_{EG}\end{bmatrix} \otimes \mathbf{1}_6\right)\cdot\begin{bmatrix}\mathbf{v}^G \\ \mathbf{v}^E\end{bmatrix}$$

where $\mathbf{v}_{G/E}$ is the magnetization matrix, $\mathbf{R}_{G/E}$ is the relaxation matrix, $R_{1G/E}$ is the ^{15}N longitudinal relaxation rate, $R_{2G/E}$ is the ^{15}N transverse relaxation, $R_{1HNG/E}$ is the ^{15}N-^{1}H two-spin order relaxation rate, $R_{2HNG/E}$ is the ^{15}N antiphase relaxation rate, $\eta_{zG/E}$ is the N-H dipolar-dipolar/nitrogen CSA cross-correlated relaxation between the ^{15}N longitudinal and two-spin order elements, $\eta_{xyG/E}$ is N-H dipolar-dipolar/nitrogen CSA cross-correlated relaxation between ^{15}N transverse and antiphase magnetizations, $\omega_{NG/E}$ is the offset of the applied ^{15}N B_1 field with a strength of ω_1, and $J_{NHG/E}$ is the ^{15}N-^{1}H scalar coupling for the ground (G) and excited (E) states, and k_{GE} and k_{EG} are forward and backward exchange rates as defined by $k_{GE} = p_E\, k_{ex}$ and $k_{EG} = p_G\, k_{ex}$. Here, $k_{ex} = k_{GE} + k_{EG}$ is the rate of exchange, p_G and p_E are populations of ground and excited states, respectively, and $\omega_G = \Omega_{obs}$ and $\omega_E = \omega_G + \Delta\omega$, where $\Delta\omega$ is the chemical shift difference between the ground and excited states. Ground state and excited state magnetizations at the

beginning of the T_{EX} period are the two-spin order ($2N_zH_z$) along Z and are set to be at populations of p_G and p_E. The fitting parameters are $\Delta\omega$, k_{ex}, p_E, $R_2 = R_{2G/E}$, $R_{1HN} = R_{1HNG/E}$, and $R_{2HN} = R_{2G/E} + R_{1HNG/E} - R_{1G/E}$ as described previously[60,79]. To simplify data fitting, $\eta_{zG/E}$ and $\eta_{xyG/E}$ were set to 0 as they have been shown not to affect the extracted $\Delta\omega$, k_{ex}, and p_E values[60,80], and $R_{1G/E}$ were set to 0, as the data does not constrain determination of R_{160}. For residues without conformational exchange, the two-state model was simplified to a one-state model by fixing all exchange parameters (rate of exchange k_{ex} and population of excited state p_E) to 0. All profiles were fitted using an in-house OriginLab® program with a Levenberg-Marquardt algorithm.

Preparation of Dicer substrates

Pre-miR-21 substrates were prepared by *in vitro* transcription, where a hammerhead ribozyme (HH) was fused to the 5'-end of pre-miR-21s to generate 5'-OH substrates with wild-type 5'-end nucleotide. Sequences of the pre-miR-21 substrates are as follows,

HH-Pre-miR-21 WT:
CUAAUACGACUCACUAUAGGAGCUACUGAUGAGGCCGAAAGGCCGAAACCCGAAA
GGGUCUAGCUUAUCAGACUGAUGUUGACUGUUGAAUCUCAUGGCAACACCAGUC
GAUGGGCUGUC

HH-Pre-miR-21 GC Mutant:
CUAAUACGACUCACUAUAGGAGCUACUGAUGAGGCCGAAAGGCCGAAACCCGAAA
GGGUCUAGCUUAUCAGACUGAUGUUGACCGUUGAAUCUCACGGCAACACCAGUCG
AUGGGCUGUC

HH-Pre-miR-21 G38U Mutant:
CUAAUACGACUCACUAUAGGAGCUACUGAUGAGGCCGAAAGGCCGAAACCCGAAA
GGGUCUAGCUUAUCAGACUGAUGUUGACUGUUGAAUCUCAUGUCAACACCAGUC
GAUGGGCUGUC

After *in vitro* transcription, samples were exchanged into H_2O using Amicon filtration units (Millipore) with 10K Da MW cut-off membranes. The resulting substrates were subsequently subject to 5'-end ^{32}P-labeling with T4 Polynucleotide Kinase (T4PNK) (New England Biolabs Inc.). 50 µl of RNA at 400 nM was incubated in a solution containing T4PNK buffer, 20 U T4PNK,

and γ-^{32}P ATP at 35°C for 30 minutes before being gel purified by 20% denaturing polyacrylamide gel. The ^{32}P-labeled RNAs were extracted from the gel by crushing and soaking in 250 mM NaCl in 1xTBE buffer for 24 hours before filtering and exchanging multiple times into H_2O using an Amicon filtration unit (Millipore) with a 10K Da MW cut-off. ^{32}P-labeled substrates were stored at -20°C prior to dicer processing assay.

Dicer processing assays

^{32}P-labeled pre-miR-21 substrates were heated to 95°C for 5 minutes and then placed on ice for 5 minutes to anneal the precursor hairpin. 6 μl substrate was mixed with 2 μl of 5x dicing buffer (12 mM HEPES, 1 M NaCl, and 0.02 mM EDTA at pH 7.8) and 2 μl of 25 mM $MgCl_2$. In a separate tube, concentrated Dicer enzyme (Genlantis, Inc.) was diluted in 1x dicing buffer (2.4 mM HEPES, 200 mM NaCl, 4 μM EDTA) and mixed with an equal volume of 10 mM ATP. 2 μl of ATP/Dicer mixture (10 mM ATP and 0.2U Dicer) was added to each tube of 6 μl substrate and incubated at 35°C overnight. Reactions were quenched by adding 1.6 U proteinase K (New England Biolabs Inc.), 2 μl of 0.5 M EDTA and incubated for 45 minutes at 35°C. Samples were denatured in formamide with trace bromophenol blue and xylene cyanol, ran on a 15% denaturing polyacrylamide gel, visualized using Amersham Typhoon 5 Biomolecular Imager (GE Healthcare) and analyzed with Image Quant (GE Healthcare) using peak area and rubberbanding for background deletion. All assays were carried out in quadruplicate to estimate experimental errors.

RNA secondary structure prediction

All RNA secondary structures were predicted based on sequences using programs Mfold[81] and MC-Fold[64] using standard input options.

CHAPTER 3 - MIR-21 PRECURSOR AND DICER

Introduction

The human version of Dicer, DICER1, is 1,922 amino acids long, and contains multiple domains, including: ATP binding, helicase, DUF283, PAZ, dsRBD, and two RNaseIII catalytic domains[82]. Dicer's PAZ domain has two basic pockets that bind to the 5' and 3' ends of the pre-miRNA. This domain interacts with the 5' phosphate and the binding pocket is arranged such that Dicer prefers a two-nucleotide overhang at the 3' end. Dicer itself functions as a "molecular ruler,"[83] and measures two helix turns – 22 nucleotides away – from the 5' and 3' ends, thus reliably cleaving to produce the ~22 (typically 21-25) nucleotide dsRNA product from which a strand is selected and loaded into the Argonaute protein to form the active RNA-induced silencing complex.

The dicing step has a number of effector molecules, most notably the 366 amino acid protein TARBP[84], which contains two dsRNA binding domains and a Dicer binding domain. Although not essential for precursor microRNA processing, TARBP has been shown to increase dicing efficiency. KSRP[52] has also facilitates Dicer processing of specific precursor microRNAs, including miR-21, through its interaction with the terminal loop. Human DICER1 has been the subject of numerous studies, and a number of important Dicer structures have been published to date, originating with the structure of the Piwi Argonaut Zille (PAZ) and RNase III domains of Giardia Dicer by Jennifer Doudna in 2006[82]. Since then, deeper structural interrogation has been performed, including mutations in the PAZ domain[83], and finally a structure of the entire Dicer

protein by cryogenic electron microscopy (Cryo-EM)[85,86]. These studies, which have been performed in the absence of substrate, have brought to light the dynamic nature of Dicer's interaction with pre-miRNAs and propose multiple structural rearrangements that are necessary for the regulation and activation of Dicer. However, recent structural studies[39] have shown that Dicer continues to be dynamic after substrate loading, and that conformational sampling likely occurs for regulation of Dicer loading, but that substrate rearrangement may enable access to a dicing competent state. Meaning that dicing is not only a multi-step process, but that protein and RNA structural rearrangement is necessary for productive dicing to occur. After elucidating the pH dependent structural rearrangement of pre-miR-21, here we set out to determine in more depth, the contribution of each precursor structure on dicing activity. We show that pre-miR-21 and the corresponding ES mutant are preferentially processed by Dicer when compared with a GS locked mutant. While confirming the importance of this structural rearrangement, we also uncover a unique pH dependent dicing phenomenon, that likely involves pH dependent interactions between the RNA substrate and dicer. We also show that this substrate-dicer interaction is RNA dependent and may serve as another mechanism for dicer mediated cleavage of pre-miRNA substrates.

Results

Recombinant Dicer differentially cleaves each state of miR-21

As the miR-21 precursor adopts two structures in solution, we wanted to determine the effect of these structures on Dicer turnover rates at various pH points. Although typically performed around pH 8, Dicer has been shown to remain active at lower pH points[87], with reduced activity at pH as low as 5.5. To begin, we designed precursor miRNA substrates (Fig 3.1a), adopted from previous studies on human Dicer[84]. We then tested commercially available Dicer from Genlantis at various pH points to see if there were pH dependent changes on Dicer activity (Fig

3.1b,c). As Let-7g is known to undergo a structural rearrangement in solution[39] that is Dicer dependent but should not respond to pH, we included it as a control. Monitoring both miR-21 and Let-7g substrates, enzymatic activity was greatly reduced at lower pH, stopping product formation even after an extended incubation at 37˚C. Genlantis enzyme decreased in activity almost uniformly, but activity was more noticeable for Let-7g at low pH, even though this substrate had been processed less efficiently at higher pH. Although this suggests an inhibitory mechanism for the excited state of miR-21 at lower pH, the lowest pH point tested, 6.0, is above the pK_a of the protonated adenine in miR-21 (~5.8). It is important to note that Genlantis enzyme is known to process microRNAs very inefficiently, taking hours to observe the activity shown in these assays. This contrasts with cellular conditions, where these precursors are processed in minutes[84].

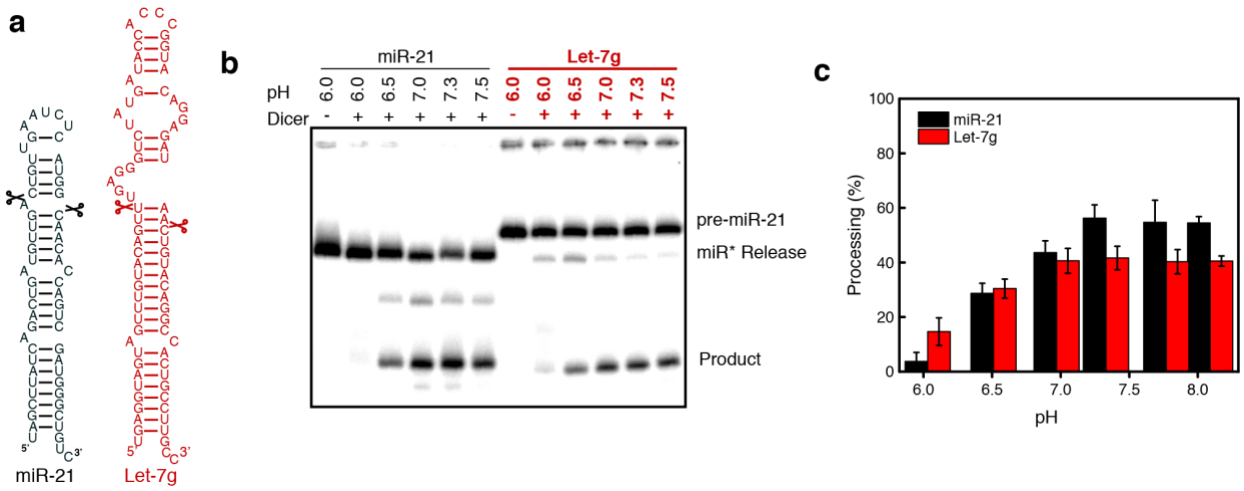

Figure 3.1 - Genlantis Dicer assays at varying pH

(**a**) Precursor construct for miR-21 and Let-7g used in Dicer assays. Note: an additional 3'C was added to create a 2-nucleotide overhang, which is a more ideal substrate, and has been used instead of the native sequence in previous studies[84]. (**b**) Dicer assays for miR-21 and Let-7g precursor performed at typical Dicing pH: 7.7 down to pH 6.0. First lane of each RNA loaded without Dicer, and respective products labeled for Let-7g shown on the right. (**c**) ImageQuant quantification of mature product from Dicer assays performed for miR-21 and Let-7g, referenced to total ^{32}P counts in each lane.

Recombinant Dicer is active at low pH and differentially cleaves each state of miR-21

In order to more accurately recreate the dicing that occurs in cell, we repeated Dicer assays using a recombinantly expressed and purified enzyme. As shown by previous research, highly purified recombinant Dicer, which complexes with TRBP, processes precursors similar to the enzymatic turnover that occurs in cells. For this reason, recombinant Dicer enzyme was prepared and used to perform a duplicate set of Dicer assays (Fig 3.2a). Unlike Genlantis Dicer, this recombinantly expressed enzyme not only required significantly reduced incubation time (~30 minutes), but it also yielded higher amounts of mature product and continued to be active at lower pH points. Again, we initially tested cleavage of miR-21 in comparison to Let-7g. This time, we noted significantly higher amounts of cleavage for Let-7g as the pH decreased, while miR-21 dicing slowed in response to changes in pH. Most importantly, our recombinant enzyme stock was more capable of cleaving at pH points that approached the pK_a of the protonated adenine in miR-21, meaning that the RNA substrate should be equally populated at the lowest pH tested. In order to directly confirm the contribution of both states, we tested GS and ES mutants as described previously (Figure 2.5). Notably, the ES mimic is preferentially cleaved at high pH, which mimics the data shown for Genlantis Dicer. At lower pH, the ES mimic is still preferentially processed, but overall Dicer activity is decreased (Fig 3.2c). At low pH, the GS mimic is poorly cleaved, with overall mature production stagnating to almost zero. Wild type miR-21 closely resembles the plotted population average, which accounts for contributions in substrate fitness from both the GS and ES. To further confirm the processing efficiencies of these substrates through multiple turnovers, we also performed time course assays for recombinant Dicer at high (Fig 3.2c) and low pH (Fig 3.2d). At high pH of 7.77, the ES mimic is preferentially cleaved, producing mature product at a much faster rate in initial turnover conditions for the first point at 2 minutes, and

creating more product throughout the rest of the time course, reaching almost 100% cleavage. At high pH, we expect precursor miR-21 to mimic the ground state, as it occupies ~99% ground state in these conditions. We observe similar levels of initial turnover for miR-21 and the GS mimic, as well as overall processing efficiency throughout the experiment, confirming that the GS mimic matches the substrate fitness of wild type precursor miR-21. At low pH of 6.51, the ES mimic is still preferentially cleaved over the wild type substrate, producing more product in initial turnover and through multiple turnovers. The GS mimic is much less processed and shows slower rates of product formation than the ES. At this lower pH, NMR experiments observe preE-miR-21 occupying ~15% ES and 85% GS, respectively. Time course data confirms that the wild type precursor cleaves at an intermediate level, with rates in between the GS and ES mimics. miR-21 matches a population-adjusted dicer time course plot, which combines the NMR observed population data with the processing rates from the GS and ES mimics. Despite the downward trend in processing for all of the miR-21 constructs, the contribution of the folding states to processing is still distinct. This confirms that miR-21 has differential substrate fitness embedded in its conformational landscape, and that the distinct structural ensemble encoded in the terminal loop has the potential to serve as an additional layer of regulation for microRNA biogenesis.

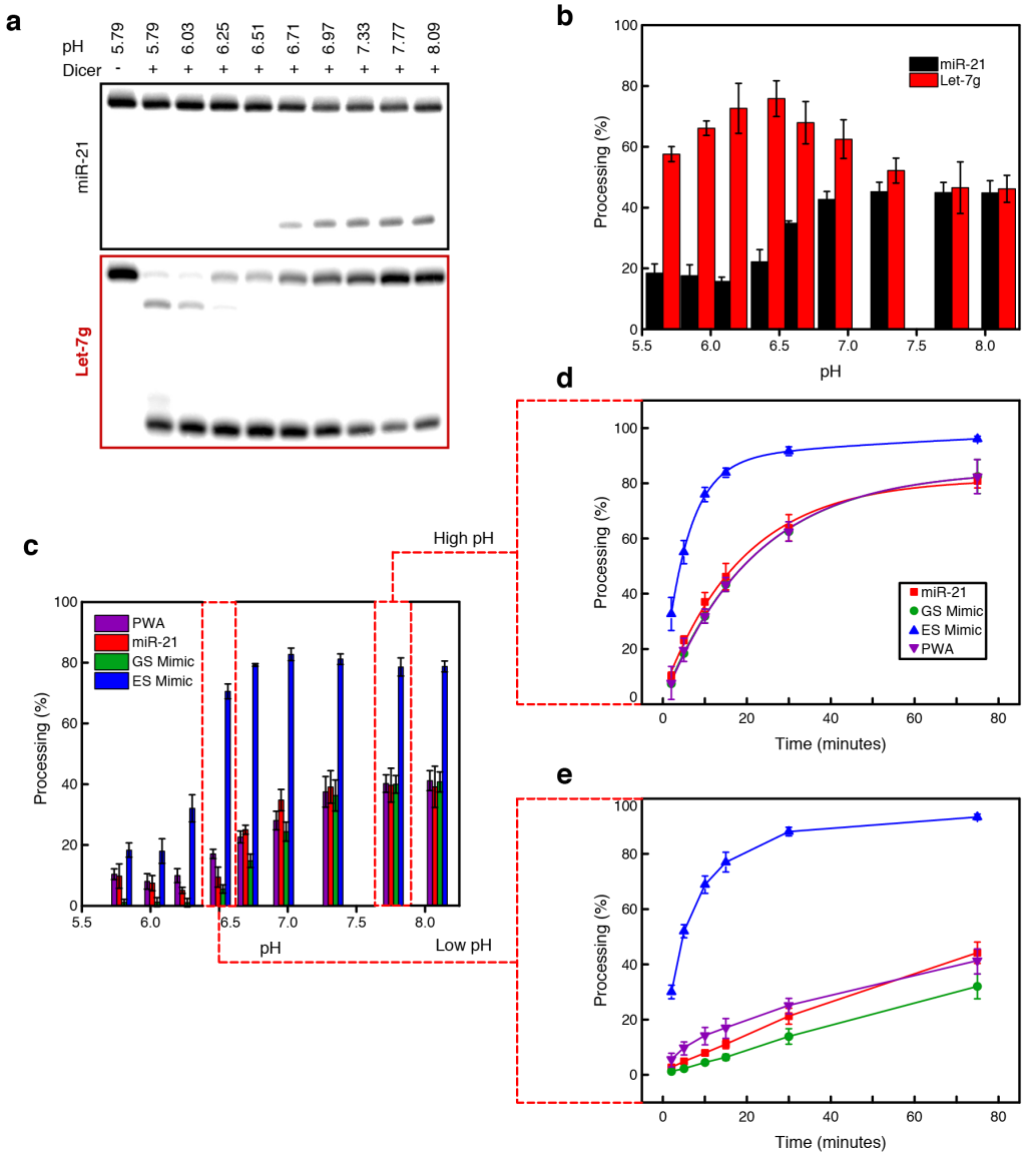

Figure 3.2 - Recombinant Dicer assays at extended pH range

(a) Dicer assays for miR-21 (top) and Let-7g (bottom) precursor performed at an expanded pH range: 8.09 down to pH 5.79. First lane of each RNA loaded without Dicer, and respective products labeled for Let-7g shown on the right. (b) ImageQuant quantification of mature product from Dicer assays performed in triplicate for miR-21 and Let-7g, referenced to total ^{32}P counts in each lane. (**c**) Image Quant Dicer assay quantifications for miR-21 and GS/ES mimics performed at an extended pH range. PWA-Population Weighted Average calculated from NMR Data (**d**) High pH (7.77) Dicer time course assays. Mature product formation measured using ImageQuant, referenced to total ^{32}P counts in each lane. Shown are means and standard deviations (s.d.) from n = 3 independent assays. (**e**) Low pH (6.51) Dicer time course assays. Mature product formation measured using ImageQuant analysis, referenced to total ^{32}P counts in each lane. Shown are means and standard deviations (s.d.) from n = 3 independent assays. Note: Time course assays were performed on a separate, repeated Dicer preparation, resulting in slight changes in overall processing due to enzyme concentration.

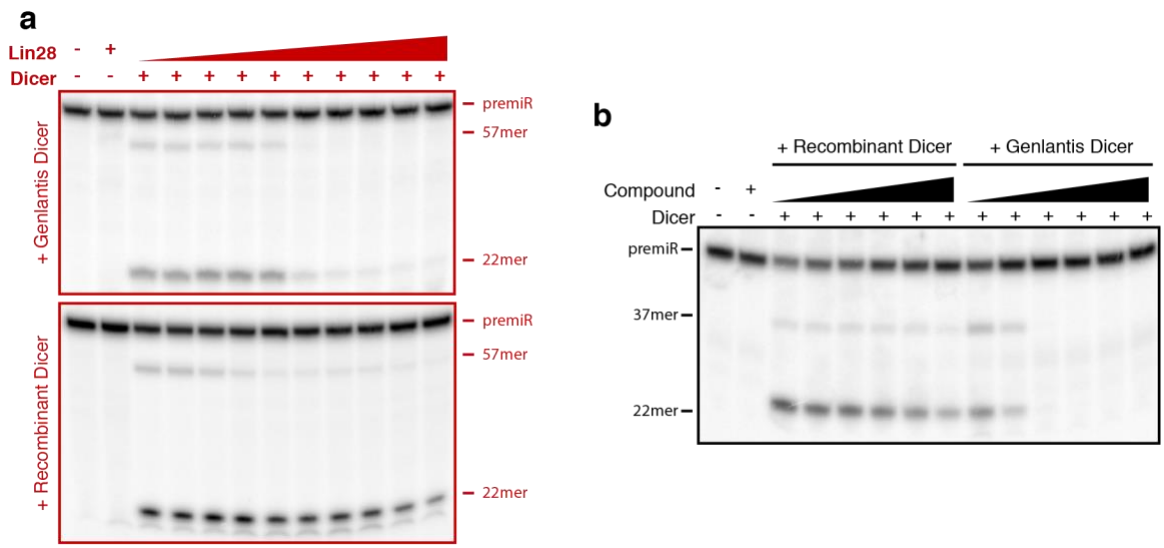

Figure 3.3 - Recombinant Dicer inhibition assays

(**a**) Lin28 titration confirms inhibition of biogenesis of Let-7g for Genlantis Dicer (top). In recombinant dicer (bottom), Lin28 is less efficient at inhibition. (**b**) Small molecule inhibitor from Amanda Garner at Michigan was shown to inhibit processing of miR-21 when using Genlantis Dicer, but does not inhibit recombinant dicer.

Discussion

With this work, we confirmed that the intrinsic conformational plasticity of the miR-21 precursor can serve as a new layer of regulation for miRNA biogenesis. Dicer is intrinsically dynamic, with multiple folding states involved in regulation of substrate binding, and efficient substrate cleavage. While confirming the improved substrate fitness of the ES of pre-miR-21, we uncovered distinct differences between commercially available Dicer and highly concentrated and purified recombinant Dicer. Dicer processing of Let-7g and miR-21 varied in response to pH changes in distinct ways. As the improvement for dicing at low pH was only observed in recombinantly expressed dicer, we performed an additional set of experiments to characterize recombinant Dicer with protein and small molecule inhibitors of the dicing step. Lin-28, a well-known inhibitor of Let-7 biogenesis, blocks the RNase III cleavage site in the RNA through binding. Indeed, this protein inhibits biogenesis when tested with commercial dicer (Fig 3.3a) but did not inhibit the recombinantly expressed enzyme. A small molecule inhibitor for miR-21 was provided for testing by Amanda Garner, PhD. at University of Michigan. This inhibitor, which targets the miR-21 precursor, inhibits Dicer processing for commercial Dicer, but does not inhibit recombinantly expressed and purified Dicer (Fig 3.3b). In combination with the pH dependent dicer assays, these results reveal that our recombinantly expressed and purified enzyme is more active and also less likely to be inhibited when compared with its commercial counterpart. Interestingly, both enzymes are full-length, human copies of DICER1. This implies the presence of either a post-translational modification in the recombinantly expressed enzyme that greatly enhances processing, or a co-factor that was co-purified with DICER1. Although both are possible, Dicer was greatly over-expressed, meaning that co-factors would be present as sub-stoichiometric levels. Regardless, either of these possibilities alludes to the presence of a Dicer activator, and the

differential activation of Let-7g over miR-21 may involve RNA specific interactions with this new co-factor. This discovery is subject to further investigation.

Methods

Preparation of Dicer substrates

Pre-miR-21 substrates were prepared by *in vitro* transcription, where a hammerhead ribozyme (HH) was fused to the 5'-end of pre-miR-21s to generate 5'-OH substrates with wild-type 5'-end nucleotide. Sequences of the pre-miR-21 substrates are as follows,

HH-Pre-Let-7g WT:
CCTAATACGACTCACTATAGGGCTACTACCTCACTGATGAGGCCGAAAGGCCGAAA
CCCGAAAGGGTCTGAGGTAGTAGTTTGTACAGTTTGAGGGTCTATGATACCACCCGG
TACAGGAGATAACTGTACAGGCCACTGCCTTGCC

HH-Pre-miR-21 WT:
CUAAUACGACUCACUAUAGGAGCUACUGAUGAGGCCGAAAGGCCGAAACCCGAAA
GGGUCUAGCUUAUCAGACUGAUGUUGACUGUUGAAUCUCAUGGCAACACCAGUC
GAUGGGCUGUC

HH-Pre-miR-21 GC Mutant:
CUAAUACGACUCACUAUAGGAGCUACUGAUGAGGCCGAAAGGCCGAAACCCGAAA
GGGUCUAGCUUAUCAGACUGAUGUUGACCGUUGAAUCUCACGGCAACACCAGUCG
AUGGGCUGUC

HH-Pre-miR-21 G38U Mutant:
CUAAUACGACUCACUAUAGGAGCUACUGAUGAGGCCGAAAGGCCGAAACCCGAAA
GGGUCUAGCUUAUCAGACUGAUGUUGACUGUUGAAUCUCAUGUCAACACCAGUC
GAUGGGCUGUC

After *in vitro* transcription, samples were exchanged into H_2O using Amicon filtration units (Millipore) with 10K Da MW cut-off membranes. The resulting substrates were subsequently subject to 5'-end ^{32}P-labeling with T4 Polynucleotide Kinase (T4PNK) (New England Biolabs Inc.). 50 μl of RNA at 400 nM was incubated in a solution containing T4PNK buffer, 20 U T4PNK, and γ-^{32}P ATP at 35°C for 30 minutes before being gel purified by 20% denaturing polyacrylamide gel. The ^{32}P-labeled RNAs were extracted from the gel by crushing and soaking in 250 mM NaCl

in 1xTBE buffer for 24 hours before filtering and exchanging multiple times into H_2O using an Amicon filtration unit (Millipore) with a 10K Da MW cut-off. ^{32}P-labeled substrates were stored at -20°C prior to dicer processing assay.

Dicer processing assays

^{32}P-labeled pre-miR-21 substrates were heated to 95°C for 5 minutes and then placed on ice for 5 minutes to anneal the precursor hairpin. 6 μl substrate (4 μl for inhibition assays) was mixed with 2 μl of each corresponding 5x dicing buffer (12 mM HEPES, 1 M NaCl, and 0.02 mM EDTA at pH 8.09/7.77/7.33/6.97) (or 12mM BisTris, 100mM NaCl, and 0.02 mM EDTA aat pH 6.71/6.51/6.25/6.03/5.79) and 2 μl of 25 mM $MgCl_2$. For inhibition assays, 2 μl of increasing amounts of Lin-28 (156 nm up to 40μM final concentration) or small molecule inhibitor was added to each tube, pipetted up and down to mix, and then incubated at 37°C for 15 minutes before the addition of Dicer in the next step. In a separate tube, concentrated Dicer enzyme (Genlantis, Inc. or recombinantly expressed and purified) was diluted in 1x dicing buffer (2.4 mM HEPES, 200 mM NaCl, 4 μM EDTA) and mixed with an equal volume of 10 mM ATP. 2 μl of ATP/Dicer mixture (10 mM ATP and 0.2U Dicer, Genlantis or 0.1U recombinant Dicer) was added to each tube of 6 μl substrate and incubated at 35°C (overnight for Genlantis, or 30 minutes for recombinant Dicer). Reactions were quenched by adding 1.6 U proteinase K (New England Biolabs Inc.), 2 μl of 0.5 M EDTA and incubated for 45 minutes at 35°C. Samples were denatured in formamide with trace bromophenol blue and xylene cyanol, ran on a 15% denaturing polyacrylamide gel, visualized using Amersham Typhoon 5 Biomolecular Imager (GE Healthcare) and analyzed with Image Quant (GE Healthcare) using peak area and rubberbanding for background deletion. All assays were carried out in triplicate to estimate experimental errors. Population weighted average was calculated using the following equation:

$$PWA = p_{ES} * (Mature\ production\ with\ ES\ mimic - G38U) + p_{GS} * (Mature\ production\ with\ GS\ Mimic - CCMut)$$

where p_{ES} and p_{GS} are the respective ground and excited state occupation in each folding state as determined in Chapter 2.

CHAPTER 4 - DYNAMICS IN PRIMARY MICRORNA

Introduction

The mGHG mismatch includes a U-C mismatch sandwiched between two G-C base pairs in the hairpin of pri-miRs. Strikingly, the strand arrangement of G-C base pairs, i.e. 5'G and 3'C versus 5'C and 3'G has been shown to alter substrate recognition by Drosha. 5' sequences of CUC confer the highest cleavage efficiencies, and this single motif is able to encode single-nucleotide precision for Drosha, outweighing contributions from the other basal and apical junction motifs. Alternate arrangements of these nucleotides have been shown to serve as a recognition site for cleavage but result in less efficient recognition and inaccurate cleavage products. As the Drosha step determines the 5' seed sequence of the mature microRNA, precise cleavage is essential for microRNA biogenesis. Lastly, the dsRBD of Drosha has been proposed to bind this motif after allosteric rearrangement from binding between Drosha and DGCR8. Although the sequence of essential motif has been explored through systematic modification and measurement, the exact RNA structure has yet to be determined. Due to the conservation of a U-C mismatch and the flanking sequence, we hypothesize that this motif adopts a very unique structure in solution. We also propose that dynamics between the U-C mismatch may confer differential fitness for recognition by the microprocessor, enabling differential recognition despite the conservation of nucleotide composition.

Here, we set out to study the structure and dynamics of this essential mGHG motif. We showed that this small motif, which contains a conserved set of six nucleotides, indeed has a unique structure that explains its ability to provide single-nucleotide measurement for Drosha. We also

show that that changes in the arrangement of these nucleotides encodes intrinsic conformational flexibility, including a stabilized protonation state at the U-C mismatch, that confers differential recognition to the microprocessor. Like previously discussed with pre-miR-21, the observation of this dynamic ensemble serves as an additional layer of regulation for miRNA biogenesis that is driven solely by the dynamic ensemble of the substrate.

Results

The mGHG motifs adopt similar ground state structures

The mGHG recognition motif consists of a U-C mismatch sandwiched between two G-C base pairs. Alterations to the arrangement of these nucleotides confers differential recognition by Drosha, with 5' CUC enabling single nucleotide recognition for the microprocessor. First, to determine the structure and dynamics of the arrangement of this motif we designed a set of abbreviated NMR Constructs (Fig 4.1a), that include the mGHG motif, flanked by stable A-U and G-C base pairs to mimic a stable hairpin. The small hairpin constructs are capped by a UUCG tetraloop, which stabilizes the hairpin structure and prevents dimerization. A simple NMR ^{1}H-^{1}H NOESY experiment on the imino region provides an excellent characterization of RNA secondary structure, as one imino resonance is expected per canonical Watson-Crick base pair. The mGHG constructs designed for this study adopt stable conformations, where imino resonances (Fig 4.1b) can be observed for the lower and upper stems. The U-C wobble does not form in the ground state of any construct, as the central U in the U-C does not show any resonances in the ^{1}H-^{1}H NOESY experiments (Fig 4.2). Using a ^{13}C^{1}H HSQC NMR experiment that probes nonsolvent-exchangeable signals, we detected all 26 resonances for the 26 expected NMR resonances from each of the mGHG constructs (Fig 4.1c, Fig 4.3 and 4.4). However, the resonance for the mismatch cytosine has weak peak intensity, which resembles the typical NMR phenomenon of exchange

broadening. Like in pre-miR-21, this is caused by interconversion between multiple states. Again, using NMR R1ρ RD spectroscopy, we can characterize this conformational change, making it possible to gain insight into an RNA ESs that is too low-populated and short-lived to be detected by conventional structural biology approaches. Here, we carried out both on-resonance and off-resonance low spin-lock field R1ρ RD experiments to identify any dynamics that occur in this relatively small and apparently simple motif. We observed flat RD profiles for all of the base (C5 and C8) and sugar (C1') carbons resonances studied in the mGHG96 construct (Fig 4.7). A number of base (C5, and C8) and sugar (C1') carbons flanking the U-C mismatch in the mGHG35 construct also have flat RD profiles (Fig 4.6). However, resonances studies at and near the U-C mismatch in mGHG35 display power and offset dependent RD profiles. These RD profiles can be globally-fitted to a two-state (GS ↔ ES) exchange process, where the ES has a population of (10.7 ± 0.6%). These data show that the mGHG35 construct dynamically interconverts between two structurally and kinetically distinct states, while the construct with better recognition (mGHG96) adopts a single, stable conformation.

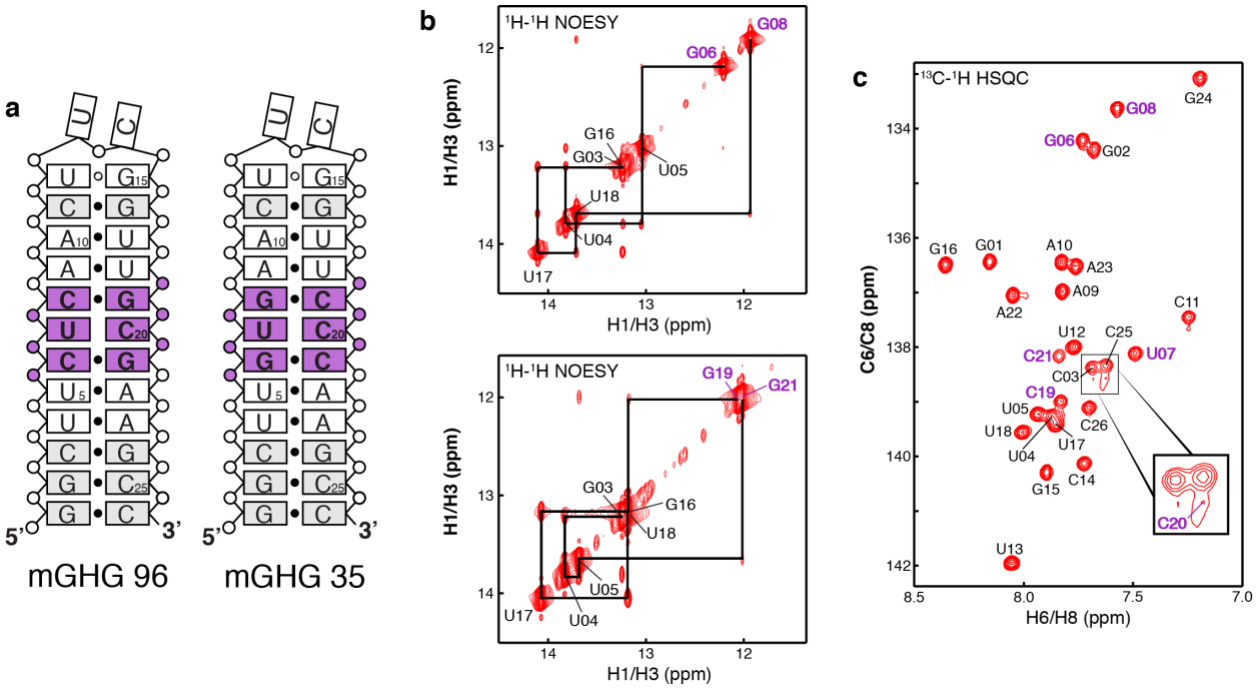

Figure 4.1-NMR characterization of mGHG motifs.
(**a**) Secondary structure of mGHG96 and mGHG 35 constructs, where the mGHG motif is highlighted with purple nucleotides, and $_1$H-$_1$H Imino NOESY observed Watson-Crick base pairs are highlighted with dark circles. (**b**) NMR 2D $_1$H-$_1$H NOESY spectrum of the imino proton region of mGHG35 (top) and mGHG96 (bottom) reveal a well folded secondary structures, with the G-C base pairs highlighted with purple assignments on the diagonal. Note: The characteristic UUC**G**-G15N1 was present in these experiments, but are outside the shown spectral width. (**c**) NMR 2D $_{13}$C-$_1$H HSQC spectra of base carbon (C6 and C8) region of mGHG35 shows exchange broadening of C20H8, shown inset.

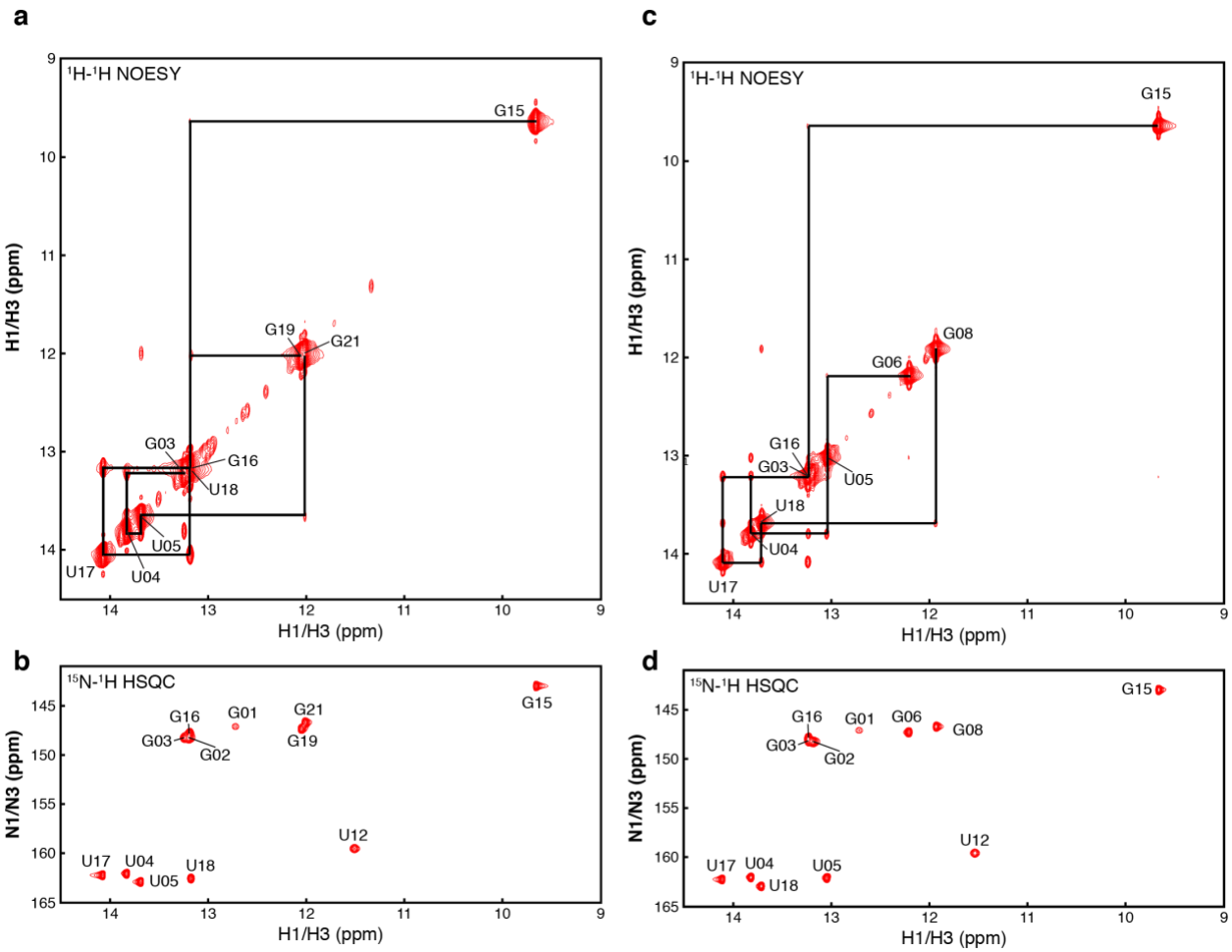

Figure 4.2 - NMR characterization of exchangeable protons

(**a**) ^{1}H-^{1}H Imino NOESY and ^{15}N-^{1}H HSQC (**b**) of mGHG96 at pH 6.5 and 10°C reveal well folded secondary structure of mGHG96, with a gap in base pairing at the U-C mismatch. (**c, d**) The same set of experiments recorded on mGHG35 reveal similar secondary structure, with a similar gap in base pairing at the U-C mismatch.

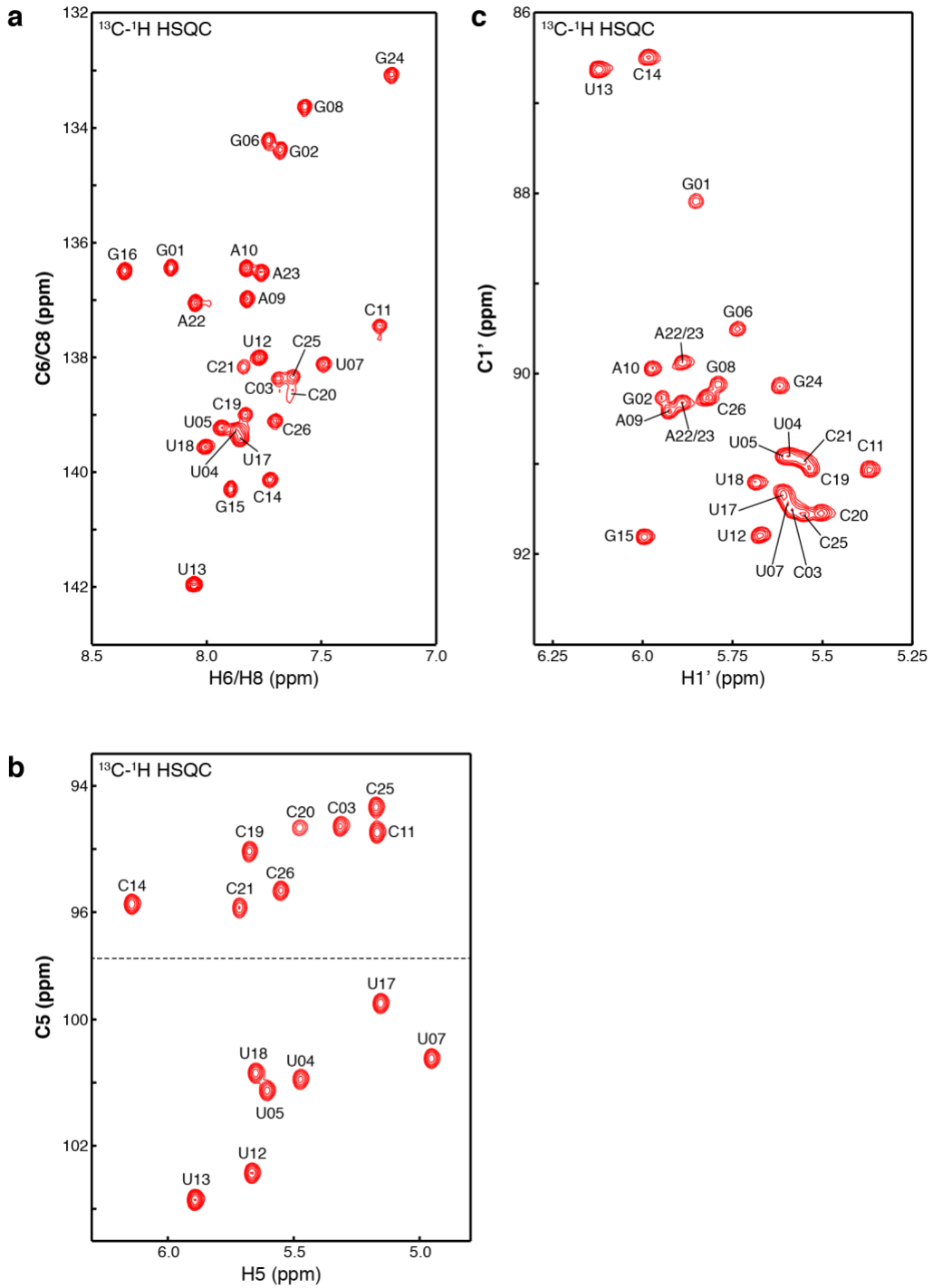

Figure 4.3 - NMR assignments for mGHG35

Shown are ^{13}C-^{1}H HSQC spectra of base carbons (a- C6/H6 and C8/H8, b- C5/H5) and sugar carbons (c-C1'/H1') for mGHG35 along with corresponding assignments for each set at pH 6.5, 35°C.

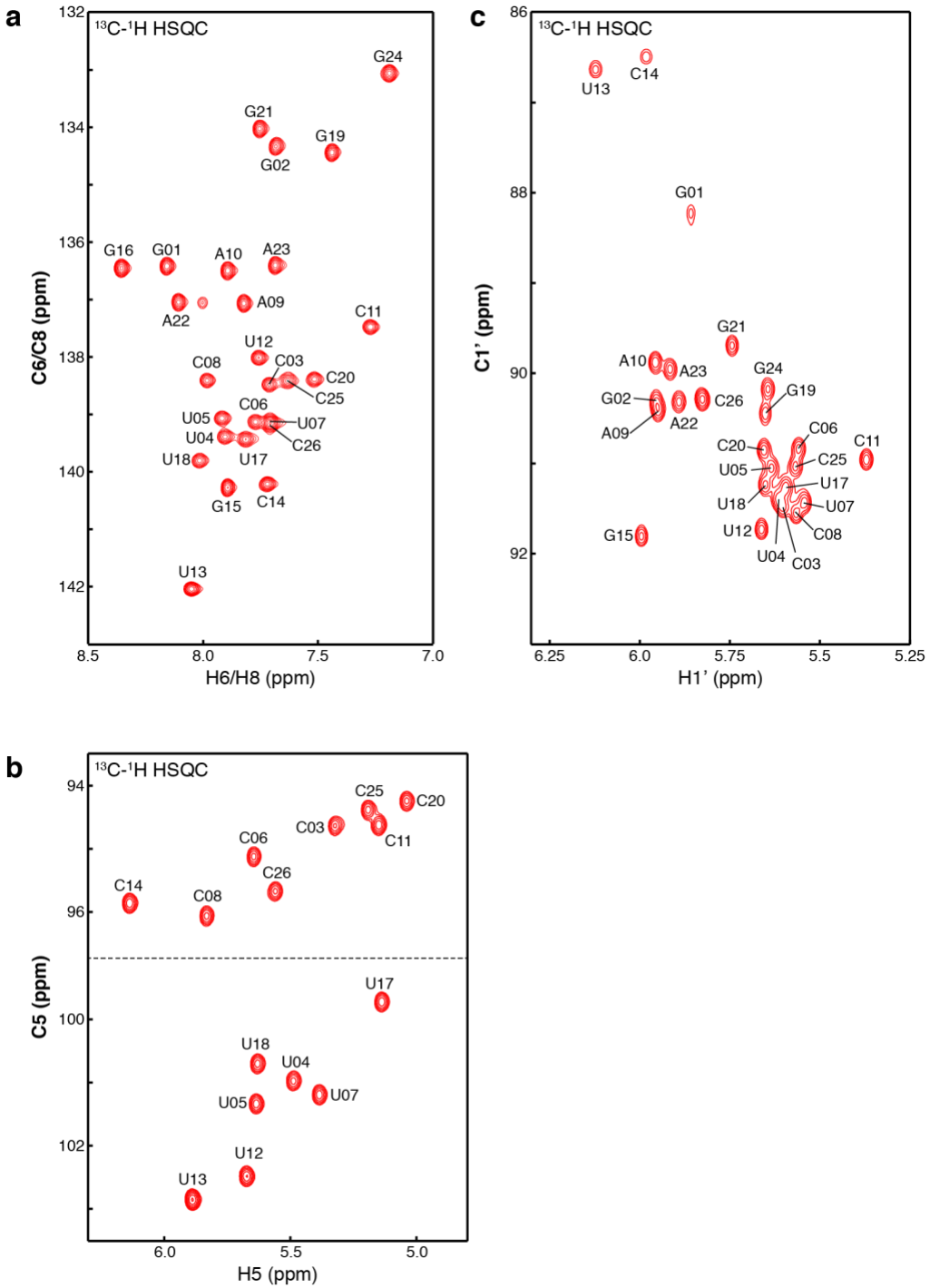

Figure 4.4 - NMR assignments for mGHG96

Shown are ^{13}C-^{1}H HSQC spectra of base carbons (**a-** C6/H6 and C8/H8, **b-** C5/H5) and sugar carbons (**c**-C1'/H1') for mGHG96 along with corresponding assignments for each set at pH 6.5, 35°C.

The low recognition motif encodes an ES with transient protonation at the N3 of cytosine in the C-U mismatch

To gain structural insights into the ES, we utilized NMR chemical shifts, which are one of the most sensitive measurements for probing local chemical environments. We examined ES carbon chemical ($\varpi_{ES} = \varpi_{GS} + \Delta\varpi$), where $\Delta\varpi$ is the chemical shift difference between ES (ϖ_{ES}) and GS (ϖ_{GS}) extracted from the two-state analysis of an R1ρ RD profile (Fig 4.5c). The largest chemical shift deviation ($\Delta\varpi$) was observed at C20, the critical cytosine in the U-C mismatch. The calculated chemical shift change for this residue was 2.86ppm downshifted from its GS position. This downfield change in chemical shift is characteristic of transient protonation in unpaired cytosines. The magnitude and direction of this change in chemical shift, coupled with the fast rate of exchange (9677±1851s^{-1}) fit out from the R1ρ RD profiles, implies transient protonation in the ES at the cytosine N3 of the critical U-C wobble in the mGHG motif. To further examine whether the exchange process is due to a protonation event, we increased the pH of the sample from 6.45 to 8.04. By increasing the pH, we aimed to shift the equilibrium towards the deprotonated state. The ^{13}C-^{1}H HSQC spectrum recorded at pH 8 is of substantially higher quality, where the exchange broadening on the C and U resonances is significantly reduced (Fig 4.8).

The high population (~10.7%) of C20N3+ at pH 6.45, which otherwise would be less than 1% based on the intrinsic pKa (~4.45) of the free cytosine N3 site, further suggests that C20 has a distinct propensity for protonation when compared with other unpaired cytosines. Upon further investigation using various pH points and chemical shift analysis, the N3 of C20 has a neutrally shifted pKa of 5.41 ± 0.01. This remarkable observation – that such a significant neutral shift in pKa is possible with such a simple structural motif – is very unique, as protonation in RNA is usually only observed in higher order structures.

In order to verify that protonation is the single driving factor behind the conformational change in the mGHG35 construct, we compared the extracted chemical shift difference between the excited state (ϖ_{ES}) and the chemical shift observed at low pH (ϖ_{lowpH}). Largely, changes in chemical shift that were detected in the R1ρ RD profiles match the chemical shifts that respond to pH (Fig 4.5c). For 6 of the 7 residues studied, U07C6, C19C5/C6, and C20C1'/C5/C6, the chemical shift changes match exceptionally well, especially for residues with larger chemical shift perturbations that are the result of protonation and a change in base pairing. However, the deviations for G08C1' could be due to its relatively small chemical shift differences; the nature of this difference is subject to further study by additional R1ρ techniques outlined in the discussion. Taken together, these results unambiguously revealed that the cytosine in the GHG mismatch motif undergoes a pH-dependent conformational transition, where the bulge cytosine is specifically protonated at N3 in the ES.

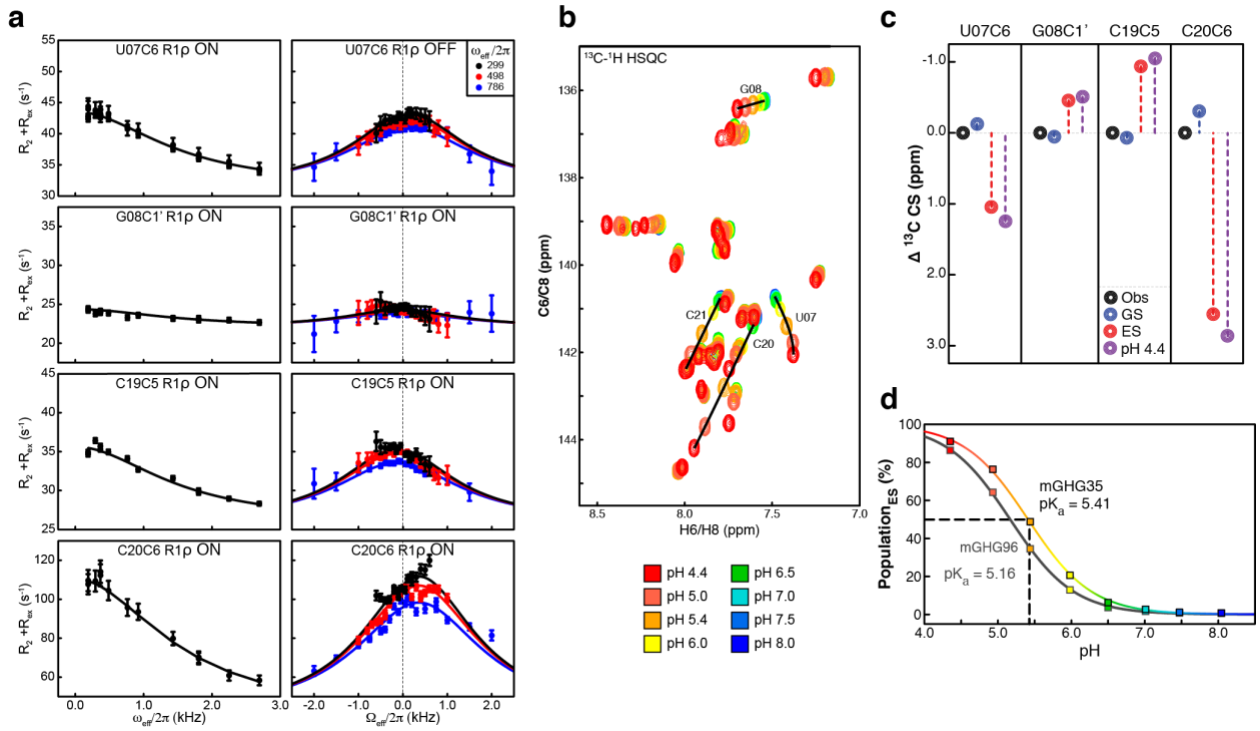

Figure 4.5- mGHG35 populates a protonated excited state with a neutral shifted pK$_a$

(a) Representative ^{13}C on-resonance and off-resonance RD profiles showing dependence of R_2+R_{ex} on spin-lock power ($\omega_{eff}/2\pi$) and offset ($\Omega/2\pi$), respectively, where Ω is the difference between the spin-lock carrier frequency and the observed resonance frequency. RD profiles are fit to a global two-state model using the Bloch-McConnell equation. Error bars are experimental uncertainties (s.d.) estimated from mono-exponential fitting of $n = 3$ independently measured peak intensities (**b**) NMR 2D ^{13}C-^{1}H HSQC spectra of base carbon (C6 and C8) region of mGHG35 performed at various pH points reveals distinct chemical shift changes for residues in the mGHG motif. (**c**) Comparison of carbon chemical shifts for the observed, GS, ES, and low pH NMR conditions for mGHG35 (**d**) The pH-dependent population of the excited state based on U07C6 and C20C6 chemical shifts: apparent pK$_a$ value for C20 based off C6/H6, C5/H5, and C1'/H1' in mGHG35 and mGHG96 shown in pH specific colors, and grey, respectively.

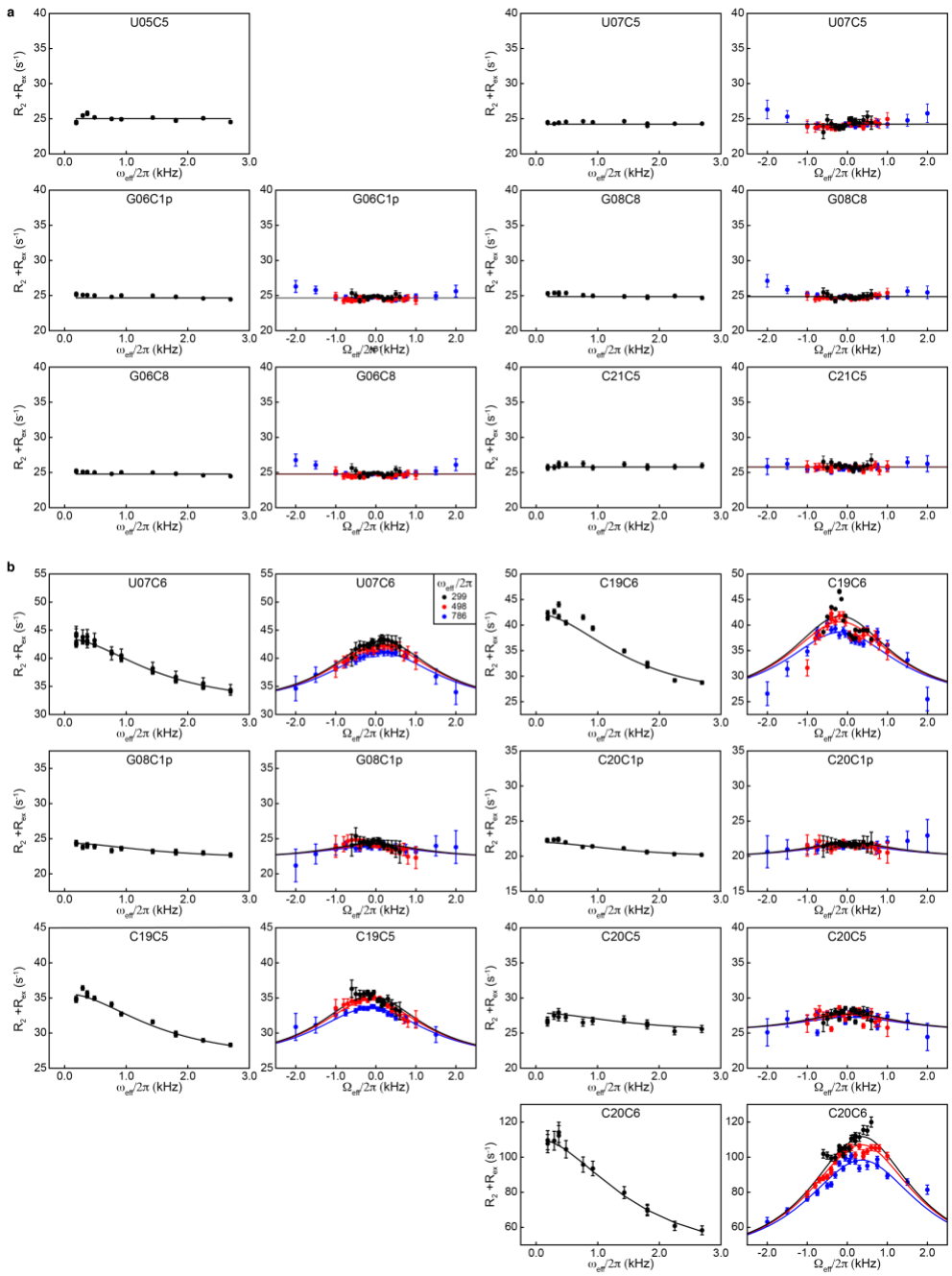

Figure 4.6 - ^{13}C $R_{1\rho}$ Relaxation Dispersion for mGHG35

(**a**) Profiles of mGHG35 residues without apparent chemical exchange. On- and off-resonance ^{13}C RD profiles depicting spin-lock power ($\omega_{eff}/2\pi$) and offset ($\Omega/2\pi$) dependence of $R_2 + R_{ex}$ measured at pH 6.5. Solid lines represent the best fits to a single-state model using the Bloch-McConnell equation. (**b**) ^{13}C $R_{1\rho}$ RD profiles of mGHG35 residues that undergo chemical exchange at pH 6.45. Solid lines represent the best fits to a global two-state exchange model (k_{ex} = 9677 ± 1851 s^{-1} and p_E = 10.7 ± 0.6%) using the Bloch-McConnell equation. Error bars, experimental uncertainties (s.d.) estimated from mono-exponential fitting of n = 3 independently measured peak intensities.

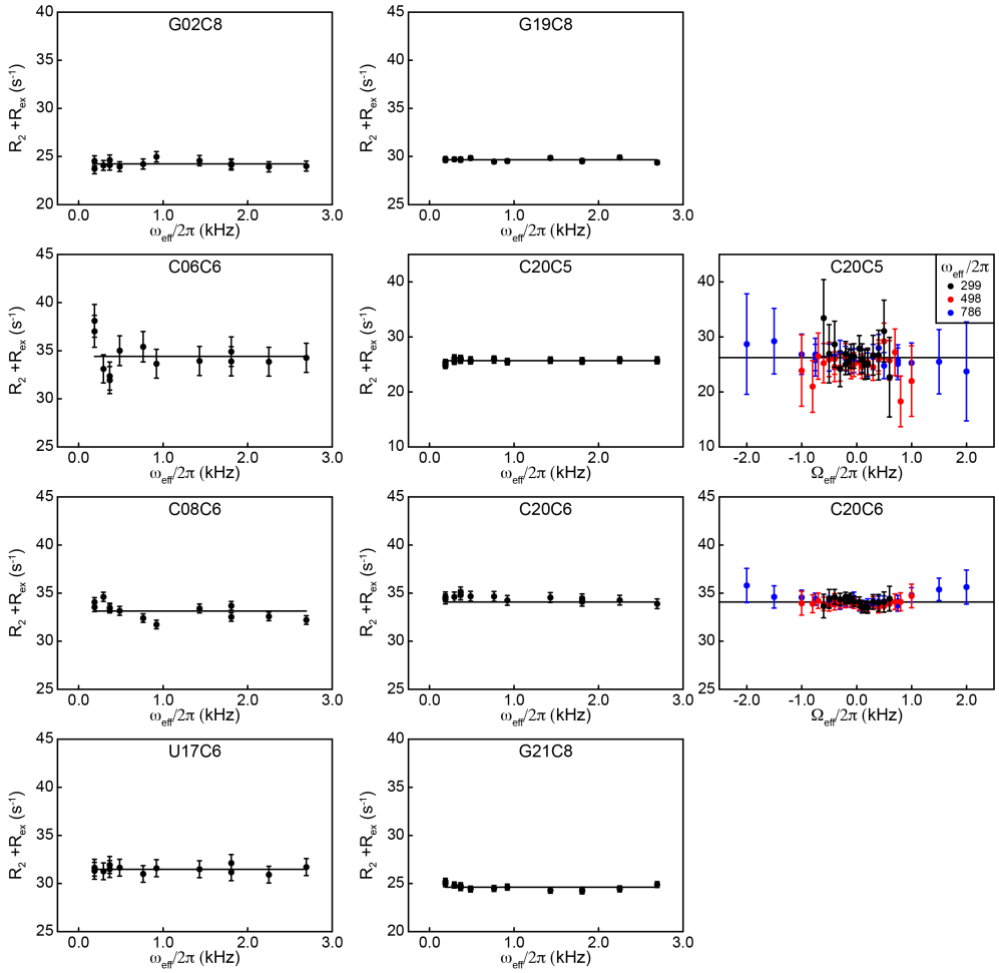

Figure 4.7 - ^{13}C $R_{1\rho}$ Relaxation Dispersion for mGHG96

Profiles of mGHG35 residues without apparent chemical exchange. On- and off-resonance ^{13}C RD profiles depicting spin-lock power ($\omega_{eff}/2\pi$) and offset ($\Omega/2\pi$) dependence of $R_2 + R_{ex}$ measured at pH 6.5. Solid lines represent the best fits to a single-state model using the Bloch-McConnell equation.

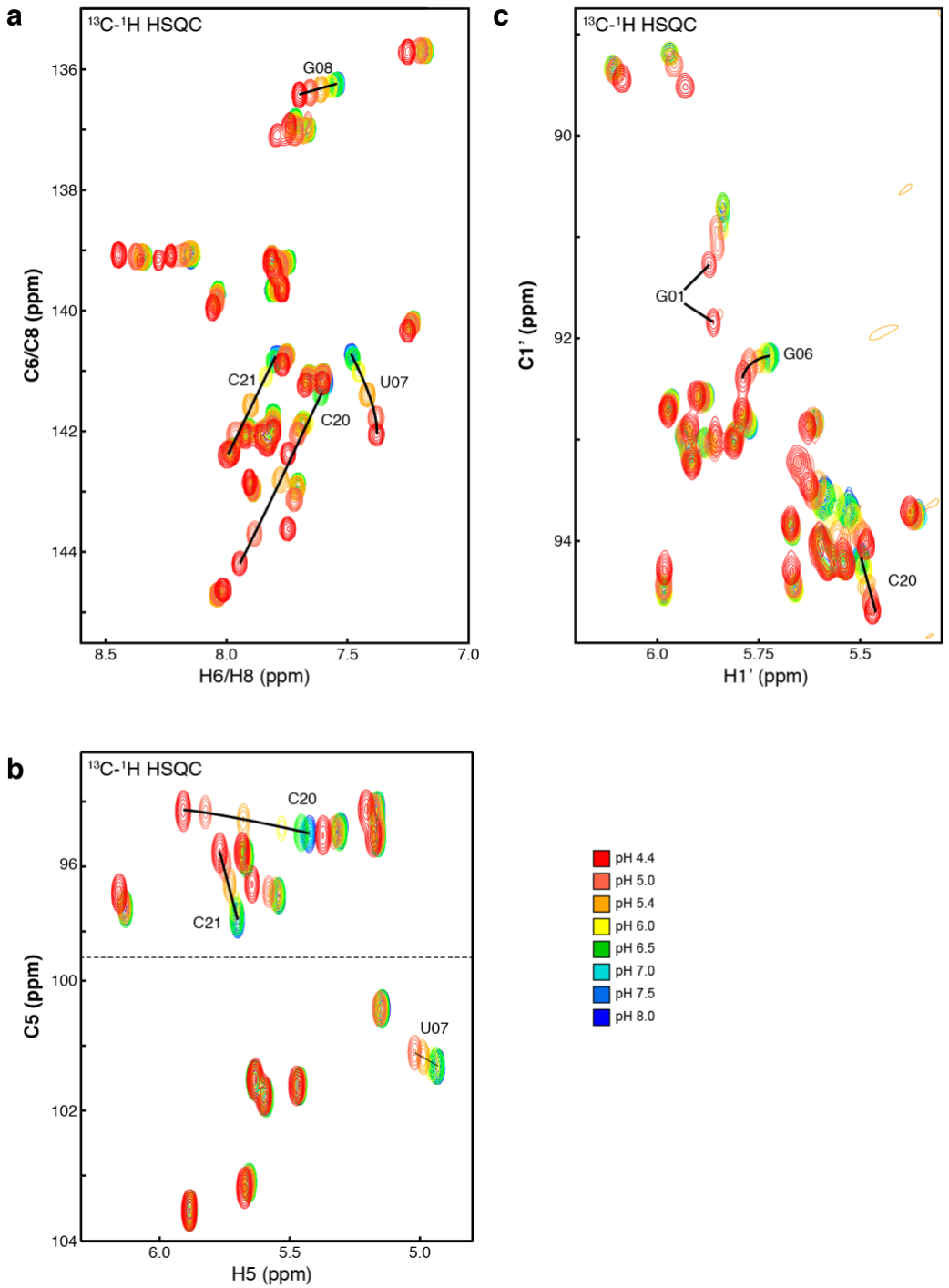

Figure 4.8 - pH Titration for mGHG35

Shown are ^{13}C-^{1}H HSQC spectra of base carbons (**a-** C6/H6 and C8/H8, **b-** C5/H5) and sugar carbons (**c**-C1'/H1') for mGHG35 across various pH points ranging from 4.4 to 8.0 at 35°C at 35°C. Peaks with distinct changes in chemical shift are assigned, with direction of chemical shift change traced in black.

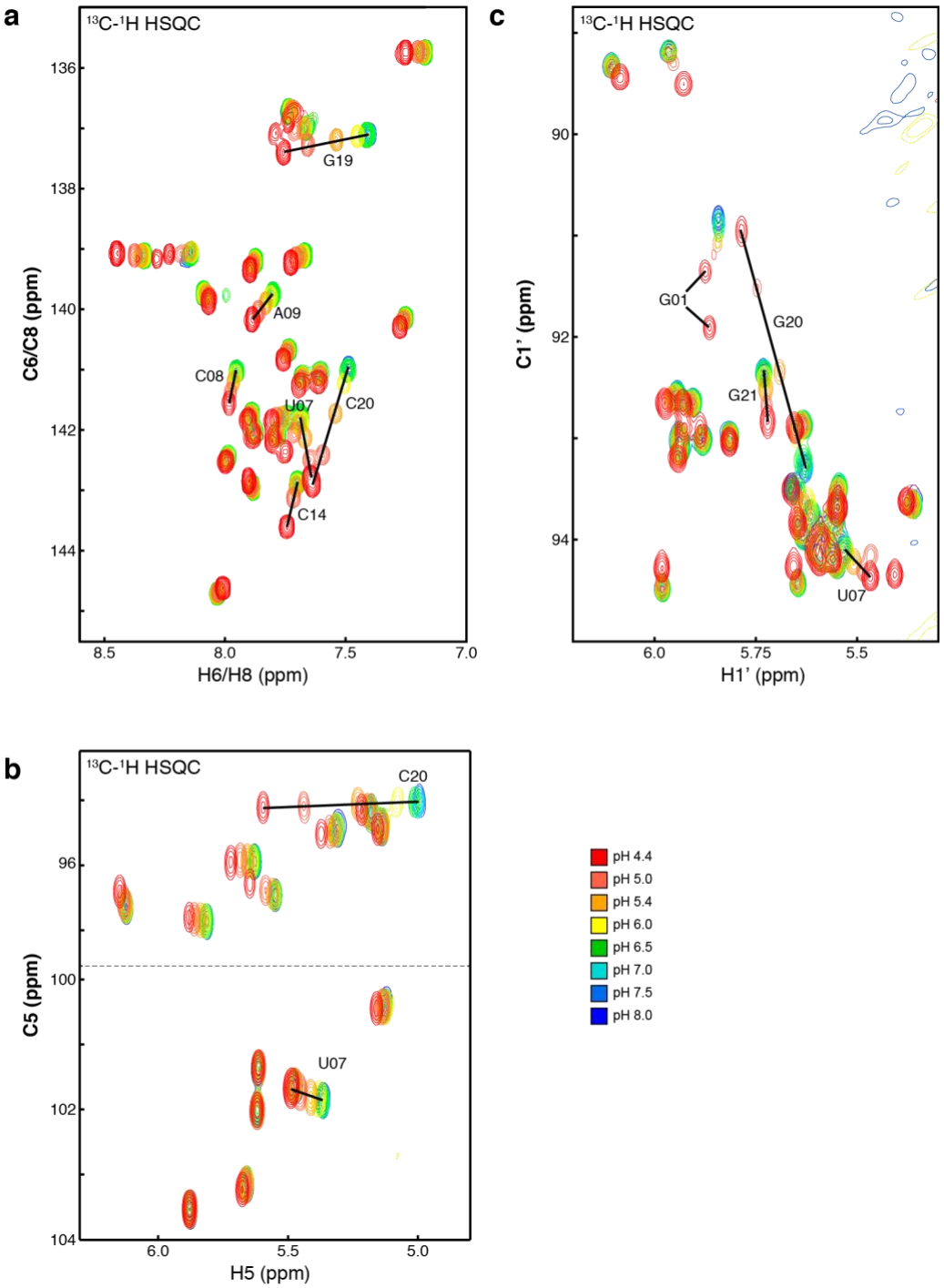

Figure 4.9 - pH Titration for mGHG96

Shown are ^{13}C-^{1}H HSQC spectra of base carbons (a- C6/H6 and C8/H8, b- C5/H5) and sugar carbons (c-C1'/H1') for mGHG96 across various pH points ranging from 4.4 to 8.0 at 35°C. Peaks with distinct changes in chemical shift are assigned, with direction of chemical shift change traced in black.

The protonated ES is stabilized through a C+U wobble base pair

In order to stabilize the cytosine protonated state in the mGHG35 construct, we propose that the uridine across the helix forms a C+U wobble base pair (Fig 4.10c). In this base pair conformation two hydrogen bonds would stabilize the interaction: the protonated N3 of C20 would hydrogen bond with the O4 of U07, and the N3 of U07 would hydrogen bond with the O2 of C20. To detect the presence of the proposed C+U wobble pair, a set of low-pH ^{1}H-^{1}H through space and ^{15}N-^{1}H through bond NMR experiments were performed. Even though this interaction was too transient to be detected by ^{1}H-^{1}H NOESY, the ^{15}N-^{1}H HSQC spectrum revealed a new peak that was absent in experiments run above the pKa (Fig 4.10a). The ^{15}N and ^{1}H chemical shift for this new peak are in a region of the NMR spectra typical for uridines in U-U and G-U wobble base pairs, residing close to the capping U-G interaction in the UUCG tetraloop. Through bond J-^{15}N-^{15}N experiments revealed that the low pH N-H peak does not have a base-pairing partner (Fig 4.10b), which matches the proposed NH-O hydrogen bonding configuration. The lack of detection for the C20H3 peak is likely due to the short-lived ES of 115μs, which is unable to be detected, due to its transient nature. This phenomenon matches the transient protonation in pre-miR-21, which was also too transient to be detected by NH-HSQC.

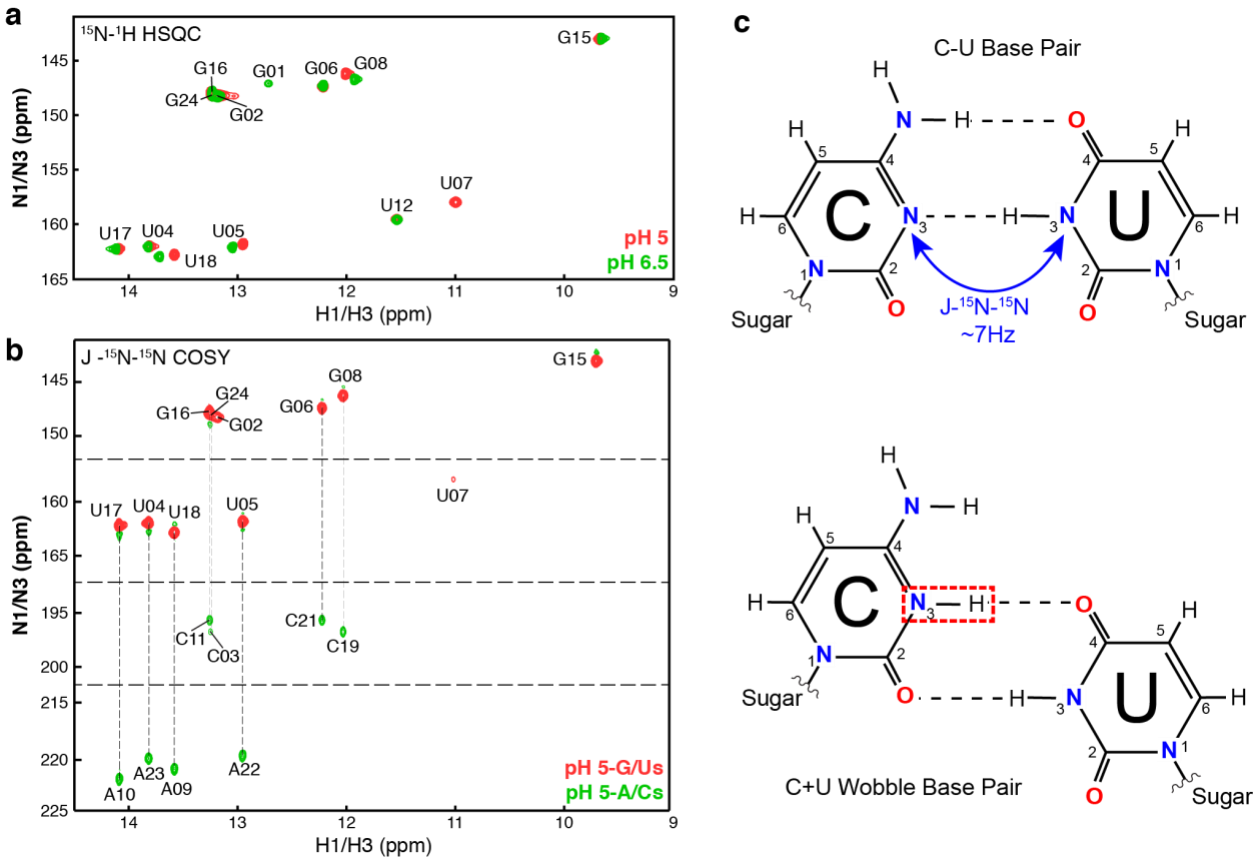

Figure 4.10 - A protonated C+U wobble base pair is the ES

(**a**) NMR ^{15}N-^{1}H HSQC spectra on exchangeable imino protons detects peaks for N_3 of uridines and N_1 of guanines shown at typical NMR conditions (green) and low pH (red). (**b**) NMR J-^{15}N-^{15}N COSY experiment performed at pH 5 detects base-pairing imino protons using J coupling (~7Hz) between hydrogen bonding base pairs. Guanine and Uridine iminos at the top of the spectra share a proton with their base adenine and cytosine base pairing partner. Conserved ^{1}H chemical shift connection highlighted with a dashed line. (**c**) Canonical C-U base pair (top) has a hydrogen bond between N3 of uridine and cytosine. Proposed C+U wobble base pair (bottom), with additional proton at N3 highlighted in red, preserves the imino base pairing between U and C, and alters the hydrogen bond receptors to oxygen, altering the base pair geometry.

Discussion

Here, by integrating structural and dynamic analyses on the mGHG motif, we showed that the intrinsic propensity for protonation - and the corresponding changes to the dynamic ensemble of this simple RNA, can greatly alter the substrates ability to be recognized and accurately processed by DROSHA. mGHG35 undergoes rapid protonation in solution at physiological pH due to its neutrally shifted pKa, while mGHG96, the most ideal motif for recognition, is not protonated. This mGHG96 construct has a less shifted pKa of 5.16 ± 0.01. The difference in pKa between these two constructs extrapolates to differences in the population of the ES at typical experimental and physiological pH. The highly populated ES in mGHG35 construct is more sparsely populated in the mGHG96 construct (<1%), escaping the limit of detection by our R1ρ analysis. As the mGHG96 motif is preferred by the microprocessor, the most simplistic model would involve protonation as an inhibitor of recognition by Drosha. In the ground state, the uridine and cytosine sandwiched in the mismatch motif may be stacked, or a specific nucleotide could flip out for interaction with the double-stranded RNA Binding Domain (dsRBD) of DROSHA. For mGHG35, this structure competes with a C+U wobble base pair, where intrinsic protonation at the N3 is stabilized by a hydrogen bond with O3 on the uridine across the helix. This stabilized conformation creates a sheared C+U wobble base-pair, which is not easily recognized by Drosha. The fast interconversion between the ideal, unpaired substrate, and the sheared C+U is responsible for the change in substrate fitness that was observed in previous work. While our observation represents the first example of an excited state in primary microRNAs, the conservation of this motif throughout evolution, from sea anemone to sponges and humans[48], strongly suggests the preservation of this dynamic ensemble as an additional mechanism for regulation in the biogenesis of primary microRNAs.

Methods

Sample preparation.

Uniformly $^{13}C/^{15}N$-labeled mGHG samples were prepared as previously described (Chapter 2). Samples were then concentrated and exchanged into NMR buffers with 50 mM KCl and 50 μM EDTA, where 10 mM acetate buffers were used for pHs at 4.35, 4.93, and 5.44, and 10 mM sodium phosphate buffers were used for pHs at 5.98, 6.5, 7.01, 7.47, and 8.04. For H_2O samples, and D_2O samples were prepared as previously described in Chapter 2.

NMR Spectroscopy

Exchangeable proton spectra were recorded using H_2O samples at 283 K, and nonexchangeable proton spectra were recorded at 298 K and 308 K using H_2O and D_2O samples. The assignments for mGHG35 and mGHG96 were obtained with samples at pHs 6.45 using 2D NOESY, 2D TOCSY, ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, 2D HCCH-COSY, and HCN experiments on unlabeled and uniformly $^{13}C/^{15}N$ labeled samples with standard protocols. The apparent pK_a values of cytosines were obtained by fitting pH-dependent excited-state population and chemical shift to the Henderson-Hasselbalch equation as described previously.

^{13}C $R_{1\rho}$ relaxation dispersion measurements and data analysis.

The on- and off-resonance relaxation dispersion profiles were measured using the 1D selective $R_{1\rho}$ experiment developed by Al-Hashimi and co-workers and a constant-time approach described by Kay and co-workers, where $R_{1\rho}$ values were obtained from a single delay period (T_{EX}). For on-resonance experiments, the relaxation delay was set to T_{EX} = 32 ms; for off-resonance experiments, the relaxation delay was set to T_{EX} = 40 ms, except T_{EX} = 16ms for C20C6. Eleven ^{13}C spin-lock fields ($\omega/2\pi$) of 100, 199 (x2), 299, 399 (x2), 498, 786, 982, 1474, 1965 (x2), 2456, and 2947 Hz were employed, (x2) indicates performed in duplicates. Due to large C-C couplings,

the lowest ^{13}C spin-lock field ($\omega/2\pi$) of 100 Hz was not used in measuring on-resonance C5/C6/C1' RD profiles. For off-resonance experiments, three ^{13}C spin-lock fields ($\omega/2\pi$) of 299, 498, and 786 Hz were used. For $\omega/2\pi$ = 299 Hz, the ^{13}C offsets ranged between -600 and 600 Hz with a spacing of 100 Hz and a smaller spacing of 50 Hz between -200 and 200 Hz; for $\omega/2\pi$ = 498 Hz, the ^{13}C offsets ranged between -1000 and 1000 Hz with a spacing of 200 Hz from -1000 to -800 Hz and from 800 to 1000 Hz, a spacing of 100 Hz from -800 to -100 Hz and from 100 to 800 Hz, and single points at -50 and 50 Hz; for $\omega/2\pi$ = 786 Hz, the ^{13}C offsets ranged between -2000 and 2000 Hz with a spacing of 500 Hz from -2000 to -1000 Hz and 1000 to 2000 Hz, a spacing of 250 Hz from -1000 to -500 Hz and 500 to 1000 Hz, and a spacing of 100 Hz from -500 to -100 Hz and from 100 to 500 Hz, and single points at -50 and 50 Hz. ^{13}C spin-lock powers were calibrated according to the 1D approach by Guenneugues *et al.* as previously described. Relaxation dispersion profiles were analyzed as described in chapter 2, with the extracted $\Delta\omega$ used for mutate and chemical shift analysis as described by[62]. Low pH chemical shifts were derived from ^{1}H-^{13}C HSQCs recorded at pH 4.35 and compared with ^{1}H-^{13}C HSQCs recorded in $R_{1\rho}$ conditions at pH 6.5 in H_2O and 5% D_2O.

CHAPTER 5 - PROBING EXCITED CONFORMATINOAL STATES OF NUCLEIC ACIDS BY NITROGEN CEST NMR SPECTROSCOPY

Introduction

Many non-coding RNA functions depend on their intrinsic conformational flexibility to dynamically interconvert between different structures upon recognition of specific cellular cues[26,88]. As a powerful tool for characterizing biomolecular structures and dynamics, NMR has played critical roles in uncovering RNA conformational dynamics, leading towards a deeper mechanistic understanding of RNA functions[89–92]. In particular, recent developments in nucleic-acid NMR, including conventional $R_{1\rho}$ RD[93–95], Carr-Purcell-Meiboom-Gill (CPMG) RD[95–100], low spin-lock field $R_{1\rho}$ RD[53–55,62,101–104], and CEST spectroscopy[69,75,105–107], have opened avenues for characterizing RNA excited states that are often too sparsely populated and transiently lived to be detected by conventional structural biology approaches. These NMR RD methods have enabled discovery and identification of excited states across diverse non-coding RNAs, reinforcing an emerging view of RNA excited states as a 'hidden' layer of regulation[26,91].

Dynamic transition to an RNA excited state often involves formation and/or rearrangement of canonical and non-canonical base pairs at tertiary and secondary structural levels, which are mediated by various distinct hydrogen-bond interactions. The current nucleic-acid NMR RD techniques have provided a variety of probes for characterizing RNA excited states, including protons[94,107], proton-bonded carbons[55,69,98–100,106,108], and proton-bonded imino and amino nitrogens[102,103]. However, no current techniques employ non-protonated nitrogens, which serve as key hydrogen-bond acceptors in nucleic acids, to characterize RNA excited states. Hence, it is of

interest to utilize these unique probes to complement existing RD measurements and directly characterize ES base pairing interactions in RNA. Developed by Forsen and Hoffman in the early 1960s[109], the saturation transfer type NMR experiment has become a powerful and versatile approach for studying excited states in proteins and nucleic acids[75,110–112]. In particular, by avoiding complications due to homonuclear and heteronuclear scalar couplings[79,80,113], CEST NMR spectroscopy has demonstrated accurate characterization of excited states using complex spin systems, such as base and sugar carbons from uniformly $^{13}C/^{15}N$ labeled RNA samples[69,106] that can be difficult to study with other NMR RD techniques. Here, we present two CEST approaches that utilize protonated and non-protonated nitrogens as probes for characterizing RNA excited states. We demonstrated these methods on the *B. Cereus* fluoride riboswitch[114] and showed that ^{15}N CEST profiles[72] provide complementary information to previously reported ^{13}C CEST profiles in depicting a potential pathway of ligand-dependent allosteric regulation of the excited state of the fluoride riboswitch.

Results

^{15}N CEST Pulse Sequence

The pulse sequences for measuring ^{15}N CEST profiles in nucleic acids are based on the conventional ^{1}H-^{15}N HSQC scheme and feature different nitrogen magnetization preparations prior to the saturation transfer period (Fig. 5.1 A,B). These schemes are used to meet the distinct spectroscopic needs of two types of nitrogens in nucleic acids: protonated nitrogen, such as G-N1 and U-N3, and non-protonated nitrogen, such as A-N1/N3/N7 and G-N7 (Fig. 5.1C). Shown in Fig. 5.1A is the pulse sequence for probing protonated nitrogen, which is similar to the original CEST scheme by Kay and co-workers[75]. Here, a flip-back scheme is used to provide a simple way of water suppression with sufficient sensitivity for our measurements. The sensitivity enhancement

scheme in the original CEST pulse sequence[75] can be further implemented to study larger RNAs. Given $^{1}J_{NH}$ couplings being ~ 90–100Hz, ^{1}H magnetization can be efficiently converted to pure ^{15}N magnetization (N_z) via refocused INEPT prior to the saturation period, which we refer to hereinafter as the $^{1}J_{NH}$-based approach. During the subsequent mixing period with a weak ^{15}N B1 field, a 2.35 kHz $90_x240_y90_x$ composite pulse train[115] is used for ^{1}H decoupling to suppresses N-H cross relaxation, dipolar-dipolar/nitrogen CSA cross-correlated relaxation, and the ^{15}N multiplet structure in the CEST profile[75].

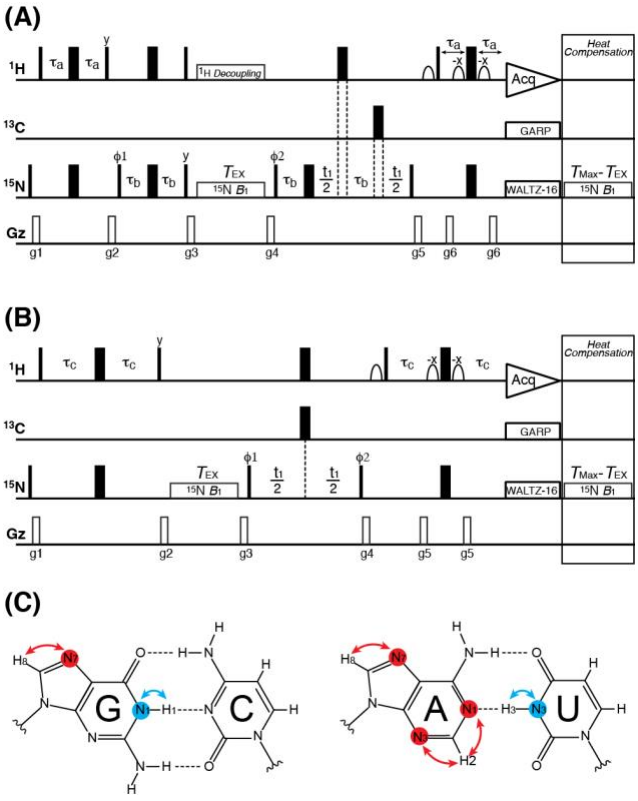

Figure 5.1 - 2D ^{15}N CEST pulse schemes.
Using (A) protonate nitrogen and (B) non-protonated nitrogen as probes for characterizing slow chemical exchange in nucleic acids**.** Highlighted in (C) are schematic magnetization transfers from protons to G-N1 and U-N3 via $^{1}J_{NH}$ couplings and to A-N1/N3/N7 and G-N7 via $^{2}J_{NH}$ couplings. The sequences are based on conventional ^{1}H-^{15}N HSQC pulse scheme with a simple water flip-back approach. Narrow (wide) rectangles are hard 90° (180°) pulses, and open shapes are 1-ms selective 90$_o$ pulses on water. All pulses are applied along *x*-axis unless indicated otherwise. (A) For measuring protonated nitrogens, a 90$_x$240$_y$90$_x$ composite pulse[115], as previously described by Kay and co-workers[75], is used for ^{1}H decoupling to suppress N-H cross relaxation and N-H dipolar-dipolar/nitrogen CSA cross-correlated relaxation during the T_{EX} period with a weak ^{15}N B_1 field. The ^{1}H carrier is kept on water resonance throughout the experiment except during the T_{EX} period, where it is shifted to the center of the region of interest. The ^{15}N carrier is also kept on-resonance throughout the experiment and is shifted to a desired offset during the T_{EX} period. Inter-pulse delays are set to τ_a = 2.4 ms and τ_b = 2.77 ms. The phase cycle used is ϕ_1 = {*x*, -*x*}, ϕ_2 = {*y*}, receiver = {*x*, -*x*}. Gradients with smoothed-square shape (SMSQ10.100) profile are applied with the following strength (G/cm)/duration (ms): g_1 = 19.8/1.0, g_2 = 29.7/1.0, g_3 = 6.6/1.0, g_4 = -26.4/1.0, g_5 = 29.7/1.0, g_6 = 33.0/1.0. ϕ_2 and the receiver phase are incremented in a States-TPPI manner. ^{13}C and ^{15}N decoupling during acquisition are achieved using 2.5 kHz GARP and 1.25 kHz WALTZ-16, respectively. (B) For measuring non-protonated nitrogens, only a weak ^{15}N B_1 field is applied during the T_{EX} period. The ^{1}H carrier is kept on water resonance throughout the experiment. The ^{15}N carrier is kept on-resonance throughout the experiment except being shifted to a desired offset during the T_{EX} period. Inter-pulse delay is set to τ_c = 9 ms. The phase cycle used is ϕ_1 = {*y*}, ϕ_2 = {*x*, -*x*}, receiver = {*x*, -*x*}. Gradients with smoothed-square shape (SMSQ10.100)

profile are applied with the following strength (G/cm)/duration (ms): g_1 = 19.8/1.0, g_2 = 29.7/1.0, g_3 = -26.4/1.0, g_4 = 29.7/1.0, g_5 = 33.0/1.0. ϕ_1 and the receiver phase are incremented in a States-TPPI manner. ^{13}C and ^{15}N decoupling during acquisition are achieved using 2.5 kHz GARP and 1.25 kHz WALTZ-16, respectively. To ensure uniform heating for experiments with variable lengths of T_{EX}, a heat compensation scheme is employed after the acquisition with length of T_{MAX} - T_{EX}, where T_{MAX} is the maximum relaxation delay time, and far off-resonance for both ^{1}H and ^{15}N channels.

For probing non-protonated nitrogens that can serve as hydrogen-bond acceptors, we employed a slightly different HSQC-based strategy to obtain ^{15}N CEST profiles (Fig. 5.1B) that is similar to recent approaches for measuring protein amide ^{1}H CEST profiles[80,116]. Here, unlike protonated nitrogens, transferring magnetization of a carbon-bonded proton to pure N_z magnetization of a nitrogen of interest is significantly less efficient due to small $^{2}J_{NH}$ scalar couplings (~ 10-15 Hz), which requires a lengthy refocused INEPT module that results in significant signal losses. Instead of generating pure N_z magnetization, the pulse scheme shown in Fig. 1B converts ^{1}H magnetization of a non-exchangeable proton to longitudinal two-spin order ($2N_zH_z$) prior to the saturation period, which we refer to hereinafter as the $^{2}J_{NH}$-based approach. However, due to the small $^{2}J_{NH}$ scalar couplings, we found that an inter-pulse delay of $1/4J \sim 17–25$ ms during INEPT remains too long to obtain adequate sensitivity. Hence, in practice, the length of the inter-pulse delay is varied to optimize sensitivity and achieve an optimal balance between the proton-nitrogen magnetization transfer and the loss of magnetization due to relaxation. For the riboswitch studied here, an inter-pulse delay of τ_c = 9 ms was found to give optimal sensitivity in the 2D ^{1}H-^{15}N HSQC spectrum, and any residual inphase transverse proton magnetization is dephased via Z gradient prior to the T_{EX} period. Applying a selective ^{15}N pulse together with nucleus-specific τ_c could further improve sensitivity for measuring site-specific CEST profiles, such as for N1 and N3 in As. In contrast to the $^{1}J_{NH}$-based approach (Fig. 5.1A), only a weak ^{15}N B_1 field is applied during the mixing period (Fig. 5.1B), as any inhomogeneity in the ^{1}H B_1 field can lead to additional signal losses of the two-spin order. Contributions from in-phase/anti-phase relaxation, longitudinal relaxation and cross-correlated relaxation to CEST profiles can be taken into account during the stage of data analysis with Bloch-McConnell equations (see Material and Methods). It is worth noting that the longitudinal two-spin order ($2N_zH_z$) can relax faster than pure

N_Z magnetization. Therefore, a shorter mixing time T_{EX} is typically used, which, consequentially, also limits the overall sensitivity of the $^2J_{NH}$-CEST approach in detecting excited states. While the triple resonance HCN scheme[117] could provide a potential pathway to transfer magnetizations from H8 to N7 in As/Gs and from H2 to N1/N3 in As, its application is severely limited due to extremely small scalar couplings ($^1J_{CN}$ < 3 Hz) for C8-N7 and C2-N1/N3 spin systems[118]. Alternatively, pure N_Z magnetization can be prepared using 1D selective Hartmann−Hahn polarization transfer scheme[119,120], as demonstrated in obtaining residue-specific $R_{1\rho}$[54,55] and carbon CEST[69,79] profiles. While it would still require a lengthy period of magnetization transfer (~1/J), the CEST profile of a single nitrogen of interest could be measured with a longer T_{EX} within a reasonable experimental time. Finally, given the large range of N1/N3/N7 chemical shifts, multi-frequency irradiation techniques, such as DANTE[121,122] (D-CEST)[123] and cosine-modulated (cos-CEST)[124] excitation schemes, can be implemented to expedite data collection.

Probing N1, N3 and N3 in the Fluoride Riboswitch

To demonstrate the two ^{15}N CEST methods, we carried out measurements on the *Bacillus cereus* fluoride riboswitch, a transcriptional riboswitch that regulates gene expression of fluoride transporters (Fig. 5.2A)[114]. Riboswitches are a class of non-coding RNAs that regulate transcription and/or translation via ligand-dependent conformational changes[125]. However, we recently showed that the aptamer domain of the *B. cereus* fluoride riboswitch adopts essentially identical tertiary structures with and without ligand[72]. Instead of changing its GS structure, the fluoride riboswitch utilizes a novel switching mechanism for riboswitch in which the ligand allosterically regulates dynamic access to a functional excited state (Fig. 5.2B). In the absence of ligand, the apo aptamer undergoes a GS ↔ ES conformational exchange, where the aptamer transiently unlocks the highly conserved reverse Hoogsteen base pair A37-U45, a linchpin that

resides at the interface between the aptamer and expression platform of the riboswitch. This linchpin gating process was shown to provide an efficient path for strand invasion to terminate transcription[72]. By contrast, ligand binding allosterically suppresses this ES transition, resulting in a single stable conformation that outcompetes the terminator during the co-transcriptional event to ensure continued gene transcription.

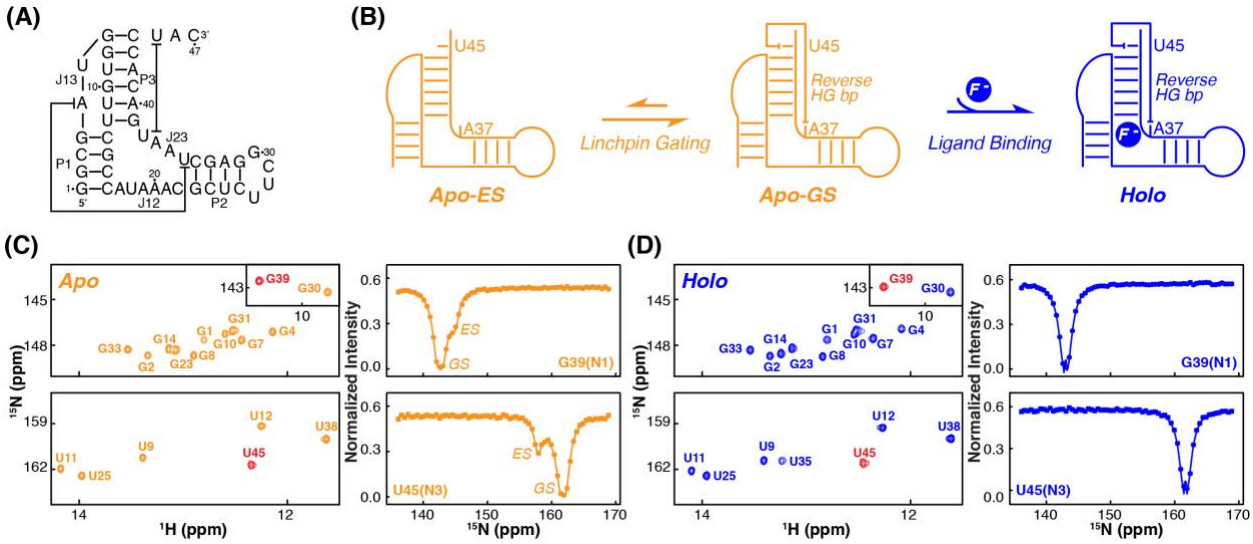

Figure 5.2 - Quantification of the excited state in the *B. cereus* fluoride riboswitch by $^{1}J_{NH}$-based 2D ^{15}N CEST.

(A) Sequence and secondary structure of the *B. cereus* fluoride riboswitch aptamer. (B) Schematic representation of ligand-dependent conformational transitions in the fluoride riboswitch aptamer. The apo aptamer undergoes an exchange between the excited state (ES) and the ground state (GS). Upon ligand binding, the holo aptamer adopts a single stable state. (C) ^{15}N CEST measurements of protonated nitrogens in the apo state. Shown on the left panel is ^{1}H-^{15}N HSQC spectrum of N1–H1 (Gs) and N3–H3 (Us) of the apo riboswitch, where colored in red are residues whose ^{15}N CEST profiles are shown on the right panel. Solid lines represent the best fits to a two-state exchange process using the Bloch-McConnell equation. (D) ^{15}N CEST measurements of protonated nitrogens in the holo state. Shown on the left panel is ^{1}H-^{15}N HSQC spectrum of N1–H1 (Gs) and N3–H3 (Us) of the holo riboswitch, where colored in red are residues whose CEST profiles are shown on the right panel. Shown are representative profiles with a ^{15}N B_1 field ($\omega/2\pi$) of 27.21 Hz for a duration of T_{EX} = 0.4 s.

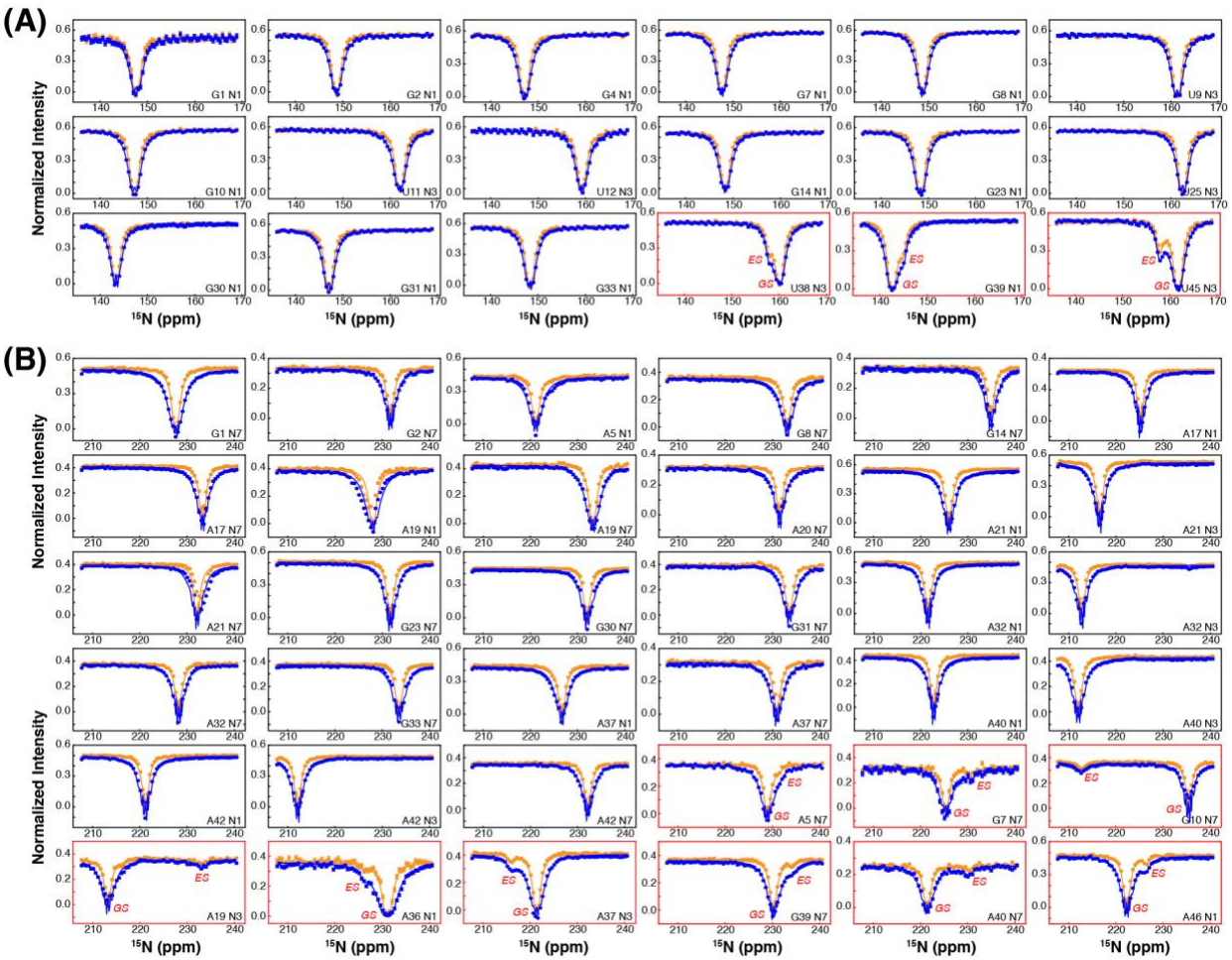

Figure 5.3 - Quantifying chemical exchange in the apo B. cereus fluoride riboswitch aptamer by 15N CEST spectroscopy.

(A) 15N CEST profiles depicting 15N B1 field strength and carrier (in ppm) dependence of protonated nitrogen intensities. Colored in orange and blue are profiles obtained with 15N B1 fields of 27.21 Hz and 37.17 Hz, respectively. (B) 15N CEST profiles depicting 15N B1 field strength and carrier (in ppm) dependence of non-protonated nitrogen intensities. Colored in orange and blue are profiles obtained with 15N B1 fields of 27.21 Hz and 52.1 Hz, respectively. Highlighted in red are profiles with second and asymmetrically broadened intensity dips.

We first applied the $^{1}J_{NH}$-based approach to obtain ^{15}N CEST profiles for the protonated G-N1 and U-N3 nitrogens of the fluoride riboswitch (Fig. 5.2 C,D). The ^{1}H-^{15}N HSQC spectrum of the imino region of the apo riboswitch is well resolved, and we were able to obtain a total of 18 ^{15}N CEST profiles, including 12 G-N1 profiles and 6 U-N3 profiles (Fig. 5.2C 5.3A). Consistent with our previous ^{13}C CEST characterizations[72], the imino ^{15}N CEST profiles also revealed the presence of an excited state in the apo state. While most Gs and Us display single dips in their intensity profile that match peak positions in the ^{1}H-^{15}N HSQC spectrum, second and asymmetrically broadened intensity dips that correspond to the apo ES can be clearly seen for U38, G39, and U45 (Fig. 5.2C and 5.3A). Global fitting of these three ^{15}N CEST profiles to a single two-state model gave an exchange rate (k_{ex}) of 184 ± 10 s^{-1} and an ES population (p_{ES}) of 1.6 ± 0.1%, resulting in an ES lifetime ($\tau_{ES} = 1/k_{EG}$) of 5.5 ± 0.5 ms. These ES parameters are similar to values (p_{ES} = 1.4 ± 0.1% and τ_{ES} = 3.2 ± 0.3 ms) from our previous ^{13}C CEST measurements[72]. The observed discrepancy is largely due to differences in sample conditions, where the ^{15}N CEST experiments were carried out on ^{15}N-labeled samples in H_2O and the ^{13}C CEST measurements were carried out on ^{13}C/^{15}N-labeled samples in D_2O. As a control, we measured imino ^{15}N CEST profiles for the holo aptamer (Fig. 5.2D). Due to spectral overlap between G10 and G31 as well as the appearance of U35 resonance, we were able to obtain a total of 17 ^{15}N CEST profiles in the holo state, including 10 G-N1 profiles and 7 U-N3 profiles (Fig. 5.2D and 5.5A). The uniform single-intensity dip of these ^{15}N CEST profiles is consistent with the holo state being in a single stable conformation.

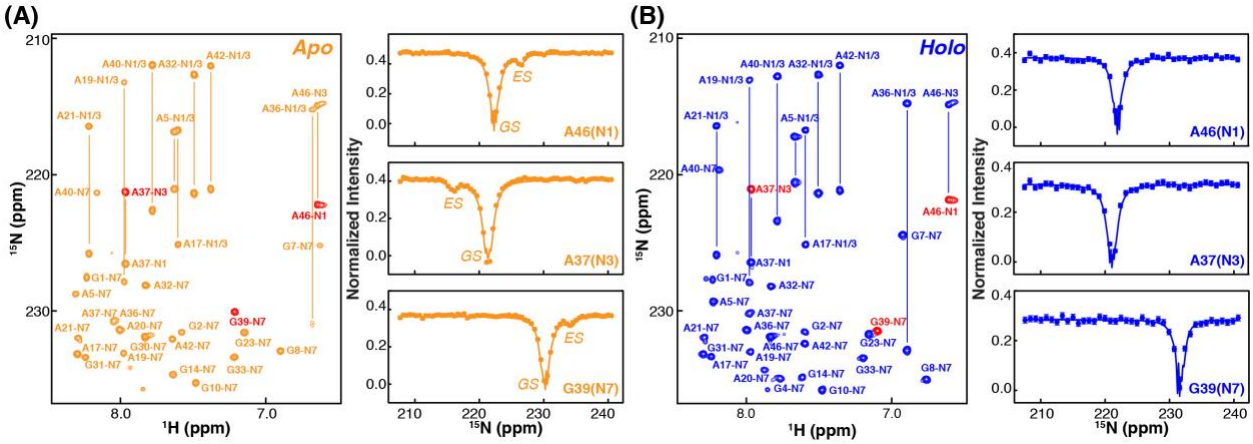

Figure 5.4 - Quantification of the excited state in the *B. cereus* fluoride riboswitch by $^{2}J_{NH}$-based 2D ^{15}N CEST.

(A) ^{15}N CEST measurements of non-protonated nitrogens in the apo state. Shown on the left panel is ^{1}H-^{15}N HSQC spectrum of N7–H8 (As/Gs) and N1/3–H2 (As) of the apo riboswitch, where colored in red are residues whose ^{15}N CEST profiles are shown on the right panel. Solid lines represent the best fits to a two-state exchange process using the Bloch-McConnell equation. (B) ^{15}N CEST measurements of non-protonated nitrogens in the holo state. Shown on the left panel is ^{1}H-^{15}N HSQC spectrum of N7–H8 (As/Gs) and N1/3–H2 (As) of the apo riboswitch, where colored in red are residues whose ^{15}N CEST profiles are shown on the right panel. Shown are representative profiles with a ^{15}N B_1 field ($\omega/2\pi$) of 27.21 Hz for a duration of T_{EX} = 0.1 s.

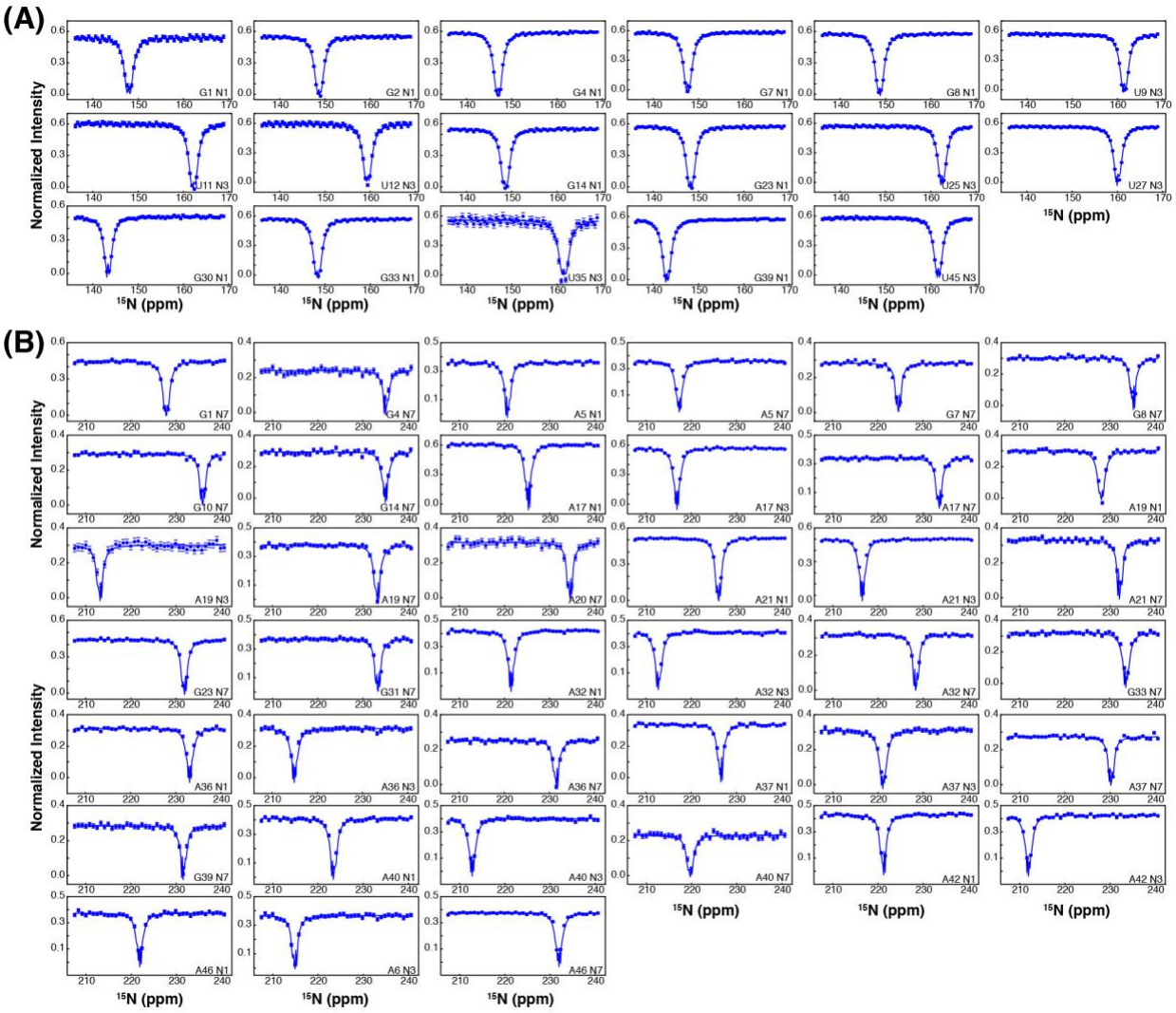

Figure 5.5 - Quantifying chemical exchange in the holo *B. cereus* fluoride riboswitch aptamer by ^{15}N CEST spectroscopy.

^{15}N CEST profiles depicting carrier (in ppm) dependence of (A) protonated and (B) non-protonated nitrogen intensities in the holo state with a ^{15}N B_1 fields of 27.21 Hz.

Relative to carbon-based approaches, measurable imino nitrogens can be too spatially clustered around structured regions of RNA to provide a comprehensive set of CEST profiles. This is largely due to imino protons being solvent exchangeable, and only structured or highly protected imino resonances are observable for direct measurement. In contrast to imino nitrogens, non-protonated nitrogens can provide probes distributed across structured and unstructured regions, as their magnetizations are transferred from non-exchangeable, carbon-bonded protons. In addition, while a single ^{1}H-^{15}N HSQC spectrum detects one non-protonated nitrogen (N7) of guanine, each adenine has up to three of these probes (N1, N3 and N7), further expanding the number of probes available for characterizing conformational exchange. Shown in Fig. 3 are ^{1}H-^{15}N HSQC spectra of these non-protonated nitrogens, and their corresponding ^{15}N CEST profiles are measured using the $^{2}J_{NH}$-based approach. In the apo state, we were able to obtain a total of 36 ^{15}N CEST profiles from the non-protonated nitrogens, 11 from G-N1 and 25 from A-N1/N3/N7 (Fig. 3A and Fig. 5.3B). While most of the non-protonated nitrogen CEST profiles display single intensity dips at corresponding resonance positions in ^{1}H -^{15}N HSQC spectrum, 9 of the 36 profiles exhibit second and asymmetrically broadened intensity dips. These non- protonated nitrogen profiles can be globally fitted to extract k_{ex} = 171 ± 31 s^{-1} and p_{ES} = 2.0 ± 0.2%, agreeing well with parameters obtained from imino nitrogen CEST profiles. Hence, we globally fit all ^{15}N CEST profiles to a single two-state model, and the resulting k_{ex} = 192 ± 15 s^{-1}, p_{ES} = 1.7 ± 0.1%, and τ_{ES} = 5.3 ± 0.5 ms are similar to ES parameters from previous ^{13}C CEST characterization of the apo fluoride riboswitch. We also measured ^{15}N CEST profiles for non-protonated nitrogens in the holo aptamer, in which all 39 profiles uniformly displayed single intensity dips (Fig. 2D and Fig. S5.5B).

With protonated and non-protonated nitrogen probes, we have obtained a total of 54 and 56 ^{15}N CEST profiles for the apo and holo fluoride riboswitches, respectively. These data, together

with our previously reported 61 apo-state and 65 holo-state ^{13}C CEST profiles[72], provide a comprehensive map of the ligand-dependent conformational tuning of the fluoride riboswitch (Fig. 5.5). Upon ligand binding, the fleeting dynamics to the ES is suppressed, resulting in one kinetically stable conformation as strongly supported by the uniform single-intensity dips across all holo-state ^{13}C/^{15}N CEST profiles. Consistent with our previous results, except A17 and A19 that are stacked on the end of P1, residues experiencing the chemical exchange in the apo state are clustered around the junction of P3, J13, J23, and the 3' tail (Fig. 5.6). In particular, the N7 ^{15}N CEST profile of A40 unveiled the presence of GS ↔ ES transition at this location, which was not detected in our previous ^{13}C CEST measurements. The ^{15}N CEST profiles have also unveiled changes in hydrogen bonds that cannot be directly characterized using ^{13}C CEST experiments.

For example, the N3 ^{15}N CEST profile of U45, which base pairs with A37 in the GS, displays an upfield-shifted ES chemical shift (Fig. 5.2C), indicating that U45-N3H3 may not be base paired in the ES. Interestingly, we could not detect dispersion on the hydrogen-bond receptor A37-N7, which could be due to limited difference between GS and ES chemical shifts of A37-N7. Similar to U45, the N3 ^{15}N CEST profile of U38 also displays an upfield-shifted ES chemical shift (Fig. 5.3A), suggesting that the hydrogen bond between U38-N3H3 and C41-O2P may be absent in the ES. Furthermore, the N7 ^{15}N CEST profiles of G7 and A40 reveal that the G7-N7 to A37-2'OH and A40-N7 to U38-2'OH hydrogen bonds are possibly broken in the ES. Due to the unique N–HO hydrogen bond, N7 chemical shifts of G7 and A40 are substantially upfield-shifted in the GS (Fig. 3A). In contrast, their ES chemical shifts are both downfield-shifted to the chemical shift range of non-hydrogen-bonded N7s in As and Gs (Fig. 5.3B). Together, ^{15}N/^{13}C CEST profiles of the apo riboswitch depict a trail of spatially continuous residues that bridge the ligand binding pocket and the linchpin gating site, suggesting a potential pathway for how ligand-binding

allosterically regulates dynamic access to the functional excited state.

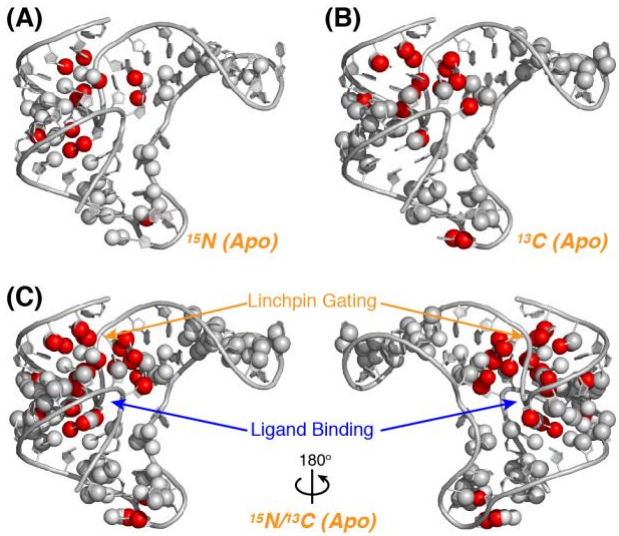

Figure 5.6 - Nitrogen and carbon CEST mapping of the excited state in the apo *B. cereus* fluoride riboswitch.
Spheres shown on the apo aptamer structure are the sites where (A) ^{15}N, (B) ^{13}C, and (C) ^{15}N/^{13}C CEST data were measured. Gray spheres are probes fit to a single-state model and red spheres are probes fit to a two-state exchange model.

Discussion

In summary, we have presented two ^{15}N CEST experiments to characterize slow chemical exchange in nucleic acids, which add to the growing list of NMR RD methods that have played critical roles in discovering and identifying excited states across diverse non-coding RNAs. The protonated and non-protonated nitrogens employed in these methods not only complement current probes for NMR RD measurements, but further enable direct characterization of hydrogen-bond donors and acceptors, which provide key interactions for canonical and non-canonical base pairs. Recently, it was shown that the network of these base pairs defines the overall topology of RNA tertiary structures[126,127], where even sparsely populated, imino-based NMR constraints can provide sufficient experimental input for computational modeling of high-resolution RNA structures[128]. With the ability to directly probe ES base pairing interactions, the ^{15}N CEST methods presented here further promise an opportunity for high-resolution structural modeling of functional RNA excited states.

Methods

Sample Preparation

Uniformly ^{15}N-labeled *Bacillus cereus* fluoride riboswitch samples were prepared as previously described[69]. Briefly, after *in vitro* transcription, RNA samples were ethanol precipitated, gel purified, electro-eluted with the Elutrap system (Whatman), anion-exchange purified with a Hi-Trap Q column (GE Healthcare), and desalted by exchanging into H_2O. The ligand-free (apo) samples were prepared by exchanging the desalted RNA into a Mg^{2+} saturated condition with a final RNA concentration ~ 0.5 – 1mM with 2mM free Mg^{2+} in NMR buffer (10 mM sodium phosphate with pH 6.4, 50mM KCl, 50μM EDTA). The fluoride-bound (holo)

samples were prepared by the addition of 10mM sodium fluoride directly to the apo RNA samples. 5% D_2O (Sigma) was added into all NMR samples.

NMR spectroscopy

All NMR experiments were carried out on a Bruker Avance III 600 spectrometer equipped with 5mm triple-resonance cryogenic probes at 303K. For protonated imino (G-N1 and U-N3) ^{15}N CEST experiments on the apo riboswitch, ^{15}N B_1 fields ($\omega/2\pi$) of 27.21 Hz and 37.17 Hz were used during the T_{EX} = 0.4 s period. The ^{15}N carrier was set to 152.4 ppm with a spectral width of 25 ppm, and the ^{15}N offsets ranged from -990 to 990 Hz with spacing of 30 Hz. For non-protonated (A-N1/N3/N7 and G-N7) ^{15}N CEST experiments on the apo riboswitch, ^{15}N B_1 fields ($\omega/2\pi$) of 27.21 Hz and 52.1 Hz were used during the T_{EX} = 0.1 s period. The ^{15}N carrier was set to 224 ppm with a spectral width of 26 ppm, and the ^{15}N offsets ranged from -990 to 990 Hz with spacing of 30 Hz. For protonated imino (G-N1 and U-N3) ^{15}N CEST experiments on the holo riboswitch, a ^{15}N B_1 field ($\omega/2\pi$) of 27.21 Hz was used during the T_{EX} = 0.4 s period. The ^{15}N carrier was set to 152.4 ppm with a spectral width of 25 ppm, and the ^{15}N offsets ranged from -990 to 990 Hz with spacing of 30 Hz. For non-protonated (A-N1/N3/N7 and G-N7) ^{15}N CEST experiments on the holo riboswitch, a ^{15}N B_1 field ($\omega/2\pi$) of 27.21 Hz was used during the T_{EX} = 0.1 s period. The ^{15}N carrier was set to 224 ppm with a spectral width of 26 ppm, and the ^{15}N offsets ranged from -1000 to 1000 Hz with spacing of 50 Hz. These spin-lock powers were calibrated according to the 1D approach by Guenneugues et al.[74] as previously described[69,75]. For all CEST measurements, three spectra with T_{EX} = 0 s were recorded for reference in data fitting and error estimation.

Data analysis

NMR spectra were processed and analyzed using NMRPipe/NMRDraw[70], NMRView[71], and Sparky 3.110. (University of California, San Francisco, CA). All CEST profiles were obtained

by normalizing the peak intensity as a function of spin lock offset Ω to the peak intensity recorded at T_{EX} = 0 s, where $\Omega = \omega_{rf} - \Omega_{obs}$ is the difference between the spin-lock carrier (ω_{rf}) and the observed peak (Ω_{obs}) frequencies. Measurement errors were estimated based on both triplicates at T_{EX} = 0 and the baseline of CEST profiles. The CEST profiles of residues displaying conformational exchange were fit to a two-state exchange between ground (G) and excited (E) states based on the Bloch-McConnell equation[76] that describes magnetization evolution in a coupled two-spin ^{15}N-^{1}H system[77,78]. For individual state (i), the evolution of its magnetization ($\mathbf{v}_i$) as a coupled two-spin ^{15}N-^{1}H system is described by[77],

$$\frac{d}{dt}\mathbf{v}^i = -\mathbf{R}^i\mathbf{v}^i = \begin{pmatrix} R_2^i & \omega_N^i & 0 & \eta_{xy}^i & \pi J_{NH}^i & 0 \\ -\omega_N^i & R_2^i & \omega_1 & -\pi J_{NH}^i & \eta_{xy}^i & 0 \\ 0 & -\omega_1 & R_1^i & 0 & 0 & \eta_z^i \\ \eta_{xy}^i & \pi J_{NH}^i & 0 & R_{2HN}^i & \omega_N^i & 0 \\ -\pi J_{NH}^i & \eta_{xy}^i & 0 & -\omega_N^i & R_{2HN}^i & \omega_1 \\ 0 & 0 & \eta_z^i & 0 & -\omega_1 & R_{1HN}^i \end{pmatrix} \begin{pmatrix} N_x^i \\ N_y^i \\ N_z^i \\ 2H_zN_x^i \\ 2H_zN_y^i \\ 2H_zN_z^i \end{pmatrix}$$

where R_{1i} is the ^{15}N longitudinal relaxation rate, R_{2i} is the ^{15}N transverse relaxation, R_{1HNi} is the ^{15}N-^{1}H two-spin order relaxation rate, R_{2HNi} is the ^{15}N antiphase relaxation rate, η_{zi} is the N-H dipolar-dipolar/nitrogen CSA cross-correlated relaxation between the ^{15}N longitudinal and two-spin order elements, η_{xyi} is N-H dipolar-dipolar/nitrogen CSA cross-correlated relaxation between ^{15}N transverse and antiphase magnetizations, ω_{Ni} is the offset of the applied ^{15}N B_1 field with a strength of ω_1 from state i, and J_{NHi} is the ^{15}N-^{1}H scalar coupling. The evolution of GS and ES magnetizations in a two-state exchange model can be described by,

$$\frac{d}{dt}\boldsymbol{\sigma}(\mathrm{t}) = -\mathbf{L}\cdot\begin{bmatrix}\mathbf{v}^G \\ \mathbf{v}^E\end{bmatrix} = \left(\begin{bmatrix}\mathbf{R}^G & \mathbf{0}_6 \\ \mathbf{0}_6 & \mathbf{R}^E\end{bmatrix} + \begin{bmatrix}-k_{GE} & k_{EG} \\ k_{GE} & -k_{EG}\end{bmatrix} \otimes \mathbf{1}_6\right)\cdot\begin{bmatrix}\mathbf{v}^G \\ \mathbf{v}^E\end{bmatrix}$$

where $v_{G/E}$ and $R_{G/E}$ are magnetization and relaxation matrices for ground and excited states as detailed above, 0_6 and 1_6 are 6x6 null and identity matrices, respectively, and k_{GE} and k_{EG} are

forward and backward exchange rates as defined by $k_{GE} = p_E k_{ex}$ and $k_{EG} = p_G k_{ex}$. Here, $k_{ex} = k_{GE} + k_{EG}$ is the rate of exchange, p_G and p_E are populations of ground and excited states, respectively, and $\omega_G = \Omega_{obs}$ and $\omega_E = \omega_G + \Delta\omega$, where $\Delta\omega$ is the chemical shift difference between the ground and excited states. Ground state and excited state magnetizations at the beginning of the T_{EX} period are along Z and are set to be at populations of p_G and p_E. For G-N1 and U-N3 ^{15}N CEST profiles obtained using the $^1J_{NH}$-based HSQC CEST pulse sequence, the initial magnetization is N_z magnetization; for A-N1/N3/N7 and G-N7 ^{15}N CEST profiles obtained using the $^2J_{NH}$-based HSQC CEST pulse sequence, the initial magnetization is the two-spin order ($2N_zH_z$). During analysis of $2N_zH_z$ CEST profiles, the fitting parameters are $\Delta\omega$, k_{ex}, p_E, $R_2 = R_{2G/E}$, $R_{1HN} = R_{1HNG/E}$, and $R_{2HN} = R_{2G/E} + R_{1HNG/E} - R_{1G/E}$ as described previously[79,106]. To simplify data fitting, $\eta_{zG/E}$ and $\eta_{xyG/E}$ were set to 0 as they have been shown not to affect the extracted $\Delta\omega$, k_{ex}, and p_E values[80]. In addition, we set $R_{1G/E}$ to 0, as the data does not constrain determination of R_1, and varying R_1 (0.0–2.0) s^{-1} minimally effects extracted $\Delta\omega$, k_{ex}, and p_E values. Since G-N1 and U-N3 ^{15}N CEST profiles were measured in the presence of ^{1}H decoupling, their data analysis can be simplified by setting all two-spin relaxation parameters (R_{1HNi}, R_{2HNi}, η_{zi}, η_{xyi}, and J_{NHi}) to 0, which essentially turns the individual-state evolution from a 6x6 matrix to a 3x3 matrix. While the non-^{1}H-decoupled $^2J_{NH}$ CEST profiles are split into doublets, the small couplings ($^2J_{NH}$ ~ 10–15 Hz) are not resolved with the applied B_1 fields. For residues without conformational exchange, the two-state model was simplified to a one-state model by fixing all exchange parameters (rate of exchange k_{ex} and population of excited state p_{ES}) to 0. All profiles were fitted using an in-house MATLAB® program with a Levenberg-Marquardt algorithm.

CHAPTER 6 - FUTURE DIRECTIONS

pH as a Widespread Mechanism for RNA Regulation

It has become increasingly clear that many non-coding RNAs (ncRNAs) do not fold into single static structures, instead, they dynamically interconvert between different conformational states for function[26]. Recent developments in NMR techniques have opened new avenues to probe RNA structural transitions involving alternative conformational states that often evade detection from conventional biophysical and biochemical methods due to their low populations and/or transient lifetimes[56,57,62,91,129]. Indeed, these technical advances have unveiled the presence of a diverse set of excited states in non-coding RNAs. Our discovery of the transiently protonated state in pre-miR-21 was also made possible by these new techniques. In particular, the development of nucleic-acid-optimized ^{15}N NMR CEST method has enabled, for the first time, direct identification of transient protonation in nucleic acids, illuminating the crucial chemical basis for delineating the relationship between structure, dynamics, and function. With the emerging view of ESs as a 'hidden' layer of regulation[26], the detection of an another functional protonation state, the mGHG motif of primary microRNAs, make a compelling case for protonation as a prolific and essential contributor to the dynamic ensemble of RNA. Unlike other, higher order structures that involve protonation[130–133], these new observations of protonated RNA states involving less shifted pK_as reveal the potential for protonated states to exist as low-populated structures in the RNA ensemble at physiological conditions. Due to their transient nature, these states haven't been widely observed in these more simplistic, hairpin-like RNA structures. This discovery, in combination with the growing repertoire of functional RNA ESs with distinct structural features, promise potential for

the discovery of additional functional protonated states across RNA as a general mechanism for function. Although very information rich, the NMR techniques discussed in this work are very labor and time intensive. To investigate the large number of potential protonation sites across the transcriptome, a higher throughput technique, such as chemical probing could be adapted for use across multiple pH points. Dimethyl sulfate (DMS) or trimethyloxonium (TMO) probing, in combination with high-throughput sequencing, could provide hits across wide classes of RNA, which would be promising candidates for deep interrogate of structure, dynamics, and function by NMR.

NMR method development

Recently, it was shown that the network of base pairs defines the overall topology of RNA tertiary structures[126,127], where even sparsely populated, imino-based NMR constraints can provide sufficient experimental input for computational modeling of high-resolution RNA structures[128,134]. The protonated and non-protonated nitrogens employed in our new ^{15}N CEST method not only complement current probes for NMR RD measurements, but further enable direct characterization of hydrogen-bond donors and acceptors, which provide key interactions for canonical and non-canonical base pairs. Furthermore, we plan to expand the NMR-RNA toolkit to include ^{1}H and ^{15}N $R_{1\rho}$, increasing the overall number of atomic probes and timescales of motion in which excited states can be detected[93–95]. With the ability to directly probe ES base pairing interactions, our ^{15}N CEST methods can directly identify transient folding conformers in the RNA ensemble. Complimented with information about the ES from additional probes, this toolkit further promises an opportunity for high-resolution structural modeling of functional RNA ESs, enabling developments in RNA-targeted therapeutics that require such structures.

REFERENCES

1. Friedman, R. C., Farh, K. K. H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* (2009).

2. Siomi, H. & Siomi, M. C. Posttranscriptional Regulation of MicroRNA Biogenesis in Animals. *Molecular Cell* (2010).

3. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology* (2014).

4. Gebert, L. F. R. & MacRae, I. J. Regulation of microRNA function in animals. *Nature Reviews Molecular Cell Biology* (2019).

5. Treiber, T., Treiber, N. & Meister, G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nature Reviews Molecular Cell Biology* (2019).

6. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* (2004).

7. Cai, X., Hagedorn, C. H. & Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* (2004).

8. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* (2003).

9. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* (2004).

10. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* (2003).

11. Hutvágner, G. *et al.* A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA. *Science (80-.).* (2001).

12. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev.* (2001).

13. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics* (2010).

14. Esteller, M. Non-coding RNAs in human disease. *Nature Reviews Genetics* (2011).

15. Lin, S. & Gregory, R. I. MicroRNA biogenesis pathways in cancer. *Nature Reviews Cancer* (2015).

16. Zhang, X. & Zeng, Y. The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic Acids Res.* (2010).

17. Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H. & Tomari, Y. Recognition of the pre-miRNA structure by Drosophila-Dicer-1. *Nat. Struct. Mol. Biol.* (2010).

18. Auyeung, V. C., Ulitsky, I., McGeary, S. E. & Bartel, D. P. Beyond secondary structure: Primary-sequence determinants license Pri-miRNA hairpins for processing. *Cell* (2013).

19. Nguyen, T. A., Park, J., Dang, T. L., Choi, Y. G. & Kim, V. N. Microprocessor depends on hemin to recognize the apical loop of primary microRNA. *Nucleic Acids Res.* (2018).

20. Gu, S. *et al.* The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell* (2012).

21. Contrant, M. *et al.* Importance of the RNA secondary structure for the relative accumulation of clustered viral microRNAs. *Nucleic Acids Res.* (2014).

22. Van Kouwenhove, M., Kedde, M. & Agami, R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nature Reviews Cancer* (2011).

23. Liu, Z., Wang, J., Li, G. & Wang, H. W. Structure of precursor microRNA's terminal loop regulates human Dicer's dicing activity by switching DExH/D domain. *Protein Cell* (2015).

24. Pandolfini, L. *et al.* METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. *Mol. Cell* (2019).

25. Dallaire, P. *et al.* Structural dynamics control the MicroRNA maturation pathway. *Nucleic Acids Res.* (2016).

26. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* (2019).

27. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science (80-.).* (2001).

28. Sekuklu, S. D., Donoghue, M. T. A. & Spillane, C. miR-21 as a key regulator of oncogenic processes. in *Biochemical Society Transactions* (2009).

29. Pfeffer, S. R., Yang, C. H. & Pfeffer, L. M. The Role of MIR-21 in Cancer. *Drug Dev. Res.* (2015).

30. Krichevsky, A. M. & Gabriely, G. miR-21: A small multi-faceted RNA. *Journal of Cellular and Molecular Medicine* (2009).

31. Si, M. L. *et al.* miR-21-mediated tumor growth. *Oncogene* (2007).

32. Asangani, I. A. *et al.* MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* (2008).

33. Zeng, Y., Yi, R. & Cullen, B. R. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* (2005).

34. Diaz, J. P. *et al.* Association of a peptoid ligand with the apical loop of pri-miR-21 inhibits cleavage by Drosha. *RNA* (2014).

35. Shortridge, M. D. *et al.* A Macrocyclic Peptide Ligand Binds the Oncogenic MicroRNA-21 Precursor and Suppresses Dicer Processing. *ACS Chem. Biol.* (2017).

36. Bose, D. *et al.* Selective inhibition of miR-21 by phage display screened peptide. *Nucleic Acids Res.* (2015).

37. Chen, Y. *et al.* Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. *Nat. Chem. Biol.* (2016).

38. Velagapudi, S. P. *et al.* Approved Anti-cancer Drugs Target Oncogenic Non-coding RNAs. *Cell Chem. Biol.* (2018.

39. Liu, Z. *et al.* Cryo-EM Structure of Human Dicer and Its Complexes with a Pre-miRNA Substrate. *Cell* (2018).

40. Chirayil, S., Wu, Q., Amezcua, C. & Luebke, K. J. NMR characterization of an oligonucleotide model of the MiR-21 Pre-Element. *PLoS One* (2014).

41. Das, R., Karanicolas, J. & Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods* (2010).

42. Chen, Y. *et al.* Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. *Nucleic Acids Res.* (2016).

43. Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* (2004).

44. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* (2004).

45. Han, J. *et al.* Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell* (2006).

46. Ma, H., Wu, Y., Choi, J. G. & Wu, H. Lower and upper stem-single-stranded RNA junction.

47. Fang, W. & Bartel, D. P. The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. *Mol. Cell* (2015).

48. Kwon, S. C. *et al.* Molecular Basis for the Single-Nucleotide Precision of Primary microRNA Processing. *Mol. Cell* (2019).

49. Zeng, Y. & Cullen, B. R. Sequence requirements for micro RNA processing and function in human cells. *RNA* (2003).

50. Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U. S. A.* (2006).

51. Kumarswamy, R., Volkmann, I. & Thum, T. Regulation and function of miRNA-21 in health and disease. *RNA Biology* (2011).

52. Trabucchi, M. *et al.* The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* (2009).

53. Massi, F., Johnson, E., Wang, C., Rance, M. & Palmer, A. G. NMR R1ρ Rotating-Frame Relaxation with Weak Radio Frequency Fields. *J. Am. Chem. Soc.* (2004).

54. Korzhnev, D. M., Orekhov, V. Y. & Kay, L. E. Off-resonance R1ρ NMR studies of exchange dynamics in proteins with low spin-lock fields: An application to a fyn SH3 domain. *J. Am. Chem. Soc.* (2005).

55. Hansen, A. L., Nikolova, E. N., Casiano-Negroni, A. & Al-Hashimi, H. M. Extending the range of microsecond-to-millisecond chemical exchange Detected in labeled and unlabeled nucleic acids by selective carbon R NMR spectroscopy. *J. Am. Chem. Soc.* (2009).

56. Sekhar, A. & Kay, L. E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proceedings of the National Academy of Sciences of the United States of America* (2013.

57. Palmer, A. G. Chemical exchange in biomacromolecules: Past, present, and future. *J. Magn. Reson.* (2014).

58. Houck-Loomis, B. *et al.* An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* (2011).

59. Chu, C. C., Plangger, R., Kreutz, C. & Al-Hashimi, H. M. Dynamic ensemble of HIV-1 RRE stem IIB reveals non-native conformations that disrupt the Rev-binding site. *Nucleic Acids Res.* (2019).

60. Zhao, B., Baisden, J. T. & Zhang, Q. Probing excited conformational states of nucleic acids by nitrogen CEST NMR spectroscopy. *J. Magn. Reson.* (2020).

61. Thaplyal, P. & Bevilacqua, P. C. Experimental approaches for measuring pKa's in RNA and DNA. in *Methods in Enzymology* (2014).

62. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-Hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* (2012).

63. Lee, J., Dethoff, E. A. & Al-Hashimi, H. M. Invisible RNA state dynamically couples distant motifs. *Proc. Natl. Acad. Sci. U. S. A.* (2014).

64. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* (2008).

65. Pan, B., Mitra, S. N. & Sundaralingam, M. Crystal structure of an RNA 16-mer Duplex R(GCAGAGUUAAAUCUGC)2 with nonadjacent G(Syn)·A+(Anti) mispairs. *Biochemistry* (1999).

66. Leonard, G. A. *et al.* Crystal and molecular structure of r(CGCGAAUUAGCG): an RNA duplex containing two G(anti)· A(anti) base pairs. *Structure* (1994).

67. Muth, G. W., Ortoleva-Donnelly, L. & Strobel, S. A. A single adenosine with a neutral pK(a) in the ribosomal peptidyl transferase center. *Science (80-.).* (2000).

68. Nam, Y., Chen, C., Gregory, R. I., Chou, J. J. & Sliz, P. Molecular basis for interaction of let-7 MicroRNAs with Lin28. *Cell* (2011).

69. Zhao, B., Hansen, A. L. & Zhang, Q. Characterizing slow chemical exchange in nucleic acids by carbon CEST and low spin-lock field R 1ρ NMR spectroscopy. *J. Am. Chem. Soc.* (2014).

70. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* (1995).

71. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* (1994).

72. Zhao, B., Guffy, S. L., Williams, B. & Zhang, Q. An excited state underlies gene regulation of a transcriptional riboswitch. *Nat. Chem. Biol.* (2017).

73. Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. pK(a) values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry* (1999).

74. Guenneugues, M., Berthault, P. & Desvaux, H. A Method for Determining B 1 Field Inhomogeneity. Are the Biases Assumed in Heteronuclear Relaxation Experiments Usually Underestimated? *Journal of Magnetic Resonance* (1999).

75. Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying 'invisible' excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* (2012).

76. McConnell, H. M. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* (1958).

77. Allard, P., Helgstrand, M. & Härd, T. The Complete Homogeneous Master Equation for a Heteronuclear Two-Spin System in the Basis of Cartesian Product Operators. *J. Magn. Reson.* (1998).

78. Helgstrand, M., Hard, T. & Allard, P. Simulations of NMR pulse sequences during equilibrium and non-equilibrium chemical exchange. *J. Biomol. NMR* (2000).

79. Hansen, A. L. & Kay, L. E. Measurement of histidine pKa values and tautomer populations in invisible protein states. *Proc. Natl. Acad. Sci. U. S. A.* (2014).

80. Sekhar, A., Rosenzweig, R., Bouvignies, G. & Kay, L. E. Hsp70 biases the folding pathways of client proteins. *Proc. Natl. Acad. Sci. U. S. A.* (2016).

81. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* (2003).

82. MacRae, I. J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science (80-.).* (2006).

83. MacRae, I. J., Zhou, K. & Doudna, J. A. Structural determinants of RNA recognition and cleavage by Dicer. *Nat. Struct. Mol. Biol.* (2007).

84. Lee, H. Y. & Doudna, J. A. TRBP alters human precursor microRNA processing in vitro. *RNA* (2012).

85. Lau, P. W. *et al.* The molecular architecture of human Dicer. *Nat. Struct. Mol. Biol.* (2012).

86. Taylor, D. W. *et al.* Substrate-specific structural rearrangements of human Dicer. *Nat. Struct. Mol. Biol.* (2013).

87. Zhang, H., Kolb, F. A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* (2002).

88. Dethoff, E. A., Chugh, J., Mustoe, A. M. & Al-Hashimi, H. M. Functional complexity and regulation through RNA dynamics. *Nature* (2012).

89. Rinnenthal, J. *et al.* Mapping the landscape of RNA dynamics with NMR spectroscopy. *Acc. Chem. Res.* (2011).

90. Bothe, J. R. *et al.* Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nature Methods* (2011).

91. Zhao, B. & Zhang, Q. Characterizing excited conformational states of RNA by NMR spectroscopy. *Current Opinion in Structural Biology* (2015).

92. Rangadurai, A., Szymaski, E. S., Kimsey, I. J., Shi, H. & Al-Hashimi, H. M. Characterizing micro-to-millisecond chemical exchange in nucleic acids using off-resonance R1ρ relaxation dispersion. *Progress in Nuclear Magnetic Resonance Spectroscopy* (2019).

93. Hoogstraten, C. G., Wank, J. R. & Pardi, A. Active Site Dynamics in the Lead-Dependent Ribozyme. *Biochemistry* (2000).

94. Steiner, E., Schlagnitweit, J., Lundström, P. & Petzold, K. Capturing Excited States in the Fast-Intermediate Exchange Limit in Biological Systems Using1H NMR Spectroscopy. *Angew. Chemie - Int. Ed.* (2016).

95. Palmer, A. G., Kroenke, C. D. & Loria, J. P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. in *Methods in Enzymology* (2001).

96. Korzhnev, D. M. & Kay, L. E. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Accounts of Chemical Research* (2008).

97. Lundström, P., Hansen, D. F. & Kay, L. E. Measurement of carbonyl chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy: Comparison between uniformly and selectively 13C labeled samples. *J. Biomol. NMR* (2008).

98. Johnson, J. E. & Hoogstraten, C. G. Extensive backbone dynamics in the GCAA RNA tetraloop analyzed using 13C NMR spin relaxation and specific isotope labeling. *J. Am. Chem. Soc.* (2008).

99. Kloiber, K., Spitzer, R., Tollinger, M., Konrat, R. & Kreutz, C. Probing RNA dynamics via longitudinal exchange and CPMG relaxation dispersion NMR spectroscopy using a sensitive 13C-methyl label. *Nucleic Acids Res.* (2011) .

100. Wunderlich, C. H. *et al.* Synthesis of (6- 13C)pyrimidine nucleotides as spin-labels for RNA dynamics. *J. Am. Chem. Soc.* (2012) .

101. Nikolova, E. N. *et al.* Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* (2011).

102. Nikolova, E. N., Gottardo, F. L. & Al-Hashimi, H. M. Probing transient Hoogsteen hydrogen bonds in canonical duplex DNA using NMR relaxation dispersion and single-atom substitution. *J. Am. Chem. Soc.* (2012).

103. Szymanski, E. S., Kimsey, I. J. & Al-Hashimi, H. M. Direct NMR Evidence that Transient Tautomeric and Anionic States in dG·dT Form Watson-Crick-like Base Pairs. *J. Am. Chem. Soc.* (2017).

104. Kimsey, I. J. *et al.* Dynamic basis for dG· dT misincorporation via tautomerization and ionization. *Nature* (2018).

105. Bouvignies, G. & Kay, L. E. Measurement of proton chemical shifts in invisible states of slowly exchanging protein systems by chemical exchange saturation transfer. *J. Phys. Chem. B* (2012).

106. Zhao, B. & Zhang, Q. Measuring Residual Dipolar Couplings in Excited Conformational States of Nucleic Acids by CEST NMR Spectroscopy. *J. Am. Chem. Soc.* (2015).

107. Chen, B., LeBlanc, R. & Dayie, T. K. SAM-II Riboswitch Samples at least Two Conformations in Solution in the Absence of Ligand: Implications for Recognition. *Angew. Chemie - Int. Ed.* (2016).

108. Blad, H., Reiter, N. J., Abildgaard, F., Markley, J. L. & Butcher, S. E. Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. *J. Mol. Biol.* (2005).

109. Forsén, S. & Huffman, R. A. Study of moderately rapid chemical exchange reactions by means of nuclear magnetic double resonance. *J. Chem. Phys.* (1963).

110. Fawzi, N. L., Ying, J., Ghirlando, R., Torchia, D. A. & Clore, G. M. Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature* (2011).

111. Vallurupalli, P., Sekhar, A., Yuwen, T. & Kay, L. E. Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. *J. Biomol. NMR* (2017).

112. Sekhar, A. & Kay, L. E. An NMR View of Protein Dynamics in Health and Disease. *Annu. Rev. Biophys.* (2019) .

113. Vallurupalli, P., Bouvignies, G. & Kay, L. E. A computational study of the effects of 13c-13c scalar couplings on 13C CEST NMR spectra: Towards studies on a uniformly 13C-labeled protein. *ChemBioChem* (2013).

114. Baker, J. L. *et al.* Widespread genetic switches and toxicity resistance proteins for fluoride. *Science (80-.).* (2012).

115. Levitt, M. H. Symmetrical composite pulse sequences for NMR population inversion. I. Compensation of radiofrequency field inhomogeneity. *J. Magn. Reson.* (1982).

116. Wu, Q., Fenton, B. A., Wojtaszek, J. L. & Zhou, P. Probing the excited-state chemical shifts and exchange parameters by nitrogen-decoupled amide proton chemical exchange saturation transfer (HNdec-CEST). *Chem. Commun.* (2017).

117. Sklenář, V., Peterson, R. D., Rejante, M. R. & Feigon, J. Two-and three-dimensional HCN experiments for correlating base and sugar resonances in 15N, 13C-labeled RNA oligonucleotides. *J. Biomol. NMR* (1993).

118. Ippel, J. H. *et al.* Heteronuclear Scalar Couplings in the Bases and Sugar Rings of Nucleic Acids: Their Determination and Application in Assignment and Conformational Analysis. *Magn. Reson. Chem.* (1996).

119. Pelupessy, P., Chiarparin, E. & Bodenhausen, G. Excitation of Selected Proton Signals in NMR of Isotopically Labeled Macromolecules. *Journal of Magnetic Resonance* (1999).

120. Pelupessy, P. & Chiarparin, E. Hartmann - Hahn polarization transfer in liquids: An ideal tool for selective experiments. *Concepts Magn. Reson.* (2000).

121. Bodenhausen, G., Freeman, R. & Morris, G. A. A simple pulse sequence for selective excitation in Fourier transform NMR. *J. Magn. Reson.* (1976).

122. Morris, G. A. & Freeman, R. Selective excitation in Fourier transform nuclear magnetic resonance. *J. Magn. Reson.* (1978).

123. Yuwen, T., Kay, L. E. & Bouvignies, G. Dramatic Decrease in CEST Measurement Times Using Multi-Site Excitation. *ChemPhysChem* (2018).

124. Yuwen, T., Bouvignies, G. & Kay, L. E. Exploring methods to expedite the recording of CEST datasets using selective pulse excitation. *J. Magn. Reson.* (2018).

125. Serganov, A. & Patel, D. J. Metabolite Recognition Principles and Molecular Mechanisms Underlying Riboswitch Function. *Annu. Rev. Biophys.* (2012).

126. Bailor, M. H., Sun, X. & Al-Hashimi, H. M. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science (80-.).* (2010).

127. Chu, V. B. *et al.* Do conformational biases of simple helical junctions influence RNA folding stability and specificity? *RNA* (2009).

128. Williams, B. *et al.* Structure modeling of RNA using sparse NMR constraints. *Nucleic Acids Res.* (2017).

129. Marušič, M., Schlagnitweit, J. & Petzold, K. Front Cover: RNA Dynamics by NMR Spectroscopy (ChemBioChem 21/2019). *ChemBioChem* (2019).

130. Keller, H., Weickhmann, A. K., Bock, T. & Wöhnert, J. Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitches. *RNA* (2018).

131. Smith, K. D., Lipchock, S. V., Livingston, A. L., Shanahan, C. A. & Strobel, S. A. Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch. *Biochemistry* (2010).

132. Ren, A. *et al.* Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. *Cell Rep.* (2015).

133. Wolter, A. C. *et al.* A Stably Protonated Adenine Nucleotide with a Highly Shifted pKaValue Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer. *Angew. Chemie - Int. Ed.* (2017).

134. Ebrahimi, P., Kaur, S., Baronti, L., Petzold, K. & Chen, A. A. A two-dimensional replica-exchange molecular dynamics method for simulating RNA folding using sparse experimental restraints. *Methods* (2019).