ABSTRACT

Title of Dissertation: THE CONFORMATIONAL LANDSCAPE OF RNA TRANSLATIONAL REGULATORS AND THEIR POTENTIAL AS DRUG DISCOVERY TARGETS

Regan Michael LeBlanc, Doctor of Philosophy, 2016

Dissertation directed by: T. Kwaku Dayie, Associate Professor, Biochemistry

RNA is an underutilized target for drug discovery. Once thought to be a passive carrier of genetic information, RNA is now known to play a critical role in essentially all aspects of biology including signaling, gene regulation, catalysis, and retroviral infection. It is now well-established that RNA does not exist as a single static structure, but instead populates an ensemble of energetic minima along a free-energy landscape. Knowledge of this structural landscape has become an important goal for understanding its diverse biological functions. In this case, NMR spectroscopy has emerged as an important player in the characterization of RNA structural ensembles, with solution-state techniques accounting for almost half of deposited RNA structures in the PDB, yet the rate of RNA structure publication has been stagnant over the past decade. Several bottlenecks limit the pace of RNA structure determination by NMR: the high cost of isotopic labeling, tedious and ambiguous resonance assignment methods, and a limited database of RNA optimized pulse programs. We have addressed some of these challenges to NMR characterization of RNA structure with applications to various RNA-drug targets. These approaches will increasingly become integral to designing new therapeutics targeting RNA.

THE CONFORMATIONAL LANDSCAPE OF RNA TRANSLATIONAL REGULATORS AND THEIR POTENTIAL AS DRUG DISCOVERY TARGETS

by

Regan Michael LeBlanc

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:

Professor T. Kwaku Dayie, Chair
Professor Dorothy Beckett
Professor Jonathan Dinman
Professor David Fushman
Professor George Lorimer

Acknowledgements

The road to achievement is built through the support and companionship of your true champions for success. This has been a long and challenging journey, yet it has been worth it because of the support I have been given by so many people. I hope to serve them well in the future and give them my sincerest gratitude here.

First, I would like to thank all of my teachers. My first chemistry class was a practice in futility and I tried to drop the course on multiple occasions, but every time I tried my teacher, Michael Schaab, convinced me to keep at it. I did not like chemistry in high school. Well, Mr. Schaab, thank you for not letting me give up then – it turns out that I like chemistry. After high school, I had the opportunity to work with a prolific researcher at Furman University, Dr. Moses Lee, and that was when I first gained an appreciation for the journey of discovery. I will forever be grateful for that opportunity and his guidance in my early research career.

I would also like to thank all of my colleagues over the years who have challenged me and inspired me to be a better scientist. Andrew Longhini, thank you for your tremendous enthusiasm and incredible work-ethic. It's been a true pleasure to work with someone passionate about our efforts to push RNA discovery. And of course, a heartfelt thanks to the best lab mascot ever, Jake. My Le, thank you for sharing the graduate school journey with me and helping me finish my thesis by sharing your resources and knowledge. To Owen Becette and Hyeyeon Nam, you are the future of the Dayie group, so I know we are in

good hands; thank you for all your hard-work. And Luigi Alvarado, thank you for laying the foundation for which much of our labeling technologies have been built. You left our lab a better place than when you arrived and we are all grateful for your work.

To my advisor, T. Kwaku Dayie, I continue to appreciate the wisdom and guidance you have brought to my work. Your willingness to let me experiment and discover my own passions led me to truly appreciate NMR. I came to NMR kicking-and-screaming but now I could not imagine myself any place else. Thank you for your support and I hope to continue our fruitful collaborations. To my other NMR mentor, David Fushman, I cannot express how grateful I am that you were always willing to field my questions and offer your expertise. Your class was instrumental to helping me conquer some of my fears about NMR.

To my committee members, thank you for your patience and more importantly for thoughtfully challenging me. You were always concerned that I endeavored to do too many things and you were right. You pushed me to focus my work and it has made me a better scientist. To Bin Chen, my office-mate, we worked well together and your initial concepts and progress with SAM-II enabled us to make important contributions to the field. That work was integral to my success, thank you.

To the many collaborators that I have had the great fortune to work with throughout my doctoral career, thank you. In particular, to Bruce Johnson who continuously updated his excellent NMRViewJ software to help streamline the new resonance assignment strategy I developed, thank you. To Carolina

Salguero and Victoria D'Souza at Harvard University, thank you for collaborating on the SARS coronavirus frame-shifting project. I am excited to get all of our hard work published soon. To Chandar Thakur, my first doctoral lab mentor, thank you for teaching me the ways of RNA wet-lab research.

I would also like to take this opportunity to thank all the professors I have had the privilege to teach with during my time at the University of Maryland. In particular, I want to thank Bonnie Dixon, Lee Friedman, and Jason Kahn. I learned more teaching for these professors than I would have ever learned on my own; they are some of the best instructors you will find at any university.

To my family, my friends, and my wife, none of this is possible without the unconditional support you have given me. To my parents, thank you for always believing in me and most importantly for always being there. You were always there to cheer on my accomplishments on the field and have continued to do so through-out my life. To my brother and sister, we have collected enough degrees, no more school for any of us! To my in-laws, Robin and Sharyn, you got to watch all of this from the beginning and you have been tremendously supportive through the whole process. I cannot thank you enough. To my wife, Kyle, it has been a long journey. We are here, finally, and I owe every last bit of it to you. Also, Happy Birthday ☺

Finally, I must thank the only person who might actually read this whole thesis. Cousin Justin, thank you.

Table of Contents

List of Tables

List of Figures

Abbreviations

AIDS	auto-immune deficiency syndrome
AMR	anti-microbial resistance
ATP	adenosine triphosphate
CDC	Centers for Disease Control and Prevention
CEST	chemical exchange saturation transfer
COSY	correlation spectroscopy
CPMG	Carr-Purcell Meiboom-Gill
CTP	cytidine triphosphate
cryo-EM	cryo-electron microscopy
ddH2O	double-distilled water
DNA	deoxyribonucleic acid
DQ	double quantum
EDTA	ethylenediaminetetraacetic acid
FBDD	fragment-based drug design
FDA	Federal Drug Administration
FMN	flavin mononucleotide
FPLC	fast pressure liquid chromatography
FRET	Forster resonance energy transfer
FWHH	full width at half-height
GTP	guanine triphosphate
HCV IRES	hepatitis C virus internal ribosome entry site
HMQC	heteronuclear multiple quantum coherence
hNOE	heteronuclear Overhauser effect
HSQC	heteronuclear single quantum coherence
HTS	high-throughput screening
IRE	iron responsive element
IRP	iron regulatory protein
mRNA	messenger RNA

MRSA	Methicillin-resistant Staphylococcus aureus
MWCO	molecular weight cut-off
ncDNA	non-coding DNA
NDP	nucleotide diphophosphate
NMP	nucleotide monophosphate
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NTP	nucleotide triphosphate
nt	nucleotide
NUS	non-uniform sampling
PAGE	polyacrylamide gel electrophoresis
PDB	Protein Databank
RCSA	residual chemical shielding anisotropy
RDC	residual dipolar coupling
RNA	ribonucleic acid
rRNA	ribosomal RNA
SAM	S-adenosyl-L-methionine
SAXS	small angle X-ray scattering
SQ	single quantum
TCA	tricarboxylic acid
TOCSY	total coherence transfer spectroscopy
tRNA	transfer RNA
TROSY	transverse relaxation optimized spectroscopy
UTP	uridine triphosphate
UTR	untranslated region
VS	virtual screening
WHO	World Health Organization
ZQ	zero quantum

1 Introduction

1.1 The rise of antibiotic resistance

The development of antimicrobial agents was one of the crowning achievements of the 20th century. Prior to the antibiotic era, infectious diseases such as cholera, the plague, and tuberculosis (TB) were generally untreatable, frequently fatal, and often widespread or pandemic. Infectious diseases, however, are still the number one cause of premature death (determined as deaths between 0-49 years of age) worldwide – more than cancer, heart disease, chronic obstructive pulmonary disease (COPD), diabetes, and diarrheal disease combined (Figure 1.1)[1]. The problem is compounded by increasing incidences of antibiotic and antiviral drug resistance.

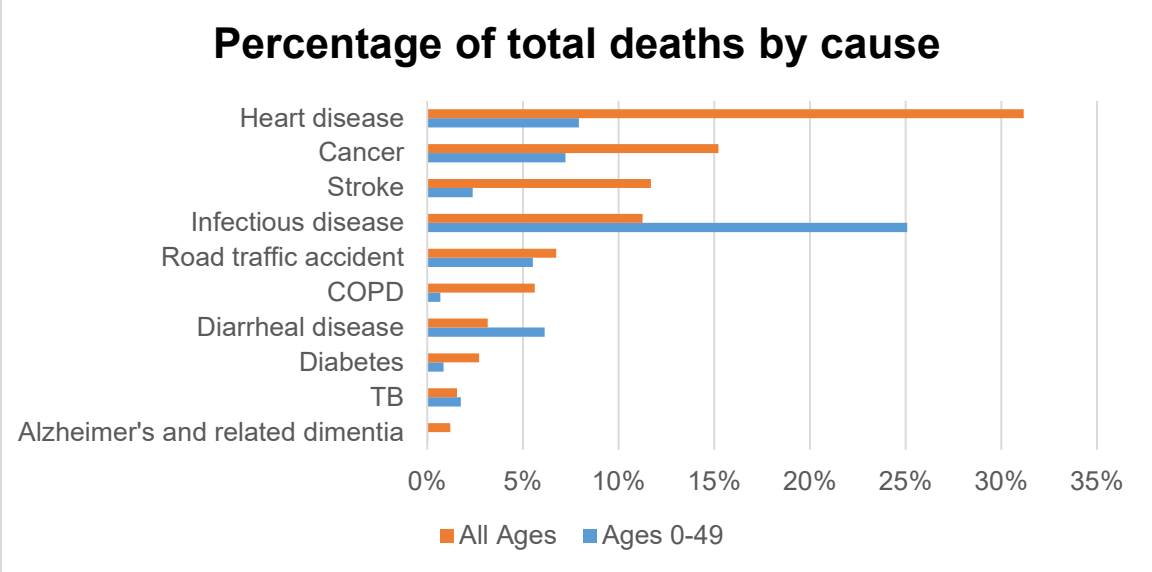

Figure 1.1. WHO statistics on total deaths by cause. Death by cause are given as a percentage of total deaths for all age groups (orange) and ages 0-49 (blue)[1]. Infectious disease is the number one cause of death amongst the 0-49 age group.

The rise of antibiotic resistance over the past few decades has necessitated the development of novel drugs to circumvent current drug-resistance. Recent estimates from the Centers for Disease Control and Prevention (CDC) puts the annual cost of antibiotic resistant bacteria at $21-34 billion annually in excess health care cost as well as $35 billion in other societal costs[2,3]. An estimated 5-10% of hospital patients will develop infections with approximately 90,000 of these patients dying each year – up from 13,300 infection related patient deaths in 1992[2]. The incidence rate of antibiotic resistant bacteria such as methicillin resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *enterococci* (VRE), and fluoroquinolone-resistant *pseudomonas* (FQRP) has risen from <10% to more than 50, 25, and 30%, respectively, over the past 20 years[4]. MRSA is now responsible for more death than HIV/AIDS[3,4]. The startling increase in patient deaths related to infections is most closely link to this rise in antibiotic resistant bacteria.

An analysis by economist Jim O'Neill determined that at current rates of drug resistance, by the year 2050 the global cost of antimicrobial resistance (AMR) could spiral to $100 trillion[5]. O'Neill's analysis, based on projections by RAND and KPMG, assumed that present day rates of resistance will increase without intervention. An unchecked global trend projects to increase the attributable deaths from AMR to over 8 million per year by 2050, which would be equivalent to current cancer death totals (Figure 1.2)[5]. African and Asian countries are projected to suffer the most and are already seeing the challenges of tuberculosis drug resistance[6–8]. This disturbing trend led the CDC to declare

in 2014 that the human race is now in the “post-antibiotic era” with the World Health Organization (WHO) warning that the crisis is becoming dire[9]. New strategies and solutions are needed now.

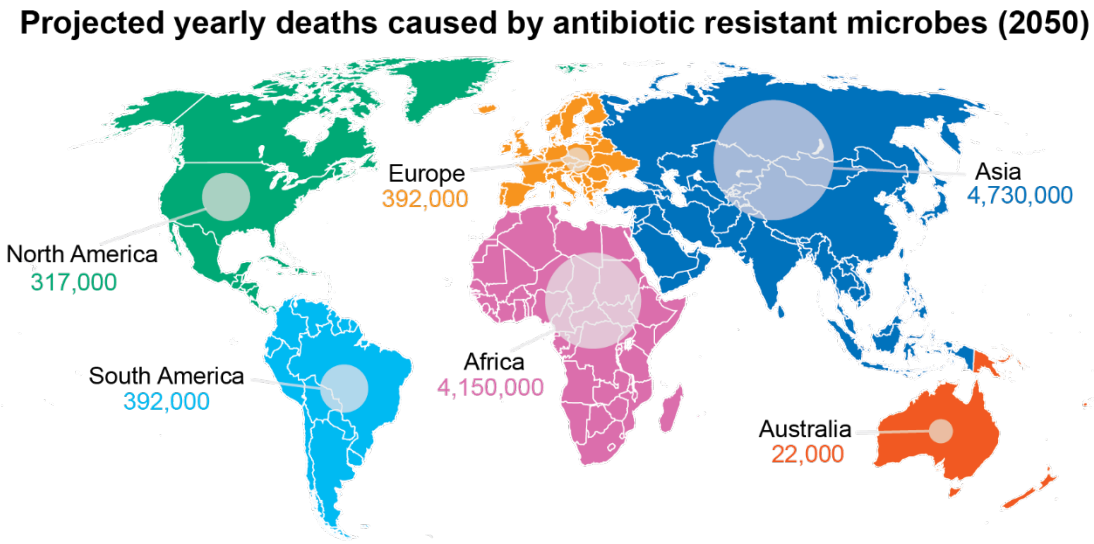

Figure 1.2. The projected yearly deaths attributable to antimicrobial resistance (2050). Figure adapted by author from Review on Antimicrobial Resistance[5]. Projections are grouped by continental region.

Pharmaceutical companies, however, have shifted focus to developing drugs for chronic diseases that require longer treatment causing the antibiotic drug development pipeline to dry up[10]. Of the top 18 pharmaceutical companies worldwide, 15 have abandoned the search for new antibiotics[11]. Bristol-Myers Squibb, Eli Lilly, Wyeth, and Procter & Gamble have all discontinued their antibiotic discovery research. GlaxoSmithKline has downsized its research of anti-infective agents, yet we see a need for continued research and discovery for antibiotics. The lack of emphasis for new anti-infective drug discovery has ultimately led to a steady decline in the number of new antibiotics approved for sale by the FDA (Figure 1.3)[2,12]. In fact, over the past three decades, only four

new classes of antibiotics have been approved by the FDA[13]: oxazolidinones[14], cyclic lipopeptides[15], fidaxomicin[16], and bedaquiline[6,7].

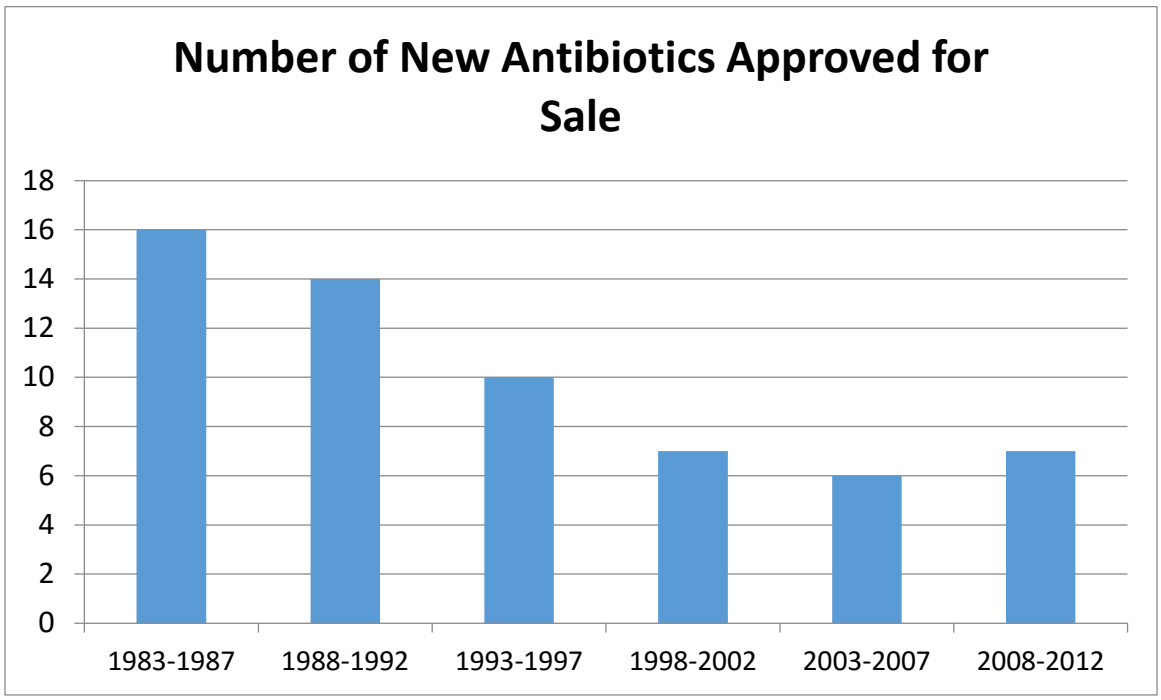

Figure 1.3. New antibiotics approved by the FDA over the last three decades. A steady decline in newly approved antibiotics occurred until the early 21st century. Recently, renewed interest in anti-infective agents has still not bounced back the rate of discovery to pre-2000s levels[2,17].

The CDC has in light of reduced private sector research provided its own plan for combating rising antibiotic resistance[2]. In this plan published as part of a 2013 report on drug resistance, four core actions were suggested: prevent infection and spread of resistance, track resistant strains, improve and reduce the use of current antibiotics, and promote the development of new antibiotics and diagnostic tests for resistant bacteria.

Preventing infections and the spread of resistant strains has its own challenges. Bacterial infections are often the product of either systemic immune problems or injury. Preventing injury is a non-trivial endeavor. Treating underlying immune system diseases will reduce the rates of infections and drug discovery advancements for chronic immune disease such as AIDS are major improvements in this area. Tracking resistant strains will allow researchers to better understand and reduce the spread of infection.

Reducing the use of current antibiotics could curb the spread of antibiotic resistance. It has been estimated that 80% of total antibiotics sold in the United States are used in animals to promote growth and prevent infection[4,11,18]. These antibiotics can be spread to humans through the food we consume. With up to 90% of the antibiotics used on animals secreted through urine and stool, antibiotics also spread to the soil and groundwater affecting the environmental microbiome[2,11]. The voluminous and unnecessary use of antibiotics in the agricultural industry has partly led to an ecology-wide spread of resistance given that bacteria can exchange genetic information to confer resistance to other strains[19].

Whether through over-use or inevitable natural selection, since the first introduction of antibiotics into clinical use in the early 1940s[20], there has been a strong selective pressure for bacteria to evolve drug resistance. Strains of bacteria resistant to penicillin were observed in the mid-1940s[18]. MRSA strains were discovered in 1962 just two years after the introduction of methicillin[21]. On average, resistance to antibiotics has been seen within eight years of widespread

use in the clinic[2]. Hence, there is a narrow window from when a new antibiotic is introduced before pathogenic bacteria acquire resistance, leading to an endless quest to find new antibiotics.

1.2 Challenges in drug discovery

From the nineteenth century synthesis of Aspirin to the twentieth century discovery of penicillin[20] and paclitaxel[22], modern medicine has changed the way we live and die. Supplemented with discoveries in biology and chemistry, the pharmaceutical industry continues to provide new diagnostic tools and medicine to improve human health. Even with these enormous successes, the pharmaceutical industry faces substantial scientific and financial challenges.

The cost of developing and bringing a single drug to market can approach more than $1.5-4.2 billion[23–25]. On average, the process of successful drug discovery takes 13.5 years and an estimated 32% of the total cost comes from the pre-clinical drug discovery phase[24,26]. Approximately 11 new drugs are needed to yield just a single successfully approved drug[24]. Given the average rate of antibiotic resistance takes 8 years to hit clinically relevant levels from the time an antibiotic is first introduced, there is a major challenge to staying ahead of drug resistance. New opportunities and targets are needed to invigorate the antibiotic drug discovery pipeline.

1.3 RNA as a drug target

Ribonucleic acid (RNA) is an under-utilized target for drug discovery. In fact, less than 1% of all approved drugs do not bind protein targets and more

than 80% of drugs target only two classes of proteins: enzymes and receptors[27,28]. Yet, RNA plays a critical role in essentially all aspects of biology including signaling, gene regulation, catalysis, and retroviral infection[29–35]. Regulatory, structural, and catalytic RNAs unique to bacteria and virus also present an opportunity as highly specific anti-infective drug targets.

Not so surprisingly, there are already antibiotics that target RNA. In fact, half of all antibiotics target the ribosomal RNA (rRNA) complex. Recent high-resolution structures have elucidated the binding sites of many rRNA targeting antibiotics and facilitated the design of novel antibiotics[36,37]. The bacterial rRNA is comprised of three distinct RNAs, the 5S, 16S, and 23S rRNA. The 16S rRNA is an approximately 1,500 nucleotide (nt) continuous RNA that binds about 20 different proteins and forms an intersubunit bridge with the 23S rRNA. Many antibiotics bind to regions of the 16s rRNA in which mutations have been found to be deleterious to function[38]. The decoding site of the 16S rRNA subunit, in particular, is a target of aminoglycosides such as paromomycin, neomycin, gentamicin, tobramycin, and kanamycin (Figure 1.4). Upon binding, two conserved residues (A1492 and A1493) “flip-out” to base-pair with the mRNA/tRNA cognate duplex in the A-site preventing distinction between cognate and non-cognate tRNA leading to missense translation and ultimately cell-death. Since the bacterial ribosome has evolutionary differences from eukaryotic ribosomes, antibiotics that target the ribosome have less potential for cross-reactivity.

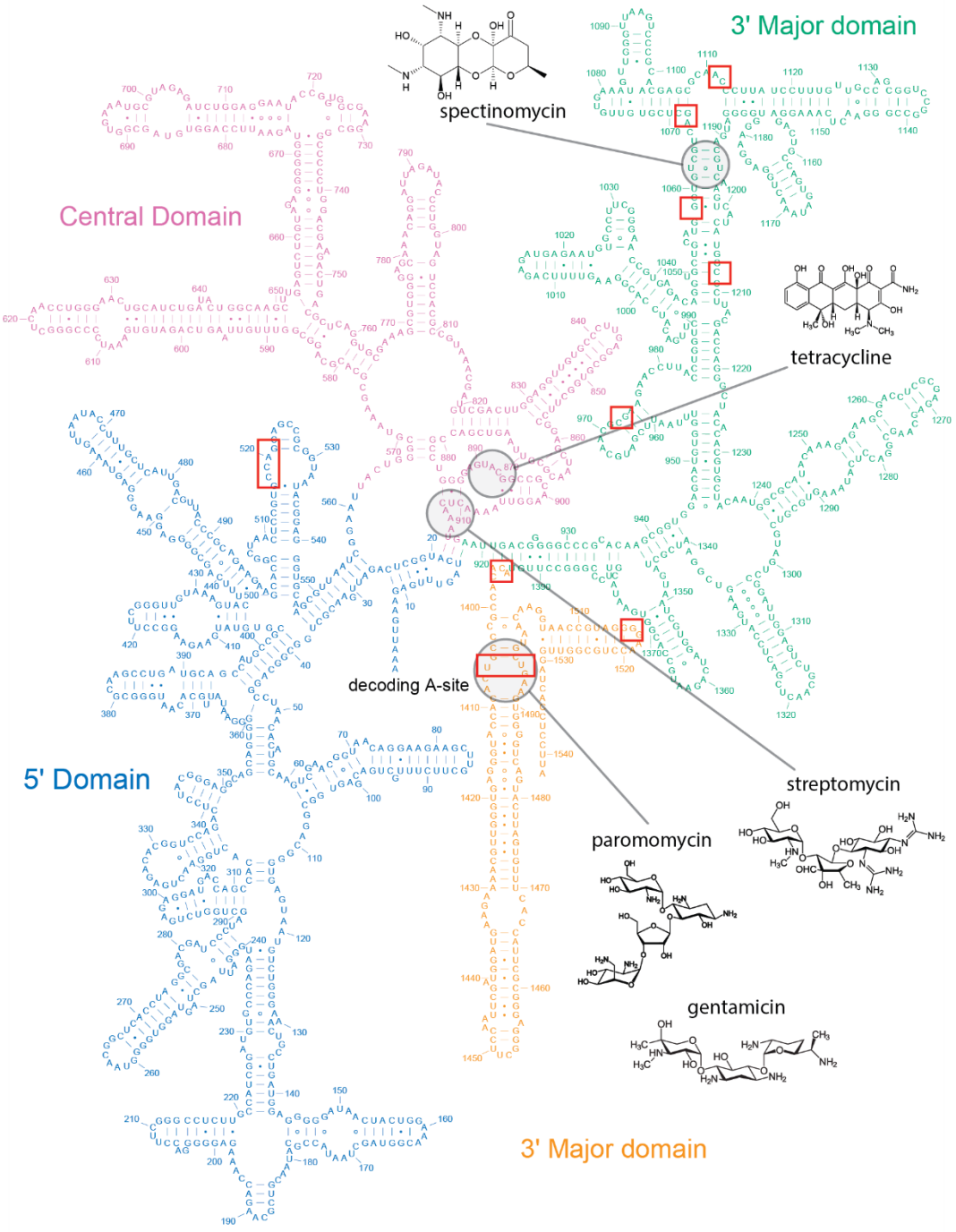

Figure 1.4. The secondary structure of the 16S rRNA and its antibiotic target sites. The secondary structure of the small ribosomal RNA subunit is shown with the four distinct regions color coded. Antibiotic binding sites are highlighted and labeled by the compound name and structure. Deleterious mutations are highlighted with red boxes and pose potential as targets for new antibiotic drug development[38].

Our emergent understanding of how rRNA functions through observations of drug binding and structure has led to repurposing antibiotics to treat chronic disease. The realization that aminoglycosides bound to the ribosome to cause missense read-throughs during protein translation led to new strategies for treating both Duchenne muscular dystrophy (DMD)[39] and cystic fibrosis (CF)[40,41]. These genetic disorders are caused in part by premature stop codons that prevent complete protein synthesis. In the case of DMD, gentamicin, an aminoglycoside antibiotic, was shown to promote read-through, increase expression of dystrophin, and reduce DMD symptoms[39]. It was ultimately an understanding of the ribosome structure and function that led to these potential treatments for DMD and CF.

Riboswitches, ligand binding aptamers and expression platforms in the 5'-untranslated region (5'-UTR) of mRNA that can regulate either transcription or translation, are also unique to prokaryotes and therefore present a specific target with reduced cross-over to human targets. The complexity of ligand-riboswitch interactions matches those of protein-ligand interfaces. For example, the SAM-1 riboswitch forms interactions with every functional group of its cognate ligand, S-adenosyl methionine (SAM) ligand[42]. *Bacillus anthracis*, the causative agent of anthrax, has 17 SAM-1 riboswitches that collectively regulate 36 genes[43]. This redundancy, also seen with rRNA[44], within the bacterial genome implies that mutations are less apt to convey drug resistance (Figure 1.5). And for riboswitches indicates a single drug could potentially target multiple riboswitches or the same type.

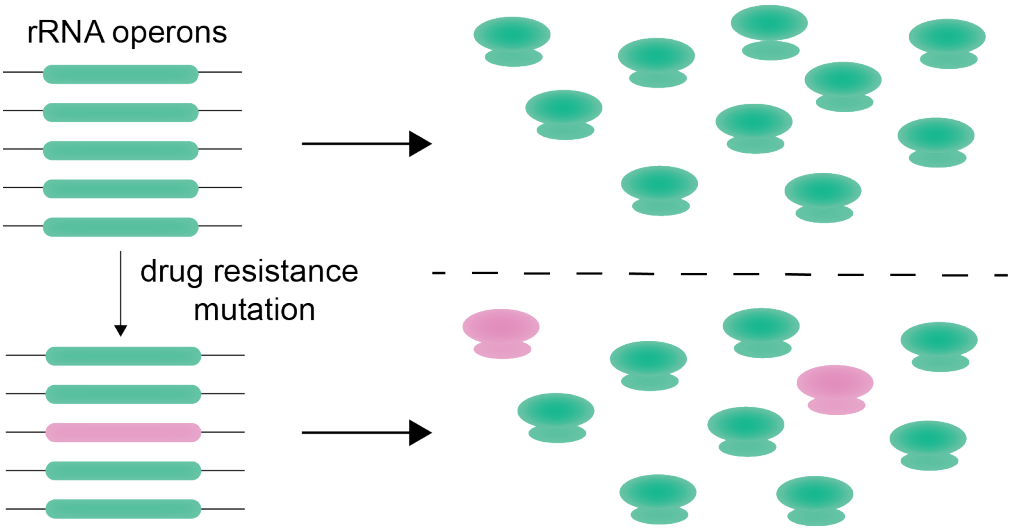

Figure 1.5. Redundancy of rRNA operons limits adaptation to antibiotics. Bacterial genomes display up to 15 redundant operons that code for rRNA. As seen in this Figure, a single mutations (purple) that presents drug resistance will only effect a small portion of ribosomes and limits the ability to quickly adapt or spread resistance from rRNA mutations.

RNA is not just a potential target for antimicrobial drug discovery. With more than 95% of the human genome coding for non-protein products, there are an abundance of potential non-coding RNAs (ncRNA) that can be specifically targeted for drug discovery in cancer[45,46], Alzheimer's disease[47,48], heart disease[49,50], and other chronic conditions[49]. It is estimated that only 600 of the more than 25,000 human proteins are targetable[28]. Therefore, the focus in target selection has recently shifted to other macromolecules such as ncRNA.

For example, recent publications suggest a small population of mRNA that bind the 13-subunit, 800-kilodalton eukaryotic initiation factor 3 (eIF3) and are up-regulated in cancer cells[45]. With about 500 out of more than 10,000 mRNAs (<3%) in the cell specifically carrying eIF3 binding regions in the 5'-UTR of the mRNA – the potential for targeting this interaction provides new targets for

cancer drug discovery in tumors over-expressing eIF3. Targeting of two-cancer related genes by tweaking the 5'-UTR of mRNA for c-*JUN* and B-cell translocation gene 1 (BTG1) to prevent eIF3 binding was shown to successfully prevent tumor cells from becoming invasive. RNAs with potential anti-viral therapeutic potentials have also been well-studied including the HIV-1 dimerization initiation site (DIS)[51,52], the HIV-1 transactivation response element (TAR)[53–55], and the hepatitis C virus internal ribosome entry site (HCV IRES)[56,57].

Through evolutionary pressures, nature has revealed small molecules that can selectively bind RNA structures to disrupt their function as is seen with the naturally occurring antibiotics that target rRNA. Riboswitches have also evolved to recognize small molecules with high specificity and affinity. In the laboratory, researchers have mimicked this selection process by designing RNA aptamers to bind small molecules by a method called systematic evolution of ligands by exponential enrichment (SELEX)[58,59], which can provide insights into how RNA and ligands recognize each other[60,61] and yield powerful biochemical tools for probing RNA function *in vivo*[62]. So why too can't we design small molecules that will bind RNA with high specificity and affinity in the hopes of treating disease? To address this challenge will require a thorough understanding of the binding target coupled with a tried-and-true drug discovery strategy.

1.4 RNA drug discovery methods

Just as most current drug targets are proteins, a majority of the daunting drug discovery statistics and strategies we have mentioned apply to our understanding of small molecule inhibitors of protein. For RNA drug discovery,

the challenges are even greater. Many of the tools used for protein drug discovery, however, have migrated to the field of RNA drug discovery. RNA researchers have implemented some of these strategies including high-through put screening (HTS)[63–67], virtual screening (VS)[53,68–70], fragment-based drug design (FBDD)[64,69,71,72], and rational drug design[73,74]. The methods, however, have proven uniquely challenging for RNA.

Of these methods, VS, FBDD, and rational drug design all require a detailed understanding of the target RNA structure. Al-Hashimi and co-workers successfully adopted VS methods to identify six novel inhibitors of HIV-1 TAR[53]. NMR and molecular dynamics data were required to create an ensemble of TAR structures that each bound a different ligand. As evidenced by riboswitches, ligand binding can either select sparse populations[75–78] or induce large conformational changes to RNA[79–82], therefore care must be taken to sample enough conformational space for VS drug discovery.

FBDD utilizes small more chemically and structurally diverse fragments than the larger molecules used in HTS, which allows for more efficient exploration of chemical space and increased likelihood of hit discovery[83]. In the protein world, FBDD methods have been successfully employed to develop high affinity inhibitors for multiple disease models. The drug discovery process of a BACE1 inhibitor, currently in phase III trials, for treating Alzheimer's has been well documented that combined both X-ray and NMR to screen over 50,000 fragments and guide the combinatorial synthesis of lead compounds to yield nM binding hits[84–87]. RNA FBDD has not yet resulted in similar success stories due

to the limitations of protein optimized fragment libraries. New fragment libraries are being designed specifically for RNA drug discovery to deal with the challenges RNA possess as a drug target[64,71,72].

Rational drug design requires an understanding of either the binding mode of a ligand-RNA target or the function of the RNA target. For example, potential antibiotics that target riboswitches have been designed based on the ligand partner of the riboswitch. The lysine riboswitch (L-box) senses L-lysine concentration and regulates the expression of lysine biosynthesis and transport genes[88]. L-lysine analogs, L-4-oxalysine and L-aminoethylcysteine, were reported to selectively bind better than L-lysine to L-box and inhibit *Bacillus subtilis* cell growth[63]. Structures of the bound guanine riboswitch[89,90], which regulates genes involved in purine metabolism and transport, suggested that specific substitutions to guanine could be tolerated. Derivatives of guanine designed based on the binding pocket of the guanine riboswitch bound with pM affinity and reduced bacterial growth by greater than 50%[73].

1.5 Biophysical methods to study RNA

No matter the drug discovery strategy employed, there must be a solid foundational understanding of the drug target including its structure and function. A combination of biophysical methods have been utilized to capture the structural landscape of RNA. These methods include, but are not limited to, single molecule FRET[77,91–94], optical tweezers[95–97], SAXS[98–101], X-ray crystallography, cryo-EM[102–105], and NMR[106–109]. Each method has its own limitations.

FRET can characterize tertiary structure, inform on flexibility, quantify kinetics, and yield information on conformational changes at single molecule levels, but requires chemical modifications that can perturb the structure, is limited to short timescales (<200 ms), and has specific distance restraints (2-10 nm)[110]. Another single molecule method, optical tweezers, can identify and quantify secondary and tertiary architecture through folding kinetics and thermodynamics, but does not provide atomic resolution information and requires customized and expensive instrumentation[111].

SAXS, X-ray crystallography, cryo-EM, and NMR are bulk molecule methods that average information over all conformational space. SAXS provides information on global shape, but cannot by itself give dynamic or atomic structural information and is often used to supplement NMR data[112,113]. X-ray crystallography yields high resolution structural data, but is limited to homogeneous RNAs that are readily crystallized, which are routinely RNAs that have been mutated to bind U1A protein[114–116] or shortened constructs selected from thermophiles[101,117–122] to reduce dynamics and improve crystallization[123,124]. Cryo-EM can provide atomic resolution structures of large complexes[102–105] without the need for crystallization, but cannot yet provide information on smaller RNAs since the theoretical size limit is 38 kDa[125] and no structure has yet been reported below 170 kDa[126]. Improvements in cryo-EM methods, however, promises to revolutionize structural biology of large RNA and has recently been used to solve the structure of the spliceosome[105], resolve HCV IRES bound to

the ribosome[127], and elucidate the structural basis of stop codon recognition by the ribosome[102].

NMR uniquely overlaps with many of the individual strengths each of these previous methods provides. Atomic resolution, structural and dynamic, local and global information can be obtained in an unmodified near native condition. NMR does have its limitations; however, it is insensitive, time consuming, costly, and size limited[128–131]. Even with all of its limitations, NMR has remained a staple in the biophysical toolset for studying RNA structure and dynamics.

1.6 NMR of RNA

The extensive functional diversity of RNA derives from its ability to fold into complex architectures and is further enhanced by motions (thermal fluctuations) that are intrinsic to these structures[31,32,34,132]. Yet there is a dearth of solved RNA structures to aid us in better understanding RNA function. This is reflected in the inverse 98:2 correlation between the genomic outputs of RNAs and proteins and the number of solved 3D structures of RNA and protein structures (2:95) deposited in the protein databank (PDB)[133,134]. The combined difficulties of crystallizing RNA for X-ray crystallography and the size limitations of cryo-EM along with the rapid signal decay and spectral crowding associated with NMR studies of RNA makes structural studies challenging. Furthermore, the average size of the ~300 deposited RNA NMR structures in the PDB is only 25 nt. While half of the deposited RNA structures in the PDB (Figure 1.6) were solved by NMR methods[130], the usefulness of NMR is still limited by the high cost

of sample preparation, difficulties in resonance assignment, and lack of freely available RNA optimized NMR pulse programs.

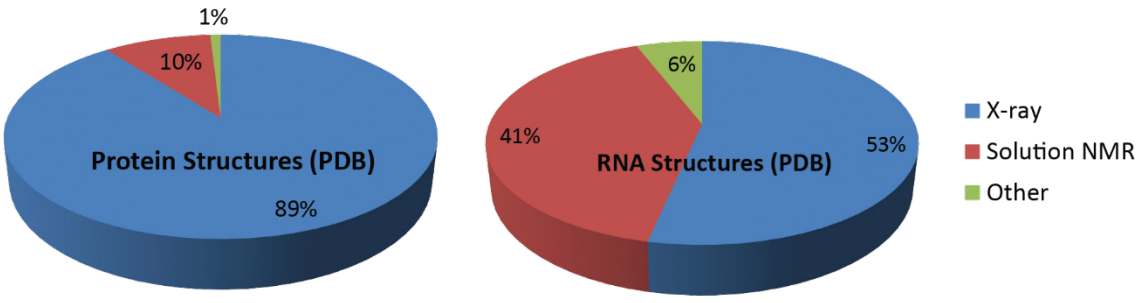

Figure 1.6. PDB statistics for protein and RNA structure methods. The method used to solve PDB structures for protein and RNA structures is given in a pie chart. NMR accounts for 10% of protein structures and 41% of RNA structures to date.

RNA NMR is built on the same principles and techniques as protein NMR, but has several important differences. First, RNA has a smaller percentage of hydrogen atoms, which are the most commonly observed NMR signal. The higher density of protons in proteins along with its ability to fold back on itself allows for rapid structure calculations using through-space distance constraints like nuclear Overhauser effect spectroscopy (NOESY), which is essentially a molecular ruler that measures and identifies protons within 5 Å of each other. For RNA, this method of structure determination is more difficult since RNA tends to be less globular without many residues folding-back on itself.

Two-dimensional NMR methods such as correlation spectroscopy (COSY) and total coherence transfer spectroscopy (TOCSY) that allow for almost full assignment of resonances in small proteins are also difficult to implement with RNA since residues are connected by the poorly dispersed phosphorous nuclei.

Even with the difficulties of studying RNA by NMR, many groups have pushed through the barriers of RNA NMR by developing new technologies and strategies. The improvements to pulse programming by implementing so-fast techniques[135–137] for rapid data acquisition and non-uniform sample (NUS)[138] have the potential to drastically reduce instrumentation time. NUS also may permit the reasonable acquisition of 4D NMR spectrum[138], which could eliminate ambiguity in resonance assignment. Additional advances in sample preparation such as segmental labeling[139] combined with isotopic filtering/editing strategies[140,141] provide new tools to conquer challenging RNA systems.

Chemical shift prediction for RNA structure will also increase the ease and speed of resonance assignment. When 3D structural data is available, predictive software can help assign the NMR resonances[142]. Combined with automated resonance assignment tools, NMR of RNA is becoming more accessible[143]. Once assignment is achieved, underlying questions about how structure and dynamics can effect RNA function can begin to be unraveled. Indeed, recent improvements to structural modeling software such as Xplor-NIH[144] should allow the incorporation of more NMR data for mapping the structural landscape of RNA.

NMR tends to provide multiple conformers that fit NOE distance constraints since NMR is an averaged ensemble method. In order to best identify the structures that describe any averaged NMR data, an ensemble based approach has become popular for selecting a small representative population of structures that best describes the NMR data[145]. A similar approach has been

used for proteins called sparse ensemble sampling to refine structures and determine the distribution of structures[146]. Ultimately, the flexibility and dynamic nature of RNA will require a much broader understanding of the conformational landscape than can be provided by a single structure. Here is where NMR has the opportunity to shine.

With all these developments in NMR methodologies, however, there has been little increase in the rate of RNA structures deposited in the PDB (Figure 1.7). The amount of RNA structures deposited by year has remained stagnant due to the numerous roadblocks present for current methodologies. These challenges, which we introduced earlier, include among others high cost of sample preparation, tedious and non-trivial resonance assignment strategies, and limited RNA specific pulse programs.

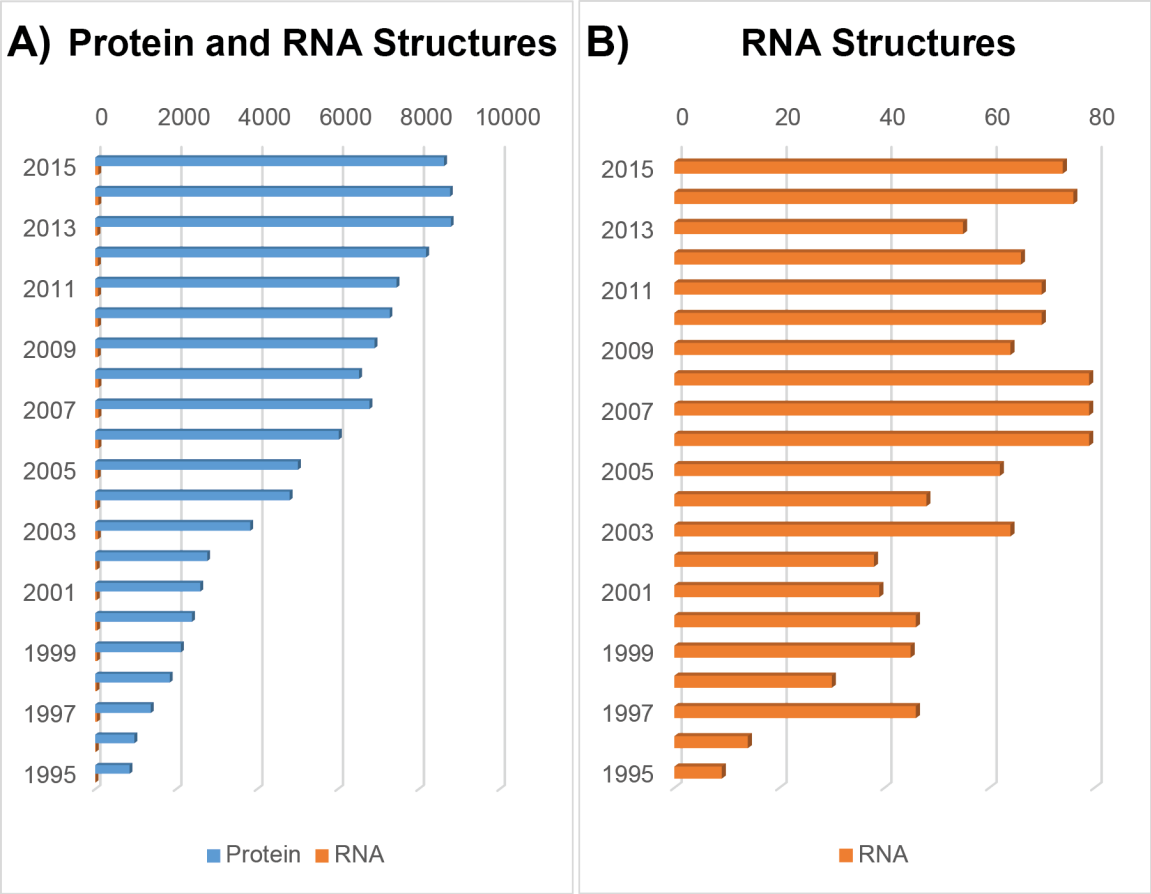

Figure 1.7. Protein and RNA structures deposited in the PDB by year. A) Protein (blue) and RNA (orange) structures deposited in the PDB by year for all methods. RNA is barely visible given the disparity between protein and RNA structures deposited. B) RNA structures deposited per year only. X-axis is scaled to 80 compared to 10,000 for part A. The median average of RNA structures over the last decade is 70 per year.

1.7 What's next?

We have presented a case for the immediate need to invigorate the antibiotic drug discovery pipeline. Yet there is a stunning lack of momentum from the private sector. To address this problem, we have suggested targeting RNA for novel drug discovery. The challenges presented by RNA drug discovery are cumbersome indeed, but at the heart of the drug discovery movement is the

need for a solid understanding of RNA structure and dynamics and it is here that progress is being made. To this end, we will implement new technologies to improve the way we study RNA by NMR spectroscopy.

The need for low-cost isotopic labeling strategies (Chapter 2), a stream-lined and unambiguous resonance assignment strategy (Chapter 3), and optimized NMR pulse sequences for RNA (Chapter 4) will all be addressed through the development of new technologies.

We will apply these new methods to better understand the structure and function of translational regulator RNAs. In particular, we will investigate potential antibiotic targets such as the bacterial 16S rRNA decoding site (Chapter 4) and the SAM-II riboswitch (Chapter 5). We will also provide insight into the underlying dynamics of iron regulation by the iron responsive element RNA (Chapter 4).

2 Isotopic enrichment strategies

The first thing we must understand when using NMR is that it is a spectroscopic technique that measures a bulk average signal from a mixture of chemical states in solution. The signals obtained in NMR are nucleus specific spin transitions between unequal energy states (owing to the spectroscopic nature of the technique). The transitions between these energy states (spin states) occur within the radiofrequency wave range. By pulsing the spins within a magnetic field with radiofrequencies, the transition between the spin states can be observed as a decaying sinusoidal signal or free induction decay (FID). This signal decays due to relaxation of the signal arising from magnetic field fluctuations caused by molecular motions; these motions contain important structural and dynamic information about the macromolecule. The information content of a NMR signal for RNA studies is vast and well-reviewed[130,147–149].

The energy difference between the low energy, alpha (α), and higher energy, beta (β), spin states is determined by the external magnetic field and gyromagnetic ratio of the nuclei as given by equation 2.1.

$$\Delta E = hv = h(B_0 - B_e)\gamma \qquad (2.1)$$

Where ΔE is the energy difference between the α and β spin states, h is Planck's constant, v is the frequency corresponding to the energy difference (in Hz), B_0 is the strength of the external magnetic field (in Tesla), B_e is the small magnetic field generated by the local electronic environment of the nuclei, and γ is the gyromagnetic ratio constant which is an intrinsic property for the nucleus of

interest. The energy difference between spin states is directly proportional to the strength of the magnetic field and therefore higher field spectrometers provide greater sensitivity.

As given from equation 2.1, the typical energy difference for a ^{1}H nucleus in a 600 MHz (14.1 T) magnet is 2.86 x 10^{-5} kcal mol^{-1}, which is a relatively small difference. At 25 °C this energy difference yields a population distribution of approximately 97 more α spin states for every 10^{6} β spin states based on the Boltzmann distribution. The small difference in population distributions owes to the relative insensitivity of NMR spectroscopy for the study of bio-macromolecules. It is typical to require mg quantities of sample for comprehensive NMR studies.

Not only does NMR require large quantities of homogenous RNA, it also relies on spin-active nuclei. For most bio-macromolecular studies, the nuclei of interest are ^{1}H, ^{13}C, ^{15}N, ^{31}P, and ^{19}F. In particular, proton NMR spectroscopy has been a valuable tool given the ~100% natural abundance of ^{1}H in bio-macromolecules. Extensive signal overlap, however, limits ^{1}H only NMR to RNAs under 20 nt long[133]. Isotopic enrichment strategies have extended the usefulness of NMR spectroscopy to ^{13}C, ^{15}N, and ^{19}F spin probes thereby allowing heteronuclear multi-dimensional NMR methods to reduce the extensive chemical shift overlap problems in order to study larger RNAs.

There are four commonly used strategies for obtaining $^{13}C/^{15}N$ isotopically labeled RNA; the benefits, limitations, and the improvements from our work to these methods will be reviewed in this section. The first strategy is to purchase

commercially available $^{13}C/^{15}N$-labeled ribonucleotide triphosphates (NTPs) for use in T7 *in vitro* transcription of RNA from DNA templates. The second strategy is to grow cell-cultures with well-defined chemical sources of isotopically labeled carbon and nitrogen (^{13}C and/or ^{15}N-enriched) sources and harvest the isotopically enriched total cellular RNA, digest this RNA to ribonucleotide monophosphates (NMPs), and finally phosphorylate to NTPs[133,150–152]. The third strategy is a complete *de novo* biosynthesis of the NTPs, followed by *in vitro* RNA transcription[153,154]. Finally, the fourth strategy is a hybrid chemo-enzymatic method that combines chemical synthesis of the nucleobase with enzymatic ribose coupling and phosphorylation to prepare NTPs for *in vitro* transcription[155,156].

We will review the methods for obtaining isotopically enriched NTPs for NMR studies of RNA and provide insights into the benefits and drawbacks for each strategy. A comprehensive guide to the labeling patterns available through each method will be given with a cost analysis. Methods for obtaining mg quantities of RNA from NTPs, such as T7 *in vitro* transcription, are also briefly reviewed. At the end of the chapter, we will provide a summary for the cost of each method and outline which methods are appropriate depending on the researcher's needs.

2.1 Commercial isotopes

To incorporate NMR active stable isotopes into RNA for detailed structure and dynamic studies, it is critical to obtain sources of $^{13}C/^{15}N$-enriched nucleotides from the previously mentioned synthesis methods. The easiest route

available is to purchase NTPs with the desired $^{13}C/^{15}N$-isotopic labeling pattern from commercial sources. As of the spring of 2016, Cambridge Isotope Laboratories and Isotec (Sigma-Aldrich) are the major suppliers of isotopically enriched nucleotides. Unfortunately, these commercially available nucleotides are limited to select isotopic labeling schemes and are usually prohibitively expensive (Table 2.1).

Table 2.1. Commercial NTP cost analysis. The price for commonly used and commercially available NTPs is given in $ per 100 µmol of NTP. The labeling pattern is provided in the columns with the indicated isotope type. Dash line indicates no enrichment and U=uniformly enriched (all possible positions).

Nucleotide Type	^{2}H	^{13}C	^{15}N	$/100 µmol
U/CTP	5,3',4',5',5"	-	-	595
A/GTP	5,3',4',5',5"	-	-	595
NTP	-	U	U	675
NTP	-	-	U	675

On average, one can expect to spend $600-800 per 100 µmol of NTP. An *in vitro* transcription reaction (20 mL scale) typically requires 5 mM (100 µmol) of total NTPs (1.25 mM of each NTP) and yields roughly 0.2-1 mM of RNA (250 µL in Shigemi tube) depending on the length and transcription efficiency. The cost per sample from commercial NTPs alone is $1000+. Since it often takes multiple samples for assignment of NMR spectra of RNA, total costs to study the structure and dynamics of RNA by NMR can easily reach tens of thousand dollars based on isotopic labeling costs alone. This estimate does not include the expense of NMR instrumentation time, sample purification, and the tedious data analysis process.

The isotopic labeling, sample preparation, and instrument time expenses help to explain the slow rate of PDB publications for RNA structures, which has maintained a constant pace of about 70 new structures per year despite the rapid growth of newly discovered RNAs of biological interest. It is therefore of critical importance to reduce the cost of obtaining isotopically labeled RNA samples in order to expand the accessibility of NMR to the study of RNA structure and dynamics. At the present time, commercial sources of isotopically enriched nucleotides are prohibitively expensive, which limits the field of research to well-funded laboratories.

2.2 Cell-based labeling strategy

To reduce the overall cost of obtaining isotopically enriched nucleotides, our lab optimized the previously reported cell-based labeling strategies[133,157–159]. Building upon protein isotopic labeling strategies, various mutated strains (K12[160], K10zwf[161], and DL323[162]) of *E. coli* were grown in minimal media using lower cost isotopically labeled precursors such as acetate[163,164], glycerol[158,165], and pyruvate[166–169] as ^{13}C sources and ammonium chloride/sulfate as ^{15}N sources. These site-selectively ^{13}C and/or ^{15}N stable isotopically enriched NTPs are desirable for tackling the structure and dynamics of large RNAs[133,170].

Given the high cost of most ^{13}C-labeled precursors, it was also important to obtain optimal growth conditions for these *E. coli* strains in various commonly used media[166] as well as investigate alternative species. Within this section we will review the standard protocol and updates for obtaining NTPs from cell-based enrichment methods, test the application of succinate for labeling pyrimidines,

address complications with the separation of NMPs into individual nucleotides, scale-up production with fermentation, and apply enrichment strategies to yeast cell lines. We will also critically evaluate the cost benefits of cell-based labeling strategies.

2.2.1 Cell-based labeling protocol

The standard protocol for obtaining isotopically enriched NTPs from *E. coli* is shown in Figure 2.1 as previously described by Batey et al[157]. In brief, a starter culture of *E. coli* at the exponential growth phase is used to inoculate a minimal media where the only source of nitrogen is from ^{15}N-labeled ammonium chloride/sulfate and the only source of carbon is from ^{13}C-labeled precursors (acetate, lactate, glycerol, pyruvate, or glucose). Following growth of the cells to pre-stationary phase, the cultures are pelleted by centrifugation, lysed, and total nucleic acid (DNA and RNA) is extracted with SDS and phenol/chloroform. The total nucleic acid is digested into NMPs and dNMPs with P1 nuclease. The NMPs and dNMPs are separated using a *cis-diol* boronate affinity chromatography column and individual NMPs are separated by C18 reverse phase FPLC column chromatography. Finally, individual NMPs are phosphorylated to NTPs using kinases specific for each NMP in the presence of a coupled thioredoxin-dithiothretol redox system[171].

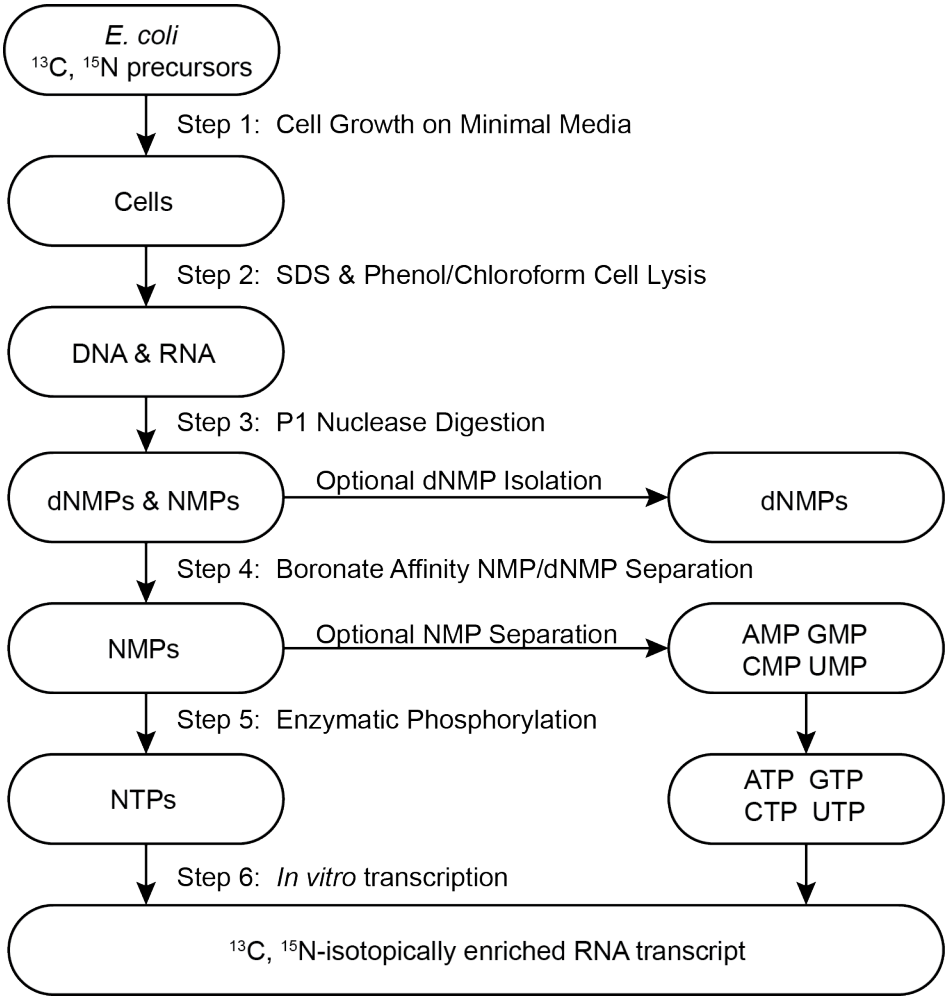

Figure 2.1. Cell-based isotopic labeling scheme. A schematic of the cell-based labeling protocol is given with each step. Optionally dNMPs can be collected from the boronate step and NMPs can be separated into each individual nucleotide before phosphorylation. RNA is synthesized from NTPs through T7 *in vitro* transcriptions reactions using a DNA template.

Two major improvements were made to this method during the purification phases. First, we experimented with high pressure chromatography for separation of NMPs and dNMPs on the boronate affinity column. Past protocols required manual chromatography methods performed in a cold-room, which would take up to 6 hours to perform with constant monitoring in uncomfortable conditions. We packed a FPLC column by Pharamacia with boronate resin in

order to automate the flow-rate, UV-monitoring, and fraction collection process. The Pharamacia column, however, was rated for a maximum pressure of 6 bar. Under the high-pressure conditions of the boronate column, flow-rates of 1 mL/min were required. With total volumes >800 mL needed to elute the NTPs, this process took up to 20 hours. We upgraded the FPLC column to an Econoline column which extended the maximum operating pressure to 50+ bar. Boronate separation was achieved in less than 4 hours, which allowed multiple samples to be processed in a day with no loss in quality.

The second major improvement came in the separation of individual NMPs. The original protocols published relied on C18 hydrophobic columns to separate AMP, CMP, GMP, and UMP[152]. In our hands and under various buffer conditions and flow-rates, CMP and UMP co-eluted, preventing complete separation. Previous publications utilized anion-exchange chromatography to achieve cleaner separation of NMPs[172,173]. We adopted these methods into an automated FPLC system. Complete separation of large quantities of NMPs (>200 mg) was achieved (Figure 2.2).

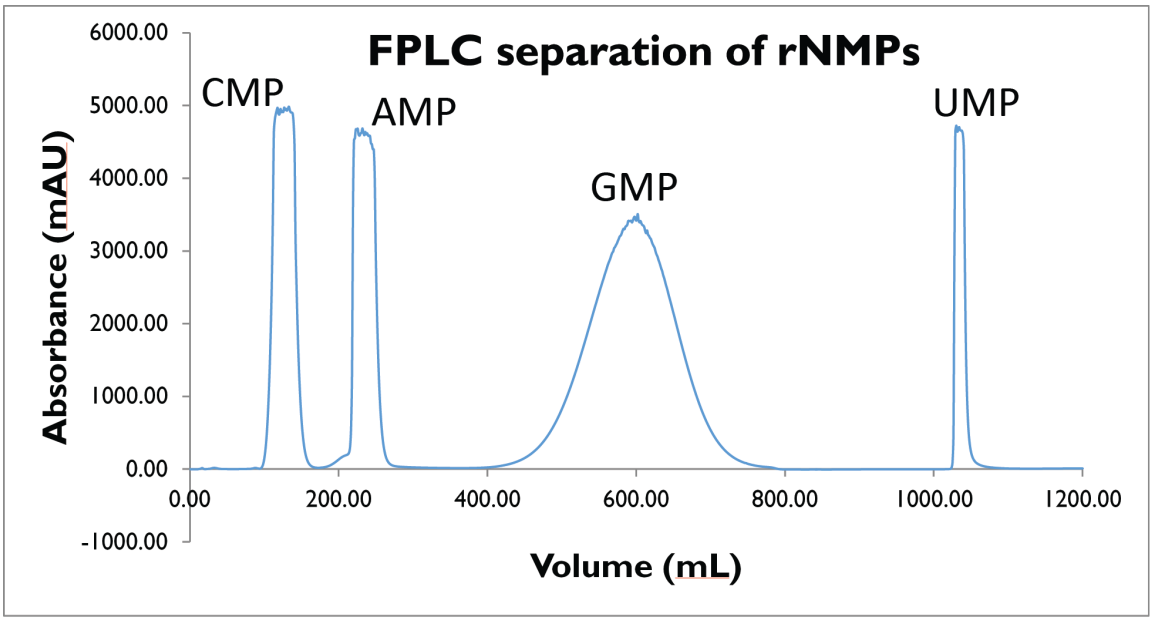

Figure 2.2. Anion exchange separation of NMPs. Full separation of NMPs was achieved with an AGX1 resin (Bio-RAD) packed FPLC column. The following buffer conditions were utilized at a flow-rate of 10 mL/min: increasing concentration of acetic acid (0.1-2 mM). NMPs dissolved in water were loaded directly onto the column.

2.2.2 Labeling schemes from cell-based culture

Depending on the carbon source and *E. coli* strain used to produce NTPs, a unique set of ^{13}C-labeling schemes are obtained[158,163,167–169]. In particular, deletions to succinate dehydrogenase (*sdh*) and malate dehydrogenase (*mdh*) in the TCA cycle (DL323) are important for reducing the mixing of isotopic labeling in the purine and pyrimidine ring systems[169]. K10zwf (a glucose-6-phosphate dehydrogenase deficient *E. coli* mutant) also prevents mixing of the C5 and C6 positions in the pyrimidine ring as well as reducing isotopic mixing in the ribose ring[166]. A transketolase (*tktA*) mutant was used to investigate ribose scrambling[168].

As part of an additional study into isotopic mixing for the pyrimidine C5/C6 positions, we grew DL323 and DL10[174] *E. coli* in minimal media supplemented with ^{12}C-aspartate which is utilized in the biosynthesis of pyrimidine. Analysis of the NTP products showed that all pyrimidines had reduced ^{13}C-labeling schemes indicating that aspartic acid was utilized in the biosynthesis of pyrimidine over the ^{13}C lactate carbon source. Supplementing *mdh*/*sdh* deficient *E. coli* mutants (DL323/DL10) with aspartate could allow for increased control of the isotopic labeling of pyrimidines and also enable site-selective deuteration of the C5/C6 positions, however, the high cost of isotopically enriched aspartate precludes this from being a useful method.

At the time of our work, lactic acid was a cheaper alternative to pyruvate, so we utilized the DL10 cell line to test ^{13}C-enrichment from lactic acid precursors. The results match those obtained from pyruvate and DL323. The complete results of ^{13}C-labeling based on cell line and precursor is given in Table 2.2. Cost are calculated assuming bulk purchases of ^{13}C precursors (>20 g).

Table 2.2. Cell-based labeling cost analysis. The cost in $/100 µmol of NTP for a typical cell-labeling yield is given for each cell line type and ^{13}C precursor. The level of enrichment for each atom position is listed by percentage.

Cell line	**tktA**	**tktA**	**K10**	**DL323**	**DL323**	**DL10**	**K12**
^{13}C source ->	1-glucose	2-glucose	2-glycerol	3-pyruvate	1,3-glycerol	3-lactate	2-acetate
Purine							
Ade C2	75 ± 1	11 ± 1	14 ± 3	99 ± 1	88 ± 4	85 ± 3	>95
C8	44 ± 1	2 ± 1	10 ± 1	93 ± 1	92 ± 7	93 ± 5	>95
Pyrimidine							
C5	43 ± 4	28 ± 4	45 ± 1	98 ± 4	88 ± 4	14 ± 2	>95
C6	28 ± 1	28 ± 4	45 ± 1	15 ± 1	<5	15 ± 4	>95
Ribose							
C1'	20 ± 4	93 ± 2	<1	42 ± 4	58 ± 8	36 ± 2	90 ± 5
C2'	<2	7 ± 1	90 ± 10	4 ± 1	50 ± 4	<2	80 ± 10
C3'	<2	<1	<1	<2	93 ± 1	<2	25 ± 10
C4'	<2	19 ± 1	90 ± 10	3 ± 1	<2	<2	>90
C5'	87 ± 12	<3	<1	95 ± 5	94 ± 3	95 ± 5	>90
$/100 µmol NTP	56	148	135	455	582	287	22

With RNA making up roughly 20% w/w of growing *E. coli*[175] and site-selective isotopic enrichment now possible through mutated cell lines, cell-based isotopic labeling seemed to be a strong candidate for providing low-cost $^{13}C/^{15}N$ NTPs. However, many challenges and additional costs in the purification stages led to low overall yields and prevented complete separation of NTPs. Growth of *E. coli* in minimal media conditions can sometimes results in long lag phases or inconsistent exponential growth phases. Cell cracking was also a tedious step and required blending the cells under harsh conditions with repeated extractions using phenol/chloroform that over time damages commercial blenders. Alternative cell-cracking methods such as sonication and French press resulted in lower overall recovery of RNA (<50%) when compared to the blending method. The final separation of NTPs into ATP, GTP, UTP, and CTP did not yield complete separation of UTP and CTP using previously published

protocols[157,168,171]. Long anion-exchange chromatography methods using AGX1 resin did provide better separation. However, after each NMP separation numerous regeneration steps were required in order to restore the anion-exchange column for future use[176].

The cell-based labeling strategy afforded 20-60 mg (60-180 μmol) of mixed NMPs from 1 L of cell culture. On average 0.2-0.4% w/v of ^{15}N and 1-2% w/v of ^{13}C sources provided optimal growth for various strains of *E. coli* in the presence of minimal media. At a cost of \$50/g of ^{15}N-$(NH_4)_2SO_4$ and \$150-1000/g for ^{13}C-glucose, glycerol, acetate, or pyruvate, the average cost per 100 μmol NTPs is \$22-600 depending on the ^{13}C-labeling source with acetate being the cheapest option. Although new labeling patterns that are commercially unavailable were now obtainable from cell-based isotopic enrichment strategies, the overall cost advantage was minimal and the purification steps were time consuming.

For ^{15}N-isotopic enrichment, cell-based labeling is still a strong candidate for lowering overall cost of NTPs. At a price of \$675/100 μmol for commercial ^{15}N-NTPs and with ^{15}N-$(NH_4)_2SO_4$ available at \$50/g, the net savings using the standard *E. coli* cell-based enrichment protocol is approximately \$575/100 μmol ^{15}N-NTP based on isotope costs. As part of fermentation startup award from the Institute for Bioscience and Biotechnology Research (IBBR), we were able to scale-up the production of NTPs using industrial scale fermentation. A 2 L culture of *E. coli* fed 15 g of ^{15}N-$(NH_4)_2SO_4$ along with constant oxygenation, pH monitoring, and agitation in minimal media yielded 26 g of wet cells and a total of

315 mg of nucleic acid. These results are in-line with expected scaled-up amount based on 1 L cell cultures. Large scale production does not circumvent the tedious cell-cracking and purification steps, but it did maximize the cell-growth using less total volume of media. The constant feeding required to maintain dense cell-growth limits the cost benefit for ^{13}C enrichment, but does reduce growth time.

2.2.3 Alternative species for cell-based labeling

Since the fermentation scale-up did not result in an increase in NTP/g of isotope, we investigated the usefulness of alternative cell types. The complexity of an organism has been shown to correlate with the percentage of non-protein coding DNA (Figure 2.3). It would, therefore, make sense that increased organism complexity yields a higher total RNA content. As a test case, we investigated the use of yeast cells (JD16 – provided kindly by Arturas Meskauskas and Jonathan Dinman) in producing isotopically enriched NTPs. It was immediately clear that much higher percentages of carbon sources (1% w/v compared to 0.2% w/v for *E. coli*) were needed for cell growth, which ruled out yeast as a potential ^{13}C-enrichment factory. There was still a potential to use yeast for ^{15}N-enrichment of NTPs. Indeed, a 500 mL cell culture resulted in 3.6 g of wet yeast cells and 79 mg of total RNA (232 µmol of NTPs). Interestingly, the amount of NTPs isolated from yeast cells was roughly 3 fold that of *E. coli*, which correlates well with the ratio of ncDNA in each species (0.124 for *E. coli* and 0.301 for yeast). This trend has been seen before with *Methylophilus methylotrophus*, which has half the ncDNA of *E. coli*, and when used for cell-

based labeling produced approximately 58% less NTPs than *E. coli*[150]. The total isotopic cost to produce 100 µmol of ^{15}N-NTPs was approximately $18 from yeast cell cultures, a 2-3 fold reduction compared to *E. coli* production for ^{15}N-NTPs.

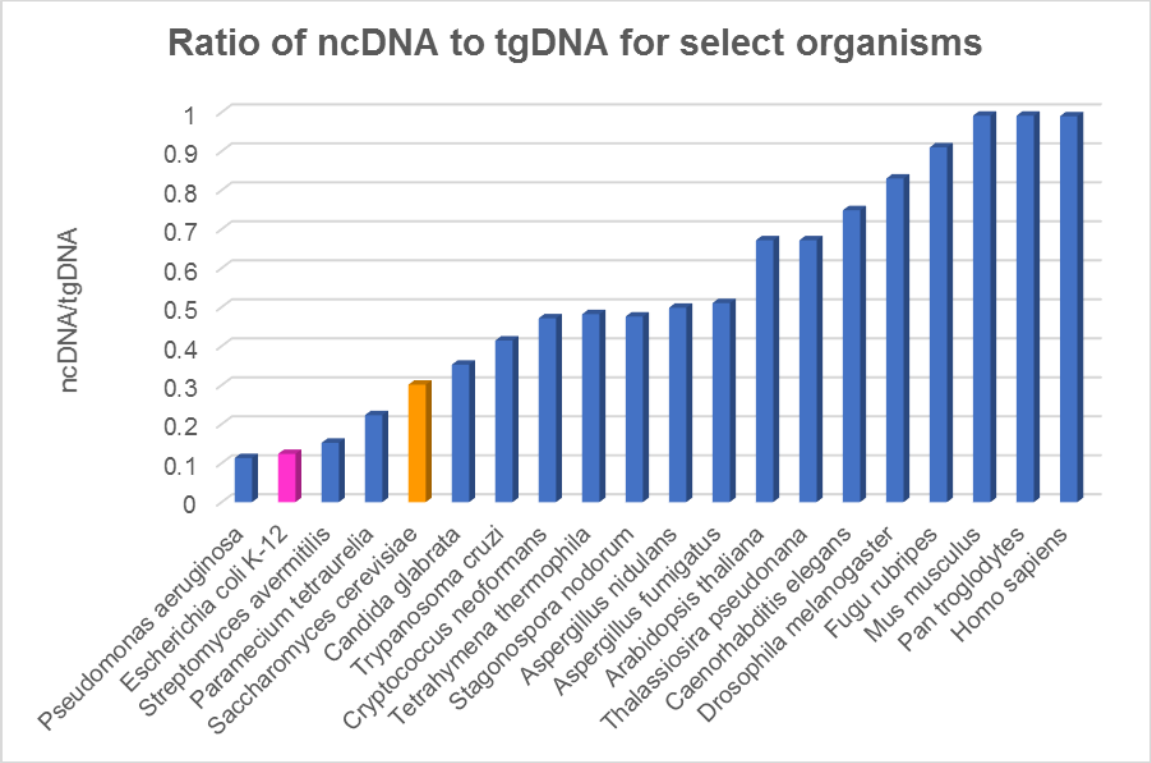

Figure 2.3. Ratio of non-coding DNA to total genomic DNA for select organisms. The ratio of ncDNA to tgDNA is given for various species as an indication for the amount of probable ncRNA present. *E. coli* (purple) and yeast (orange) are highlighted for quick reference[177].

Another benefit of using yeast cells to produce NTPs is that the cell-cracking procedure utilized glass bead agitation and did not require harsh phenol/chloroform blending conditions. However, given the greater complexity of yeast cells, the degree of post-transcriptional modification of isolated RNA may be greater than that of *E. coli* and could result in undesired modified NTPs. We

did not quantify the degree of post-transcriptional modifications in the RNA product and this should be investigated further. Yeast remains a strong alternative candidate for the production of ^{15}N-NTPs and also indicates that alternative species should be investigated for producing site-selective isotopically labeled NTPs.

2.3 *De Novo* biosynthesis of nucleotides

The first total *de novo* biosynthesis of pyrimidine[153] and purine[154] nucleotides was published by the Williamson group at The Scripps Research Institute. The benefits of enzymatic synthesis includes reduced reaction time, increased product yield and specificity, and reduced environmental impact compared with traditional chemical synthesis methods[178–184]. After exhaustive protein cloning, preparation, and purification, a *de novo* biosynthesis of either purine or pyrimidine nucleotides was achieved by the researchers from a set of isotopically enriched commercially available precursor compounds. Reasonable overall yields were reported with respect to the precursor isotopes for both pyrimidine (25-45%) and purine (24-66%) nucleotides using the *de novo* synthesis method[153,154].

A variety of isotopic labeling patterns for NTPs are amenable to the *de novo* synthesis method depending on the isotopic enrichment of the precursor compounds. Multiple cofactor regeneration systems reduce the need for multiple amino acids typically involved in the biosynthesis of ribonucleotides. The *de novo* synthesis of the pyrimidines only requires isotopically enriched D-glucose, sodium bicarbonate, ammonium chloride, and L-aspartic acid. The *de novo*

synthesis of purines require isotopically enriched D-glucose, sodium bicarbonate, ammonium chloride, and L-serine. We will briefly summarize the steps for the *de novo* synthesis of both the pyrimidines and the purines and indicate at which atomic positions the isotopic enrichment occurs for each nucleobase.

2.3.1 Pyrimidine de novo synthesis and labeling scheme

For pyrimidine synthesis, orotate is synthesized in a multistep reaction that starts by combining bicarbonate (C2 position of the pyrimidine ring) with ammonium (N3). The resulting carbamoyl phosphate is coupled to aspartate (C4, C5, C6, and N1), followed by ring closure to give dihydroorotate, and finally a reduction step to the aromatic orotate. Orotate is subsequently coupled to phosphoribose-pyrophosphate (PRPP – obtained from multi-step enzymatic synthesis from D-glucose) to give orotate monophosphate (OMP). Elimination of carbon dioxide from OMP yields UMP which is phosphorylated to UTP and can be further converted to CTP with CTP synthase and ammonium. In total, 15 enzymes were required for the synthesis of UTP and an additional enzyme for the conversion of UTP to CTP. The reaction details for the total *de novo* synthesis of pyrimidine can be found in the original publication by Schultheisz, Szymnczyna, Scott, and Williamson[153].

As shown in Figure 2.4, the commercially available isotopically enriched L-aspartic acid combinations dictate the levels of enrichment at the C4, C5, C6, and N1 positions of pyrimidine. Ammonium chloride controls the isotopic enrichment of N3 and for CTP the N4 position. Sodium bicarbonate determines the enrichment of the C2 position of the pyrimidines. If isotopic enrichment of C2

is important, care must be taken to prevent scrambling with atmospheric CO_2 or the eliminated C1 carbon of D-glucose by performing the reactions under inert conditions.

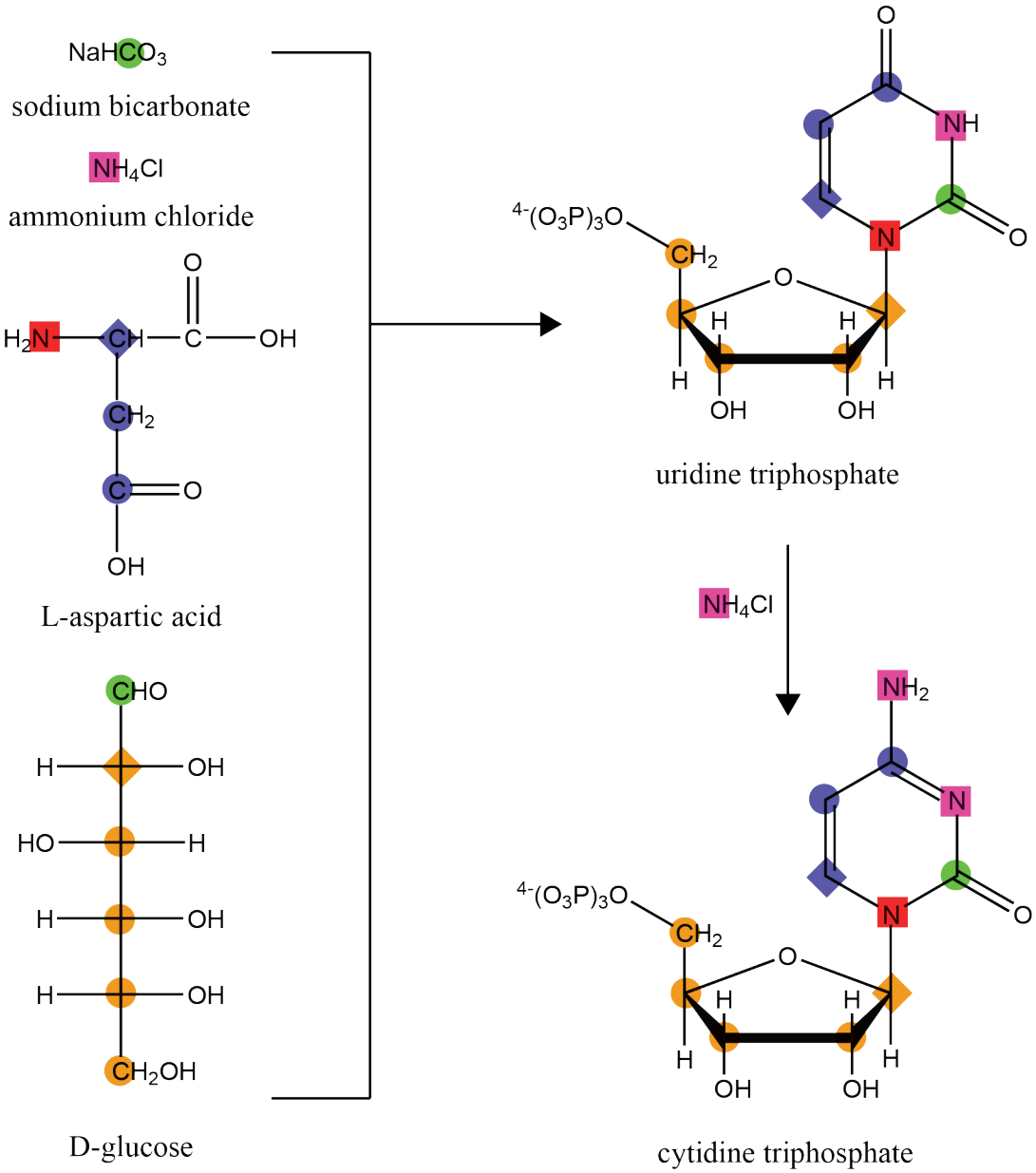

Figure 2.4. Pyrimidine de novo synthesis. The four isotopically labeled precursors needed for the *de novo* synthesis of pyrimidine are shown and the heteroatoms are color coded to match their corresponding position of enrichment on uridine triphosphate and cytidine triphosphate. UTP is converted to CTP in an additional step. For consecutive carbon atoms that are incorporated together, the first position is diamond shaped to indicate the direction of incorporation.

Based on the published yields for pyrimidine *de novo* synthesis and the current cost of isotopically enriched precursors, we have estimated the cost of production for each labeling pattern presented by the authors (Table 2.3). For CTP, the cost should be doubled due to the efficiency of conversion from UTP to CTP. Obtaining deuterated NTPs from this method is surprisingly cheap compared with commercial sources.

Table 2.3. Pyrimidine *de novo* synthesis cost analysis. The cost of synthesis for UTP is calculated based on the published yields and current cost of the precursors from CIL. The cost of CTP is double that of UTP. The labeling patterns are indicated by the positions of enrichment for either ^{2}H, ^{13}C, and/or ^{15}N. U=uniformly labeled (all possible positions).

Nucleotide Type	^{2}H	^{13}C	^{15}N	$/100 umol
UTP	5,3',4',5',5"	-	U	29
UTP	-	1',6	-	304
UTP	-	U	U	101
UTP	-	U	-	168
UTP	-	2',6	-	364

2.3.2 Purine de novo synthesis and labeling scheme

The *de novo* synthesis of the purines starts with the production of phosphoribose-amine (PRA) from PRPP via the amide nitrogen of glutamine, which provides what will become the N9 nitrogen of the purine ring system. Glycine – produced from serine starting material and a tetrahydrofolate (THF) regeneration system – is then coupled to PRA converting it to phosphoribosyl-glycinamide (GAR) providing the source for the C4, C5, and N7 atoms. The C_{β} atom of serine that was removed to produce glycine in the previous step becomes the basis for the formyl group of 10-formyl-THF in the THF regeneration

system, which is substituted onto GAR to produce phosphoribosyl-formylglycinamide (FGAR) resulting in the C8 atom of the purine nucleobase.

FGAR is converted to phophoribosyl-formylglycinamidine (FGAM) followed by ring closure to produce phosphoribosyl-aminoimidazole (AIR) incorporating the N3 atom from ammonium via a glutamine regeneration system. AIR and CO_2 from the sodium bicarbonate precursor are combined to yield phosphoribosyl-amino-carboxy-imidazole (CAIR) providing the C6 atom. The N1 atom is obtained from ammonium precursor introduced via aspartic acid coupling to CAIR yielding phosphoribosyl-amino-succinocarboxamide-imidazole (SAICAR). Fumarate elimination followed by another CO_2 addition at the C2 atom via 10-formyl-THF results in formation of phosphoribosyl-formylamido-imidazole carboxamide (FAICAR). Ring closure and removal of water yields inosine monophosphate (IMP). IMP is then converted to either GTP or ATP.

In summary and show in Figure 2.5, the *de novo* purine synthesis uses ammonium to provide N1, N2, N3, N6, and N9 atoms. The N7 atom is derived from serine. The carbon atoms at C2, C4, C5, and C8 are generated from serine while the C6 atom comes from sodium bicarbonate. By utilizing aspartic acid, glutamine, and 10-formyl-THF regeneration systems, the basic ammonium, bicarbonate, and serine precursors provide all the isotopic enrichment sources for purine. Complete reaction details are presented in the original paper by Schultheisz, Szymcyzyna, Scott, and Williamson and require 23 enzymes as well as multiple cofactors[154].

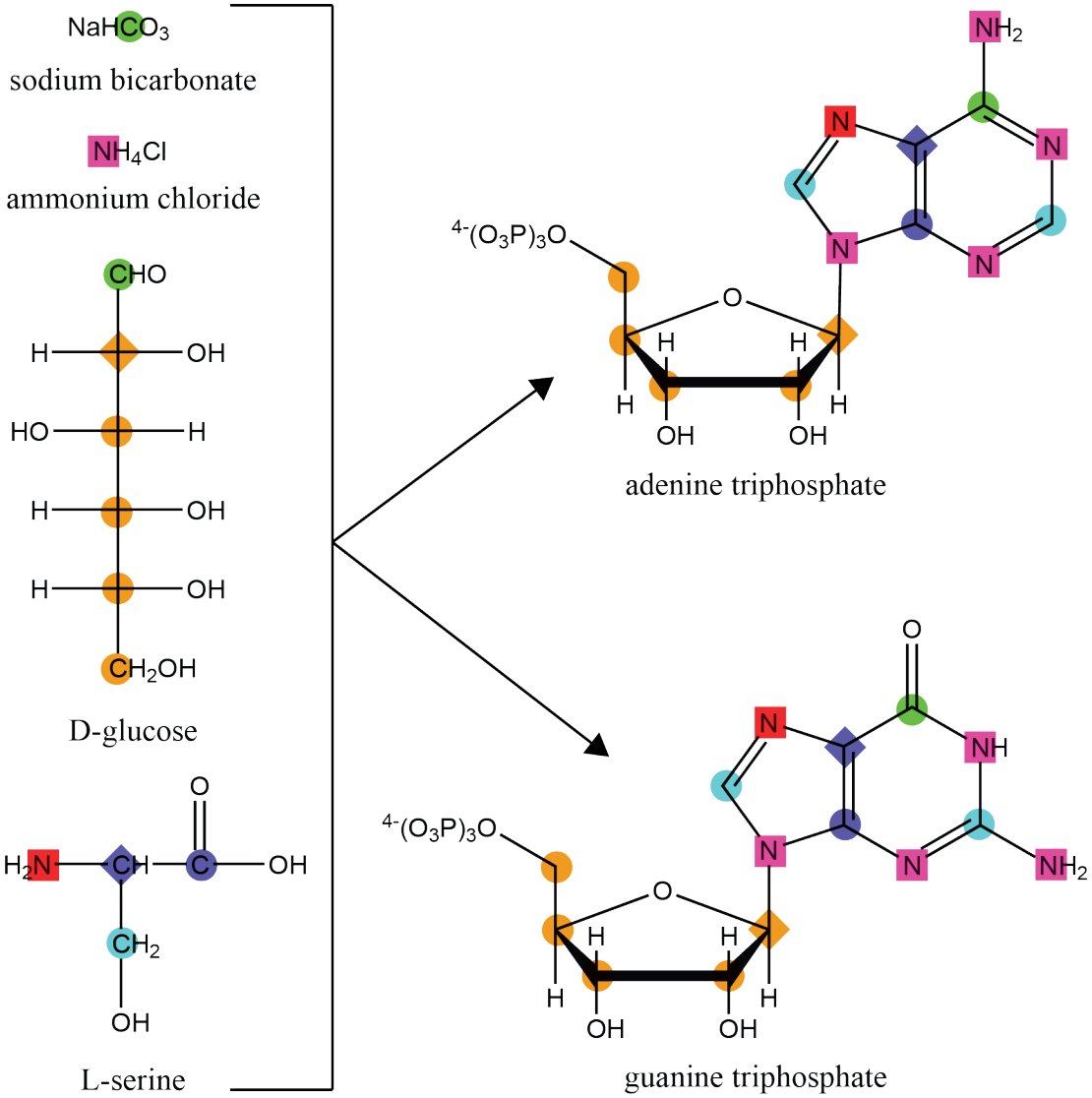

Figure 2.5. Purine de novo synthesis. The four isotopically labeled precursors needed for the *de novo* synthesis of purine are shown and the heteroatoms are color coded to match their corresponding position of enrichment in ATP and GTP. For consecutive carbon atoms that are incorporated together, the first position is diamond shaped to indicate the direction of incorporation.

Based on the published yields for purine *de novo* synthesis and the current cost of isotopically enriched precursors, we have estimated the cost of production for each labeling pattern presented by the authors (Table 2.3). For deuterated GTP, the cost was estimated based on replacing uniformly ^{13}C-

labeled D-glucose with uniformly ^{2}H-labeled D-glucose. The methods only offers slight (2 fold) reductions in cost compared to commercial NTPs. Reductions in the cost of isotopically enriched L-serine would greatly improve the cost benefits of this method.

Table 2.4. Purine *de novo* synthesis cost analysis. The cost of synthesis for ATP and GTP is calculated based on the published yields and current cost of the precursors from CIL. The labeling patterns are indicated by the positions of enrichment for either ^{2}H, ^{13}C, and/or ^{15}N. U=uniformly labeled (all possible positions). *Yield assumed to be similar to U-^{13}C GTP.

Nucleotide Type	^{2}H	^{13}C	^{15}N	$/100 umol
ATP	-	2,8	-	311
GTP	-	-	U	310
GTP	-	U	U	358
GTP	-	U	-	279
GTP*	3',4',5',5"	-	-	283

2.3.3 Analysis of de novo synthesis for isotopic labeling of nucleotides

Through the combination of several cofactor regeneration schemes, a *de novo* synthesis was reported in a one-pot reaction, reducing the overall reaction time and subsequent purification steps. The substantial upfront cost in producing the commercially unavailable enzymes (15-28 enzymes) required for the *de novo* synthesis of NTPs does limit the applicability of the method for some laboratories. The authors also note that several of the enzymes in the purine synthesis pathway are unstable and last only a few weeks in storage[154].

The regeneration schemes eliminate the need for multiple amino acids, but it also limits the site-selective labeling offered by the method. The purine C2 and C8 positions must be labeled in tandem or not at all, which limits the type of NMR strategies available for resonance assignment, which will be highlighted in

Chapter 3. The available ribose labeling schemes are limited to commercially available sources of D-glucose. If the upfront cost of enzyme production is not a limiting step, the *de novo* synthesis offers an efficient method to obtain isotopically enriched NTPs. A significant savings for deuteration labeling schemes can be realized with the *de novo* synthesis pathway, but uniform ^{13}C and ^{15}N only offers marginal savings.

2.4 Chemo-enzymatic synthesis of isotopically labeled nucleotides

Multiple methods have been published for the chemical synthesis of pyrimidine[185–187] and purine[186,188–190] bases; however, coupling the ribose system to these synthesized bases was not achieved with good efficiency until a hybrid strategy to include enzymatic synthesis of the sugar and coupling of the sugar-base was employed. In collaboration with Christoph Kreutz, we have developed and improved methods for complete chemo-enzymatic synthesis of isotopically enriched purine and pyrimidine NTPs. These $^{13}C/^{15}N$-isotopically labeled NTPs can be combined as needed for T7 transcription of RNA for NMR structural and dynamic studies.

The pyrimidines are all synthesized initially from the uracil base. Uracil is chemically synthesized from 2-bromoacetic acid, KCN, and urea (Figure 2.6). The 2-bromoacetic acid precursor provides the C4 and C5 carbons of uracil. Addition of KCN accounts for the C6 position of uracil. Both nitrogen positions

are enriched from the urea precursor, therefore they must be labeled together or not at all due to the symmetry of urea. The C2 atom is also enriched from urea.

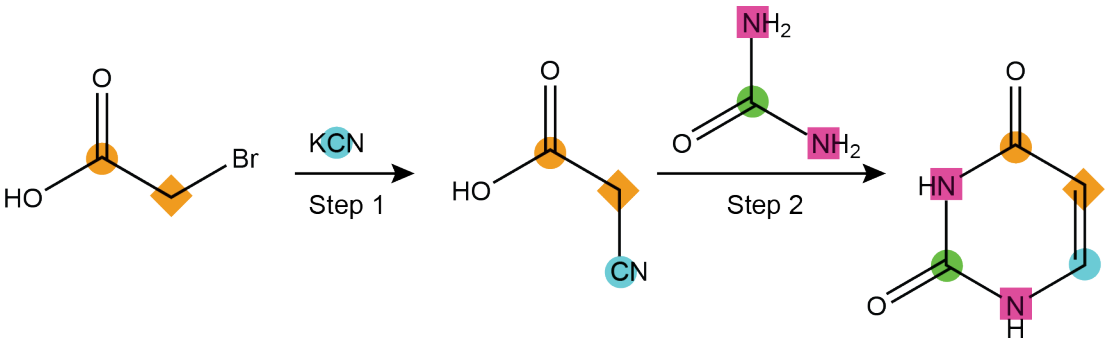

Figure 2.6. Chemical synthesis of uracil. The potential isotope labeling sites of uracil are color coded to match the source of enrichment from the precursor compounds. The two carbon position of 2-bromoacetic acid are labeled as either a circle or diamond to reflect the corresponding position in uracil. Step 1) 2-bromoacetic acid, Na_2CO_3 and KCN are refluxed in H_2O at 80 °C for 3 h while stirring and then allowed to stir at room temperature for another 20 h. Step 2) Isolated 3-cyanoacetic acid is combined with urea in diethyl ether and heated to 90 °C for 30 min then 5% $Pd/BaSO_4$ is added under a H_2 atmosphere until uracil is obtained. More detailed synthesis conditions are provided elsewhere[155].

The cost analysis for the synthesis of uracil from 2-bromoacetic acid, KCN, and urea is given in Table 2.5. All possible heteroatom positions can be easily enriched and with relatively high yields and easy scale-ability. We have listed the costs based on current CIL pricing assuming 55% yields and only taking into account the cost of isotopes. Additional price reduction can be achieved by starting with acetate and brominating to obtain 2-bromoacetic acid.

Table 2.5. Uracil synthesis cost analysis. U=uniform labeling.

Base	^{13}C	^{15}N	$/100 µmol
Uracil	U	U	15
Uracil	U	-	9
Uracil	-	U	2
Uracil	6	U	6
Uracil	5	U	6

Purine bases are also chemically synthesized separately using established methods[186]. Although many alternative methods have been proposed[189,190], we find that this simple synthesis provides high yields and suitable labeling for our needs. In brief, morpholine is combined with ^{13}C formic acid to form a morpholinium formate precipitate. Either 4-hydroxy-2,5,6-triaminopyrimidine sulfate or 4,5,6-triaminopyrimidine sulfate is added to the morpholinium formate mixture and refluxed at 200 °C for 1 hour followed by continued stirring for up to 2 hours to yield 8-^{13}C-guanine or 8-^{13}C-adenine, respectively (Figure 2.7). This chemical labeling method for purines has an average isotope cost of about $10 per 100 µmol of 8-^{13}C-guanine and $14 per 100 µmol of 8-^{13}C-adenine.

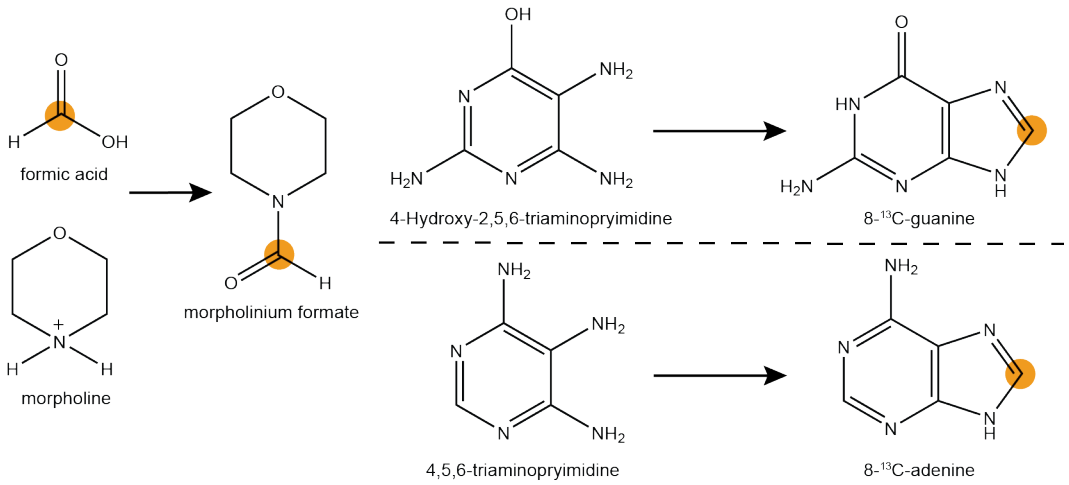

Figure 2.7. Chemical synthesis of purines. The C8 atom of adenine and guanine are 13C enriched from formic acid precursor. Typical yields are 64% for adenine and 93% for guanine. Full experimental details can be found in the supplementary material of our previously published work[191].

The overall cost of obtaining either pyrimidine or purine bases is reasonably low. There was little need to improve these methods. The

improvements from the chemo-enzymatic method comes from the coupling of D-ribose to the respective base and phosphorylating the NMP to NTP. Starting from D-ribose has many advantages over previously reported *de novo* synthesis methods including fewer enzymes and more diverse commercially available labeling patterns for D-ribose. Through rigorous optimization of enzyme conditions, Andrew Longhini and Luigi Alvarado from our laboratory were able to obtain almost quantitative yields for the enzymatic phase of the chemo-enzymatic synthesis process[191,192]. The enzymatic coupling scheme is shown in Figure 2.8.

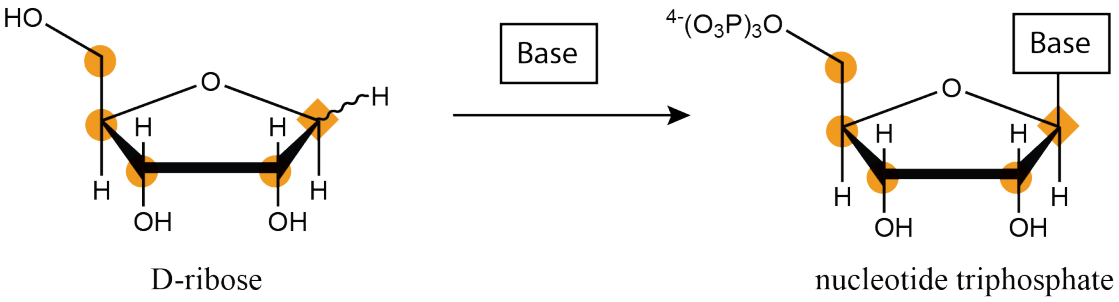

Figure 2.8. Enzymatic coupling of ribose and base. Either adenine, guanine, or uracil can be coupled to D-ribose using ribokinase, phosphoribosyl pyrophosphate, a base-specific phosphoribosyl transferase, nucleoside monophosphate kinase, and a dATP regeneration system. Carbon labeling positions on the final NTP ribose match the input ribose.

The cost of using D-ribose precursors substantially reduces the overall price of obtaining a variety of labeling patterns for the sugar moiety of the NTP (Table 2.6).

Table 2.6. Enzymatic coupling cost analysis. The cost in $ per 100 umol is given for the enzymatic coupling or D-ribose to any given base. We assumed 80% conversion for our analysis and used Omircon Isotope pricing.

D-Ribose	1'	2'	3'	4'	1',2'	Uniform
$/100 umol	6	11	45	91	24	23

For the chemo-enzymatic method, CTP is made by the enzymatic amination of the C4 carbon of UTP. This process has typical yields around 80-90% and ^{15}N-NH_4Cl can be used to enrich the amine of CTP. Overall, much less enzyme is required compared to the *de novo* synthesis method. Synthesis of variations for the purine base is still an ongoing effort that requires revisiting alternative synthesis pathways. Currently, with the chemo-enzymatic method we are limited to only C8 enrichment of either adenine or guanine. Incorporating alternative purine synthesis pathways developed by the Jones group at Rutgers University would allow specific labeling at the C2 as well as any nitrogen position[193,194]. The pyrimidines, however, have a much higher degree of flexibility and we can obtain numerous isotopic labeling schemes.

2.5 Cost analysis for all methods of isotopic labeling

A thorough analysis of the isotopic labeling cost for each method reveals that each has its strengths. The cell-based labeling method is best utilized for obtaining u-^{13}C/^{15}N purines, since the *de novo* synthesis is much more challenging and the chemo-enzymatic method does not afford flexibility in the base labeling scheme. Another major strength of the cell-based labeling method is the production of ^{15}N-labeled NTPs from yeast cell cultures, which provides a 38-fold savings compared to commercial ^{15}N-NTP sources. The *de novo* synthesis methods only provide marginal savings over commercial sources for purine NTP synthesis.

Pyrimidine NTPs were much cheaper to produce through the chemo-enzymatic method, with savings of 338-fold for the production of ^{15}N-labeled UTP

and CTP. The chemo-enzymatic method also gave new labeling schemes that are not commercially available and opens the door for new NMR experiments. We will discuss these experiments in further detail within the following chapter. A full overview of the savings associated with each method is given in Table 2.7.

Table 2.7. Overall cost analysis for isotopic labeling methods. The cost savings for obtaining isotopically enriched NTPs is given as a fold decrease in cost (i.e. half the cost is 2x). NA = no commercially available comparison. The method used is listed with yeast and *E. coli* being the cell-based methods, CE is the chemo-enzymatic method, and *de novo* is the full enzymatic synthesis.

Nucleotide	^{2}H	^{13}C	^{15}N	Method	$/100 µmol	Savings
Purine	-	-	U	yeast	18	38x
Purine	-	U	U	*e. coli*	50	14x
Purine	-	1',8	-	CE	20	NA
Purine	3',4',5',5"	-	-	*de novo*	283	2x
Pyrimidine	-	-	U	CE	2	338x
Pyrimidine	-	U	U	CE	48	14x
Pyrimidine	-	U	-	CE	32	NA
Pyrimidine	-	1',6	-	CE	12	NA
Pyrimidine	5,3',4',5',5"	-	-	*de novo*	29	21x

2.6 RNA preparation methods

Once isotopically enriched NTPs are obtained from commercial, cell-based, *de novo* enzymatic synthesis, or chemo-enzymatic methods, the synthesis of the complete RNA biopolymer is achieved through either *in vitro* transcription or solid-phase synthesis. The choice of method is dependent on the needs and the length of the RNA primary sequence. For *in vitro* transcription, any length RNA >12 nt can be readily transcribed yet specific sequence requirements can effect both the yield and heterogeneity of the resulting transcript. For solid-phase synthesis, overall yield plummets for larger RNAs

(>60 nt), but can be a fast and efficient method for obtaining position specific labeling. We have previously reviewed some of the strategies employed for obtaining mg quantity RNA transcripts from *in vitro* transcription, solid-phase synthesis, hybrid RNA ligation, and a cell-based preparation of RNA[195]. Here we will provide brief summaries of the methods.

2.6.1 T7 transcription

RNAs for NMR spectroscopy studies are commonly prepared from NTPs by *in vitro* transcription using T7 RNA polymerase (T7) enzyme (alternatively, T3 or SP6 RNA polymerase may be used for large scale transcription of RNA). The synthesis is straightforward and allows for the production of mg quantities of RNA over 12 nt in length using DNA template that contains the T7 promoter sequence (5'- CTA ATA CGA CTC ACT ATA G -3')[196]. The T7 enzyme is highly specific to its own promoter sequence and error frequency in transcripts can be as low as 6 x 10^{-5} as determined by incorrect nucleotide substitutions[197]. A DNA template that contains the reverse complimentary sequence to the RNA of interest followed by the reverse complimentary sequence to the T7 promoter can be synthesized, purchased, or PCR amplified. Any of set of complete NTPs can be used in the T7 *in vitro* transcription including the isotopically enriched NTPs discussed in this chapter.

For our purpose, we performed T7 *in vitro* transcriptions for the work in this thesis using previously described methods with minor modifications[198]. Transcriptions were performed in 40 mM Tris-HCl pH 8.0 (at 37°C), 1 mM spermidine, 0.01% Triton-10, 80 mg/mL PEG, 0.3 µM DNA template, 1 mM DTT,

2 U/μL Thermostable Inorganic Pyrophosphatase (TIPP), 5-10 mM NTPs, 5-20 mM $MgCl_2$, and 0.5-2 mg/mL T7 RNA polymerase. Reactions were allowed to proceed for 3 hours at 37°C. For each RNA transcribed, the concentrations of $MgCl_2$, T7 RNA polymerase, and NTPs were optimized at small scales (20-50 μL) before scaling up to large-scale transcriptions. We typically performed 10-20 mL of transcription for an adequate NMR sample. RNA transcripts are then purified by polyacrylamide gel electrophoresis (PAGE), excised and electro-eluted, then buffer exchanged and concentrated into the appropriate volume for NMR studies.

T7 *in vitro* transcription has been a tried-and-true method for producing large quantities of RNA, yet there are still drawbacks to the method. Chief amongst these concerns is the propensity to produce 5' and 3' heterogeneous ends to the RNA sequence. The presence of these faulty RNA sequences can create a mixture of chemical environments that leads to reduction in signal intensities and can introduce ambiguity in NMR data interpretation[130,133]. To alleviate these concerns, several groups have implemented ribozyme cleaving sequences into the DNA template to reduce heterogeneity, which has been well-reviewed[170].

2.6.2 Cell culture overexpression

Although T7 *in vitro* transcription remains the most common method for obtaining mg quantities of RNA for NMR, many groups have tried to adapt alternative strategies. Most recently, Ponchon and Dardel developed an *in vivo* recombinant RNA expression method in order to obtain isotopically enriched

RNA from cell extract[199,200]. Since RNA is typically degraded quickly in the cell, they used a tRNA-scaffold plasmid to express recombinant RNAs in the anticodon loop of a tRNA-scaffold thereby disguising the RNA from cellular nucleases. Several other groups have modified this method including the use of 5S rRNA as scaffold[201,202]. A recent publication from our group combined this cell culture RNA expression system with the cell-based labeling strategy we have discussed within this chapter to obtain site-selectively labeled *in vivo* expressed RNA[203]. Although this method allows for mg quantity expression of RNAs from *E. coli*, there should be caution when implementing the method. Not all sequences have been shown to express well and cleavage of the RNA from the tRNA scaffolding is not always efficient.

2.6.3 PLOR

A recently developed technique was published for the position-selective incorporation of NTPs into a growing RNA chain[204]. This position-selective labeling of RNA (PLOR) relies on the ability to stall T7 RNA polymerase by omitting the required NTP for processivity. This pause-restart method enables the synthesis of position- or region-specific incorporation of nucleotides by using solid-phase-coupled DNA templates and limited combination of NTPs available for T7 *in vitro* transcription. This method will most likely become a powerful tool for RNA structural biologist, but currently it requires customized robotic instrumentation that is not freely available. The PLOR method is not necessarily restricted to NMR sample preparation as fluorescently tagged NTPs can easily be incorporated at specific positions in the RNA as well.

2.6.4 Solid-phase synthesis

RNA can be synthesized through solid-phase chemical synthesis, but the phosphoramidites needed are not commercially available with stable isotopic enrichments. Working with our collaborators, we have shown that any of the NTPs obtained for T7 *in vitro* transcription can be converted to phosphoramidites for solid-phase synthesis of RNA[205,206]. A common practice with phosphoramidite synthesis is to couple a base with a protected ribose through the *Vorbruggen* condensation[207]. As we previously published, subsequent steps of protecting group removal and tritylation at 2'-OH, TOM addition at 5'-OH, base amine protection, and finally 3'-OH conversion to 3'-*O*-(2'-cyanoethyl *N,N*-diisopropylphosphoramidte) yields the final phosphoramidite ready for solid-phase synthesis[205].

Solid-phase synthesis methods have been well established[208]. The main advantage over T7 *in vitro* transcription is that there is no sequence requirements and nucleotides can be specifically labeled at any site within the RNA. When only a few residues are of interest, solid-phase synthesis can be used to only label these nucleotides thereby minimizing signal crowding and reducing resonance assignment time for NMR studies. Phan and Patel proposed a single-nucleotide isotopic enrichment strategy using solid-phase synthesis to unambiguously assign small DNA tetrads[209]. Solid-phase synthesis is size limited in that yields drop quickly as RNA length extends beyond 60 nt. Combining solid-phase synthesis with ligation methods will be a key strategy for studying large RNA systems by NMR.

2.6.5 Ligation

Current ligation strategies in RNA have become routine, but still require a bit of tinkering and pre-planning[109,210–214]. Ligation is most commonly used to reduce spectral crowding in NMR spectroscopy by only labeling a small fragment with isotopically enriched NTPs and then ligating that fragment onto the full RNA construct. Either T4 DNA or RNA ligase is used to couple two RNA fragments one of which must have a 3'-OH and the other a 5'-monophosphate. The use of either solid-phase synthesis or ribozyme cleavage of T7 *in vitro* transcripts can typically yield the desired termini[170]. T4 DNA ligase tends to be slower and results in lower yields compared to T4 RNA ligase and it also requires a DNA template to bring the termini of the two fragments in close proximity[215]. Many novel strategies have been developed for RNA ligation[170]. Within this thesis, we have used ligation to unambiguously assign a critical cytosine residue in the SAM-II riboswitch and gain insightful exchange measurements[210].

2.7 Conclusion

NMR is a powerful tool for studying RNA structure and dynamics, yet we have seen a stagnation in published RNA structures. In order to promote continued research and further progress in RNA structure discovery, we endeavored to minimize the cost associated with NMR studies of RNA. Within this chapter we have presented several methods for preparing RNA samples with cost savings upwards of 300-fold compared to commercially available NTPs. We have also provided a table (Table 2.7) to help inform the reader of which strategy is most cost-efficient depending on the labeling scheme they desire. In the

following chapter we will show a powerful application of the labeling schemes afforded through the chemo-enzymatic synthesis of NTPs. This method would not be feasible by current commercially available NTPs.

3 A novel strategy for rapid and logical NMR resonance assignment of RNA

Even though ~98% of the human genomic output is transcribed non-protein coding RNA, less than 2% of the protein data bank structures comprise RNA. This huge structural disparity stems not only from difficulties in readily crystallizing but also assigning the resonances of RNAs due to extensive chemical shift overlap and broadened line-widths. Here we propose a novel strategy for resonance assignment that combines new strategic ^{13}C labelling technologies with filter/edit type NOESY experiments to greatly reduce spectral complexity and crowding. With this new strategy, we were able to assign important non-exchangeable resonances of proton and carbon (1',2',2,5,6, and 8) using only one sample and less than 24 hours of NMR instrument time for a 27 nucleotide model RNA system.

3.1 Background

Resonance assignment of RNA is non-trivial with no consensus strategy or software to aid in automation. To begin to tackle this gap in knowledge using NMR, a large suite of multi-nuclear NMR experiments have been developed to overcome this bottleneck and various strategies proposed to assign resonances in RNA molecules[108,130,131,147,170,216–227]. Chemically, RNAs have only four bases that look essentially the same, whereas proteins feature 22 chemically distinct amino acids. This profound lack of chemical diversity within RNA leads to severe spectral crowding of the information rich proton resonances. Implementation of through-bond experiments has enabled automated sequential assignment of

resonances in proteins, but these experiments are difficult to implement for RNAs[224–226,228]. Therefore, sequential assignment by through-space magnetization transfer using nuclear Overhauser effect spectroscopy (NOESY) has become the main tool for assigning RNA proton resonances.

An unambiguous starting point is a critical step in the chemical shift assignment strategy for RNAs, and traditionally this begins with NOE-based sequential assignment of the exchangeable resonances[229,230]. Here we propose another approach that starts with non-exchangeable resonances, but still relies on through-space NOESY experiments. However, we now combine isotope filtering and editing with our new $^{13}C/^{12}C$ alternative labeling patterns within the sugar and base moiety of the nucleotide building blocks[156,192,231] to systematically assign the non-exchangeable proton resonances as a powerful alternative to previously established methodologies[131,229,230]. The combination of an asymmetric labeling pattern and isotopic filtered/edited NOESY experiments[218,232] allows us to logically distinguish between intra- and inter-nucleotide NOEs for sequential resonance assignment. The filtering/editing NOESY strategy has previously been implemented using $^{15}N/^{14}N$ labeling strategies for duplex formation between labeled and unlabeled strands of RNA, for protein:RNA complexes, and for samples containing only uniformly $^{13}C/^{15}N$ labeled uridine and unlabeled AGC rNTPs[218,233,234].

As a model system we use a 27 nt RNA fragment encompassing the highly conserved aminoacyl-tRNA anticodon site (A site) of 16S rRNA on the 30S subunit[235]. This novel strategy that combines filtering and editing NOESY

techniques with our new labeling scheme enables rapid and unambiguous assignment for adenine H2 inter-nucleotide NOESY cross-peaks with sequential and cross-stand H1' resonances. We were able to successfully assign all the H1', H2', H2, H5, H6, and H8 resonances of the 27 nt A-site RNA from a single NMR sample using only three 2D ^{1}H-^{1}H NOESY experiments. We anticipate that this improved resonance assignment strategy for RNAs will allow more researchers access to the growing field of RNA biophysics.

3.2 Experimental methods

3.2.1 NTP and oligonucleotide synthesis

We strategically synthesized labeled NTPs from selectively labeled commercially available ribose and chemically synthesized base for an in vitro T7 polymerase based transcription. A schematic of the one-pot synthesis of our model RNA from the NTP precursors is shown in Figure 3.1. Reactions were carried out as previously described[156,191,192,231].

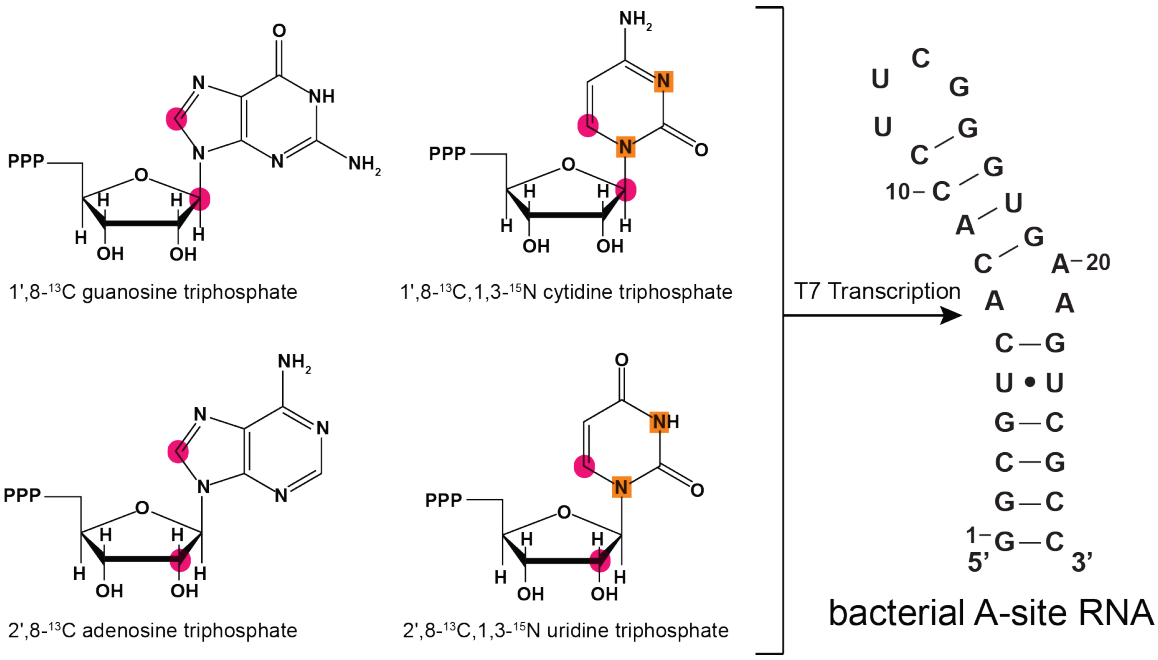

Figure 3.1. Alternating ^{13}C isotopic labeling of bacterial A-site RNA. Following previously published methods, D-ribose-1-^{13}C was enzymatically coupled separately either to 8-^{13}C-guanine[191] or 6-^{13}C-1-^{15}N-cytosine[192,231], and then phosphorylated to their respective rNTPs. D-ribose-2-^{13}C was coupled separately to either 8-^{13}C–adenine or 6-^{13}C-1-^{15}N-uracil, and phosphorylated. Atoms enriched with ^{13}C are highlighted with purple circles and atoms enriched with ^{15}N are indicated by orange squares. Triphosphates are abbreviated as PPP. The purified NTPs were used to synthesize an alternatively labeled RNA sample (27 nt bacterial A-site RNA) for NMR resonance assignment.

RNA was transcribed from DNA template (Integrated DNA Technologies) with T7 RNA polymerase using standard methods[131,236]. The transcribed RNA was purified by denaturing PAGE and electro-eluted from gel slices. The electro-eluted RNA was buffer exchanged in 3K MWCO centrifuge tubes (Millipore) against EDTA followed by extensive washes with ddH_2O. The sample was then lyophilized to dryness after addition of NMR buffer (final concentration of 10 mM sodium phosphate, 0.1 mM EDTA, 0.1% NaN_3, and 0.1 mM DSS at pH 6.4) and

re-suspended in ~300 µL of 99.98% D_2O for a final RNA concentration of 0.3-1 mM.

3.2.2 NMR experiments

All NMR experiments were performed on an Ultrashield Plus 600 MHz Avance III Bruker spectrometer with a room temperature TXI 600SB H-C/N-D-05 Z BTO probe. NMR experiments in D_2O were collected at 25°C and exchangeable proton experiments in H_2O were collected at the 15°C (optimized from several 1D imino experiments at various temperatures from 5-35°C, see Figure A.1). Data were collected and analyzed using TopSpin 3.2, NvFX, and NMRViewJ[237].

For the sequential assignment of the non-exchangeable proton resonances (1', 2', 2, 5, 6, 8), we implemented three different filter/edit type NOESY experiments as well as a ^{13}C-HSQC. The filter/edit type NOESY experiments were classified by the dimension (F1=indirect dimension and F2=direct detected dimension) as well as the proton chemical shift observed (a=all protons, e=edited, only protons attached to ^{13}C, and f=filtered, only protons attached to ^{12}C). The F1eF2e NOESY experiment was a standard NOESY-HSQC with an initial double purge module to dephase ^{1}H-^{12}C magnetization adapted from previously published methods[218]. The F1eF2f and F1fF2e NOESY experiments were standard Bruker pulse sequences (*hsqcgpnowgx33d* and *noesyhsqcgpwgx13d*, respectively). The F1eF2e, F1eF2f, F1fF2e NOESY and a standard ^{1}H-^{13}C HSQC were collected on the alternatively labeled A-site RNA sample in less than 24 hours of total instrument time.

Since there are no published or deposited peak lists available for the 27 nt bacterial A-site RNA, we ran imino NOESY and through-bond heteronuclear correlation experiments on a uniformly $^{13}C/^{15}N$- labeled A-site RNA to confirm our resonance assignment. The following experiments were run in H_2O for a total of 52 hours of NMR instrument time: 1D ^{1}H imino with Watergate, Imino NOESY with jump-return or Watergate, ^{1}H-^{15}N-HSQC, 2-bond ^{1}H-^{15}N HSQC, and HNN-COSY[238–244]. Imino resonance assignment was confirmed with the automated assignment method, RNA-Pairs[224]. We also collected heteronuclear correlation data for the same sample in D_2O with the following NMR experiments for a total of 16 hours of NMR instrument time: ^{13}C-HSQC, HCN, and HCCH-COSY[245–247]. The parameters for all the NMR experiments are given in the supplementary data (Table A.1). All ^{1}H chemical shifts were calibrated to DSS and ^{13}C chemical shifts were calibrated indirectly to ^{1}H by the gyromagnetic ratio using previously established protocols for RNA[248].

3.3 Selective labeling strategy for logical assignment of RNA NMR resonances

Based on previously published methods from our laboratory, we devised a new labelling scheme using isolated ^{13}C labels within ribose and base moieties to eliminate ^{13}C-^{13}C spin-spin and edit or filter signals within NOESY experiments[192,231]. Our strategy uses alternate ^{13}C-labelling of the ribose C1' and C2' positions. This allowed a nice separation of all peaks in the H2' and C2' dimension of a ^{13}C-HSQC. The remaining G and C residues were ^{13}C-labelled at the C1' position with minimal overlap in the H1' dimension. For the base

residues, pyrimidine C6 and purine C8 were ^{13}C-labelled, and Ade C2 and pyrimidine C5 were unlabeled to minimize the usual crowding in the NOESY spectra.

The power of filtering/editing the signal of the NOESY spectra was immediately evident when compared with a standard 2D ^{1}H-^{1}H NOESY, F1aF2a NOESY (Figure 3.2). By separating half of the nucleotides into the H1' or H2' region, the spectral crowding of the “fingerprint” region is essentially reduced to half for the F1eF2e NOESY. The spectral crowding was significantly reduced to the point that full resonance assignment could be accomplished from a combination of three 2D ^{1}H-^{1}H NOESY spectra. Specific regions of the various NOESY spectra contain valuable structural information and our labeling strategy allowed us to develop a logical resonance assignment scheme based on the unique information content of each region. In Figure 3.2, we have highlighted these regions in a color coded manner which is discussed in further detail below.

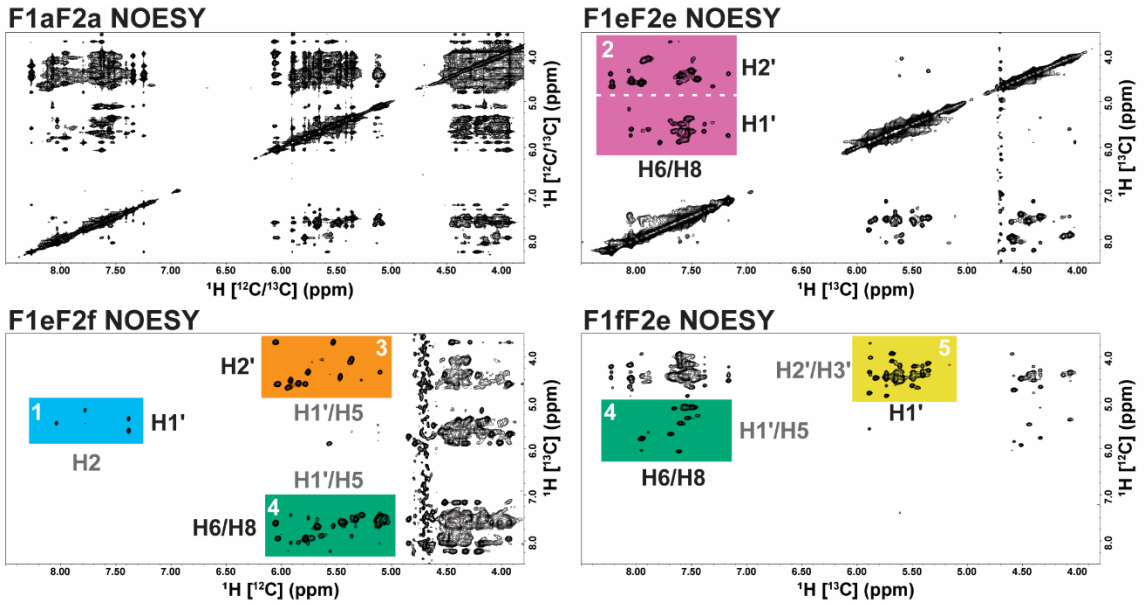

Figure 3.2. 2D NOESY spectra of site-selective alternately labeled A-site RNA. Four separate NOESY spectra from the alternatively labeled bacterial A-site RNA with color coded regions labeled by expected proton cross-peaks (^{13}C-label in black and ^{12}C-label in gray) on the edge of each color coded region. Each spectrum is labeled by the dimension (F1=indirect and F2=direct) followed by the type of proton observed (a=all [^{13}C-label and ^{12}C-label], e=edited [^{13}C-label], and f=filtered [^{12}C-label]). The axis of each spectrum is also labeled by the type of observed proton resonance with the attached carbon isotope given in brackets. For example, the F1eF2e NOESY in the top right panel contains only cross-peaks from the ^{13}C-labeled non-exchangeable protons. The color coded regions are as follows (white number): 1=sky blue, 2=reddish purple, 3=orange, 4=bluish green and 5=yellow. The utility of these regions are described in detail within the text and the color and numbering of the highlighted regions correspond to the steps for logical resonance assignment.

3.4 Logical assignment of non-exchangeable proton resonances

A logical sequential assignment strategy that arises from the alternative labeling of RNA is represented in Figure 3.3 alongside a schematic with the NOE cross-peaks expected for three stacked base-pairs in the upper-stem of the 27 nt bacterial A-site RNA. The various edited/filtered NOESY spectra (F1eF2e,

F1eF2f, and F1fF2e) provide useful separation of cross-peaks into identifiable regions that enables a logical assignment as well as simplification of spectral analysis. The colors and numbers depicted in Figure 3.3 correspond to the regions of the spectra shown in Figure 3.2. Step 1 is essential to the assignment process by identifying starting points for the NOESY “walk” and our unique labeling simplifies the assignment of the H1' inter-nucleotide cross-peaks to adenine H2 resonances. Step 2 is the typical NOESY “walk” of the fingerprint region, but with our labeling scheme H2' cross-peaks are easily assigned by removing crowding from H3', H4', H5', and H5'' resonances. Step 3 is useful for assigning the unlabeled (^{12}C) H1' of adenine and uridine which is further used for step 4 to confirm intra- and inter-nucleotide NOEs with labeled (^{13}C) H6/H8 resonances. Step 5 is the final process of assigning the remaining unlabeled (^{12}C) H2' resonances of guanine and cytidine. We demonstrate this assignment process using the A9-C10 stacked nucleotides within the same upper-stem of A-site RNA.

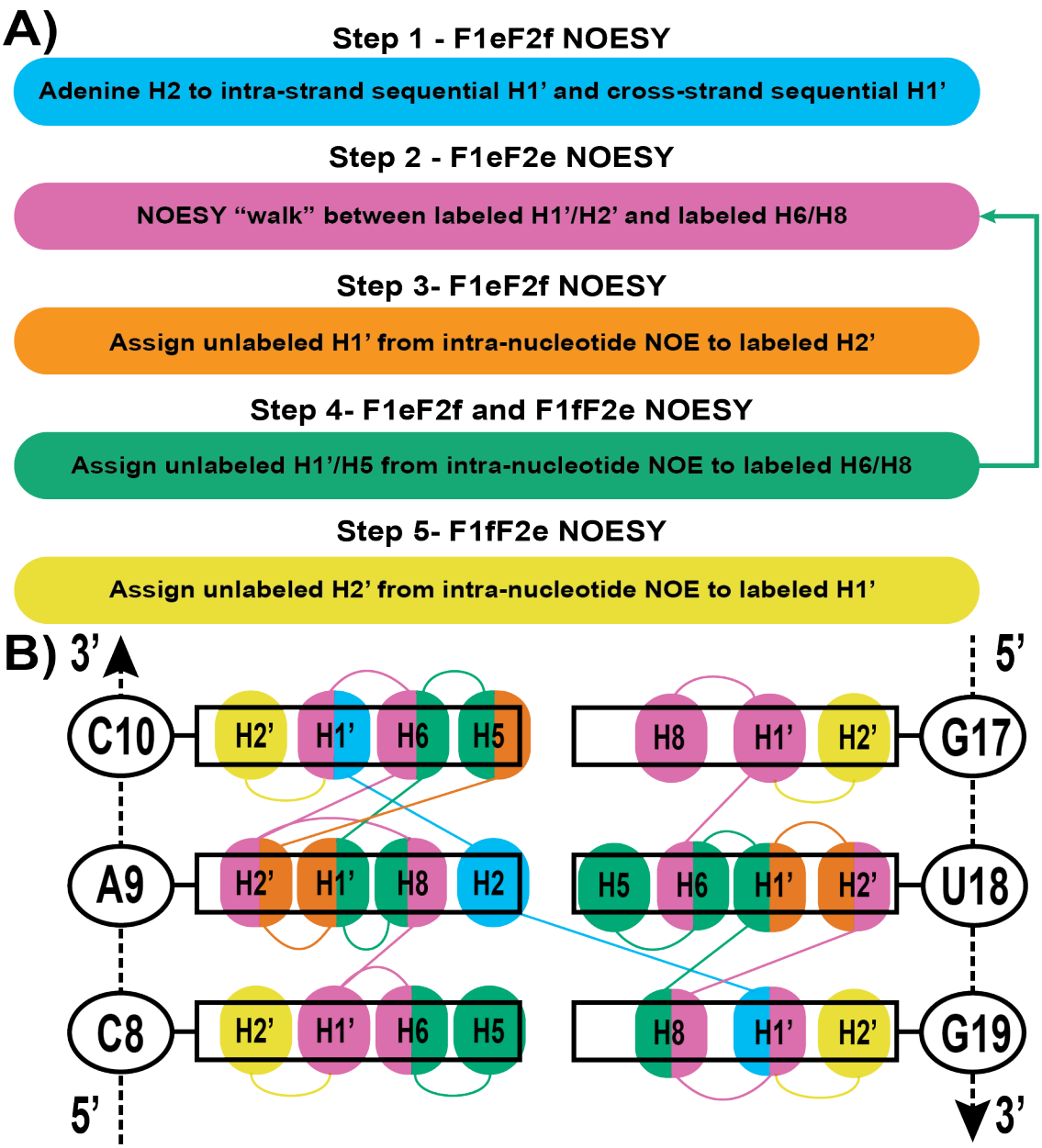

Figure 3.3. Resonance assignment schematic. A) The sequential NOESY "walk" strategy is given as a step-by-step process color coded to match the regions of the NOESY spectra in Figure 3.2. Step 4 allows for a cross-validation of the assignments from the sequential NOESY "walk" in step 2. B) A three base-pair region of the upper-stem of the 27 nt bacterial A-site RNA demonstrates the NOE cross-peaks expected from each of the NOESY regions in Figure 3.1. To simplify the NOE cross-peaks, only the non-exchangeable H1', H2', H2, H5, H6, and H8 atoms are labeled within this schematic figure and are color coded by the spectral regions from which they were assigned. Lines connect protons with expected NOEs (<5 Å in distance from each other) and are color coded by the NOESY experiment that provides the NOE cross-peak.

An unambiguous starting point for the NOESY “walk” is a critical step for resonance assignment and this is Step 1 of our logical assignment process. For this step, we analyzed the sky blue region from the F1eF2f NOESY experiment (Figure 3.4). The base region of the filtered dimension (6.5-8.5 ppm) reveals only cross-peaks arising from adenine H2 resonances since the remaining base proton resonances in this region are filtered by the removal of proton signals attached to ^{13}C labeled nuclei. The crowded “fingerprint” NOESY region is reduced to five cross-peaks between H2 (^{12}C-labeled) resonances and ^{13}C-labeled H1’ resonances (Figure 3.4C). By editing the second dimension for ^{13}C-labeled protons, only H1' resonances from guanine and cytidine are observed and both unlabeled H5 and H1' resonances are rejected. Thus, no intra-nucleotide NOEs are observed for adenine H2 to H1' since only the C2' position of the adenines are labeled with ^{13}C, and thus we do not see any cross-peaks for A20-H2. Even though an F1fF2f NOESY experiment is sufficient to identify these cross-peaks, it is not necessary. The F1eF2f NOESY provided adequate starting points for beginning our NOESY “walk”.

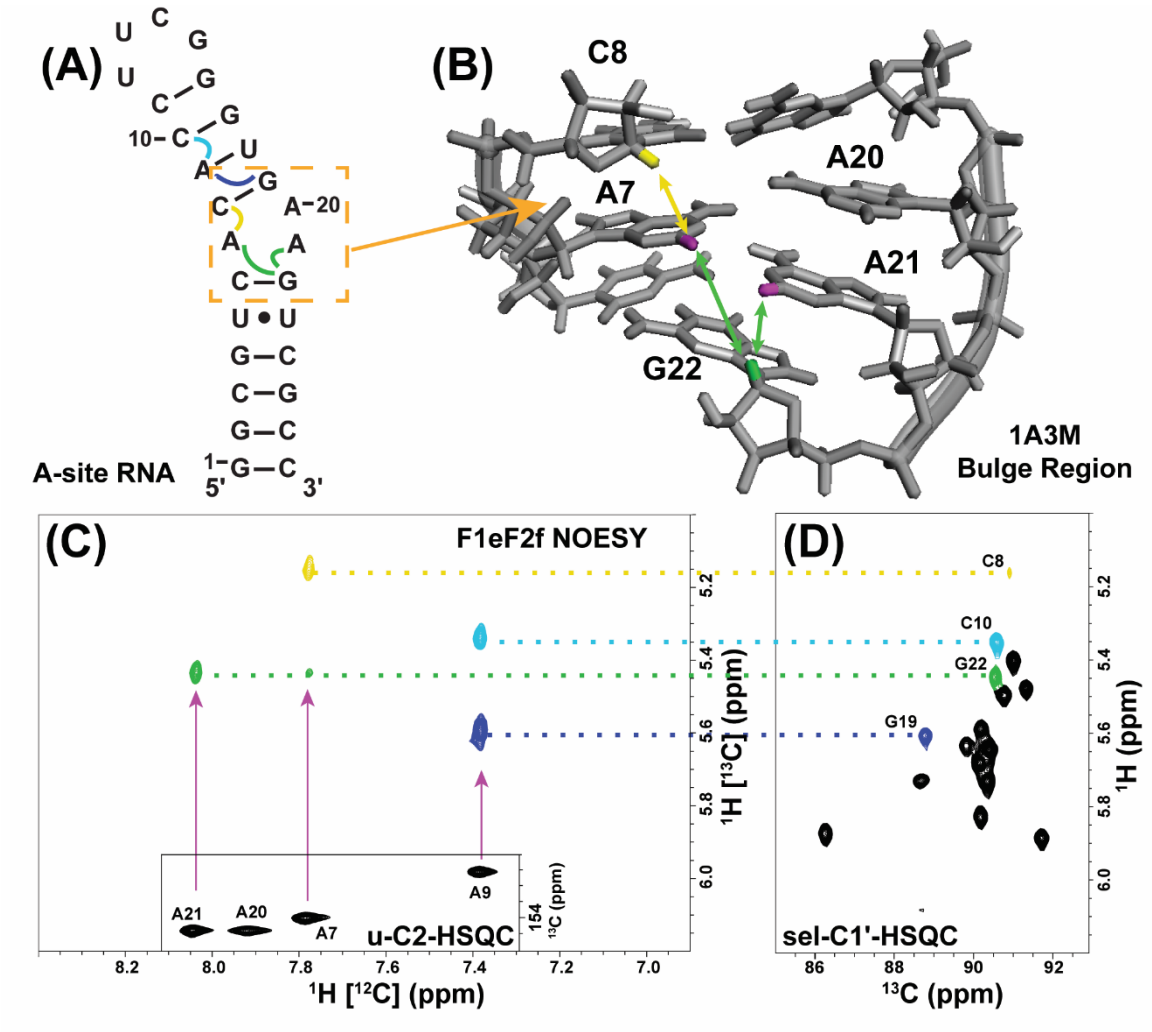

Figure 3.4. Identification of adenine H2 inter-nucleotide NOEs. (A) The secondary structure of A-site RNA with inter-nucleotide NOEs expected in the F1eF2f NOESY indicated with curly colored lines. The bulge region is highlighted with an orange dashed box. (B) The 3D structure of the bulge region with the A7-H2 and A9-H2 atoms (magenta) and their NOEs with C8-H1' (yellow) and G22-H1' (green) marked with colored lines and arrows. (C) The sky blue region of the F1eF2f NOESY spectrum is depicted with an inset of the C2-H2 HSQC region of the uniformly labeled A-site RNA. The cross-peaks are colored to correlate with their C1'-H1' partner in the C1'-H1' HSQC of the alternatively labeled A-site sample (sel-C1'-HSQC). (D) Labeled peak assignments are connected by dotted color-coded lines to the corresponding ones in (C).

In addition to a sequential inter-nucleotide NOE transfer between the adenine H2 (i) proton and the H1' proton of the next nucleotide (i+1) in the 5'→3'

direction, adenine H2 protons involved in base-pairing or stacking in A-helical regions, will also “talk” with H1' protons on the nucleotide proceeding (j+1) its cross-strand base-pair partner (j) in the 5'→3'. This unique cross-strand NOE is seen in Figure 3.4C for A7-H2 to G22-H1'. Both of these types of inter-nucleotide NOE interactions provide excellent starting points for resonance assignment as they are the only cross-peaks observed in the sky blue highlighted region of the F1eF2f NOESY spectrum in Figure 3.2. A7-H2 and A21-H2 share a cross-strand or same-strand NOE contact, respectively, with G22-H1'. While this is the only NOE observed in this region for A21-H2, as expected, A7-H2 has an additional same-strand NOE cross-peak with C8-H1' making the assignment of these resonances straightforward. The two cross-peaks to A9-H2 are ambiguously assigned as either G19-H1' or C10-H1' – its cross-strand and same-strand partners – in the F1eF2f NOESY spectrum. However, this ambiguity is removed when we compared the F1eF2e and F1eF2f NOESY spectral regions using the logical assignment process illustrated for representative cases in steps 2-4 (Figure 3.3). The assignment of adenine H2 inter-nucleotide NOE cross-peaks with sequential and cross-strand H1' resonances provides a powerful tool for RNA resonance assignment and our labeling scheme provides a means of simplifying the NOESY spectrum to allow for quick identification of these important cross-peaks.

We demonstrate our strategic NOESY “walk” with the sequentially stacked base pairs A9-C10. We chose residue C10 as a starting point because we had identified its H1' chemical shift using step 1 discussed above. The following

steps 2-5 are color coded to match the regions of the various NOESY spectra with corresponding information content. An overlay of the essential F1eF2f NOESY region (sky blue), which identifies the A9-H2 to sequential C10-H1' resonance, with the expanded "fingerprint" region of the F1eF2e NOESY region (reddish purple) illustrates this sequential walk (Figure 3.5). The starting point residue C10-H1' is identified as NOESY cross-peaks 2, 3, and 4 (reddish purple) and the strong cross-peak between A9-H2' and C10-H6 as peak 3 (Figure 3.5). This strong inter-nucleotide H2' to H6/H8 NOEs is a distinctive and diagnostic signature for resonance assignment because of the short (~2Å) distance between the atoms in stacked A-form helical regions of RNA[229,230].

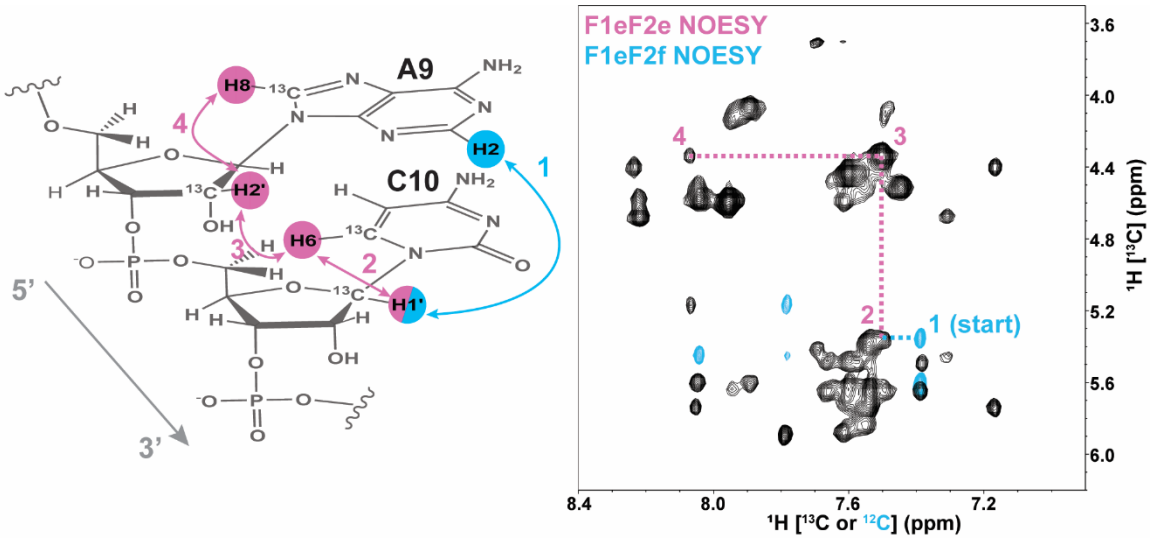

Figure 3.5. Steps 1-2 of logical resonance assignment strategy. The sequential NOESY "walk" strategy for step 1 and 2 is highlighted for the A9 and C10 stacked residues. All carbons are ^{12}C labeled unless otherwise identified. The F1eF2f NOESY spectrum (sky blue peaks) is overlaid with the reddish purple F1eF2e NOESY region (black peaks) and includes the peaks labeled for each NOE connectivity (color coded and numbered to match NOESY spectrum) expected in the sequential A9-C10 example.

To cross-validate the assignment of the ^{13}C-labeled protons from steps 1 and 2, we used filtering techniques to include cross-peaks from ^{12}C-labeled protons for steps 3 and 4 of the sequential assignment process (Figure 3.6). Starting with the previously assigned A9-H2', we could immediately identify the resonances of A9-H1' (peak 5) and C10-H5 (peak 6). In turn, the assignment of the ^{12}C-labeled A9-H1' resonance helped to identify the inter-nucleotide NOE involving C10-H6 (peak 7) and intra-nucleotide NOE with A9-H8 (peak 8). The strong intra-nucleotide NOE cross-peak between ^{13}C-labeled C10-H6 and ^{12}C-labeled C10-H5 (peak 9) served as an internal consistency check (magenta arrow) to confirm the resonance assignments from all the previous steps and the correlation between peaks 6 and 9. Peak 7 also correlated with the previously assigned peak 4 from Figure 3.4 and, therefore, served as an internal check. Thus, the different NOESY spectra contain two self-check points that allowed confirmation of the resonance assignment of the stacked residues A9 and C10.

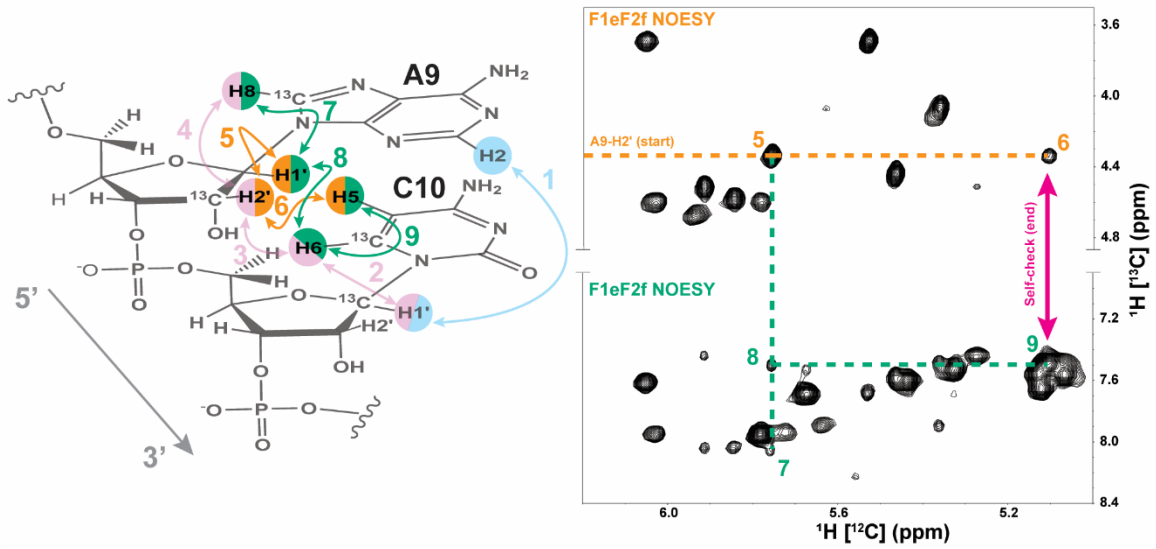

Figure 3.6. Steps 3-4 of logical resonance assignment strategy. The sequential NOESY "walk" for steps 3 and 4 are color coded and numbered to match the peaks in the F1eF2f NOESY spectrum regions. Previous assignment steps are faded within the A9-C10 example for visual reference. The A9-H2' starting point is labeled and follows the dotted orange line to peaks 5 and 6. Extension of the A9-H1' assignment is marked by the dotted bluish-green line to peaks 7 (A9-H8) and 8 (C10-H6). Peak 9 from the C10-H6 to C10-H5 serves as a self-check with peak 6 (magenta arrow).

In the final step 5 of the assignment process (Figure 3.7), the unlabeled (^{12}C) H2' resonance is unambiguously identified by a NOE contact to ^{13}C-labeled H1' cross-peaks in the yellow region of the F1fF2e NOESY spectrum. Starting from the assignment of the ^{13}C-labeled C10-H1' resonance, two strong intra-nucleotide cross-peaks are labeled as peak 10 and peak 11. The stronger NOESY cross-peak (peak 10) is identified as the C10-H2' resonance and peak 11 is assigned to C10-H3'. The yellow region of the F1fF2e NOESY contains cross-peaks from intra-nucleotide H3' and H2' resonances. Therefore, supplementing this experiment with HCCH-COSY correlations from a uniformly ^{13}C-labeled sample allows for confirmation of the H2' cross-peaks (Figure A.2).

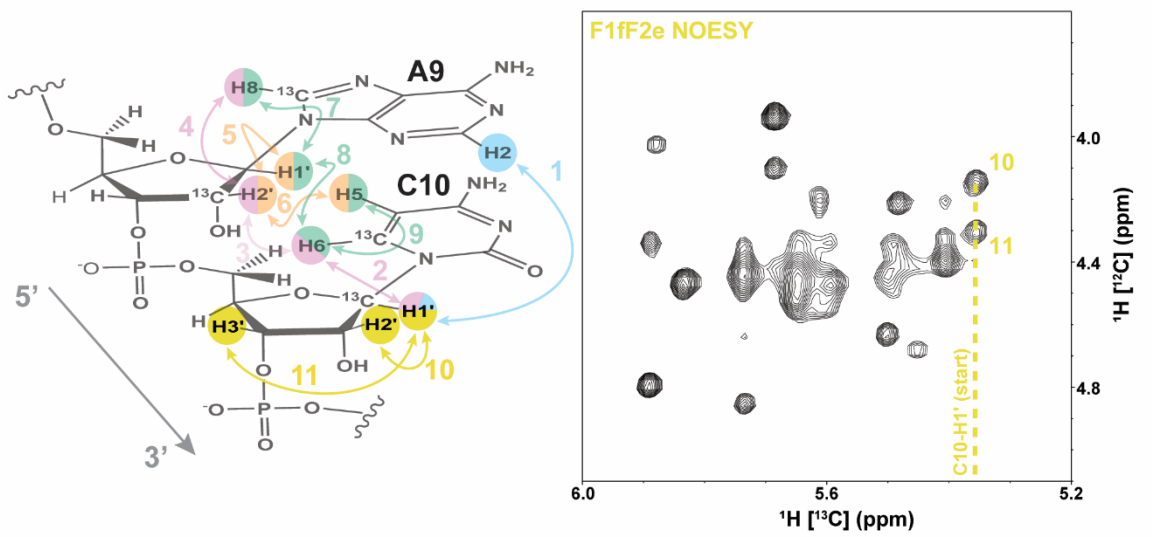

Figure 3.7. Step 5 of logical resonance assignment strategy. A schematic of the final assignment steps for unlabeled H2' and H3' from step 5 of the logical assignment process with yellow arrows and numbers. The previous assignment steps and protons are faded for visual reference. The F1fF2e NOESY (yellow region) with starting point (C10-H1') and cross-peaks to C10-H2' and C10-H3' are labeled 10 and 11, respectively.

Overall, the separation of peaks between ^{13}C-labeled (edited) and ^{12}C-labeled (filtered) regions of the NOESY spectra greatly reduced crowding, allowed logical assignment strategies, and prevented breaks in the sequential NOESY "walk" that is usually observed for selective labeling strategies. Working through Steps 1-5 of our proposed logical resonance assignment process, we were able to fully assign the non-exchangeable H1', H2', H2, H5, H6, and H8 resonances (Table 3.2). A complete overlay of steps 1-5 is provided in the Appendix to better illustrate the connectivity of the assignment process (Figure A.3). All ^{1}H resonances were referenced to DSS and ^{13}C resonances were indirectly referenced as previously described[249].

Table 3.1. Chemical shift assignment of bacterial A-site RNA. All assigned non-exchangeable resonances from the alternatively labeled bacterial A-site RNA as determined from step-by-step resonance assignment strategy. The box for each assignment in the table is color coded to correspond with the experiment necessary for initial assignment: F1fF2e (blue), F1eF2f (salmon), F1eF2e and HSQC (green). The A20-H2 resonance could not be assigned with the selectively labeled sample only and is highlighted in purple. The dashes correspond to resonances that do not exist. The blank boxes are for ^{12}C-labelled positions in the alternatively labeled sample.

Residue	H1'	H2'	H6/8	H2/5	C1'	C2'	C6/C8
G1	5.69	4.81	8.01	-	91.30		138.93
G2	5.79	4.41	7.49	-	92.88		136.30
C3	5.46	4.58	7.52	5.08	93.41		140.43
G4	5.60	4.45	7.34	-	93.05		136.03
U5	5.31	4.02	7.44	5.06		75.04	141.31
C6	5.55	4.43	7.85	5.59	92.88		141.92
A7	5.87	4.45	8.00	7.73		75.57	139.46
C8	5.12	4.14	7.40	5.23	93.58		140.51
A9	5.71	4.27	8.02	7.34		75.39	139.64
C10	5.31	4.10	7.46	5.06	93.23		140.69
C11	5.36	4.34	7.50	5.28	93.67		140.78
U12	5.48	3.64	7.64	5.63		75.74	140.51
U13	5.99	4.55	7.91	5.74		74.42	144.56
C14	5.84	3.98	7.57	6.01	88.92		142.71
G15	5.85	4.74	7.74	-	94.37		142.80
G16	4.36	4.47	8.19	-	92.79		138.76
G17	5.69	4.39	7.12	-	93.05		136.21
U18	5.42	4.39	7.51	5.02		75.21	140.77
G19	5.57	4.15	7.55	-	91.47		136.82
A20	5.80	4.52	7.99	7.87		75.92	140.60
A21	5.89	4.61	8.17	7.99		76.18	141.13
G22	5.41	4.63	7.26	-	93.23		137.35
U23	5.34	4.07	7.45	5.06		75.13	141.74
C24	5.59	4.45	7.89	5.67	92.44		141.89
G25	5.60	4.45	7.49	-	93.05		136.74
C26	5.44	4.15	7.61	5.08	94.02		141.22
C27	5.64	3.88	7.56	5.39	92.79		141.57

3.5 Confirmation of resonance assignment

3.5.1 Imino NOESY resonances.

As is good practice with any RNA system studied by NMR, we confirmed the secondary structure of our construct by analysing the imino proton resonances. For our RNA model system, there was no overlap in the imino resonances once an ideal temperature was selected from 1D experiments. All expected imino resonances could be assigned easily from a 2D ^{1}H-^{1}H NOESY experiment (Figure A.4). Our model RNA system did not require the use of ^{15}N-edited NOESY experiments to assign the imino resonances. However for more complex systems, automated imino assignment tools such as RNA-PAIRS can aid in assignment[224]. Indeed, RNA-PAIRS analysis of the ^{1}H-^{1}H NOESY and ^{1}H-^{15}N HSQC yielded a 100% probability for assignment of the upper stem, and our imino assignment results agree with previously published assignments of bacterial A-site RNA[250,251].

With the imino resonances assigned, we confirmed all our previous non-exchangeable resonance assignments through the spin-diffusion patterns observed within G-C base-pairs. In A-form helical regions of RNA, magnetization is easily transferred from G-H1 imino protons to C-H5 and to H1' of the sequential residue (5'→3') on same strand, as well as the cross-strand H1' of the sequential residue to the base-paired C (5'→3')[130]. All expected cross-peaks from this spin-diffusion pattern were observed for base-paired G-H1 resonances (Figure A.5) confirming the validity of our strategic non-exchangeable resonance assignment process.

3.5.2 Adenine H2.

To further confirm these assignments, a number of heteronuclear correlation experiments were recorded using a uniformly $^{13}C/^{15}N$-labeled bacterial A-site RNA sample. The H2 resonances for 3 of the 4 adenines were identified in a two-bond ^{15}N HSQC experiment, but the A20-H2 resonance was not visible due to line-broadening from chemical exchange as previously reported[251]. The two-bond ^{15}N HSQC experiment was combined with HNN-COSY and ^{1}H-^{1}H NOESY (Figure 3.8) experiments to confirm the resonance assignment of A9-H2. As expected, we observed a strong cross-peak between U18-H3 and A9-H2 in the ^{1}H-^{1}H NOESY spectrum. With the through-bond HNN-COSY experiment we could correlate the H3 of U18 with the hydrogen bond donor, U18-N3, and acceptor, A9-N1. With the 2-bond ^{15}N HSQC, we could correlate the N1 and N3 chemical shifts of A9 with its H2 resonance, further confirming the assignment of A9-H2.

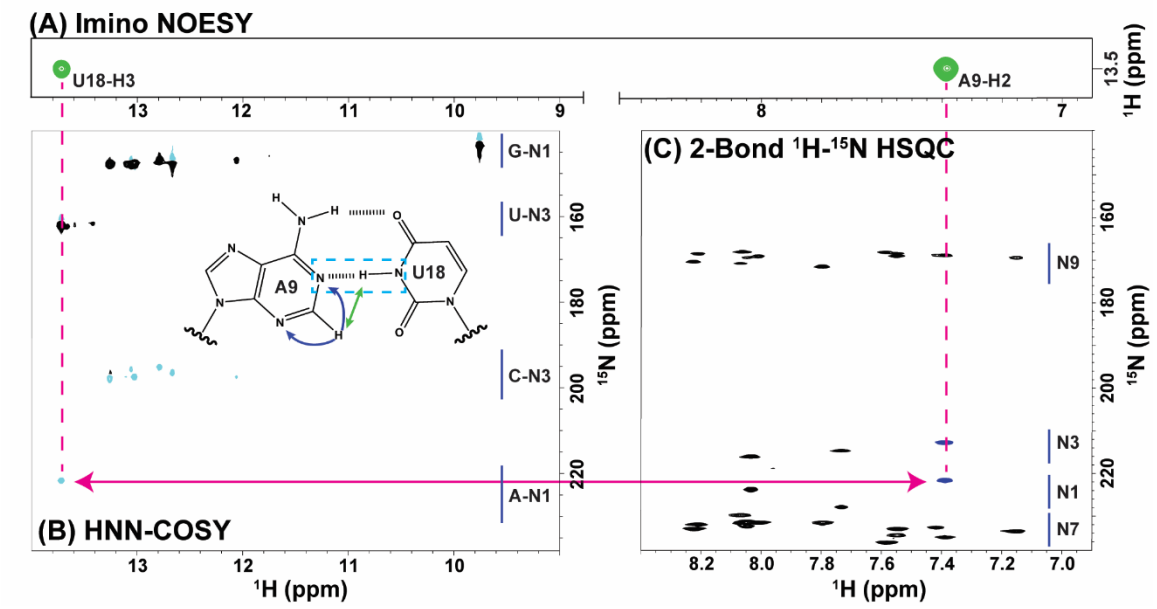

Figure 3.8. Confirmation of A9-H2 resonance. A) Imino NOESY spectra showing the through-space crosspeak between A9-H2 and U18-H3 (green peaks) as shown in the A-U base-pair cartoon (green arrow). B) The HNN-COSY hydrogen bond experiment correlates U18-N3 to A9-N1 through U18-H3 and is highlighted in the A-U base-pair by the blue dashed box. C) The 2-bond ^{1}H-^{15}H-HSQC correlates A9-H2 to A9-N3 and A9-N1 (purple arrows in A-U base-pair) and correlates with the A9-N1 resonance from the HNN-COSY experiment (magenta arrow).

For our model system there was minimal chemical shift overlap, but for larger RNAs there may be situations where three-dimensional experiments are needed to confirm base-pairing. Cases of overlap can be resolved by separating either H2 and C2 resonances or ^{1}H-^{1}H NOESY cross-peaks into a third dimension (16, 17). After assignment of adenine H2 resonances, through-bond TROSY relayed HCCH-COSY experiments can also be used to correlate H2 with H8 resonances for medium to small RNAs[253]. A9-H2 was further confirmed with the 2-bond ^{15}N HSQC identifying the N1 and N3 chemical shifts of A9. However this step is not necessary with our strategy for assignment.

3.5.3 *Pyrimidine intra-nucleotide NOEs.*

As a further verification of our assignment, we used the J_{CC} coupling splitting of the pyrimidine C6 resonances to readily identify these resonances in the HSQC spectra of the uniformly $^{13}C/^{15}N$-labeled bacterial A-site RNA sample. We ran a HCN correlation experiment on the alternatively labeled bacterial A-site RNA sample to further confirm previous assignment of the H6 resonances as belonging to either uridine or cytidine (Figure 3.9). Based on previously published and well known data from the 14 nt UUCG tetraloop RNA, we could confirm the resonance of C14[216,254]. The resonance for C24 was absent in the HCN correlation spectra most likely due to line-broadening from chemical exchange which matches previous reports[251]. The uridine C6 resonances were confirmed from the lack of corresponding resonances at C1' in the HCN spectrum and from the chemical shift of uridine N1, which is upfield of the cytidine N1 resonances. The H1' to H6 correlation of the cytidines were used to confirm intra-nucleotide cytidine H1' to H6 cross-peaks in step 2 of the logical resonance assignment process.

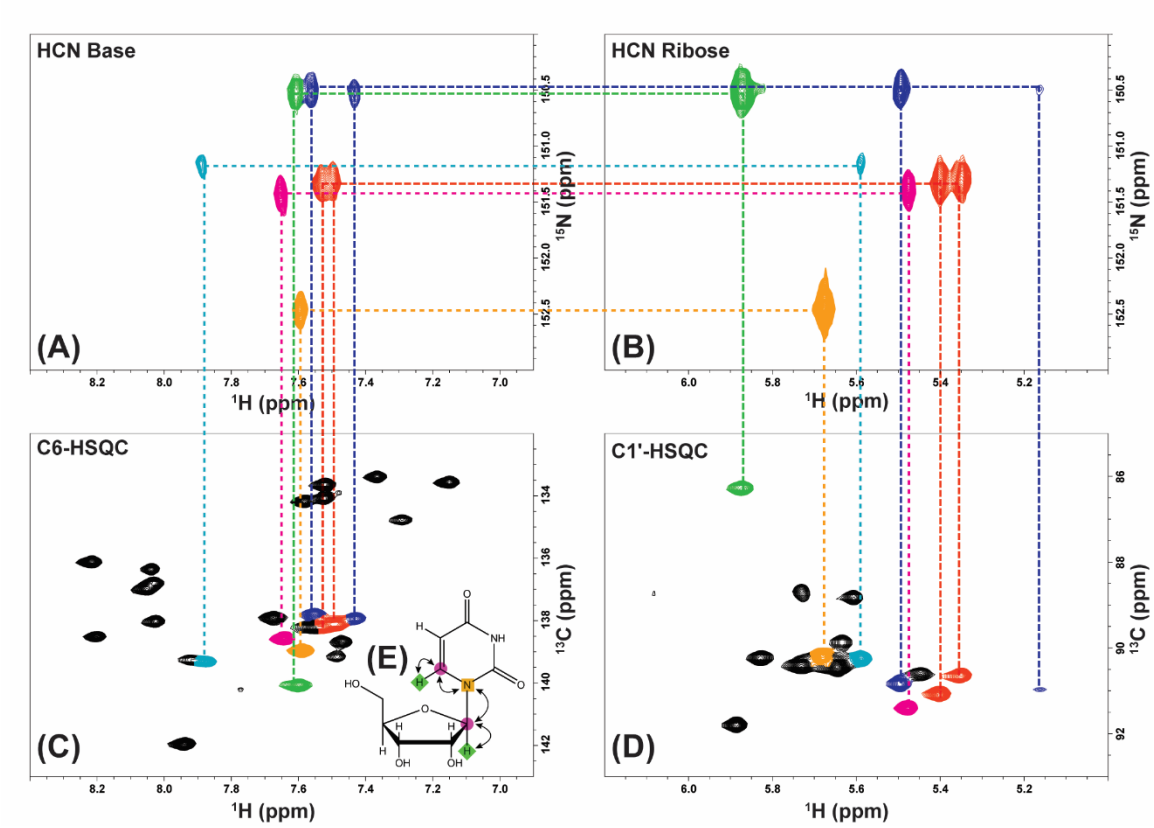

Figure 3.9. Confirmation of pyrimidine intra-nucleotide H1' and H6 resonances. The H6 and H1' intra-nucleotide pairs for cytidine were confirmed with a base HCN (A) and ribose HCN (B) along a HSQC of the base (C) and ribose (D). The transfer pathway was limited to the desired HCN transfer (E) by eliminating C-C coupling at C1' and C6 greatly enhancing the sensitivity of the experiment. The "connected" pairs were color coded for ease of viewing with similar colors indicating ambiguity due to overlap in the ^{15}N dimension as seen for purple and red peaks. This ambiguity was previously solved with the F1eF2e NOESY walk.

3.5.4 Unlabeled H2' and H5 resonances.

Finally, the unlabelled protons for H5 and H1' are easily confirmed with through-bond HCCH-COSY experiments on the uniformly ^{13}C/^{15}N-labeled bacterial A-site RNA sample. The HCCH-COSY is a fairly sensitive experiment applicable to larger RNAs and can aid in the assignment of more complex systems. By comparing the F1fF2e NOESY spectral (yellow) region with the

HCCH-COSY spectra of the uniformly labeled sample, we could determine that all the strong cross-peaks observed in this region were from intra-nucleotide H1' to H2' NOEs (Figure A.2).

3.6 Minimal NOESY model of bacterial A-site RNA

Once RNA resonances are fully assigned, solving three-dimensional structures has often been accomplished through a deterministic approach. Multiple NMR measurements such as distance restraints from NOE, dihedral angles from J-coupling or cross-correlation, and dipolar couplings from RDCs are used to obtain an energy minimized structure using a force field driven annealing protocol[216,255]. This process often leads to an ensemble of structures within 1-3 Å RMSD. Simulation of a large pool of RNA structures by programs such as Constructor, MC-SYM, 3DNA, etc. have the potential to quickly provide 3D models from minimal experimental data. Here we apply the NOE data we obtained from our rapid assignment strategy to determine if we could recapture a representative 3D model of bacterial A-site RNA

3.6.1 MC-SYM model generation

The first step in creating synthetic 3D models of RNA using MC-SYM is to obtain a representative secondary structure[256]. The primary sequence for bacterial A-site RNA was submitted to MC-Fold and resulted in 20 secondary structures with free energy folding values between -28.33 and -24.59 kcal/mol. The top 10 secondary structures are shown in Figure 3.10. Models 1-6 represent the flipping in and out of A20 and A21 as well as melting/base-pairing of the G15-U12 closing base-pair of the UUCG-tetraloop. Models 7-9 show opening of the

lower stem with Model 9 representing the formation of new base-pairs including a non-canonical A-C between A21 and C6 as well as a G-U wobble between G22 and U5. Model 10 is similar to Model 2 and 4 except with the first canonical base-pair of the upper stem between C8 and G19 are no longer Watson-Crick paired.

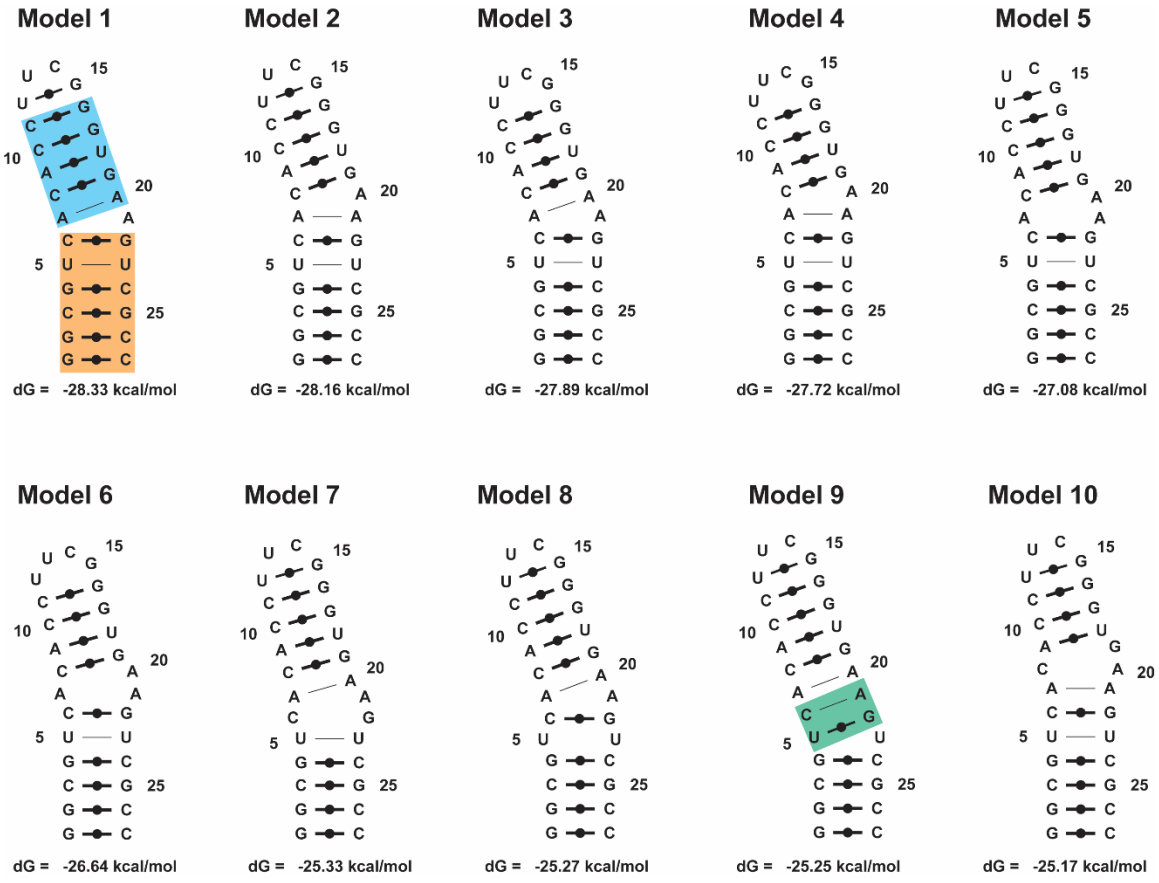

Figure 3.10. MC-Fold predicted secondary structures of bacterial A-site RNA. The top 10 secondary structures for bacterial A-site RNA as predicted by MC-Fold are shown. The free energy score for folding is given for each conformation. Non-canonical base-pairs are marked with thin lines and canonical base-pairs with a circle and solid line. The upper (sky blue) and lower (orange) stems of the RNA are labeled in Model 1. A new base-pair pattern is highlighted in Model 9 (bluish green).

Model 1 most closely resembles the published NMR solution structure (1A3M) of the 27 nt bacterial A-site RNA construct, so we submitted the secondary sequence to MC-SYM in order to generate 10,000 synthetic 3D models of the RNA. We refined the structures using previously established minimization protocols to remove steric clashes and refine the backbone angles of the 3D models[257].

3.6.2 NOE refinement of structures

Nuclear Overhauser effect spectroscopy (NOESY) provides distance restraint information based on the fact that signal intensity of the cross-peak arising from two protons close in space is inversely proportional to the sixth power of their distance as given in Equation 4.1, where r is the distance between protons in Å.

$$NOE \propto \frac{1}{r^6} \qquad (3.1)$$

It is common practice to measure NOESY cross-peaks at varying mixing times to help determine weak, medium, and strong cross-peaks. These peak intensities are then given a range of distance restraints with strong NOE cross-peaks resulting in short distances (1.8-3 Å) and weak NOE cross-peaks resulting in longer distances (4-6 Å). We binned our peak integrals to larger than normal ranges since we used long mixing times and the editing module removed H5-H6 cross-peaks which are also typically used to calibrate peak intensities and therefore distance restraints (Table A.2).

The pool of 10,000 MC-SYM structures were scored against the NOE distance restraints we obtained from cross-peaks assigned in the F1fF2e and F1eF2e NOESY spectra. In total we used 134 NOE restraints to score the synthetic structure pool (Table A.2). The top 100 structures were further refined and minimized using MC-SYM[257].

3.6.3 Analysis of best structures

The best reported structure based on NOE distance restraints in shown in Figure 3.11 overlaid with the NMR solution structure. The RMSD compared to the NMR structure is 1.82 Å. From only 134 NOE restraints, MC-SYM was able to select out a model that resembled the NMR solution structure. By starting from a pre-defined secondary structure, we biased the model selection by limiting 3D structures to hydrogen bonding patterns predicted from our imino NOESY data and the published NMR solution structure of bacterial A-site RNA.

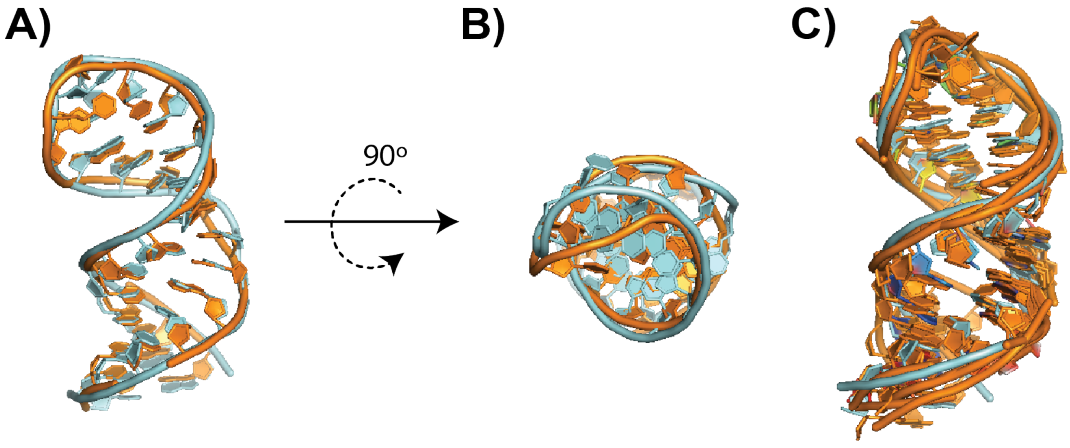

Figure 3.11. NOE refined MC-SYM models of bacterial A-site RNA. A) The best fit structure is shown (cyan) overlaid on the NMR solution structure (orange). B) A top view of the same overlay from part A. C) An overlay of the top 5 structures (orange) based on 134 NOE restraints as determined by MC-SYM with the NMR solution structure (cyan).

Two deviations from the solution structure were noticed. The backbone of the UUCG tetraloop between U13 and C14 did not overlay well with the NMR structure. This deviation led to a large shift in position for pyrimidine base of both U13 and C14. The flexibility of these residues, however, may cause fluctuations in the backbone that neither still image captures (we will address these issues of dynamics in Chapter 4). The other deviation is caused from breaks in the backbone connectivity for the MC-SYM structure. Since our work, MC-SYM has been updated to resolve errors linking sequential residues.

Overall, the minimal NOE restraints obtained from our assignment methodology was used to construct a synthetic 3D model for bacterial A-site RNA that agreed well with the NMR solution structure. NOE restraints alone, however, are indicative of local distances and can induce errors over long ranges in RNA molecules. We have not supplied dihedral angle or long range restraints in our selection process, so although we obtain a similar model to the NMR solution structure, we cannot confirm with certainty that this is indeed the correct structure for our RNA construct. Surprisingly, the helical bends and turns of our predicted structures match the NMR structure well. Using MC-SYMs we successfully recaptured a representative 3D structure of bacterial A-site RNA that matches well with previously published models in less than 3 days of NMR instrument and computational time.

3.7 Extension to larger RNA

Our model 27 nt bacterial A-site RNA system proved to be highly amenable to the alternative labelling scheme we devised. It could be argued that this method would be more challenging for larger RNA systems, so we tested the application of the F1eF2e NOESY experiment on a 62 nt viral RNA, the hepatitis B virus epsilon pre-genomic RNA (HBVε). Previous work by Wijmenga reported the resonance assignment of the apical stem loop and primer loop for a modular duck HBVε RNA[258,259]. A divide-and-conquer approach as well as selective deuteration was required because of chemical shift overlap and line-broadening issues[260,261].

We were able to greatly reduce the spectral crowding of a 2D ^{1}H-^{1}H NOESY by implementing the F1eF2e NOESY experiment on a 1',2',8-^{13}C-G labelled HBVε transcript (Figure 3.10). By labelling both the 1' and 2' carbons of GTP and no other ribose positions, we removed all the intra-residue signals from 3', 4', 5', and 5'' protons when we applied the F1eF2e NOESY experiment. Only labeling the G-residues also limited signal crowding arising from A, C, and U-residues. Working from the assigned resonances from the modular duck HBVε RNAs and using RAMSEY[142] and LarmorD[262,263] predicted chemical shifts, we were able to confirm two sequential GGG NOESY “walks” between residues G40-G42 and G51-G54 of the HBVε construct.

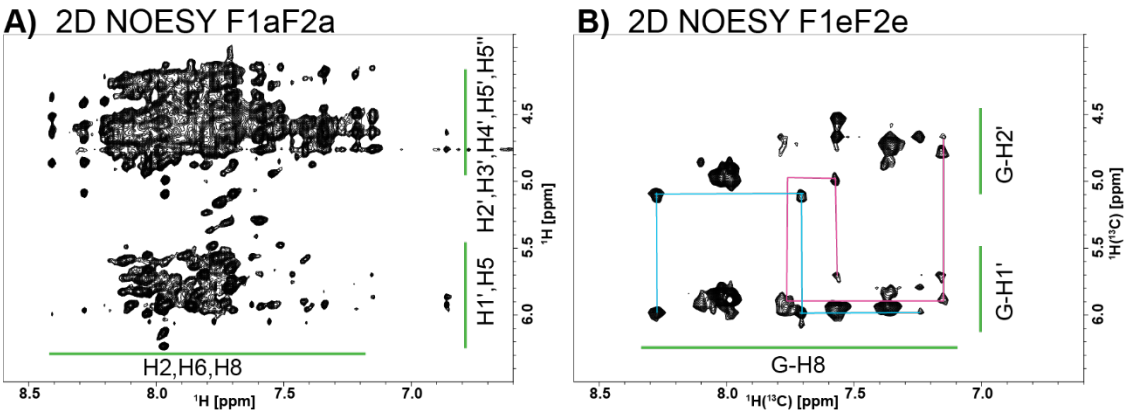

Figure 3.12. HBVε NOESY spectrum with and without ^{13}C editing. A) A full unedited ^{1}H-^{1}H NOESY of the extended "fingerprint" region of 62 nt representative HBVε RNA. B) F1eF2e NOESY showing only intra- and inter-residue NOEs between H1', H2', and H8 of G residues. Two NOESY "walks" are shown in blue and purple between 3 sequentially stacked G residues.

3.8 Conclusion

The identification of the adenine H2 resonances is of great importance when assigning NOESY peaks as they provide unique same-strand and cross-strand inter-nucleotide connectivities to H1'[229]. A similar strategy has been adopted for large RNAs in which multiple samples with and without deuteration at the H2 position of adenine are used to determine cross-peaks arising from H2 to H1' through-space transfers[107,108]. Through deuteration, this strategy provides tools for assigning large RNAs by eliminating extraneous NOESY transfer pathways, improving signal/noise, and reducing spectral crowding and complexity. Drawbacks of this approach include cost of nucleotides and the need for 6 or more deuterated samples. For large RNAs (>25kDa), this deuteration strategy may be the only feasible method for NMR resonance assignment. For small to medium RNAs (<25kDa), we recommend our new isotopic labelling strategy as a means to reduce overall resonance assignment

time. Nonetheless, combing site-specific deuteration with our selective ^{13}C-labeling might prove useful for larger RNAs.

Although complete assignment of the H1', H2', H2, H5, H6, and H8 resonances was achieved from one NMR sample, in some cases it may be necessary to confirm these assignments with heteronuclear through-bond experiments on a uniformly ^{13}C/^{15}N-labeled sample. In our work, we used imino ^{1}H-^{1}H NOESY, HNN-COSY, one- and two-bond ^{15}N HSQC, ^{13}C HSQC, HCN, and HCCH-COSY experiments to confirm all the initial non-exchangeable resonance assignments. For more complex RNAs an orthogonal sample may prove beneficial, if the labelling patterns of the ribose are switched such that GC is ^{13}C-labeled at C2' and AU is ^{13}C-labeled at C1' we could expect to achieve a NOESY "walk" complimentary to the strategy suggested in this chapter (Figure A.6). A combination of this selective labeling strategy with segmental labeling[264–268] and deuteration[106,211,269,270] will extend the usefulness of this technique to very large RNA systems. In future work, we aim to synthesize deoxyribonucleotides from our ribonucleotides using ribonucleotide reductase[271] in order to apply a similar resonance assignment strategy to DNA.

The benefit of complete resonance assignment from a minimal set of isotopically enriched RNA samples is highly desirable. It reduces both the cost and time associated with sample preparation and analysis. In terms of cost savings, this new assignment method could save upwards of 3 weeks of instrumentation time, which is billed anywhere from $10-40/hr, for a total savings of $5,000-20,000. When coupled with the savings for isotopic enrichment from

chapter 2, we think these two strategies will open the door for RNA research by NMR to new scientists and institutions.

Our unique labeling scheme also provides isolated ^{13}C-^{1}H spin systems probes, and these are ideal for measuring relaxation dispersion, RDC, and RCSA by removing ^{13}C-^{13}C coupling and reducing spectral crowding. Reasonable 3D structures can be refined from a pool of MC-SYM generated structures using the NOE restraints obtained from our rapid assignment strategy, which better guide structural biologists when designing experiments. The rapid assignment principles were also successfully extended to a 62 nt viral RNA construct.

Even though we limit the number of probe sites available compared with uniformly labelled samples, we think the utility of this labelling scheme for assignment, dynamics, and structural refinement by NMR will allow more researchers access to atomic resolution studies of RNA. We will address the usefulness of this labelling scheme for simplifying the study of RNA dynamics and structure in Chapter 4.

4 Dynamics and structure of RNA by NMR

Because of its polar nature, RNA is highly solvated and displacement of water and counterions requires extra energy compared with hydrophobic binding pockets that can be found in proteins. This could explain why RNA maintains a highly dynamic nature so that its intrinsic disorder allows it to create multiple binding pockets within the same structural region. Ligands typically interact with the polar surface and hydrogen bond donors and acceptors of unpaired RNA bases found in bulge and loop regions of RNA. For this reason, structural and dynamic information about the unpaired regions of RNA are of critical importance to drug discovery and functional analysis. It is here that NMR shows its true strength in quantifying molecular motions[82,272–278].

In Chapter 2, we made the case that isotopic enrichment increases NMR sensitivity by increasing the concentration (and, therefore, the population distribution of α and β spin states) of magnetically active nuclei. In practice, the sensitivity of the NMR signal is also affected by the exponential decay of the the bulk spin magnetization back to equilibrium. Therefore, the intensity of the NMR resonance signal is related to both the difference in population between spin-state energy levels and the rate at which bulk spin magnetization returns to equilibrium. Two major mechanisms of NMR relaxation are longitudinal (R_1) and transverse (R_2) relaxation.

The longitudinal relaxation rate, R_1, governs the rate at which the z-component of the nuclear spin magnetization vector (parallel with the external magnetic field, B_0) decays back to its thermal equilibrium value (M^0_z) and is

typically an order of magnitude slower than transverse relaxation. The transverse relaxation rate, R_2, is the rate by which the nuclear spin magnetization vector (perpendicular to B_0) M_x and M_y decays or dephases to zero (equilibrium). The main sources of relaxation for R_1 and R_2 are modulations of the dipole-dipole interactions and chemical-shift anisotropy (CSA) by Brownian rotational tumbling. The slow tumbling of macromolecules dominates the relaxation of rigid bond-vectors and therefore reports on the overall tumbling and diffusion tensor of the macromolecule.

Relaxation requires a time-dependent fluctuation of the magnetic field at instantaneous precessional frequencies (the Larmor frequency) of each spin. The time-dependence of these fluctuations originates from motions in the macromolecule such as vibrations, rotation, and diffusion. The tumbling of a molecule in solution can be characterized by a rotational correlation time (τ_c) – the time it takes to rotate 1 radian or (180°/π). The Fourier transform of this correlation function yields the spectral density function that describes the probability of finding these motions at a given angular frequency (ω) (Figure 4.1). A common representation of this spectral density function is given by the Lipari-Szabo "model free" formalism[279]:

$$J(\omega) = \left(\frac{2}{5}\right)\left(S^2\frac{\tau_c}{(1+\omega^2\tau_c^2} + (1-S^2)\frac{\tau}{(1+\omega^2\tau^2}\right) \qquad (4.1)$$

$$\frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_e} \qquad (4.2)$$

Where τ_c is the correlation time of the molecule, τ_e is the fast internal motion of the bond vector usually in ps, ω is the angular frequency in rad/s, and S^2 is the

order parameter to describe the extent of internal motion where 1 is rigid and 0 is completely flexible.

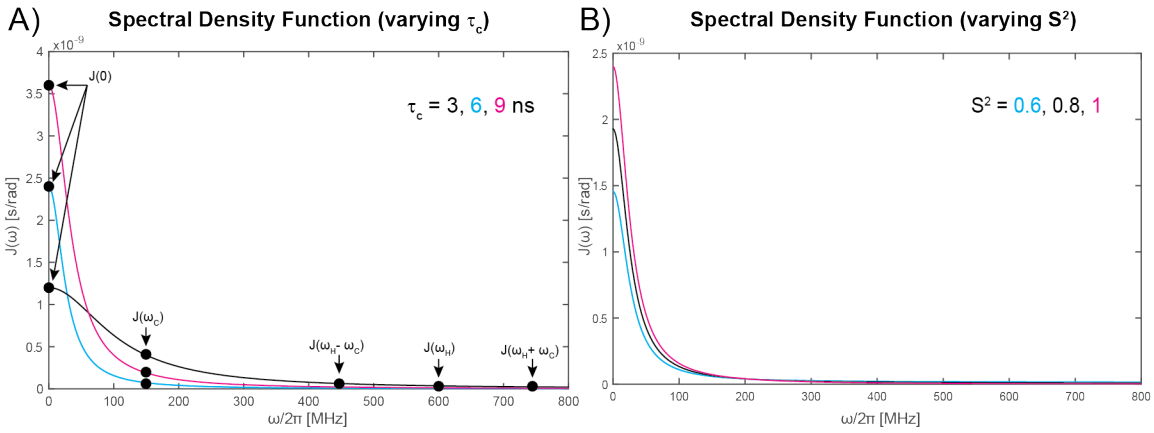

Figure 4.1. Spectral density function for isotropic rigid rotors. A) Simulation of the spectral density function assuming a completely rigid bond vector under isotropic rotation ($S^2 = 1$) for correlation times of 3, 6, and 9 ns. B) A simulation of the spectral density function for a 6 ns isotropically rotating bond vector undergoing fast internal motions ($\tau_e = 50\ ps$) with an order parameter of 0.6, 0.8, and 1.

As seen in Figure 4.1, the values of J(0) are dependent on both the overall correlation time and the order parameter. Each of the measurable relaxation rates in NMR experiments are dependent on the spectral density function at various angular frequencies. The take-home here is that relaxation rates describe molecular motions both globally through tumbling (correlation time) and locally through the order parameter. These relaxation rates, however, can be complicated by additional mechanisms such as ^{13}C-^{13}C scalar coupling and chemical exchange. Here we will describe our efforts at simplifying the study of

RNA dynamics through NMR relaxation experiments by addressing both of these challenges.

4.1 Scalar coupling effects on observed relaxation rates

Observed NMR signal (transverse magnetization) decays at a rate, R_2, that increases with the size of the macromolecule[131,280]. For ^{1}H spins, direct observation of the FID signal can be used to obtain an approximation of R_2. For hetero-atoms, R_1 and R_2 data are recorded as a series of 2D spectra in which a relaxation delay (τ) is varied between each recorded spectra. The peak intensities are measured for each 2D spectra corresponding the relaxation delay of the spectra. The data are subjected to nonlinear least squares analysis of a single exponential model to extract the decay rate.

Contributions from chemical exchange phenomenon add to the observed R_2. In order to quench this exchange contribution, an off-resonance spin-lock experiment is used to measure the so-called $R_{1\rho}$ rate[281–284]. By applying a strong B_1 field off-resonance, a high effective spin-lock field, ω_{eff}, is achieved quenching most exchange effects:

$$\omega_{eff} = \sqrt{{\omega_1}^2 + \Omega^2} \qquad (4.3)$$

For equation (4.3), ω_1 is the strength of the B_1 spin-lock field (Hz) and Ω is the spin-lock offset from the measured resonance's chemical shift in Hz. Since the magnetization is tilted along the effective field, there will be components of both R_1 and R_2 relaxation during the spin-lock period. In order to extract the

transverse relaxation, R_2, rate from the $R_{1\rho}$ measurement, we must use the following relationships:

$$R_{1\rho} = R_1 cos^2\theta + (R_2 + R_{ex}) sin^2\theta \quad (4.4)$$

$$\theta = tan^{-1}\left(\frac{\omega_1}{\Omega}\right) \quad (4.5)$$

Assuming that the high effective spin-lock field quenches R_{ex}, the relationship in equation (4.4) can be used to directly determine R_2. Using previously published pulse programs for decoupling ^{13}C-^{13}C scalar coupling during t_1 chemical shift evolution[131], we compared the relaxation rates obtained for a uniformly ^{13}C-labeled bacterial A-site RNA and the site-selectively ^{13}C-labeled bacterial A-site RNA sample used in Chapter 3 for our resonance assignment strategy (Figure 4.2).

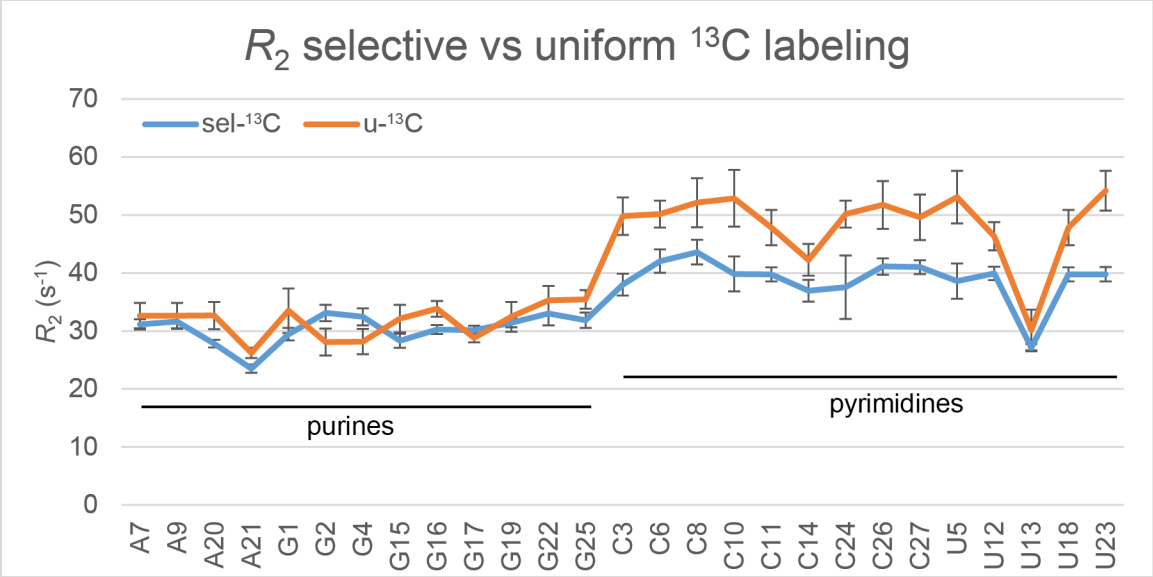

Figure 4.2. Relaxation measurements for R_2 on selective and uniformly labeled A-site RNA samples. The transverse relaxation rate (R_2) was measured for our site-selective alternatively ^{13}C labeled A-site sample (blue) and for a uniformly ^{13}C-labeled A-site sample (orange) using previously reported pulse sequences for $R_{1\rho}$ off-resonance measurements[131,283]. $R_{1\rho}$ rates were

measured at 3.3 kHz B_1 field strength at 3 kHz off-resonance. The average deviation for purines was 3 ± 1 s^{-1} and 10 ± 3 s^{-1} for pyrimidines.

We see a rather large deviation in measured transverse relaxation rates between ^{13}C selectively labeled RNA (isolated spin-system) and ^{13}C-uniformly labeled RNA (^{13}C-^{13}C scalar coupling active). The difference was greatest for pyrimidines with an average overestimation of R_2 around 24 ± 9 %. The difference for the purines was much smaller, around 8 ± 5 %. The enhanced relaxation rate for uniformly ^{13}C-labeled samples will influence the sensitivity and resolution of NMR experiments. Not only does the scalar coupling affect sensitivity, but it greatly complicates the extraction of meaningful relaxation parameters such as chemical exchange from the measured values[151,158,285–288]. With this in mind, we sought to optimize the experimental selection for our isolated ^{13}C-labeling scheme.

By removing ^{13}C-^{13}C scalar coupling effects on relaxation, we could better determine the efficiency of transverse relaxation optimized spectroscopy (TROSY), heteronuclear multi-quantum coherence (HMQC), and heteronuclear single-quantum coherence (HSQC) type NMR experiments. The precision of the calculated dynamics parameters depends critically on the quality of the primary relaxation decay, therefore optimal sensitivity and resolution are essential for thorough analysis. Empirically optimizing the selection of pulse programs will greatly aid researchers in improving the quality of NMR spectrum and data analysis.

4.2 Simulations of relaxation during NMR pulse sequences

NMR peaks are best described as Lorentzian curves having much longer tails than a Gaussian. A Lorentzian function is described by the following equation:

$$y = y_0 + \frac{2A}{\pi} * \frac{w}{4(x - x_c)^2 + w^2} \qquad (4.6)$$

In equation (4.6), y_0 is the baseline, A is the area under the curve, w is the full width at half-height (FWHH), x_c is the center of the curve. For NMR purposes, we translate these mathematical variables into physical meanings. The baseline is equivalent to the noise in the spectrum. The area under the curve represents the integral of the curve which is determined by the total number of spins associated with the NMR peak; it is common practice to integrate NMR peaks to determine the ratio of protons for each associated NMR peak. The FWHH is directly related to the signal decay by transverse relaxation properties.

The FWHH for NMR peaks determined by the total T_2 of the observed spin-system yields the following relationship:

$$\mathrm{w} = \frac{1}{\pi \mathrm{T}_2} \qquad (4.7)$$

Because the maximum peak height is found at the center of the peak ($x = x_c$), we can calculate the expected peak height (y_c) from information about the area (# of spins) and FWHH (transverse relaxation properties) by simplifying (4.6):

$$y_c = y_0 + \frac{2A}{\pi w} \qquad (4.8)$$

Combining equations (4.7) and (4.8) allows us to calculate the peak height based on the expected area and transverse relaxation:

$$y_c = y_0 + 2A\mathrm{T}_2 \qquad (4.9)$$

From equation (4.9), we see the peak height is directly proportionate to the transverse relaxation time, T_2. Assuming that the baseline is negligible and the area is equal for the spin systems we are comparing, we can use the effective transverse relaxation rate within various NMR experiments to determine the expected sensitivity and resolution.

By encoding the chemical shifts of coupled spins during an indirect chemical shift evolution (t_1) period, heteronuclear NMR has expanded the usefulness of NMR spectroscopy for macromolecules[170,289,255]. A 2D spectrum is often used to correlate proton chemical shifts with their coupled spin partner (through-space NOE partner or through-bond J-coupled nuclei, for example). As we have shown from equation (4.9), the sensitivity and resolution of the NMR spectrum for these experiments is governed by the transverse relaxation rate of the nuclei measured in the indirect dimension. For this reason, we found it prudent to characterize the relaxation rates of ^{13}C spin states classically measured during t_1 evolution periods. We simulated the effects of relaxation during the chemical shift evolution period for ^{13}C-^{1}H isolated spin-systems using known RNA parameters such as bond length and chemical shielding anisotropy (CSA) values. Although there is no consensus for RNA CSA values, we used an

average of all reported values (160 ppm for base and 23 ppm for ribose C-H spin systems was used for our calculations)[272,290–294].

The Lipari-Szabo model-free treatment is used to determine the spectral density function frequencies[279]. Here τ_c is estimated on the basis of a spherical isotropic tumbling protein. Although this estimate is not that best for RNA, which tends to be elongated and anisotropic, it should give us a starting point to simulate the expected relaxation rates for various size RNAs. Every 2.6 kDa of molecular weight for a protein is equivalent to 1 ns τ_c based on the approximation for correlation time, $\tau_c = \frac{4\pi\eta a^3}{3kT}$, where a is the radius of the protein, k is the boltzman constant, T is the temperature, and η is the viscosity of the solution. Each nucleotide in RNA is approximately 330 daltons. So using the estimate for proteins, every 2.6 kDa is equivalent to a τ_c of 1 ns, and every 7.88 nucleotides corresponds to a correlation time of 1 ns. This is an extreme underestimate corresponding the UUCG tetraloop has a correlation time of approximately 2.4 ns, but would be less than 2 ns based on the assumption for protein correlation time[273,291].

Using the derived equation for transverse relaxation of the HSQC antiphase ($R_{2,AP}$), TROSY ($R_{2,TR}$), and MQ ($R_{2,MQ}$) rates, we can begin to simulate the expected linewidths and therefore sensitivity of the corresponding NMR experiments. The previously derived rate constants for $R_{2,AP}$, $R_{2,TR}$, and $R_{2,MQ}$ relaxation rates are given as follows (note – we have ignored the contributions from ^{1}H CSA in these derivations since it is relatively small)[295–297]:

$$R_{2,AP} = \frac{1}{36}d^2\left[2J(0) + \frac{3}{2}J(\omega_C) + \frac{1}{2}J(\omega_H - \omega_C) + 3J(\omega_H + \omega_C)\right] + \frac{1}{3}c^2\left[\frac{2}{3}J(0) + \frac{1}{2}J(\omega_C)\right] + \rho_H \quad (4.10)$$

$$R_{2,TR} = \frac{1}{36}d^2\left[2J(0) + \frac{3}{2}J(\omega_C) + \frac{1}{2}J(\omega_H - \omega_C) + 3J(\omega_H + \omega_C)\right] + \frac{1}{3}c^2\left[\frac{2}{3}J(0) + \frac{1}{2}J(\omega_C)\right] + \rho_H - \frac{1}{3}cdP_2(cos\psi)\left[\frac{2}{3}J(0) + \frac{1}{2}J(\omega_C)\right] \quad (4.11)$$

$$R_{2,MQ} = \frac{1}{3}d^2\left[\frac{3}{2}J(\omega_C) + \frac{1}{2}J(\omega_H - \omega_C) + \frac{3}{2}J(\omega_H) + 3J(\omega_H + \omega_C)\right] + \frac{1}{3}c^2\left[\frac{2}{3}J(0) + \frac{1}{2}J(\omega_C)\right] + \lambda_H \quad (4.12)$$

$$d = 3\left(\frac{\mu_0}{4\pi}\right)\left(\frac{\hbar\gamma_C\gamma_H}{r_{CH}^3}\right) \quad (4.13)$$

$$c = \Delta\sigma\gamma_C B_0 \quad (4.14)$$

$$\rho_H = \frac{1}{4}\left(\frac{\mu_0}{4\pi}\right)^2\left(\frac{\hbar\gamma_H^2}{r_{HHeff}^3}\right)^2[J(0) + 3J(\omega_H) + 6J(2\omega_H)] \quad (4.15)$$

$$\lambda_H = \frac{1}{8}\left(\frac{\mu_0}{4\pi}\right)^2\left(\frac{\hbar\gamma_H^2}{r_{HHeff}^3}\right)^2[5J(0) + 9J(\omega_H) + 6J(2\omega_H)] \quad (4.16)$$

where μ_0 is the permeability of a vacuum, $\hbar$ is planck's constant divided by 2π, γ_C is the gyromagnetic ratio of ^{13}C, γ_H is the gyromagnetic ratio of ^{1}H, r_{CH} is the length of the C-H bond, $\Delta\sigma$ is the chemical shielding anisotropy (CSA), ψ is the angle between the CSA and the bond vector (we used a value of 11°), $P_2(x) = (3x^2 - 1)/2$, and r_{HHeff} is the effective distance of all protons to the C-H proton of interest. These equations although cumbersome looking, comprise mostly

constants and the spectral density function value at various angular frequencies. Using our definition of the spectral density function in Equation 4.1, we can assume a rigid bond vector with $S^2 = 1$. And combining our conversion of nucleotide size to correlation time, we obtain the simulated surface plots in Figure 4.3 which clearly show that TROSY is the optimal experiment for base C-H resonances and HMQC is the preferred method for RNAs over 20 nt for ribose C-H resonances.

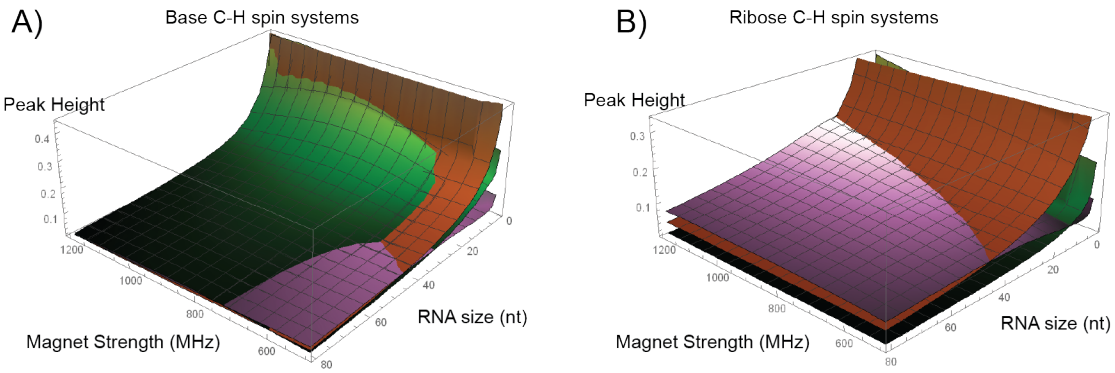

Figure 4.3. 3D surface plot of simulated peak heights for HSQC, HMQC, and TROSY. A) The simulated relative peak height for various size RNAs at magnetic field strengths ranging from 500-1200 MHz. An average CSA of 160 ppm was used for the simulation. B) The simulated relative peak heights for various size RNA ribose C-H spin-systems using 23 ppm for the CSA. Green=TROSY, Orange=HSQC, Purple=HMQC for both surface plots.

4.3 Pulse program optimization

As predicted by mathematical simulation, the TROSY experiments outperformed both HSQC and HMQC for the nucleobase ^{13}C-^{1}H spin-systems in the RNAs we studied and we have reported this in our papers on chemo-enzymatic labeling methods[155,192,195]. Interesting though for field strengths lower

than 550 MHz, HMQC is predicted to outperform TROSY and HSQC for all atomic sites within RNAs larger than 40 nt. Even though TROSY experiments essentially throw-out half of the available starting magnetization, the reduction in transverse relaxation has profound effects on the signal-to-noise ratio for residues with large CSAs[298–300]. In an attempt to recover the lost magnetization of the TROSY experiment, we modified a standard HSQC experiment by removing the decoupling ^{1}H 180° pulse during the t_1 chemical shift evolution period (Figure 4.4). This simple modification results in two observed peaks split by J_{CH} in the indirect dimension only. The downfield ^{13}C peak benefits from the cancellation of the ^{13}C CSA while retaining both components of the ^{1}H magnetization while the upfield peak is substantially weaker and typically not visible. The signal intensity and hence the signal-to-noise (S/N) is increased for base residues. On average this simple adjustment to the HSQC pulse program resulted in a 1.37-fold increase in sensitivity compared with the TROSY experiment for bacterial A-site RNA.

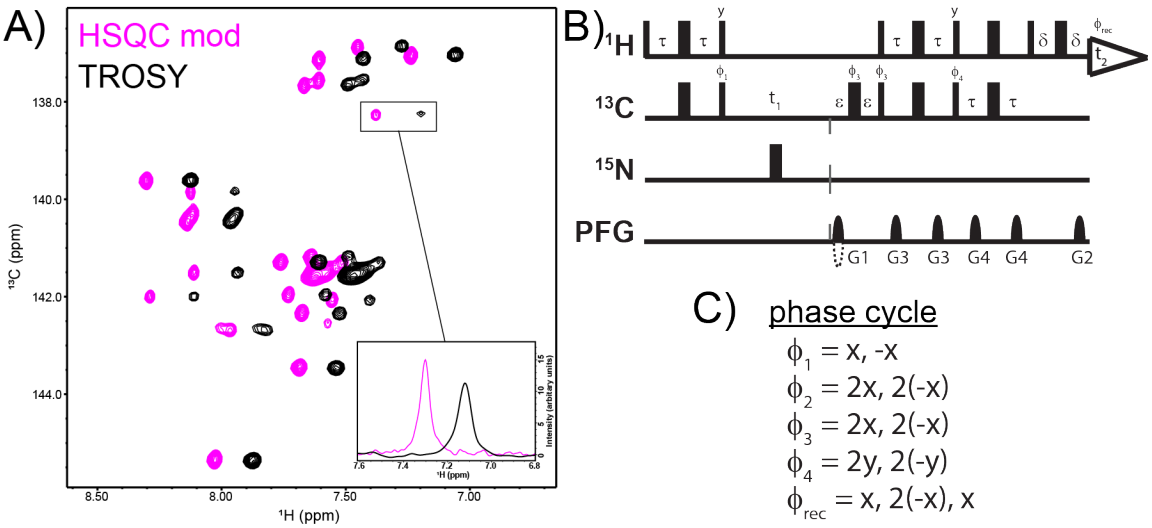

Figure 4.4. Comparison of HSQC with no t_1 decoupling and TROSY. (A) An overlay of the HSQC modified experiment with no ^{1}H 180° hard pulse during t_1 evolution and an equivalent TROSY experiment. Both experiments were run with identical parameters and acquisition time. (B) The modified HSQC pulse sequence is shown. Thin bars represent hard 90° pulses and wide bars represent hard 180° pulses. All phases are 'x' unless otherwise indicated. The delays are as follows: $\tau = 1/4J_{CH}$, $\varepsilon >= G_1$, and $\delta >= G_2$. The gradients are the same as for a typical sensitivity-enhanced, gradient-selection HSQC and are as follows (duration, power): G_1 (1 ms, 28 G/cm), G_2 (1 ms, 7.04 G/cm), G_3 (500 µs, 3.85 G/cm), and G_4 (500 µs, -1.75 G/cm). Quadrature detection and sensitivity-improvement[239]/gradient-selection[240] is implemented using Rance-Kay[241] echo/anti-echo scheme with the polarity of G_1 inverted and phase φ_4 incremented 180° for the second FID generated for each quadrature pair. Axial peak suppression is achieved with STATES-TPPI[301] by incrementing φ_1, φ_3, and φ_{rec} with t_1. (C) The phase cycle used in pulse program.

Given the sensitivity and resolution improvements for ribose residues using HMQC type experiments, we modified $R_{1\rho}$ and R_1 experiments by implementing a HMQC read-out. The basic concept is still an in-phase relaxation measurement followed by conversion to multi-quantum magnetization for t_1 chemical shift evolution and transfer back to ^{1}H for detection. The pulse program is given in Figure 4.5 with the blue line indicating the double INEPT period where

H_z magnetization is converted to C_z. After the relaxation measurement, the remaining C_z magnetization is converted to multi-quantum $2H_xC_y$ magnetization by the module highlighted with orange. The remaining pulse sequence is a standard HMQC sensitivity-improvement, gradient-selection read-out[302]. Even though the module for converting in-phase to multi-quantum magnetization adds an additional 2-3 ms transfer delay, we still expect sensitivity and resolution gains from chemical shift labeling the multi-quantum magnetization during the t_1 evolution period.

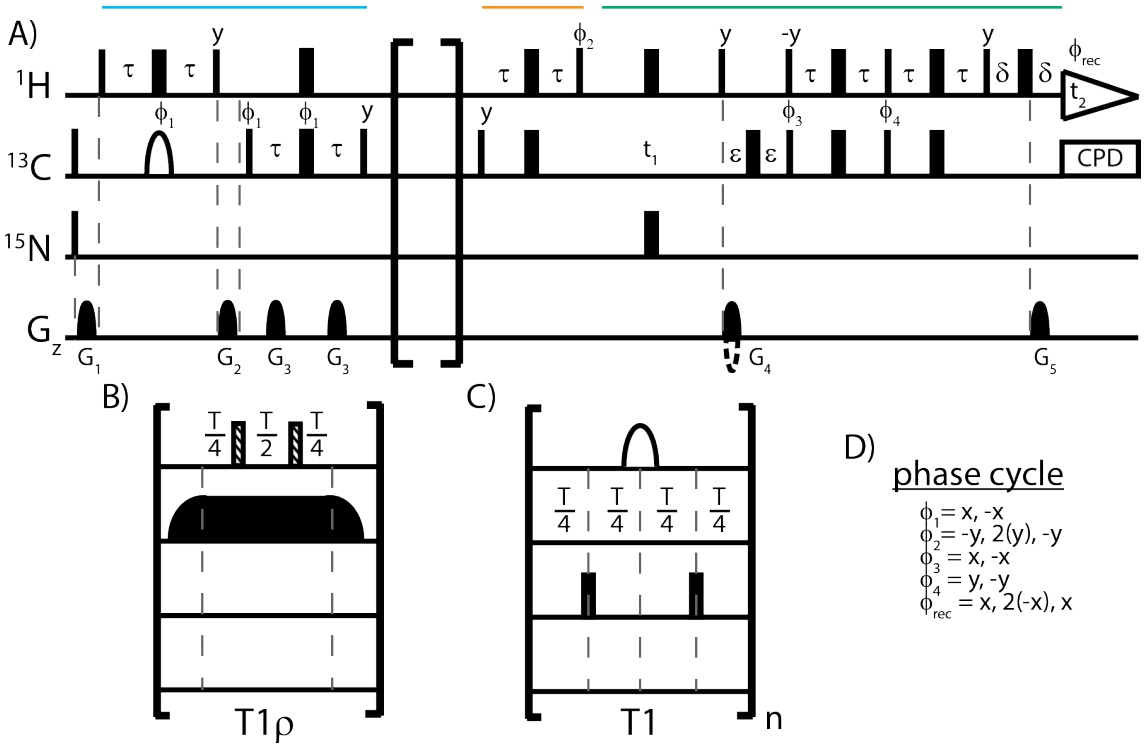

Figure 4.5. HMQC detected R_1 and $R_{1\rho}$ pulse programs. (A) The new pulse program for relaxation measurements of ribose ^{13}C nuclei in RNA. The double INEPT period (sky blue), in-phase to multi-quantum transfer block (orange), and HMQC read-out (bluish-green) are labeled above the module with color coded lines. Thin bars represent hard 90° pulses and wide bars represent hard 180° pulses. Open half-ellipsis represent selective 180° REBURP[303] pulse calibrated to cover 10 ppm bandwidth. All phases are 'x' unless otherwise indicated. Quadrature detection and sensitivity-enhanced/gradient-selection is implemented using Rance-Kay[240] echo/anti-echo scheme with the polarity of G_4 inverted and phase φ_3 incremented 180° for the second FID generated for each quadrature pair. The durations of the delays should be optimized from a 1D experiments, but are generally $\tau = (1/4J_{CH})$ or ≈ 1.5 ms, $\varepsilon \geq G_4$, and $\delta \geq G_5$. Both ε and δ delays are slightly longer than the G_4 and G_5 gradients for gradient recovery. The gradient are all sine-bell shaped with the following duration and power: G_1 (2.65 ms, 2.1 G/cm), G_2 (200 µs, -35 G/cm), G_3 (1 ms, 35 G/cm), G_4 (300 µs, 23 G/cm), and G_5 (75 µs, 23 G/cm). Gradients G_4 and G_5 are the encoding and decoding gradients for gradient selection. The open brackets indicate the module of the pulse scheme in which $T_{1\rho}$ or T_1 experiments can be performed. (B) The $T_{1\rho}$ module for measurement of $R_{1\rho}$ relaxation is shown. The curved slope up- and down-portions of the spin-lock period are to align the magnetization when off-resonance $R_{1\rho}$ experiments are performed and consist of the first and last portion of an adiabatic pulse as previously described[304,305]. The adiabatic half-passages correspond to the first and second half (3 ms each) of a tangent hyperbolic tangent (tanh/tan) adiabatic inversion pulse[282] with a total sweep width of 80 kHz and a maximum radio-frequency field strength identical to

that used for the spin-lock period. To eliminate the effect of cross-correlated relaxation, composite 90_x-180_y-90_x ^{1}H pulses (cross-hatched bar) are applied at T/4 and 3T/4 during the spin-lock period. (C) The T1 module for measuring R_1 relaxation is shown. A 180° IBURP2 shaped pulse at T/2 during the T1 period and has a bandwidth of 4 ppm. The 180° ^{15}N pulses are used to refocus ^{15}N-^{13}C cross-correlated relaxation. For the T_1 module, the entire module is looped n times (n=even number) and T can be 5, 10, or 40 ms. (D) The phase cycle used in (A) is given.

The resulting rates determined by these methods provided similar values (Figure A.7) to those from the standard methods but with improved S/N and resolution. We tested this module on several RNAs in our laboratory including A-site (27 nt), the iron responsive element (IRE, 29 nt), a fluoride riboswitch element from *Bacillus anthracis* (48 nt), a frame-shifting element from a human corona virus (SARS, 59 nt), and a long non-coding RNA element (98 nt). Overall, we observed a consistent improvement in both sensitivity (1.45-fold) and resolution (28% reduction in line-width) as determined from direct measurement and from Lorentzian fits of the peaks. The improvements found here can contribute to either a reduction in error or experimental time for the measurement of R_1 and R_2 of ribose ^{13}C resonances within RNA macromolecules.

Previous results indicated that the use of longitudinal relaxation enhanced spectroscopy improves sensitivity for the measurement of R_1 and R_2[306–308]. We did not observe these same enhancements for our RNA systems. This is reasonable given that the longitudinal relaxation enhanced spectroscopy works by maintaining ^{1}H magnetization along the z-axis throughout the pulse program. Since RNA has a much lower ^{1}H density compared to proteins, there was no significant improvement when the method was applied. Another recent

publication by Lakomek, Ying, and Bax showed that ^{1}H decoupling during T_1 measurements should be optimized in order to obtain accurate results[305]. We implemented the various modules reported in their paper but did not see appreciable effects on the measured R_1 rates obtained for our RNA systems (Figure 4.6).

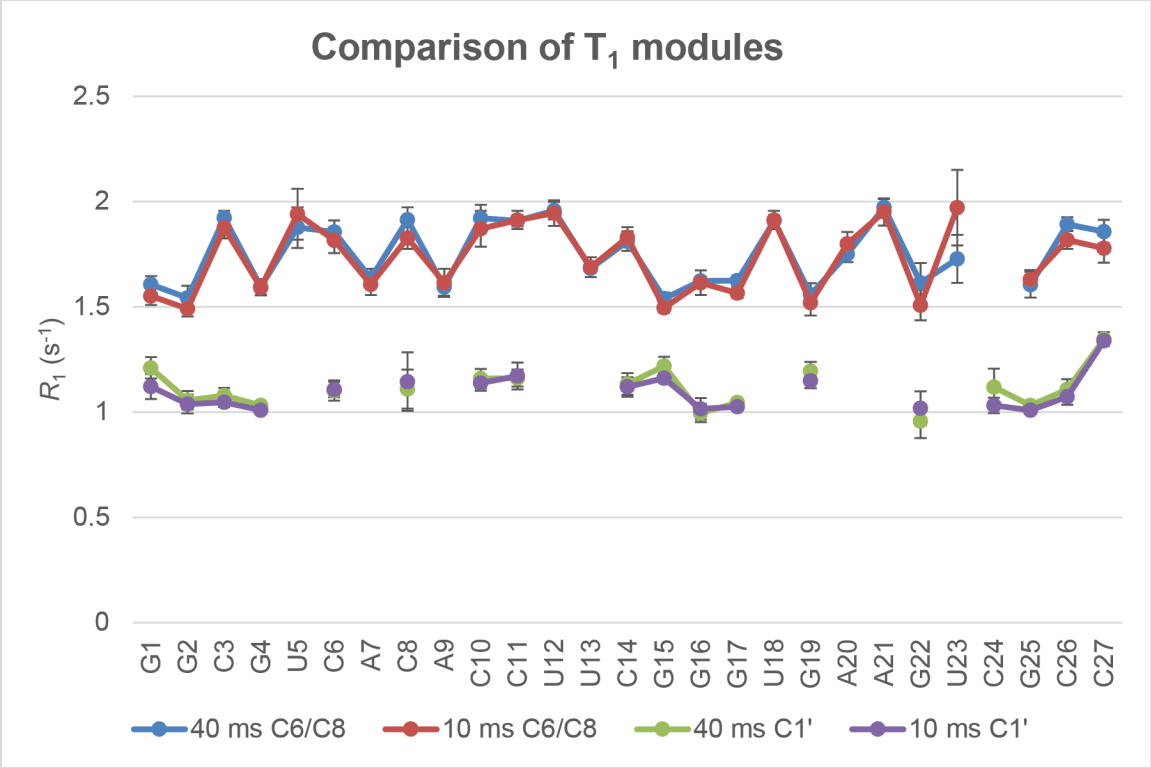

Figure 4.6. Comparison of measured R_1 values from various T_1 modules. Two separate R_1 experiments were performed using either repeated blocks of either 40 ms delay with an IBURP2 refocusing pulse on ^{1}H or 10 ms delay with a 180° hard pulse for ^{1}H refocusing[305]. Using these two modules the R_1 rates were measured for C6/C8 (blue=40 ms block, red=10 ms block) and C1' (green=40 ms block, purple=10 ms block) and only minor variations were observed. Experiments were run with identical parameters.

4.4 ROTDIF analysis of relaxation rates

Once accurate R_1, R_2, and heteronuclear Overhauser effect (hNOE) relaxation measurements are obtained, there are multiple avenues for extracting meaningful information on rotational diffusion, correlation time, internal motions, flexibility, and chemical exchange. With the exception of conformational exchange (R_{ex}) NMR relaxation measurements are not sensitive to rotational motions substantially slower than the time scale of molecular tumbling (typically on the ns time scale) or to rotational motions around the bond vector or to translational motions. Such motions make insignificant contributions to the correlation function.

We sought to investigate the fast motions and chemical exchange inherent in the bacterial A-site RNA model system. Using optimized NMR experiments for sensitivity and resolution, we measured R_1, R_2, and hNOE for the C1' and C6/8 resonances of our selectively labeled sample. We fit the measured relaxation data to obtain dynamic information using the program ROTDIF[273,276].

ROTDIF was an excellent tool for quickly determining flexible residues using the order parameter (S^2). It should be noted that ROTDIF was not intended for robust model-free analysis, but it serves our purposes for comparing differences between selective and uniform ^{13}C samples. Our analysis showed that residues U13 in the tetraloop and A21 in the bulge region are flexible residues, these result match observations from the UUCG tetraloop from both the high resolution structure[216] and relaxation data[309] (Figure 4.7). A21 of the bulge

region of A-site has also been observed to be flexible through FRET experiments using 2-aminopurine as a substitute for A21[310,311].

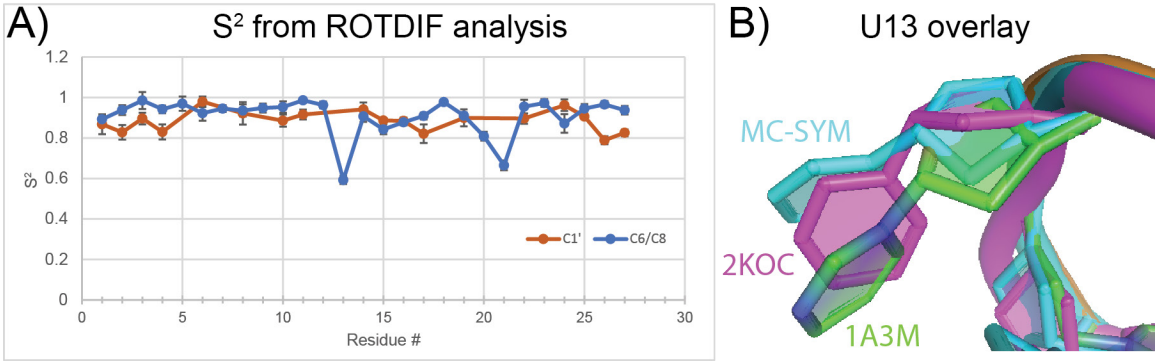

Figure 4.7. Order parameter analysis of base residues from A-site RNA. Using ROTDIF to fit R_1, R_2, and hNOE data obtained from TROSY detected relaxation experiments for base resonances and using HMQC-detected R_1, R_2, and hNOE data for the ribose, we obtained S^2 data for the C1' (brown) and C6/C8 (blue) nuclei of bacterial A-site RNA. U13 and A21 both experience a high degree of flexibility $S^2 < 0.8$. B) Overlay of U12 residue from MC-SYM selected structure (sky blue), UUCG high-resolution structure (purple – 2KOC)[216], and bacterial A-site solution structure (green/blue – 1A3M)[312].

We also observed ms-μs exchange in the lower stem of A-site RNA from the ROTDIF analysis of R_1, $R_{1\rho}$, and hNOE. We compared these results with our previous work on ^{13}C-uniformly labeled A-site RNA (Figure 4.8). The results showed a much greater difference than when we compared S^2 results between u-^{13}C and sel-^{13}C samples (data not shown). There was no clear correlation between C1' and C6/C8 resonances that experienced exchange according to the ROTDIF analysis.

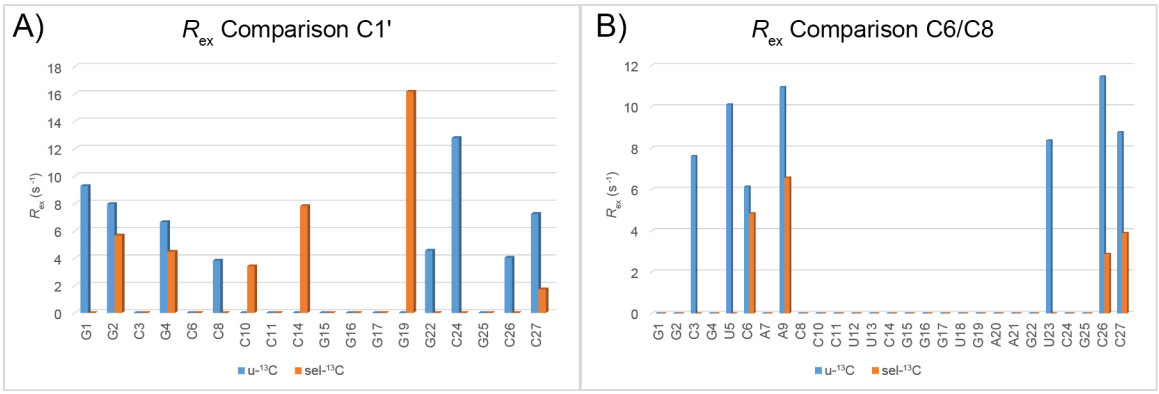

Figure 4.8. Comparison of R_{ex} values from ROTDIF. The extracted R_{ex} values for uniform (blue) and selectively (orange) labeled bacterial A-site RNA ^{13}C-^{1}H spin systems at the C1' (A) and C6/8 (B) positions. Again, the R_{ex} values obtained from ROTDIF are only used for comparison purposes as ROTDIF was not designed for robust model-free analysis.

Our analysis raises three concerns: 1) is the ^{13}C-^{13}C scalar coupling contributing to an overestimation in the R_{ex} of the ^{13}C uniformly labeled sample, 2) are the bond-vectors (the three-dimensional model) that we are fitting our values against an accurate representation of the bacterial A-site RNA construct, and 3) are helical motions on a similar timescale to overall tumbling inducing additional errors in our relaxation analysis.

We have already addressed the concern that scalar coupling overestimates the R_2 contribution to relaxation and therefore would likely cause higher R_{ex} values to be fit by ROTDIF. We can see this in the C6/C8 R_{ex} values, especially for C6 which has strong scalar coupling with C5 (approximately 40 Hz). Residues that have much higher values of R_{ex} in the uniform sample compared to the selective sample are C3, U5, A9, U23, C26, and C27. With the exception of A9, these are all pyrimidines with strong scalar coupling. This

explains the disparity between the R_{ex} values for base ^{13}C-^{1}H methine spin-systems.

If data are limited to a single magnetic field strength, as can be the case for many research laboratories and for our initial analysis using ROTDIF, the use of cross-correlated (η_{xy}) relaxation becomes critical for identifying contributions from conformational exchange line broadening[313]. To better understand the effects of chemical exchange and flexibility on the observed relaxation rates, we implemented experiments for measuring cross-correlated relaxation, η_{xy}.

4.5 Cross-correlated relaxation rate

Our isolated labeling system allowed us to adopt many pulse schemes originally designed for isolated amide residues in proteins. It was our goal initially to completely map the full spectral density function of bacterial A-site RNA, but a recent publication mapping the full spectral density function of RNA was reported first[313]. Instead, we have used multiple types of relaxation to help us determine the best route for a quick analysis of RNA dynamics and exchange in order to simplify these types of NMR studies for future investigators. In particular, the use of cross-correlated relaxation was very helpful. Cross-correlation has been previously studied for RNA systems in order to determine bond vector angles[314], dihedral angles[216,315–317], hydrogen-bond pairs[318], and even for structure refinement[319].

The anti-TROSY and TROSY coherences relax due to transverse relaxation and cross-correlated relaxation: (4.17) and (4.18). By measuring both

the anti-TROSY ($R_{2\alpha}$) and the TROSY ($R_{2\beta}$) relaxation rates, we can obtain the cross-correlated relaxation (η_{xy}) rate (4.19). In order to obtain these rates, we adapted a Hahn-Echo TROSY-select experiment for measuring transverse relaxation of the fast and slowly-relaxing components[320].

$$R_{2\alpha} = R_2^0 + \eta_{xy} + R_{ex} \qquad (4.17)$$

$$R_{2\beta} = R_2^0 - \eta_{xy} + R_{ex} \qquad (4.18)$$

$$\eta_{xy} = \frac{R_{2\alpha} - R_{2\beta}}{2} \qquad (4.19)$$

From equation (4.19) we see that η_{xy} is an exchange free parameter and represents the cross relaxation between in-phase and anti-phase magnetization (the two components that combine to give either anti-TROSY or TROSY resonances). The theoretical expression for the cross-correlation rate constant is[321]:

$$\eta_{xy} = -\frac{1}{2} cdP_2(cos\psi)[4J(0) + 3J(\omega)] \qquad (4.20)$$

in which $c = \gamma_C B_0 \Delta\sigma_C$, β is the angle between the symmetry axes of the chemical shift and dipole tensors, and $P_2(x) = (3x^2 - 1)/2$. Given that η_{xy} is exchange free and dependent on *J(0)*, it should be a strong indicator for differences in internal motion between residues in the same macromolecule (same overall correlation time) based on our previous simulations of the spectral density function. We compared the results of our η_{xy} measurements for A-site RNA with the extracted order parameter we obtained from our ROTDIF analysis (Figure 4.9).

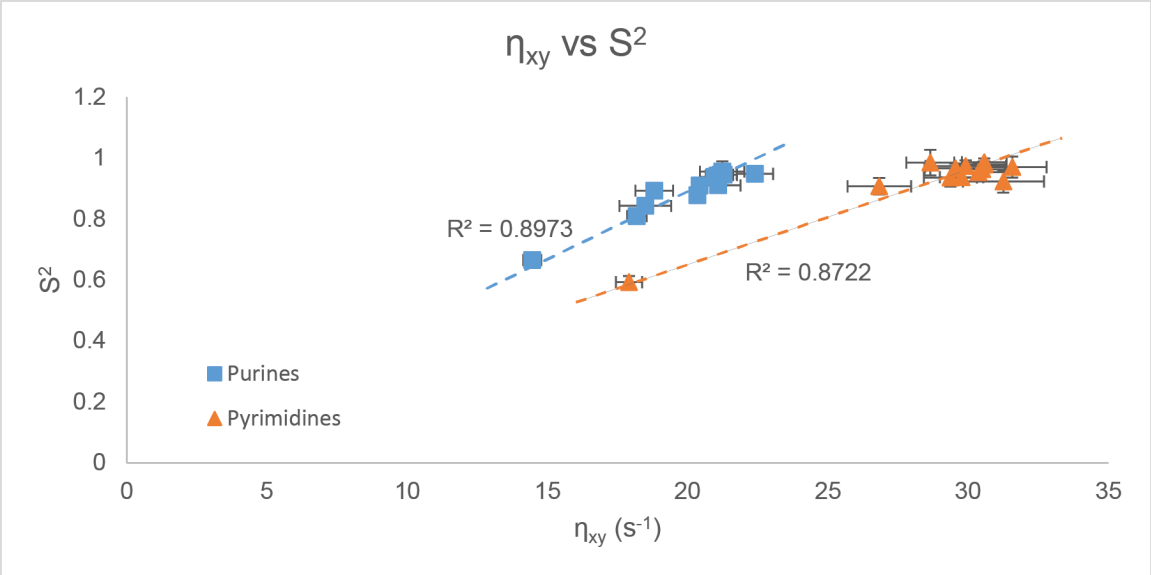

Figure 4.9. Cross-correlation and order parameter comparisons. The results comparing cross-correlated relaxation with the generalized order parameter are shown for purines C8 (blue squares) and pyridimines C6 (orange triangles) resonances. For the fit, we forced the x- and y-intercepts to the origin.

The results that the cross-correlation of ^{13}C-^{1}H aromatic spins directly correlates with the generalized order parameter, S^2, match similar results reported by Tjandra, Szabo, and Bax for amide backbone ^{15}N-^{1}H spin systems in proteins[322]. This previous report also showed that the degree of relaxation interference correlated with the isotropic ^{15}N chemical shift. For the purine C8 nuclei, we observed a weak correlation between ^{13}C chemical shift and η_{xy}. For pyrimidine C6, the observed correlation was much stronger with the most downfield shifted peak U12 (142.5 ppm) also having the lowest η_{xy} (18 s^{-1}) when compared to the average chemical shift (139.5 ppm) and η_{xy} (28 s^{-1}) for the pyrimidine C6 resonances. It is commonly understood that downfield shifted ^{13}C resonances in the base region are due to increased flexibility and typically

correlates to residues in bulge or loop regions of the RNA. The trend is reversed for ribose residues.

Our ROTDIF analysis yielded inconsistent chemical exchange contributions to relaxation, so we measured R_{ex} directly as previously reported[323] by using extracted rates from experiments used to measure R_{zz}, R_1, $R_{2\alpha}$, and $R_{2\beta}$. The results from these experiments are given in Figure 4.10. As we can see there is a clear conformational transition occurring between G19, A20, A21, G22, U23, and C24 of the bacterial A-site RNA model system.

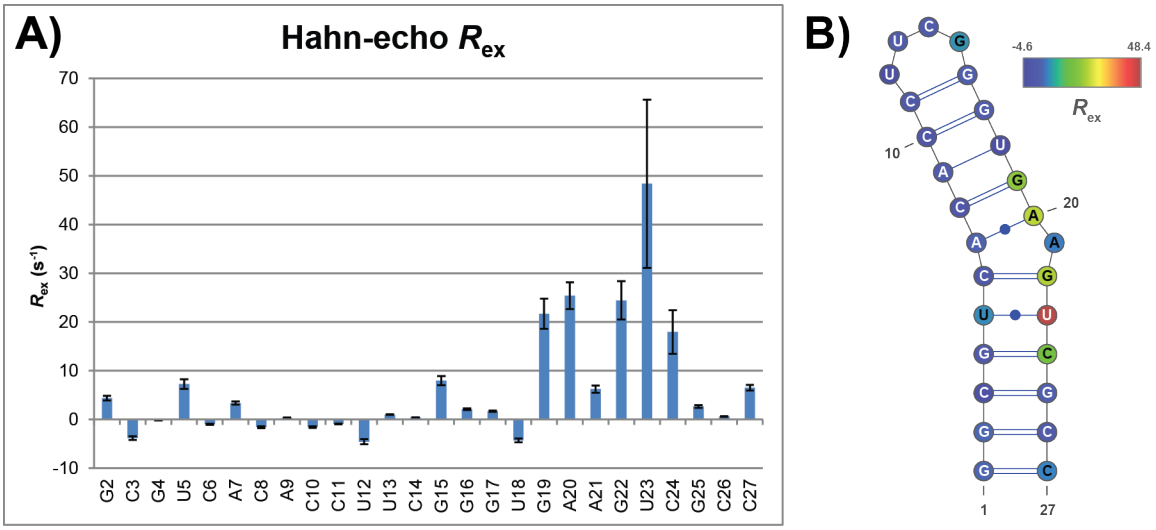

Figure 4.10. Hahn-echo Rex of Asite RNA. A) The exchange contribution to line-broadening in the Hahn-echo experiment was calculated from the equation $R_{ex}^{HE} = R_{2\beta} - \frac{1}{2}(R_{zz} - R_1) - \eta_{xy}(\kappa - 1)$ where κ is calculated from the trim mean of non-exchanging residues using the relationship for kappa, $\kappa = \langle 1 + (R_{2\beta} - \frac{R_{zz}}{2})/\eta_{xy} \rangle$. The values for $R_{2\beta}$, R_{zz}, R_1, and η_{xy} were measured independently[323]. B) A heat map of the secondary structure for bacterial A-site RNA indicating the residues experiencing R_{ex}.

We were able to successfully extract the R_{ex} contributions to R_2 for each residue without the reduced spectral density assumptions. With an eye toward minimizing the set of experiments needed to map relaxation within RNA, we took our predicted R_{ex} and S^2 values from ROTDIF and looked for correlations with more straightforward measurements based on previous observations from protein NMR[322]. We were pleasantly surprised to see that strong correlations existed between $R_{2\beta}/\eta_{xy}$ and R_{ex} (Figure 4.11) as well as S^2 and η_{xy}. Cross correlation as an exchange free measurement that includes J(0) dependence allowed us to quickly resolve flexible residues and residues undergoing μs-ms exchange.

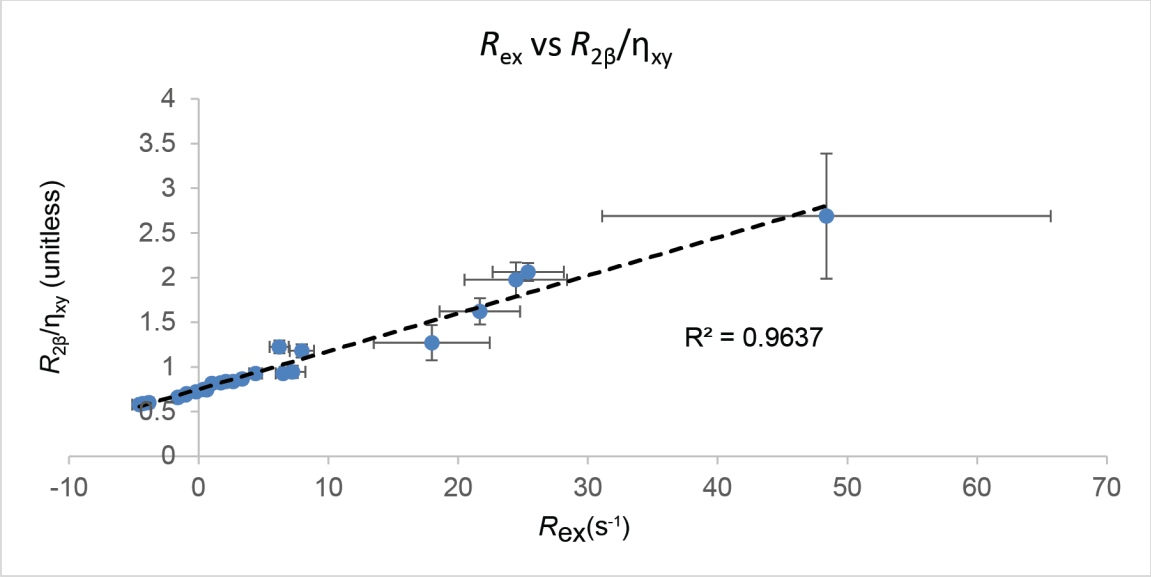

Figure 4.11. $R_{2\beta}/\eta_{xy}$ and R_{ex} correlation. The ratio of $R_{2\beta}$ and η_{xy} are plotted against the extract R_{ex} values from the Hahn echo experiment and shows strong correlation indicating that $R_{2\beta}/\eta_{xy}$ is a good indicator of chemical exchange. A similar approach comparing the correlation of R_2 and η_{xy} in order to qualitatively identify exchanging residues was first purposed by Fushman[324].

To test our theory that $R_{2\beta}/\eta_{xy}$ could indicate chemical exchange, we investigated the effects of pH on the exchanging A-site RNA system. Previous reports from Al-Hashimi using $R_{1\rho}$ and minor state stabilization through single nucleotide mutations showed that bacterial A-site RNA may sample a minor state in which U23 is extrahelical and both A20 and A21 become stacked within the A-helical region[251]. The purposed interhelical stacking of A21 introduces an A-C mismatch pair with C6.

An A-C wobble adopts the same geometry as a G-U wobble with the pKa of A-N1 shifted toward pH 6.2 perturbing the expected pKa of N1 by 2-3 pH units[325]. Previous mutagenesis and chemical probing work on the 50S ribosomal subunit of *Escherichia coli* identified a pH dependent A-C wobble base pair (A2453-C2499) crucial to maintaining correct structure of the ribosome during peptidyltransferase[326]. The importance of A-C wobbles already identified in the peptidyltransferase active site gives evidence for an evolutionary adapted RNA wobble base pair important for structure and function. Based on our MC-FOLD analysis of bacterial A-site RNA in Chapter 3 (Figure 3.10), a predicted A-C mismatch is formed in Model 9. If this type of A-C mismatch is present in our bacterial A-site RNA sample, it should be stabilized at lower pH.

In order to understand how changes in kinetics and population distributions between exchanging states affect the observed R_{ex}, we must first identify the relationship between R_{ex} and the populations, kinetic rates, and chemical shifts. When $p_a > 0.7$ and $R_{2,A} = R_{2,B}$, the relaxation contribution from chemical exchange can be shown as:

$$R_{ex} = \frac{p_a p_b k_{ex}}{\left(1 + \left(\frac{k_{ex}}{\Delta\omega}\right)^2\right)} \qquad (4.21)$$

In Equation (4.21), p_a and p_b are the populations of the major and minor state, respectively. The rate of conversion from state A to B is given as k_{AB} and the rate of conversion from state B to A is given as k_{BA}. The measured kinetic exchange parameter within R_{ex} is $k_{ex} = k_{AB} + k_{BA}$. Finally, $\Delta\omega$ is the chemical shift difference between state A and B. If we assume that the chemical shift difference between the two states remains the same, we can model the effects of stabilizing the minor state on the observed R_{ex} value as seen in Figure 4.12. Stabilizing the minor state, without perturbing the transition state, has two effects: the population of the minor state is increased relative to the major state and the rate constant k_{BA} is decreased thereby decreasing k_{ex}. Combined, these effects push R_{ex} to the maximum value as seen in Figure 4.12.

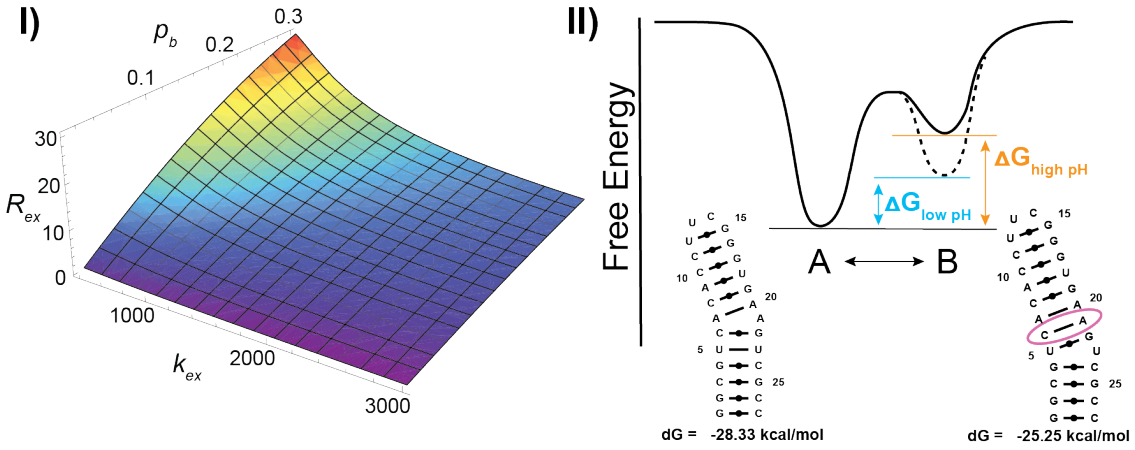

Figure 4.12. Minor state stabilization effect on R_{ex}. I) A 3D surface plot shows the effects of k_{ex} and p_b on chemical exchange line-broadening, R_{ex}. The chemical shift difference between state A and B is assumed to be 2 ppm on a 600 MHz magnet. II) A cartoon representation of the free energy landscape between state A and B. The effects of lowering pH stabilizes the C6-A21 transient base-pair thereby lowering the free energy of the minor state. The dG values under the secondary structure of state A and B are from MC-Fold[256].

In order to test our hypothesis that low pH stabilizes the A-site minor state, we measured the observed chemical exchange of the residues that undergo a secondary structure rearrangement: these residues include G19, A20, A21, G22, and U23. Using our previous observations that $R_{2\beta}/\eta_{xy}$ correlated with the calculated exchange for base residues, we conducted two sets of measurements at two pHs (6.2 and 6.5). This small pH range was enough to detect variations in the relaxation contribution from chemical exchange as seen in Figure 4.13.

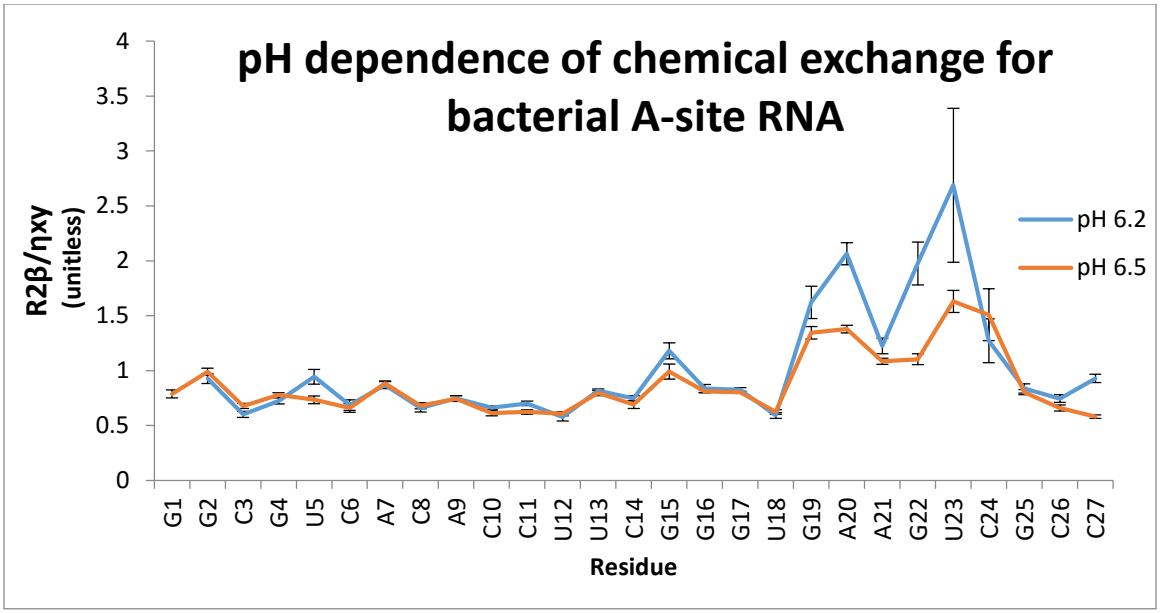

Figure 4.13. $R_{2\beta}/\eta_{xy}$ pH dependence for bacterial A-site RNA. The R2b/nxy ratio used for qualitative analysis of Rex is compared at pH 6.2 (blue) and pH 6.5 (orange) for bacterial A-site RNA C6-H6 and C8-H8 resonances.

4.6 Incorporating dynamics into structure

We used the chemical shifts we assigned from Chapter 3 to select 3D models from over 15 possible 2D structures. Chemical shifts report on the local chemical environment of each nuclei and have the potential to refine structure calculations. Currently there are multiple programs available to calculate predicted chemical shifts from 3D structures: LARMOR^{D} [263], NUCHEMICS[327], SHIFT[328], and RAMSEY[142]. LARMOR^{D} and RAMSEY both allow for the calculation of ^{13}C chemical shifts, unlike NUCHEMICS and SHIFT. These programs use quantum mechanical calculations from ring current effects as well as empirical data from RNA crystal structures (for which NMR chemical shift data are available) for determining predicted chemical shifts. As one would imagine,

given the sparse landscape of deposited RNA structures and NMR resonance assignment data, these programs are still in their infancy and will require much more empirical data to improve their use in structure refinement. With this caveat, we tested a synthetic pool of bacterial A-site RNA structures by back-calculating chemical shifts to determine whether we could recapture 3D structures representative of the chemical exchange data we had obtained using the Constructor RNA program by Nymirium[53,142].

A set of 5 minimized 3D structures of bacterial A-site were selected by using 153 chemical shifts from our rapid resonance assignment strategy and 1500 3D models based on the top 10 secondary structure folds (Figure 4.14). The experimental chemical shifts were compared with predicted chemical shifts from the 1500 3D models and the best 5 structures were ranked by the weighted mean average error (wMAE) for ^{1}H and ^{13}C chemical shifts[142]. The wMAE among the 5 selected structures is 0.439. Also, all the top 5 structures had better fits compared with the 1A3M solution structure back-calculated chemical shifts of bacterial A-site RNA. Additional model selection using residual dipolar coupling (RDC) also confirmed a multiple model ensemble for A-site RNA (Figure A.8 and Figure A.9).

Figure 4.14. Constructor models of A-site RNA compared with solution structure. Based on the top 15 predicted secondary structure, constructor selected 20 3D models from a pool of 1500 synthetic minimized structures. The solution NMR structure (1A3M) along with 3 models representative of the top 5 structures is given. The Model # corresponds to the secondary structure fold from Chapter 3.

The predicted 3D models from our chemical shift data match observations from our relaxation data. Evidence supports the notion that bacterial A-site RNA samples multiple conformations in solution, which may play key roles during the various steps of translation. Cryo-EM studies combined with X-ray crystallography and NMR helped map the translocation step of the polypeptide elongation cycle of the ribosome[329]. The A-site region of the decoding site appeared to actively participate in the translocation of the tRNA from the A- to the P-site upon GTP hydrolysis by elongation factor G, shifting approximately 8 Å toward the P-site. This implies that elongation factor G actively pushes both the decoding site and the mRNA/tRNA complex during translocation[329].

Two recent papers using Cryo-EM show a ratchet like rotation of the small 30S subunit relative to the large 50S subunit[36,330]. A two-step mechanism for

translocation is observed: first, relative rotation of the subunits and opening of the mRNA channel following binding of GTP to EF-G; and second, advance of the mRNA/(tRNA)2 complex in the direction of the rotation of the 30S subunit, following GTP hydrolysis[36,330].

Our predicted models suggest that when U23 “flips out” the RNA structure becomes elongated compared to either Model 1 or Model 3. This is also supported by an overall increase in η_{xy} values indicating a slower overall tumbling of the molecule at lower pH. Could changes in helical compaction or base-pair opening play crucial roles in the ratcheting motion necessary for translocation? Based on our data, this question merits further investigation.

To better guide chemical shift structure prediction, we would be best served by obtaining chemical shift data of the minor state and model it separately from the major state conformations. To this end, we have applied relaxation dispersion experiments to extract out chemical shift information on the minor state of chemically exchanging RNA systems.

4.7 Relaxation dispersion – CPMG for RNA

Relaxation dispersion experiments measure the changes in relaxation rates as a function of applied field strength. In a simple two-state system undergoing exchange, a major state (A) converts to a minor state (B) with a rate constant of k_{AB} and state B converts to state A with a rate constant of k_{BA}. This process of exchange can occur during NMR relaxation periods and modulate the observed relaxation rate. Relaxation dispersion experiments can be used to

study this exchange even when the minor state NMR peak is not visible. In these experiments the kinetics of the exchange process ($k_{ex} = k_{AB} + k_{BA}$) along with information about the populations of the states ($p_A + p_B = 1$) and their relative chemical shifts ($\Delta\omega = \omega_A - \omega_B$) are characterized by studying the relaxation properties of the major state (A) NMR peak.

Our selective ^{13}C isolated spin-systems allowed us to accurately extract relaxation dispersion data using Carr-Purcell Meiboom Gill (CPMG) type experiments[331,332]. RNA systems have been difficult for utilizing CPMG experiments due to the ^{13}C-^{13}C scalar couplings that leads to multi-exponential decay curves. In fact, only 4 papers have ever been published on the use of CPMG relaxation dispersion for RNA systems. In 2008, Johnson and Hoogstraten used DL323 cell-based labeling to obtain isolated C2' and C4' spin-systems and combined CPMG and $R_{1\rho}$ experiments to map chemical exchange throughout the backbone of the GCAA tetra-loop RNA motif[165]. In 2011, our collaborator, Christoph Kreutz, and colleagues published an article using ^{13}C-methyl incorporation at the C2' hydroxyl group in order to create an isolated ^{13}C spin-system for CPMG relaxation dispersion experiments[333]. Kruetz and co-workers expanded on this method by applying solid-phase synthesis to obtain an isolated C6 pyrimidine ^{13}C-^{1}H spin-system in order to map μs-ms dynamics in the HIV-1 transactivation response element RNA and the Varkud satellite stem loop V RNA motif[334]. And finally, in 2016, we exploited our chemo-enzymatic labeling method to obtain isolated ^{13}C-^{1}H spin systems in the bacterial A-site decoding RNA to acquire relaxation dispersion data from CPMG experiments[191].

Over the same time period, over 100 papers have appeared in the PubMed database using CPMG relaxation dispersion to better understand protein folding and conformational changes. The theory and application of CPMG for proteins has been well-reviewed[323,335–338]. Of particular interest are the methods that have been developed by Lewis Kay and colleagues[338–347]. CPMG experiments originally measured the relaxation of the in-phase or anti-phase hetero-atom relaxation. The work by Lewis Kay has expanded the use of CPMG to monitor multi-quantum (MQ)[348], zero-quantum (ZQ)[347,348], double-quantum (DQ)[347,348], TROSY[348,349], anti-TROSY[348,349], and ^{1}H single-quantum (SQ) relaxation dispersion profiles. The added benefit of measuring relaxation dispersion of all these components of relaxation is that they all have different dependencies on the chemical shifts of the minor states. For example, MQ, ZQ, and DQ dispersion profiles are affected by both the ^{13}C and ^{1}H chemical shift differences between the minor and major states whereas ^{13}C SQ and ^{1}H SQ are dependent on only the ^{13}C or ^{1}H chemical shift difference, respectively. Combining these methods, a more complete description of the chemical environment experienced by the minor state can be obtained.

To better understand how CPMG experiments might be applied to RNA systems, we re-wrote each published CPMG program for Bruker instruments and optimized for RNA systems. Within this section, we will describe our findings for each CPMG experiment. We applied our adapted and new CPMG experiments to our model bacterial A-site RNA system and to the iron responsive element RNA. For all CPMG experiments described in this Chapter, the relaxation

properties of transverse magnetization is monitored after a series of π pulses. A constant time CPMG module is used to obtain the effective relaxation rate, $R_{2,eff}$, as a function of the frequency at which π pulses are applied (ν_{CPMG}) during the constant time relaxation period (T_{relax}). The CPMG module is a set (N) of repeated delay-180° pulse-delay blocks (τ_{CPMG}-π-τ_{CPMG}) such that 2 x N x τ_{CPMG}= T_{relax}. A reference experiment is recorded with no CPMG module and used to calculate the effective relaxation rate at each frequency as given by the following equation, where I is the intensity of the peak measured in the CPMG experiment and I_0: is the intensity of the peak in the reference experiment:

$$R_{2,eff} = \frac{-1}{T_{relax}} \ln\left(\frac{I}{I_0}\right) \qquad (4.22)$$

The resulting $R_{2,eff}$ values are plotted against the CPMG pulse frequency as determined by the relationship:

$$\nu_{CPMG} = \frac{1}{4 * \tau_{CPMG}} \qquad (4.23)$$

The relaxation dispersion curve can be fit to various analytical solutions for general[350] or fast[351] exchange or to the Bloch-McConnell equation for describing transverse relaxation in the presence of chemical exchange[296,352].

4.7.1 A-site RNA

As our initial well-behaving RNA system in the lab, we have conducted numerous experiments and optimizations on the bacterial A-site RNA model system. In accordance with these efforts, we first experimented with TROSY CPMG pulse sequences. Adopting a TROSY-select CPMG experiment from

previous work by Weininger, Respondek, and Akke[353], we were able to successfully map chemical exchange and relaxation dispersion profiles for G19, A20, A21, and C24 as expected from our previous results (Figure 4.15). Using the publicly available ShereKahn CPMG fitting website, we obtained global fits to the analytical Luz-Meiboom equation[351] for fast two-site exchange (Equation 4.24 and 4.25)

$$R_2^{eff} = R_2 + \frac{\Phi}{k_{ex}}\left(1 - \frac{4v_{CPMG}}{k_{ex}}\tanh\left(\frac{k_{ex}}{4v_{CPMG}}\right)\right) \qquad (4.24)$$

$$\Phi = 4\pi^2 B_0^2 \varphi \qquad (4.25)$$

where R_2^{eff} is the experimentally measured relaxation rate at each given CPMG frequency (v_{CPMG}), R_2 is the fitted intrinsic transverse relaxation rate, k_{ex} is the fitted kinetic rate constant ($k_{ex} = k_{AB} + k_{BA}$ for a 2-state system), B_0 is the given field strength for the nuclei of interest in MHz, and φ is the fitted population weighted chemical shift difference in ppm^2 for a fast exchanging 2-state system. A CPMG profile for residues with no exchange on the CPMG timescale are flat as shown in Figure A.10.

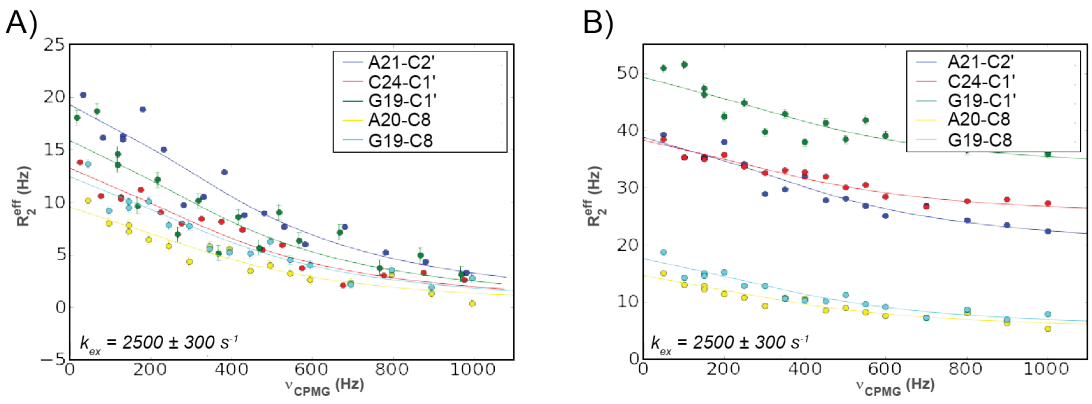

Figure 4.15. TROSY select CPMG for A-site RNA. (A) The change in R_{ex} with the CPMG pulse frequency (vCPMG). (B) The relaxation dispersion curves are shown with residues color coded to match the data and best fit line. The data was globally fit using the freely available ShereKhan software online (sherekhan.bionmr.org) and the analytical Luz-Meiboom equation for fast two-site exchange[351]. The extracted exchange rate was 2500 ± 300 s^{-1}. A 40 ms constant time period was implemented for the CPMG module using standard delays for the INEPT transfers ($1/4J_{CH}$) and a 1.5 s recovery delay.

In order to obtain chemical shift information, it is typical to measure CPMG at multiple fields, however exchange rates can be easily extracted at a single field as seen in Figure 4.15. Our obtained rates match well with those previously reported by Al-Hashimi using $R_{1\rho}$ experiments[251]. The TROSY-select experiment works by measuring the relaxation dispersion profile of either the TROSY or anti-TROSY component. The effect is less pronounced for the ribose residues with R_{2eff} rates between 30-40 s^{-1} at the highest effective field strength in the CPMG experiment (1000 Hz). For the base residues, however, the TROSY effect minimizes the decay rate during the CPMG pulse train with rates as low as 8 s^{-1} observed at 1000 Hz effective field strength. The TROSY-select experiment, therefore, is a key tool for measuring relaxation dispersion in larger RNAs in

which higher transverse relaxation rates would limit the usefulness of typical CPMG experiments.

Another benefit of the TROSY-select CPMG experiment is that the extracted Δω value determines the chemical shift of the TROSY component of the minor state. The HSQC chemical shift of the minor state is then obtained from standard in-phase CPMG experiments. The difference between these chemical shift values is equal to $J_{CH}/2$ for the minor state resonance. The ability to extract the J-coupling of the minor state C-H spin-system would allow for accurate residual dipolar coupling (RDC) measurements of the minor state thereby supplying additional structural information of the minor state[338,354]. A similar concept was applied to an exchanging protein-ligand system (Abp1p SH3 domain – Ark1p peptide) to build a 3D model of the transition state[354]. Here we have shown that these experiments can be successfully adopted for RNA systems with our isolated ^{13}C-^{1}H spin-systems.

Even with the TROSY-select CPMG experiment, we can only gain insight into the change in chemical environment for the ^{13}C nuclei and not for the ^{1}H nuclei. This limits the chemical shift information we can obtain. Similarly, ^{13}C $R_{1\rho}$ experiments are unable to determine chemical shift information for proton. Here we chose to adapt a ^{1}H CPMG experiment to test its applicability to our isolated ^{13}C-^{1}H spin-systems. Using previously published ^{1}H CPMG experiments for CHD_2 methyl groups in proteins[355], we successfully obtained CPMG relaxation dispersion curves for A-site RNA (Figure 4.16) that were fit to the full analytical MQ Carver-Richards-Jones all-timescales equation as previously

described (Equations 4.26 – 4.36) using the GUARDD Matlab CPMG fitting software[356]

$$R_2^{eff}\left(\frac{1}{2\delta}\right) = \Re(\lambda_1) - \frac{\ln(Q)}{4n\delta} \tag{4.26}$$

$$\lambda_1 = R_2^0 + \frac{1}{2}\left(k_{ex} - \left(\frac{1}{2\delta}\right)\cosh^{-1}(D_+\cosh(\eta_+) - D_-\cosh(\eta_-))\right) \tag{4.27}$$

$$D_\pm = \frac{1}{2}\left(\frac{\psi + 2\Delta\omega_X^2}{(\psi^2+\zeta^2)^{1/2}} \pm 1\right) \tag{4.28}$$

$$\eta_\pm = \sqrt{2}\delta((\psi^2+\zeta^2)^{1/2} \pm \psi)^{1/2} \tag{4.29}$$

$$\psi = (i\Delta\omega_Y + (p_A - p_B)k_{ex})^2 - \Delta\omega_X^2 + 4p_Ap_Bk_{ex}^2 \tag{4.30}$$

$$\zeta = -2\Delta\omega_X(i\Delta\omega_Y + (p_A - p_B)k_{ex}) \tag{4.31}$$

$$Q = \Re\left(1 - m_D^2 + m_Dm_Z - m_Z^2 + \left(\frac{m_D + m_Z}{2}\right)\left(\frac{p_B}{p_A}\right)^{1/2}\right) \tag{4.32}$$

$$m_D = \frac{ik_{ex}(p_Ap_B)^{1/2}}{d_+z_+}\left(z_+ + 2\Delta\omega_X\left(\frac{\sin(z_+\delta)}{\sin((d_+ + z_+)\delta)}\right)\right) \tag{4.33}$$

$$m_Z = \frac{ik_{ex}(p_Ap_B)^{1/2}}{d_-z_-}\left(d_- - 2\Delta\omega_X\left(\frac{\sin(d_-\delta)}{\sin((d_- + z_-)\delta)}\right)\right) \tag{4.34}$$

$$d_\pm = (\Delta\omega_Y + \Delta\omega_X) \pm ik_{ex} \tag{4.35}$$

$$z_\pm = (\Delta\omega_Y - \Delta\omega_X) \pm ik_{ex} \tag{4.36}$$

where $\delta = 1/(4v_{CPMG})$, $n = T_{CPMG}v_{CPMG}$ where T_{CPMG} is the length of the CPMG constant-time period in seconds, $\Re$ is the real value of the function, R_2^0 is the fitted intrinsic transverse relaxation rate, $\Delta\omega_X$ and $\Delta\omega_Y$ is the chemical shift difference between the two-states (for SQ CPMG $\Delta\omega_X$ pertains to the measured

nuclei and $\Delta\omega_Y = 0$; for MQ CPMG $\Delta\omega_X$ is the ^{13}C chemical shift difference and $\Delta\omega_Y$ is the ^1H chemical shift difference), p_A and p_B are the population of the major and minor states, respectively, and k_{ex} is the combined rate constant ($k_{ex} = k_{AB} + k_{BA}$).

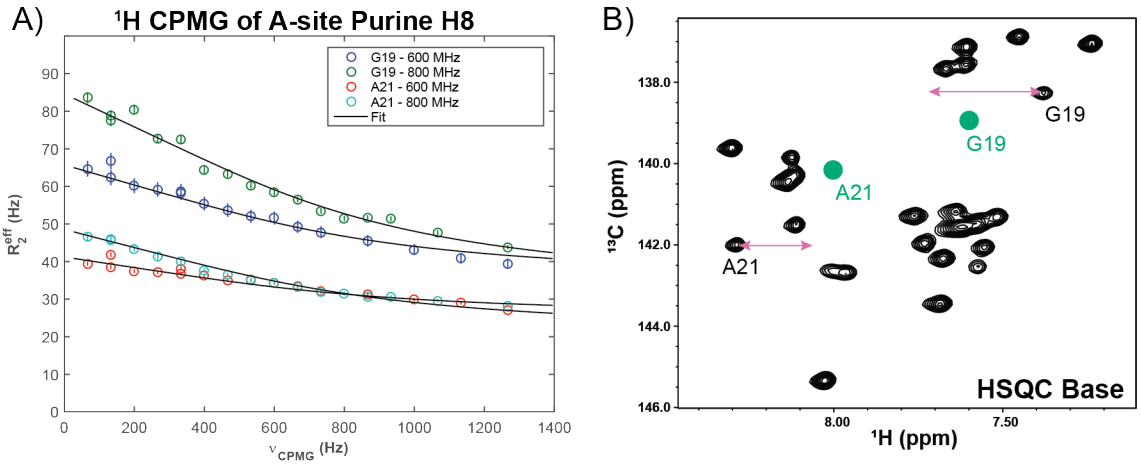

Figure 4.16. ^1H CPMG relaxation dispersion profile and chemical shifts extracted. (A) The ^1H CPMG profile for G19-H8 and A21-H8 at 600 MHz and 800 MHz. A global fit of the data was obtained using the GUARDD Matlab software for CPMG fitting[357]. A 30 ms constant time CPMG ^1H module was applied before the start of a standard TROSY experiment as previously reported[355]. (B) HSQC spectrum of A-site nucleobase region (C6-H6, C8-H8) showing the extracted chemical shift change from the ^1H CPMG (magneta arrow) and the chemical shift of a stabilized excited state reported by Al-Hashimi (green circles)[251].

The ^1H CPMG experiment was amenable to H8 protons, but gave poorly fit relaxation dispersion curves for H6 protons. This is most likely due to the ^1H-^1H through-space coupling of the nearby H5. Originally, these experiments were performed in perdeuterated protein samples[355]. Our results may be contaminated by interference from neighboring protons. The extracted k_{ex} value is similar to that obtained from our TROSY-selected CPMG experiment, except

that the error in the fit was much greater for the ^{1}H CPMG (k_{ex} = 4000 ± 1000 s^{-1}). Incorporating deuteration into our labeling strategy would provide a much more accurate measurement for ^{1}H CPMG type experiments. We do show that these experiments are feasible for RNA and could provide important data on ^{1}H chemical shift. Indeed, our results are similar to the measured chemical shifts of a stabilized minor state of A-site RNA reported by Al-Hashimi and co-workers (Figure 4.15B)[251].

4.7.2 IRE RNA

Translational regulation of ferritin (an iron storage protein) is under the control of a mRNA structural element in the 5'-untranslated region (5'-UTR) called the iron-responsive element (IRE)[358–360]. Under low cellular iron conditions, iron-regulatory proteins (IRPs) bind IRE with high affinity and sequester the translation of ferritin[361,362]. Examination of the IRE-IRP bound structure and the free IRE structure shows that two conserved C residues separated by five base-pairs are critical for binding IRP (Figure 4.17).

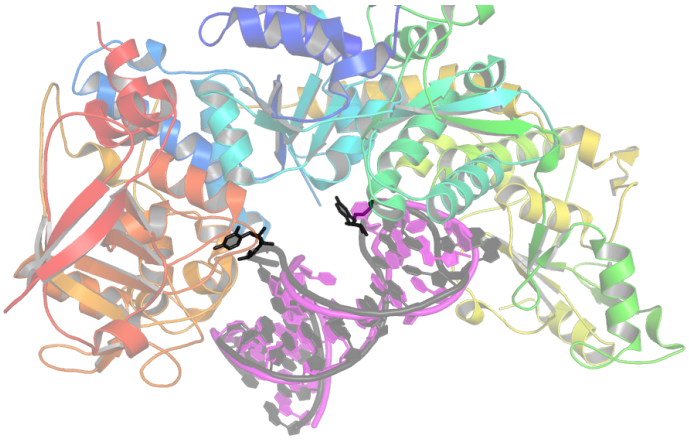

Figure 4.17. IRE free and bound to IRP overlay. The free IRE solution structure (1NBR)[363] is shown in purple overlaid with the crystal structure of the bound IRE in black with IRP (3SNP)[364] colored by domain. The critical C7 and C18 residues are drawn as bold sticks.

Previous work by Luigi Alvarado in our laboratory showed that doxorubicin could disrupt binding of the IRE-IRP complex most likely by restricting the flexibility of C7 and C18 residues needed for IRP recognition. At higher iron concentrations, the structure of IRE many undergo a transition in which these residues are hidden from IRP. Under these assumptions, we aimed to investigate the conformational landscape of IRE to identify if there might be a structural rearrangement around these critical C residues. We applied our method for fast recognition of exchanging residues by evaluating the R2B/nxy ratio (Figure 4.18).

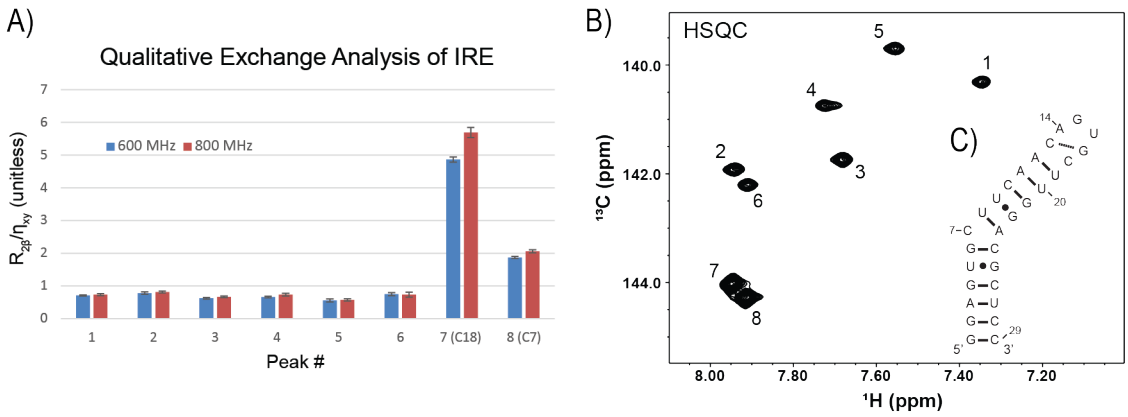

Figure 4.18. Chemical exchange mapping for IRE RNA. (A) The $R_{2\beta}/\eta_{xy}$ ratio is shown for peaks 1-8 of a ^{13}C-1',5',6-CTP labeled IRE RNA using our previously described methods for 600 and 800 MHz field strength magnets. Based on previous work peaks 7 and 8 are assigned to C18 and C7, respectively[192]. (B) HSQC of the C6-H6 region of the C-labeled IRE RNA with peaks labeled to correspond with part (A). (C) Secondary structure of IRE RNA.

Combining our selective labeling strategy along with the F1eF2e NOESY experiment we implemented in Chapter 3, we tentatively assigned the two C-residues undergoing exchange. Previous assignment of the 6 nt hairpin loop of IRE showed a distinctive downfield shift for the terminal C residue C1' (6.04 ppm) and a small splitting for the H5'/H5" peaks (48 Hz)[365], which matches well with our assignment of this residue in our selectively labeled CTP IRE sample. Since our RNA had both the 1' and 5' carbons labeled, strong intra-nucleotide NOEs confirmed our assignment of C18-H1'/H5'/H5"/H6. With this assignment in hand and a clear indication of chemical exchange within the tri-loop and bulge region of IRE, we tested previously published MQ[348] and SQ[366] CPMG experiments for extracting ^{1}H and ^{13}C chemical shifts of minor "invisible" states. It should be noted that these experiments are written for ^{15}N and must be converted for

measurement of ^{13}C CPMG relaxation dispersion. The results for our exchanging residues C7 and C18 are shown in Figure 4.19.

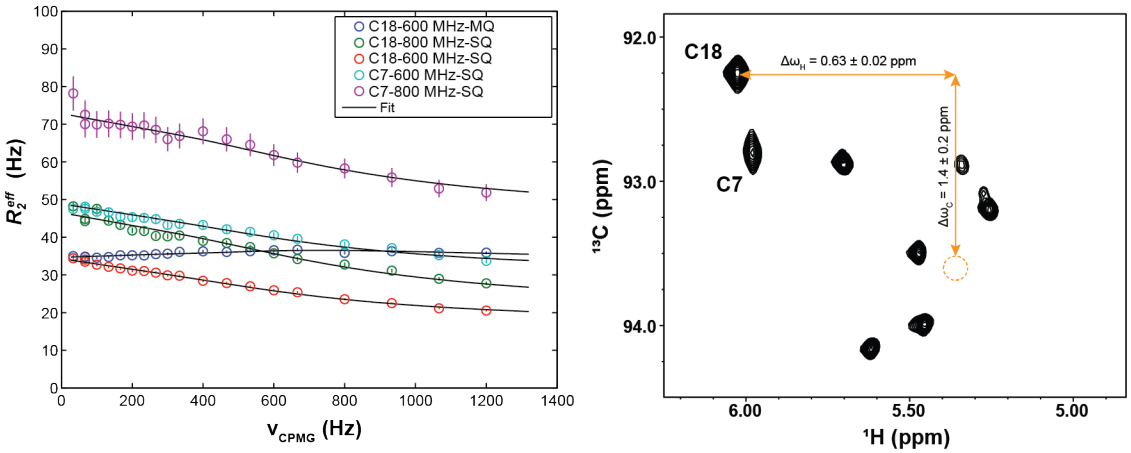

Figure 4.19. Multi-quantum and single-quantum CPMG of IRE RNA. The figure on the left is the CPMG relaxation dispersion profile for MQ and SQ data at 600 and 800 MHz magnet fields. The HMQC spectrum on the right shows the expected position of the minor "invisible" state peak as determined by the global fits of MQ and SQ CPMG data using GUARDD Matlab software to fit equations 4.26 – 4.36[357]. Residue C7 did not show any shifts in the ^1H dimension base on MQ CPMG results.

Previous relaxation experiments on ^{13}C labeled purine IRE RNA reveals flexible residues around the hairpin loop with the adenine before the 5' end of the 6 nt hairpin loop showing exchange[367]. The researchers were unable to measure exchange for the pyrimidine residues in the tri-loop and here we show that there may indeed be a global rearrangement of the tri-loop structure. The chemical shift for C18 is predicted to move from a flexible resonance to a stacked/Watson-Crick like chemical shift resonance. The extracted exchange rate for this global rearrangement is 3600 ± 300 s^{-1} with a minor state population of 1.7 ± 0.2 %. Indicating microsecond timescale exchange.

To determine whether these exchange events were merely sugar puckering or glycosidic bond rotation, we extended out analysis to the backbone of the residue via the C5' nuclei. Using our previous assignment we quickly correlated the C5' resonances to their respective C18 and C7 residues. Since our labeling strategy allowed us to obtain an isolated ^{13}C spin at the C5' position, we were able to use a new CH_2 (methylene) TROSY CPMG pulse program adapted from previous work by Miclet[368] (Figure 4.20A). We added a CPMG module to the previously reported CH_2 TROSY read-out pulse scheme and successfully measured relaxation dispersion within the backbone of the C18 residue (Figure 4.20B) indicating that the chemical exchange event most likely results in significant structural rearrangement in the tri-loop of IRE RNA. The extracted chemical shift change for the methylene C5' carbon of C18 indicates a transition from a flexible residue to a stacked A-helical like residue confirming results we saw with the C1' and C6 resonances.

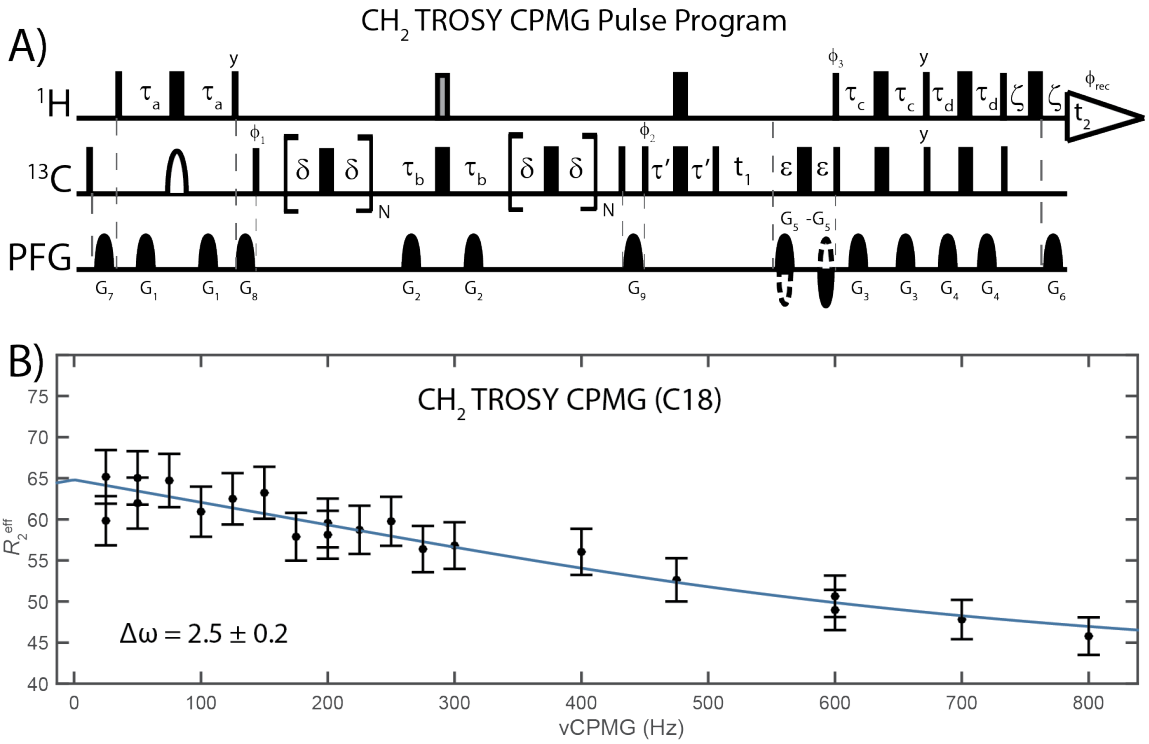

Figure 4.20. CH_2 TROSY CPMG of IRE RNA. (A) A new CH_2 TROSY CPMG pulse sequence. Thin bars represent hard 90° pulses and wide bars represent hard 180° pulses. Open half-ellipsis represent selective 180° REBURP[303] pulse calibrated to cover 10 ppm bandwidth. The gray-filled bar is a composite 90_y-180_x-90_y pulse. All phases are 'x' unless otherwise indicated. Quadrature detection and sensitivity-enchanced/gradient-selection is implemented using Rance-Kay[240] echo/anti-echo scheme with the polarity of G_6 inverted and phase φ_3 incremented 180° for the second FID generated for each quadrature pair. The durations of the delays should be optimized from a 1D experiments, but are generally τ_a = 1.6 ms (<1/4J_{CH}), τ_a = 1.66 ms (1/4J_{CH}), τ' = 0.417 ms (1/16J_{CH}), τ_c = 1.11 ms (0.34/J_{CH}), τ_d = 0.767 ms (0.23/J_{CH}), $\delta = \tau_{CPMG}$, $\varepsilon \geq G_5$, and $\zeta \geq G_6$. Both ε and ζ delays are slightly longer than the G_5 and G_6 gradients for gradient recovery. The gradient are all sine-bell shaped with the following duration and power: G_1 (350 µs, 24.5 G/cm), G_2 (400 µs, 17.5 G/cm), G_3 (400 µs, 17.5 G/cm), G_4 (350 µs, 24.5 G/cm), G_5 (500 µs, 28 G/cm), G_6 (251.5 µs, 28 G/cm), G_7 (2 ms, 24.5 G/cm), G_8 (2 ms, 24.5 G/cm), and G_9 (1 ms, 24.5 G/cm). Gradients G_5 and G_6 are the encoding and decoding gradients for gradient selection. The phase cycle is as follows: Φ_1 = x, -x; Φ_2 = 225°, 225°, 45°, 45°; Φ_3 = 4(x), 4(y); Φ_{rec} = x, 2(-x), x, -x, 2(x), -x. (B) A representative CPMG profile for C18-C5'. The data was fit to the Luz-Meiboom[351] analytical solution using the k_{ex} and p_b values obtained from global fits of the MQ and SQ CPMG data in order to obtain the chemical shift difference of the minor state.

The IRE RNA plays a crucial role in iron homeostasis through translational regulation of iron storage and transport proteins[367,369,370]. Its ability to bind IRP under low iron concentrations dictates its function in translational regulation of downstream proteins[371]. How IRE reacts to environmental stimuli such as iron concentration most likely effect its overall structure and we have shown that multiple states of IRE are likely present in solution. The population of these states is expected to be dictated by iron levels in the cell. HBVε has a similar tri-loop to IRE that shows chemical exchange (Figure A.11) indicating that these types of structural rearrangements may be ubiquitous among RNA hairpin loops and critical to binding interactions.

4.8 Conclusion

We have shown that both flexibility and chemical exchange can be quickly analyzed for RNA systems by taking advantage of isolated spin systems that allow for accurate measurement of cross-correlated relaxation rates. The cross-correlated relaxation rate, η_{xy} correlates very well with the order parameter, S^2, from Lipari-Szabo model-free type analysis. In fact, the order parameter has been used for quantitatively obtaining entropy data for folding protein systems[372–377], which could indicate a similar analysis using η_{xy} to understand the entropy of RNA folding events. We have also shown that the ratio of $R_{2\beta}/\eta_{xy}$ correlates with the measured R_{ex} values of residues within the RNA. A fast and simple approach to analyzing both flexibility and chemical exchange is critical for understanding the underlying motions and conformational changes RNA undergoes and how that may relate to its function.

We have developed improved pulse programs for measuring standard relaxation rates for ribose residues within RNA systems by switching to multi-quantum coherence during t_1 chemical shift evolution periods. We have also tested each CPMG experiment for use with RNA systems including development of a new CH_2 TROSY CPMG experiment. In addition, MQ and ^1H CPMG were shown to be effective for RNA systems and could provide chemical shift information on both the ^{13}C and ^1H nuclei. By combining chemical shift information on ^1H, we can begin to adopt chemical shift predictive tools to build models of the entire RNA conformational landscape. Chemical shifts have been successfully implemented to guide molecular simulations as well[262,378], but a continued effort to improve the quantity of structures in the PDB is ultimately needed to refine predictive chemical shift software. We are not there yet, but all of the tools we have introduced from Chapter 2-4 will be immensely useful in the ongoing quest of RNA structure and ensemble determination.

The tools we have developed shed light on RNA biology and structure, but in order to build a strong platform for drug discovery we must understand the rules of ligand binding to RNA. For this purpose, we chose to study a riboswitch system that regulates gene expression through a structural rearrangement upon ligand binding. How do small molecules reform RNA structure in order to dictate function? This is the question central to our project, discussed within the next Chapter.

5 Sparsely populated RNA states and ligand recognition

With the proper tools developed from chapters 2-4, we have laid the foundation for the investigation of “invisible” or transiently populated conformations of RNA. Unraveling the structure of sparsely populated states is essential to understanding RNA function because many of these minor states play integral roles in the function of the RNA. In particular, riboswitches are prime candidates to sample these transient states. Riboswitches typically work ‘on the fly’ –preventing RNA polymerase from continuing transcription through the formation of a terminator hairpin or the ribosome from binding to the newly-made mRNA by sequestering the Shine-Delgano (SD) sequence. The ability of the riboswitch to regulate gene expression depends almost entirely on the rate of ligand association rather than binding affinity (K_D). Therefore, we suggest that riboswitches exist in transient pre-formed bound-like states that allows for rapid ligand association of the minor state in order to finely tune gene regulation. NMR is uniquely suited for these types of studies and we have utilized new NMR experiments developed by Lewis Kay[379,380] to study slow exchange (ms-s) of the SAM-II riboswitch in work that was recently published in *Angewandte Chemie*[210].

5.1 Biological significance and background

While it is widely accepted that biomolecules such as RNAs are highly dynamic, it is only recently becoming appreciated that RNAs latently sample multiple conformations in solution as a mechanism for function, especially for the class of RNAs called riboswitches[29,75,274]. The dramatic increase in the discovery

of these riboswitches[29] has been followed by high resolution structures of the bound forms[31]. However, only four riboswitches without their cognate ligands have been structurally characterized[78]. Both high resolution x-ray crystallographic and SAXS analyses indicate that a number of riboswitches (FMN, lysine, preQ1, and SAM-1) adopt the same global conformation in the absence or presence of its cognate ligand[75,381]. For all these classes of riboswitches that have identical crystal structures for both the unbound and the bound states, how do the cognate ligands that are buried (~60-100%) within their ligand binding sites enter the binding pocket? For the riboswitches that are unstructured and adopt a structure only in the presence of a ligand, how would the ligand recognize such a disorganized binding pocket? These questions are not purely academic. Considerations of RNA dynamics and flexibility of the ligand binding site are a prerequisite for effective drug design strategies[382].

To better characterize the functional mechanism of how a riboswitch in the unliganded state eventually finds its globally minimized bound state, we probed a SAM-II riboswitch from the Sargasso Sea metagenome[383,384]. Previous x-ray crystal structural analyses indicated this RNA forms a classic H-type pseudoknot with the SAM metabolite ~64% buried in a binding pocket located in the center of helix P2b (Figure 5.1)[384], and other studies indicated the unliganded SAM-II riboswitch likely exists in multiple conformations[98,385,386]. SAXS data indicated that the RNA undergoes marked conformational changes and compaction in Mg^{2+} without the need for the SAM metabolite; yet NMR titration data indicated it is only after metabolite binds that various secondary and tertiary structural

elements form and sequester the Shine-Delgano (SD) sequence (Figure 5.1B), the ribosomal recognition site for bacteria, of the RNA[98]. This conundrum is analagous to the observation that many biological processes sample intermediate states that are only transiently and sparsely populated; therefore, they are "invisible" to traditional NMR and X-ray crystallography techniques[251,387,388].

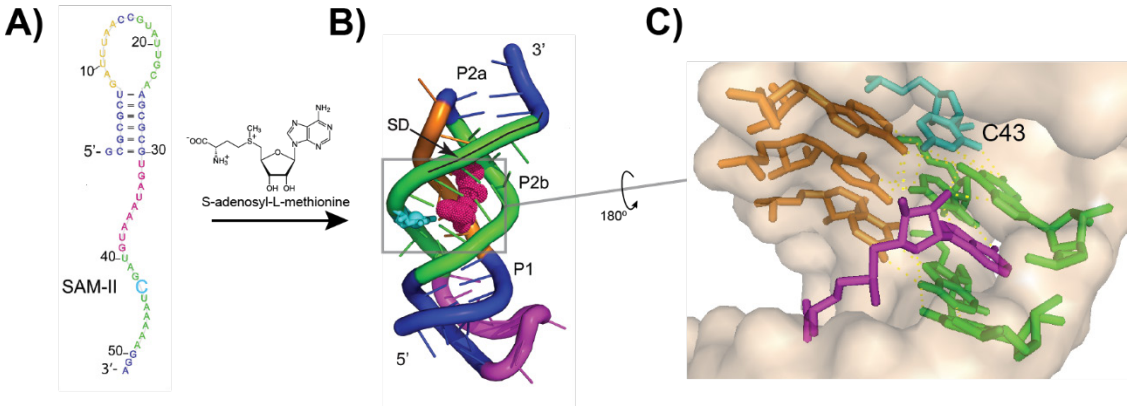

Figure 5.1. SAM-II secondary structure and 3D structure. A) The purposed secondary structure of SAM-II in the free-form. Residue C43 is highlighted in blue. B) Upon addition of S-adenosyl-L-methionine, SAM-II folds into a pseudoknot structure and sequesters the SD sequence and ligand. C) Residue C43 is labeled, which forms a critical base-triple within the binding-pocket. Each region of the SAM-II riboswitch is color-coded to match the secondary and crystal structures and residue C43 is highlighted in blue within each SAM-II representation (2QWY)[384].

To address this puzzle, we combined ligation, new chemo-enzymatic isotopic labeling strategies and newly developed NMR dynamics (Chemical Exchange Saturation Transfer, CEST) experiments that probe slow µs-ms motions to detect these otherwise invisible states[379,380,389,390]. To the best of our knowledge, this is the first study reporting the measurement of pyrimidine C6 ^{13}C-

and ^{1}H-CEST relaxation data in RNA without complications from NOE, thus extending prior purine and ribose ^{13}C-CEST studies on RNA[379,380,389,390].

5.2 CEST theory and background

Building on the concept of chemical exchange, we will introduce the NMR method CEST for measuring slow conformation exchanges. As previously explained, nuclei in the slow-exchange regime will produce two distinct NMR chemical shifts. The minor state chemical shift is sometimes visible unless the population is very small or the R_2 of the minor state is much larger than the major state. In order to identify these "invisible" states, Lewis Kay and colleagues developed the CEST experiments to investigate amide backbone transitions between major and minor chemical states[379]. We have adapted this method to the study of pyrimidine C6 and C1' resonance of RNA[210].

The basic principles of CEST are as follows: a weak B_1 field is applied to the nuclei of interest, which is aligned along the B_o field axis, at various offset frequencies and then after a constant time period the spin is returned to the transverse plan for detection. Changes in the intensity of the observable major peak are plotted against the B_1 offset to obtain a CEST graph. The CEST data are then fit to the appropriate homogeneous solution to the Bloch-McConnell equation[296,295] in order to extract k_{ex}, $\Delta\omega$, and p_b – the exchange rate, chemical shift difference, and population of the minor state, respectively.

5.2.1 NMR pulse program for CEST studies

The CEST experiment performed within this chapter is given in Figure 7.2. Based on a previously published CEST experiment for ^{13}C RNA studies by Zhang and colleagues[390], we wrote a CEST experiment that did not require selective pulses to remove ^{13}C-^{13}C scalar coupling since our isotopic labeling strategy provided us with isolated spin systems. After a d1 recovery delay and crusher gradient to remove residual ^{13}C magnetization, the pulse scheme starts with a double-INEPT period to transfer H_z magnetization completely to C_z magnetization (a→b in Figure 7.2). During the CEST period ^{13}C-^{1}H coupling is quenched by ^{1}H decoupling (b→c). A weak B_1 field is applied during the CEST period at various offsets for a time period of T_{EX}. After the CEST period, indirect measurement of the ^{13}C chemical shift occurs during t_1 evolution and the magnetization is spin-echo encoded (c→d) followed by a double INEPT transfer back to ^{1}H (d→e). Finally the gradient echo is decoded for sensitivity-enhanced coherence selection[241] and ^{1}H signal is detected with ^{13}C decoupling (e→f).

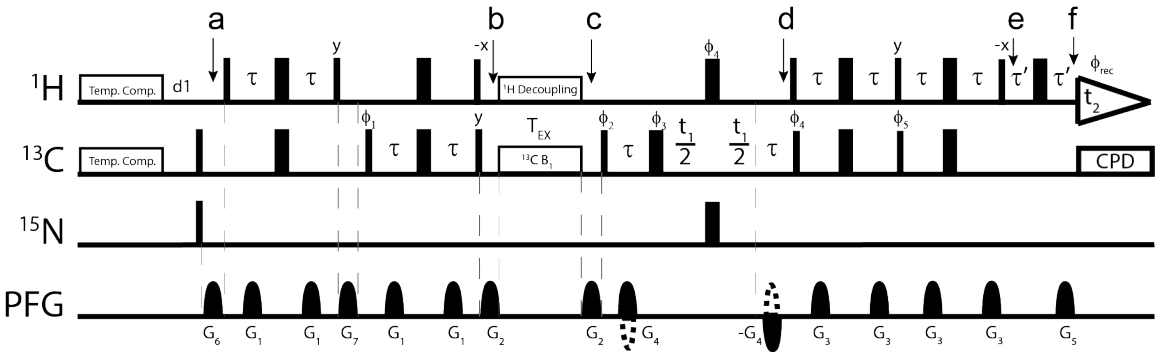

Figure 5.2. HSQC ^{13}C CEST pulse program. The ^{13}C CEST-HSQC pulse scheme adapted from Zhang and co-workers[288]. All narrow (wide) bars represent 90° (180°) hard pulses. The open rectangular pulses are constant wave decoupling pulses used for proton decoupling (^{1}H Decoupling), CEST spin-lock on the carbon channel (^{13}C B_1), and carbon decoupling during acquisition (CPD). The temperature compensation pulses (Temp. Comp.) are only used for the reference spectra, which do not include the T_{EX} period (T_{EX} = 0 s), to ensure consistent temperature between the reference and spin-lock experiments and are set to the same duration and power levels as ^{1}H Decoupling and ^{13}C B_1 pulses except that they are set far off-resonance (50 kHz). For the ^{13}C CEST experiments with broadband ^{1}H decoupling, the previously reported decoupling pulse program ($90_x240_y90_x$ composite pulse at 3.5 kHz) was implemented whereas for the weak ^{1}H decoupling experiments for determination of proton chemical shifts of the minor state, a constant wave weak field is required (1000 Hz) which can be optimized to reduce or increase the $^1J_{CH}$ splitting of the major and minor dips of the CEST profile as needed. Unless otherwise indicated, all phases are 'x'. The following phase cycle is used: φ_1 = x, -x; φ_2 = y; φ_3 = 2(x), 2(y), 2(-x), 2(-y); φ_4 = 4(x), 4(-x); φ_5 = 4(y), 4(-y); φ_{rec} = x, 2(-x), x, -x, 2(x), -x. The gradients optimized for the pulse program are as follows: G_1 = 33%, 0.8 s; G_2 = 46.2%, 0.8 s; G_3 = 46.2%, 0.8 s; G_4 = -59.4%, 0.6 s; G_5 = 29.9%, 0.6 s; G_6 = 33%, 0.8 s; G_7 = 46.2%, 0.8 s. For the echo/anti-echo gradient selection, φ_5 is incremented 180° and the G_4 encoding gradients are inverted. For States-TPPI phases φ_2 and φ_{rec} are incremented 180° with t_1. The τ delay is optimized for one-bond J_{CH} of either base or ribose residues (180-220 Hz or 160 Hz, respectively). Note that depending on the RNA size, relaxation during INEPT transfers require τ delays to be shorter than $1/(4J_{CH})$ and should be optimized for each sample. In this work, τ was set to 1.39 ms and 1.56 ms for the C6 and C1' experiments, respectively. The τ' delay in the final spin-echo is used to compensate for the decoding gradient (G_5) and was set to 600 µs. The recovery delay, d1, was set to 1.5 s and can be set to optimal signal-to-noise per unit time for a particular sample. The CEST spin-lock period (T_{EX}) was set to 200 ms and 300 ms for the C6 and C1' experiments, respectively.

5.2.2 *Effects of a weak B_1 field and offset on tilt angle*

At the beginning of the CEST period, the magnetization is aligned along the z-axis for the nuclei of interest. The application of a weak B_1 field will tilt the magnetization according the strength of the B_1 (ω_1) and the offset ($\Delta\Omega$) which is the difference between the resonance frequency of the nuclei and the frequency of the applied B_1 field. The tilt angle can be calculated from equation (5.1), the same as for $R_{1\rho}$ off-resonance experiments.

$$\theta = \tan^{-1}\left(\frac{\omega_1}{\Delta\Omega}\right) \qquad (5.1)$$

As shown in Figure 7.2A, applying a B_1 field on the x-axis will rotate the magnetization from 'z' to 'y' by an angle described by equation 7.1. The tilt angle for a weak B_1 like in Figure 5.3B (15 Hz) will only flip the spin to the y-axis very near the frequency of the spin as seen by the sharp peak at zero offset. If the graph in Figure 5.3B was inverted, this would be the CEST profile where a minimum is found at the resonance of the peak of interest.

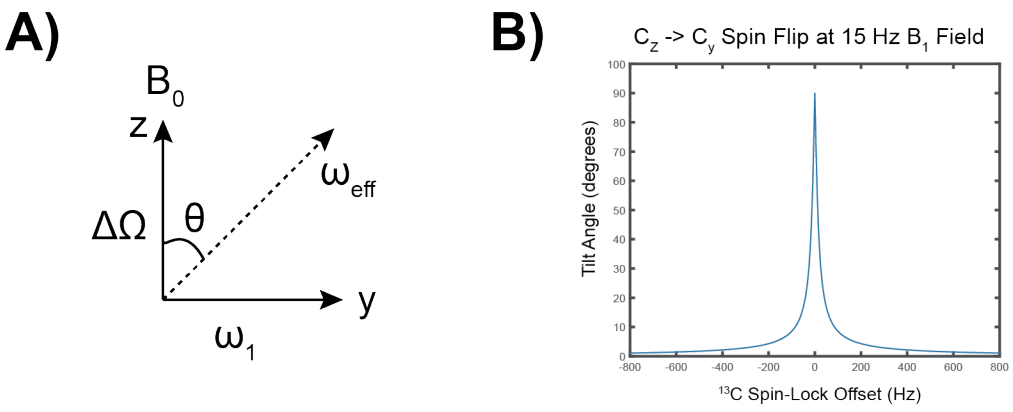

Figure 5.3. Tilt angle dependence of B_1 field and offset. A) The tilt angle for a spin along the z-axis to the y-axis is defined by the applied B_1 field (ω_1) and the offset ($\Delta\Omega$). B) The tilt angle is plotted against the spin-lock offset for a 15 Hz B_1 field. Maximum transfer to y-magnetization appears on-resonance.

5.2.3 Simulation of a CEST experiment

In order to better describe the CEST experiment, we have provided a set of simulations using a hypothetical slow exchanging system. Using the purposed exchange between Model 1 and Model 9 of the bacterial A-site RNA, we will suppose that the exchange rate is slow $k_{ex}/\Delta\omega << 1$ (Figure 5.4A). If the major state (Model 1) has a ^{13}C resonance at 91 ppm and ^{1}H resonance at 5 ppm and the minor state (Model 9) has a ^{13}C resonance at 89 ppm and ^{1}H resonance at 6 ppm, we will see one major state peak (purple – Figure 5.4B), but we will not see the minor "invisible" state (blue dashed-line circle – Figure 5.4B). If we sweep a weak B_1 field at different ^{13}C frequencies, we will detect a minor "dip" at the same frequency as the "invisible" state (Figure 5.4C). This is caused by the fact that the major state will exchange to the minor state during the CEST period and become "on-resonance" with the B_1 field. The small fraction of major state nuclei that exchange to the minor state during the CEST period will all completely flip to the transverse plane and their signal will decay from the effects of saturation. The measured intensity of the major peak, therefore, will decrease compared to a reference experiment in which no CEST period depending on if the B_1 field is off- or on-resonance, respectively, with the major or minor state. This phenomenon leads to the CEST profile as seen in Figure 5.4C.

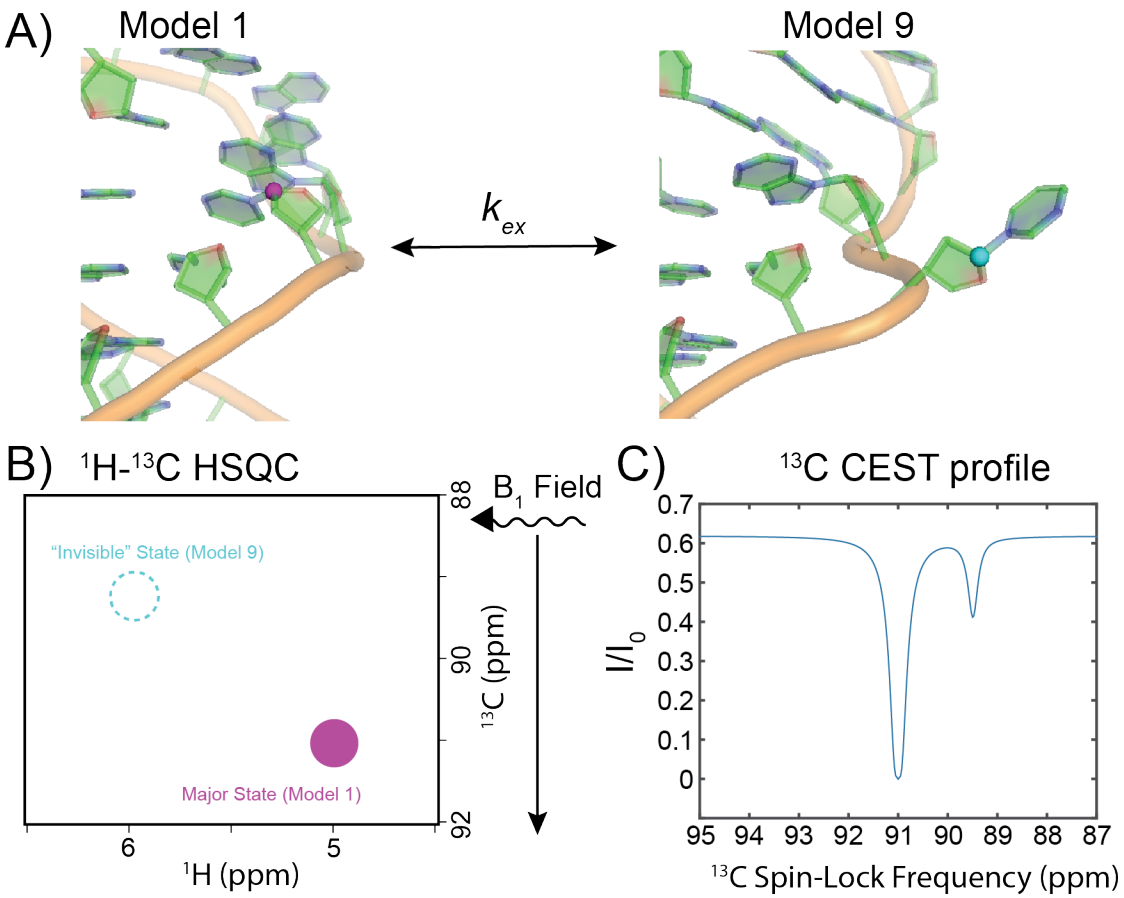

Figure 5.4. ^{13}C CEST simulation. A) Bacterial A-site model 1 and model 9 representing the closed major state and U23 "flipped out" minor state. A color coded sphere is placed at the C1' to represent the C-H spin system. B) Color-coded HSQC peaks correspond to the H1'-C1' spin system of U23 for the "invisible" state (sky blue) and the major state (purple). A squiggly line is drawn to represent the application of the weak B_1 spin-lock field which is incrementally swept over various offset frequencies. C) A representative CEST profile is given showing the relative peak intensity of the purple peak with respect to the frequency of the applied weak B_1 field. The CEST profile was simulated using the homogenous solution to the Bloch-McConnell equation assuming two-state exchange between state A, population 90%, and state B, population 10%, with a chemical shift difference of 2 ppm and an exchange constant of 30 s^{-1}.

5.2.4 Indirect measurement of ^{1}H chemical shift with ^{13}C CEST.

Direct application of CEST to H_z magnetization does not yield practical experimental data due to NOE transfers that arise from neighboring protons. Lewis Kay and colleagues, cleverly, applied a method of weak ^{1}H decoupling during the ^{13}C CEST period in order to measure changes in J_{CH} of the major and minor dips of the CEST profile. In order to illustrate the effects of weak ^{1}H decoupling, we have simulated what CEST profiles would look like for the hypothetical A-site system we purposed previously when weak ^{1}H decoupling fields are applied at different offset frequencies (Figure 5.5A). Due to the weak ^{1}H decoupling, the J_{CH} splitting is resolved for both dips when the frequency of the decoupler is far off-resonance (J_{CH} was set to 160 Hz for our simulation). As the decoupler is moved closer to the ^{1}H resonance of the corresponding state, the J_{CH} splitting of its dip in the CEST profile is quenched. By plotting the observed J_{CH} splitting against the decoupler frequency, we detect a minimum (complete quenching of the splitting) at a frequency exactly equivalent to the ^{1}H chemical shift of that state (Figure 5.5C). This clever and powerful tool for detecting ^{1}H chemical shift of “invisible” states allowed us to fully sense the chemical environment of the sparsely populated SAM-II conformation.

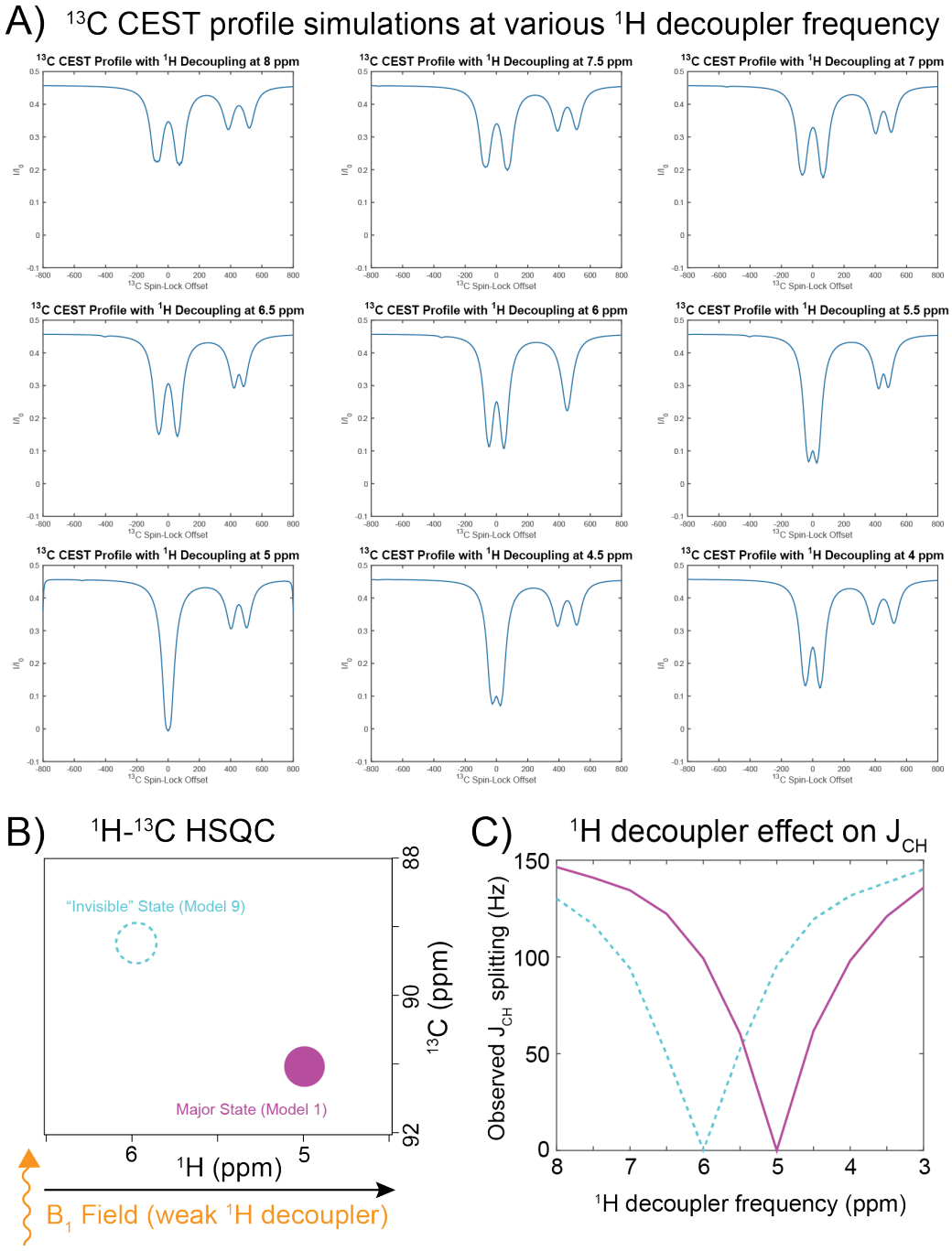

Figure 5.5. Simulated weak ^{1}H decoupled ^{13}C CEST profiles. A) Series of ^{13}C CEST profiles with the same parameters as Figure 7.4 except with a weak ^{1}H decoupling field (1000 Hz) applied at different ^{1}H offsets. The J_{CH} splitting for each peak is easily identified with the splitting quenched at 5 ppm for the major state and 6 ppm for the minor state. B) HSQC showing where the ^{1}H decoupling field is placed (orange squiggly line) for each ^{13}C CEST profile recorded. C) A plot of the measured J_{CH} of the minor and major dips versus the ^{1}H decoupler frequency. Notice the dips correspond to the ^{1}H chemical shift of each state.

5.3 Methods

5.3.1 RNA preparation and ligation

To probe the formation of the metabolite binding pocket within helix P2b, our colleague, Bin Chen, prepared a riboswitch variant using two RNA fragments prepared by *in vitro* transcription using standard T7 RNA polymerase reaction protocols[198] (Figure 5.6). An unlabeled 31 nt acceptor fragment was ligated to a 21 nt donor fragment harboring ^{13}C labels at positions 1' and 6 of residue C43. The 21 nt donor fragment was synthesized from a DNA template containing the HδV ribozyme at the 3'-end to ensure proper 3'-OH termini needed for ligation. Mass spectra confirmed the synthesis of the two homogenous RNA fragments (Figure A.12). Overall we obtained a ligated site-specific labeled product with typical ligation yields (based on the donor fragment) of ~80% (Figure 5.6B). Full experimental protocols can be found in the Appendices.

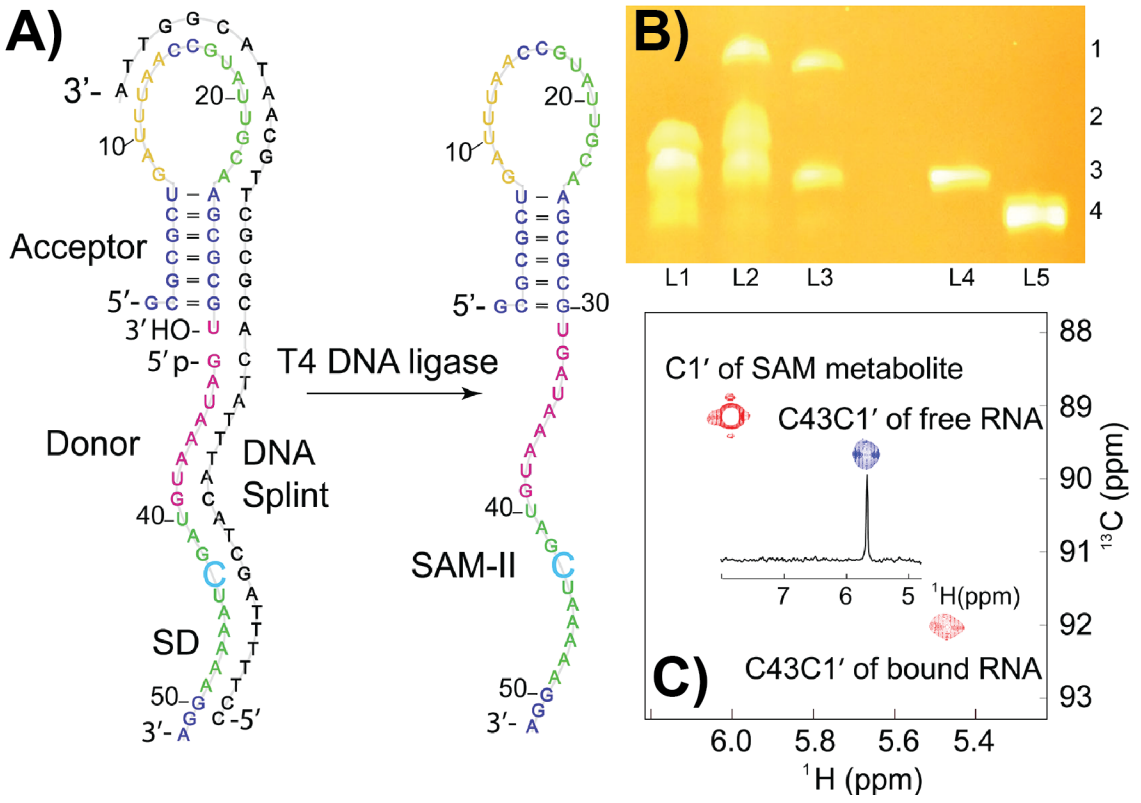

Figure 5.6. SAM-II site selective labeling through ligation. A) Scheme for the preparation of the segmentally and selectively ^{13}C- labeled, *in vitro* transcribed SAM-II RNA. The C43 residue that is selectively ^{13}C- labeled is shown in cyan. B) A 12% denaturing PAGE showing the ligation results. L1: ligation reaction at 0 h; L2: after 3 h; L3: after DNase digestion; L4: acceptor fragment; L5: donor fragment. Numbers on the right of the panel, 1: ligated product (SAM-II); 2: DNA splint; 3: acceptor fragment; 4: donor fragment. Addition of the DNA splint changed the migration rate of all RNA components in the reaction system. The selective-CTP-labeled donor fragment was almost completely converted to the ligated SAM-II RNA. (C) Overlay of 2D ^{1}H/^{13}C HSQC spectra of the site-specific labeled SAM-II riboswitch in the absence (blue) or presence (red) of the metabolite. The intense peak at ~89 ppm corresponds to a ribose C1' atom of the SAM metabolite. Ligation results kindly provided by Bin Chen[210].

This selective labeling and ligation approach provided three advantages: unambiguous assignment of C43's C6-H6 and C1'-H1' resonances without the need of tedious NMR resonance assignment strategies; direct monitoring of the base-pair interaction between G22 and C43, as C43 forms the floor of the SAM

metabolite binding pocket (Figure 5.1 and Figure 5.7); and straightforward setup and analysis of ^{1}H- and ^{13}C-CEST NMR experiments without complications from strong $^{1}J_{C5C6}$ coupling constant artefacts[379,380,389,390].

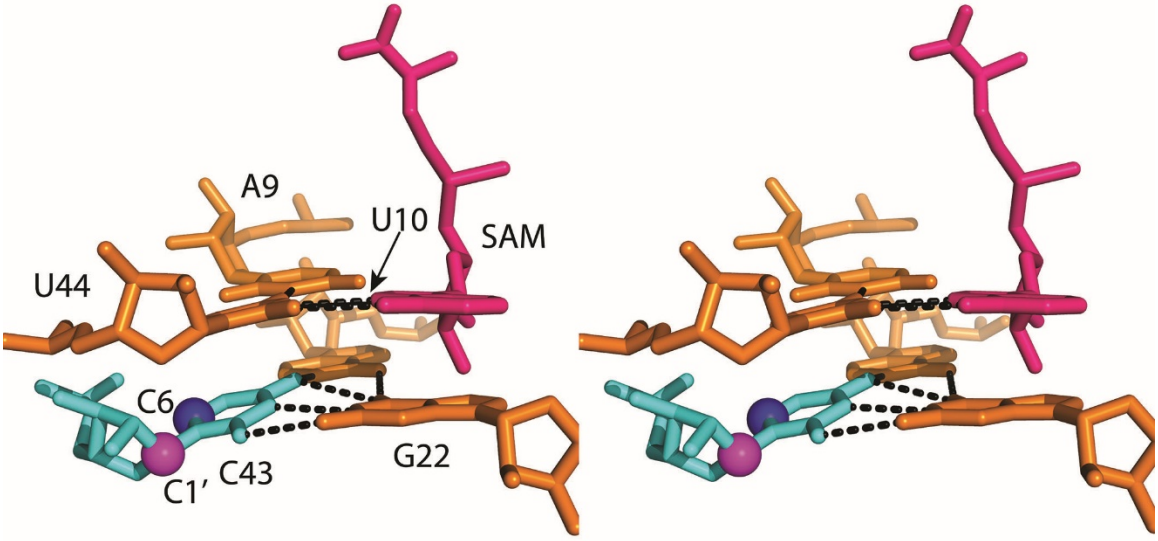

Figure 5.7. Selective label probes in SAM-II binding pocket. A zoomed in view of the SAM-II binding pocket with ligand. The placement of the isotopic labels on C43 are highlighted at the C6 (blue) and C1' (purple) positions.

5.3.2 ^{13}C CEST NMR data collection

CEST data were collected using a pseudo-3D HSQC experiment in which the B1 field offset was incremented in an interleaved manner. The 2D HSQC provided higher S/N compared to the 1D version of the experiment. A total of 1024x16 complex points with 41 spin-lock offsets (-600 to 600 Hz at 30 Hz increments) and 3 references (no CEST period) were recorded with 32 transients and a recovery delay of 1.5s for a total experimental time of approximately 12 hr for each spin-lock field. A CEST saturation period (T_{EX}) of 300 ms was used for the ribose and 200 ms for base due to the differing T_1 rates.

5.3.3 ^{1}H weakly decoupled CEST data collection

The same parameters were used for the proton off-resonance decoupling experiment, except that the proton decoupler was placed at different off-sets with a weaker field strength (1000 Hz). Since the minor dips were well resolved and known previously from the ^{13}C CEST profiles of both C1' and C6, we choose to take more offsets around the dips with increments of 20 Hz and less points at the extremes of the CEST profile. Thus, we could record 42 spin-lock offsets between -600 to 600 Hz with 40 Hz increments except between -120 to 300 Hz where 20 Hz increments were used to improve resolution of the splitting of the major and minor dips in the CEST profile.

5.3.4 B_1 field calibration

Spin-lock fields were calibrated using previously established methods for low power fields[284]. It should be noted that in this methods paper, the authors incorrectly identify the ω_1 value in Hz units rather than rad/s for the calculation of delays needed to calibrate the B_1 field. The peak intensities of a cross-polarization on-resonance spin-lock experiment recorded at various time intervals are fit to equation (5.2)

$$I(t) = I(0) * e^{(-R2*\tau)} * \cos(\omega_1 * \tau * \varphi) \qquad (5.2)$$

Where $I(t)$, $I(0)$, $R2$, τ, ω_1, and φ are the measured peak intensity, fitted intensity at time zero, the transverse relaxation rate, the spin-lock delay, the spin-lock field strength (radians/s), and the shift compensation, respectively. A non-linear least squares fitting of the parameters to 33 data points at incremented spin-lock delays provides an accurate calibration of the spin-lock field (Figure

5.8). Best fit values of B_1 are determined at multiple power levels in order to properly calibrate the B_1 fields required for accurate extraction of k_{ex}, p_b, and Δω values from the CEST profiles.

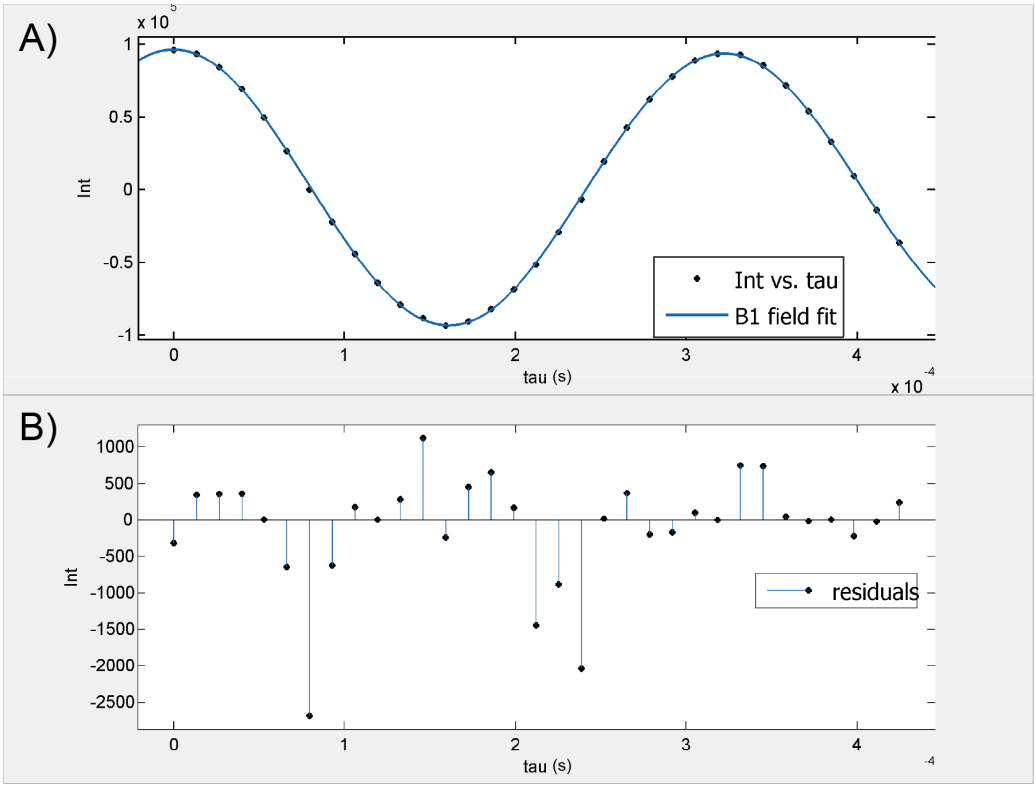

Figure 5.8. B_1 field calibration. A) The measured peaks intensities from a 1D experiment to calibrate the B_1 field. Peak intensities versus spin-lock delay time (tau) was fit using equation (5.2). B) The residuals from the best fit show that points near zero intensity were most likely underestimated. 95% confidence intervals for the fitted values gave B_1 fields with less than 3% variation.

5.4 Analysis of CEST profiles

5.4.1 Analysis of ^{13}C CEST data

To quantitatively extract the carbon chemical shift (Δω), the exchange rate, and the population of the minor state, a two-state model was fitted to each of the three profiles of the ribose carbon (C1') for C43 using the Bloch−McConnell 7x7

matrix[391]. Then global fits were performed by combining both the C1' and C6 data sets and fitting one k_{ex} and p_B value for both data sets while allowing R_2, R_1, and $\Delta\omega$ to be fit to C1' and C6 individually.

The CEST data was plotted as ***I(t)/I(0)*** versus spin-lock offset (Hz) and was fit by numerically solving the matrix exponential for the CEST spin-lock period of a 7x7 two-state Bloch-McConnell equation by non-least square linear regression as implemented previously by Zhang and coworkers[288]. The error in the fits was reported from 200 Monte Carlo simulations.

The 7x7 two-state Bloch-McConnell equation is derived from the relaxation matrix and the kinetic rate matrix for an exchanging two-state system. The addition of the equilibrium magnetization (***E/2***) provides a homogeneous solution[296] which describes the effect of a radio-frequency pulse over a period of time as given by equation (5.3):

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} E/2 \\ A_x \\ A_y \\ A_z \\ B_x \\ B_y \\ B_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -R_2^A - k_{AB} & -\omega_A & \omega_1 & k_{BA} & 0 & 0 \\ 0 & \omega_A & -R_2^A - k_{AB} & 0 & 0 & k_{BA} & 0 \\ 2R_1^A p_A & -\omega_1 & 0 & -R_1^A - k_{AB} & 0 & 0 & k_{BA} \\ 0 & k_{AB} & 0 & 0 & -R_2^A - k_{BA} & -\omega_B & \omega_1 \\ 0 & 0 & k_{AB} & 0 & \omega_B & -R_2^A - k_{BA} & 0 \\ 2R_1^B p_B & 0 & 0 & k_{AB} & -\omega_1 & 0 & -R_1^A - k_{BA} \end{bmatrix} \begin{bmatrix} E/2 \\ A_x \\ A_y \\ A_z \\ B_x \\ B_y \\ B_z \end{bmatrix} \quad (5.3)$$

In equation (5.3), $R_2^{A/B}$, $R_1^{A/B}$, $\omega_{A/B}$, and ω_1 are the transverse relaxation rate of the **A/B** state, the longitudinal relaxation rate of the **A/B** state, the offset of the B_1 spin-lock field from the peaks in the **A/B** state in radians/second, and the B_1 field strength in radians/second, respectively. The **A** state offset (ω_A) is known from the measured chemical shift of the observed peak in the **A** state whereas the **B** state offset (ω_B) is set to $\omega_B = \omega_A + \Delta\omega$ ($\Delta\omega$ is the chemical shift

difference between the **A** and **B** state). The rate for $\boldsymbol{A} \rightarrow \boldsymbol{B}$ is given as k_{AB} and the rate for $\boldsymbol{B} \rightarrow \boldsymbol{A}$ is given as k_{BA}. To calculate the evolution of magnetization for the peak in the **A** state during the CEST spin-lock period, we solve for $M(t)$ using the matrix exponential given in equation 7.4. First, we can rewrite equation (5.3) as $\frac{\mathrm{d}}{\mathrm{d}t}M(t) = L \times M(0)$ in which $M(0)$ is the initial magnetization vector $(E/2 \quad A_x \quad A_y \quad A_z \quad B_x \quad B_y \quad B_z)^+$,+ is the transpose, and L is the 7x7 Bloch-McConnell matrix.

$$M(t) = M(0) \times \exp(-L * t) \qquad (5.4)$$

The initial magnetization $M(0)$ is assumed to be the equilibrated population distribution of the spins along the z-axis as prepared before the applied spin-lock field (CEST period) and is given by $(0 \quad 0 \quad 0 \quad p_A \quad 0 \quad 0 \quad p_B)^+$. Therefore, the experimental data ***I(t)/I(0)*** corresponds to ***M(t)/M(0)*** for the non-linear least square regression.

5.4.2 Analysis of ^{1}H indirect ^{13}C CEST data

Smaller increments in offsets were taken around the minor dips in order to better distinguish the J_{CH}-splitting. The CEST profiles were fit using Matlab to numerically solve equation (5.4) with L replaced by the 31x31 two-state Bloch-McConnell matrix[380]. The 31x31 two-state Bloch-McConnell matrix is derived from the relaxation matrix for the full basis set of a two-spin (***I and S***) coupled system (equation (5.5)).

$$
R^{A/B} = \begin{bmatrix}
R_{Ixy}^{A/B} & \omega_I^{A/B} & 0 & 0 & 0 & 0 & \eta_{Ixy}^{A/B} & \pi J_{IS} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\omega_I^{A/B} & R_{Ixy}^{A/B} & \omega_{1I} & 0 & 0 & 0 & -\pi J_{IS} & \eta_{Ixy}^{A/B} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\omega_{1I} & R_{Iz}^{A/B} & 0 & 0 & \sigma^{A/B} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta_{Iz}^{A/B} \\
0 & 0 & 0 & R_{Sxy}^{A/B} & \omega_S^{A/B} & 0 & 0 & 0 & \eta_{Sxy}^{A/B} & \pi J_{IS} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\omega_S^{A/B} & R_{Sxy}^{A/B} & \omega_{1S} & 0 & 0 & -\pi J_{IS} & \eta_{Sxy}^{A/B} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma^{A/B} & 0 & -\omega_{1S} & R_{Sz}^{A/B} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta_{Sz}^{A/B} \\
\eta_{Ixy}^{A/B} & \pi J_{IS} & 0 & 0 & 0 & 0 & R_{2IxySz}^{A/B} & \omega_I^{A/B} & 0 & 0 & 0 & -\omega_{1S} & 0 & 0 & 0 \\
-\pi J_{IS} & \eta_{Ixy}^{A/B} & 0 & 0 & 0 & 0 & -\omega_I^{A/B} & R_{2IxySz}^{A/B} & 0 & 0 & 0 & 0 & 0 & -\omega_{1S} & \omega_{1I} \\
0 & 0 & 0 & \eta_{Sxy}^{A/B} & \pi J_{IS} & 0 & 0 & 0 & R_{2IzSxy}^{A/B} & \omega_S^{A/B} & 0 & 0 & -\omega_{1I} & 0 & 0 \\
0 & 0 & 0 & -\pi J_{IS} & \eta_{Sxy}^{A/B} & 0 & 0 & 0 & -\omega_S^{A/B} & R_{2IzSxy}^{A/B} & 0 & 0 & 0 & -\omega_{1I} & \omega_{1S} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & R_{IxySxy}^{A/B} & \omega_S^{A/B} & \omega_I^{A/B} & -\mu_{\mathrm{mq}}^{A/B} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \omega_{1S} & 0 & 0 & 0 & -\omega_S^{A/B} & R_{IxySxy}^{A/B} & \mu_{\mathrm{mq}}^{A/B} & \omega_I^{A/B} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \omega_{1I} & 0 & -\omega_I^{A/B} & \mu_{\mathrm{mq}}^{A/B} & R_{IxySxy}^{A/B} & \omega_S^{A/B} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \omega_{1S} & 0 & \omega_{1I} & -\mu_{\mathrm{mq}}^{A/B} & -\omega_I^{A/B} & -\omega_S^{A/B} & R_{IxySxy}^{A/B} & 0 \\
0 & 0 & \eta_{Iz}^{A/B} & 0 & 0 & \eta_{Sz}^{A/B} & 0 & -\omega_{1I} & 0 & -\omega_{1S} & 0 & 0 & 0 & 0 & R_{2IzSz}^{A/B}
\end{bmatrix} \tag{5.5}
$$

In equation (5.5), the relaxation matrix is given for the populated/sparse state (***A/B***). The diagonal represents the transverse relaxation of spin ***I***, the longitudinal relaxation rate of spin ***I***, the transverse relaxation rate of spin ***S***, the longitudinal relaxation rate of spin ***S***, the relaxation rate of the anti-phase components, the relaxation rate of multi-quantum components, and the relaxation rate of longitudinal 2-spin order. The heteronuclear spin system is also assumed to relax by dipole-dipole and chemical shift anisotropy interactions as represented by the off-diagonal terms. Equation (5.5) has been previously derived[296] and reported for weak field ^{1}H decoupled CEST experiments[380]. The full relaxation matrix is formed by expanding equation (5.5) by a direct product to a 30x30 to represent the **A/B** states as shown in equation (5.6).

$$L_{30} = \begin{bmatrix} R^A & 0_{15} \\ 0_{15} & R^B \end{bmatrix} + \begin{bmatrix} -k_{AB} & k_{BA} \\ k_{AB} & -k_{BA} \end{bmatrix} \otimes 1_{15} \qquad (5.6)$$

The equilibrium magnetization is added to create the 31x31 Bloch-McConnell matrix (L_{31}) used for fitting the ^{1}H decoupled CEST data. The magnetization of S_Z^A component is determined by solving the matrix exponential of equation 7.4 substituting in L_{31} and the corresponding initial magnetization vector, which is expanded to a 1x30 matrix to account for state **A** and state **B**:

$M^{A/B} = (I_x \quad I_y \quad I_z \quad S_x \quad S_y \quad S_z \quad 2I_xS_z \quad 2I_yS_z \quad 2I_zS_x \quad 2I_zS_y \quad 2I_xS_x \quad 2I_xS_y \quad 2I_yS_x \quad 2I_yS_y \quad 2I_zS_z)^{(A/B)+}$, and finally expanded to a 1x31 by including the equilibrium magnetization as the first term.

The Bloch-McConnell matrix (including the equilibrium magnetization terms) was fitted numerically using non-linear least squares in Matlab. The

exchange rate, minor-state population, longitudinal carbon relaxation rate, the transverse carbon relaxation rate of the most populated state (A), and the transverse carbon relaxation rate of the sparsely populated state (B) (k_{ex}, p_b, R_{Sz}, R_{Sxy}^A, R_{Sxy}^B) from the ^{13}C CEST experiments with broadband ^1H decoupling were set as 'known' values in the ^1H offset decoupled CEST fits.

5.4.3 Analysis of thermodynamics and kinetics from CEST data

To quantify the free energy landscape of SAM-II folding in the presence of magnesium ions, we applied the extracted populations and rate constants to the following equations (5.7) and (5.8), respectively. Equation (5.7) provides the free energy difference between the major and minor state. From transition state (TS) theory[284,392], the free energy difference between each state and the transition state (TS) can be obtained with Equation (5.8).

$$\Delta G = -RT \ln\left(\frac{p_b}{1-p_b}\right) \tag{5.7}$$

$$\Delta G_i^{\ddagger} = -RT \ln\left(\frac{k_i h}{T k_B \kappa}\right) \tag{5.8}$$

For equations (5.7)and (5.8), R is the gas constant, T is the temperature in Kelvin (298.1 K for all measurements in this work), k_B is Boltzmann's constant, κ is the transmission coefficient (assumed to be 1), h is Planck's constant, k_i is either the forward (i =1) or reverse rate (i =-1) for the transition from the major to the minor state, ΔG is relative free energy difference between the minor and major state, p_b is the population of the minor state, and $\Delta G_i^{\ddagger}$ is the free energy difference relative to the transition state for the major (i =1) or minor state (i =-1).

5.5 Results and discussion

A 2D ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectrum obtained for the ligated product in the absence of both Mg^{2+} ion and SAM exhibited a single cross peak that corresponds to ribose C1' of the residue C43 (Figure 5.6). In agreement with previous studies[98], the presence of Mg^{2+} ion did not cause any chemical shift change for either C1' or C6 resonances of C43. Upon addition of SAM metabolite, however, the ribose C1' carbon resonance shifted considerably from 89.5 ppm to 92.0 ppm, indicating that ligand binding alters the local chemical environment of the C43-C1' carbon atom. C43 residue, while flexible in the absence or presence of Mg^{2+} ions (similar longitudinal relaxation time T_1), becomes less flexible (increased T_1) in the presence of both Mg^{2+} ions and metabolite.

Next, application of ^{13}C-CEST to the ligand-free selective and nucleotide specific C-labeled SAM-II riboswitch, without Mg^{2+}, immediately revealed an “invisible” conformational state that is opaque to conventional NMR experiments (Figure 5.9). The ^{13}C-CEST profiles of the C43 ribose C1' atom of the riboswitch recorded at three different B_1 fields (17.5, 27.9, and 37.8 Hz) with a mixing time (T_{EX}) of 0.3 s at 298K indicate an intense dip at 89.5 ppm. This dip matches exactly the C43 labeled peak resonance position in the ^{1}H-^{13}C HSQC spectrum (Figure 5.6) consistent with the ligand-free major state (A) conformation with a characteristic CMP resonance free in solution[167]. Surprisingly, a minor dip corresponding to an “invisible” second state could be clearly identified in the profiles (Figure 5.9) at a chemical shift of 91.5 ppm ($\Delta\omega$=1.61 ppm). This second

dip matches the peak position in the ^{1}H-^{13}C HSQC spectrum (Figure 5.6C) and the chemical shifts are consistent with a ligand-bound RNA state (B) with characteristic of CMP resonances in helical RNA segments.

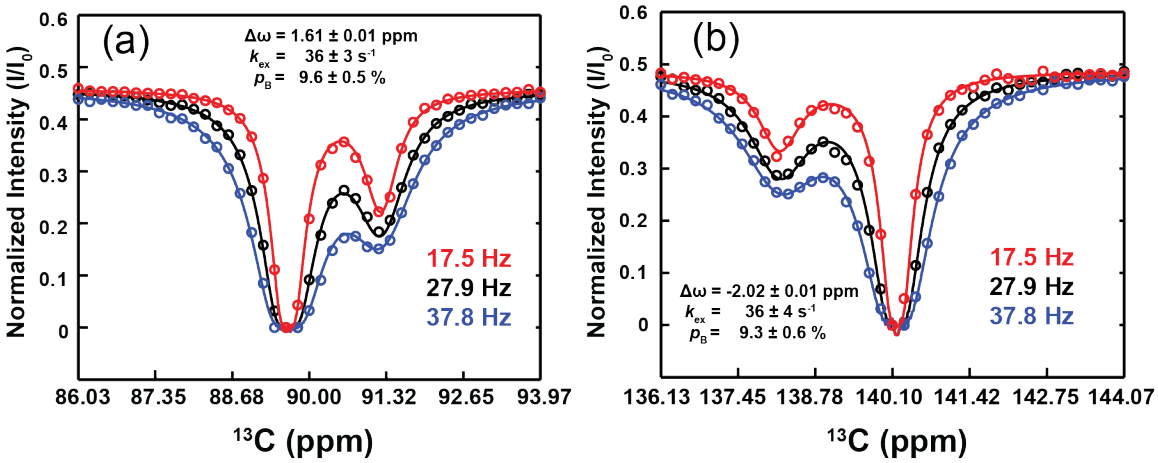

Figure 5.9. ^{13}C CEST profiles for ligand and Mg^{2+} free SAM-II. CEST experiments in the absence of Mg^{2+}. (a) ^{13}C-CEST profiles of the C43-C1' atom of the SAM-II riboswitch and (b) ^{13}C-CEST profiles of the C43-C6 atom of the SAM-II riboswitch. These were recorded under the same conditions: three different B_1 fields (17.5, 27.9, and 37.8 Hz) in NMR buffer (10 mM sodium phosphate, pH 6.2, 100 mM NaCl, 0.2 mM EDTA, 50 μM DSS, 0.1% sodium azide, 8% D_2O) at a temperature of 298K with T_{EX} = 0.3 s for C1' and T_{EX} = 0.2 s for C6.

Thus a simple visual inspection of the CEST profiles immediately provides structural insights into the nature of the invisible sparsely populated state.

To quantitatively extract the carbon chemical shift ($\Delta\omega$), the exchange rate, and the population of the minor state, a two-state model was fitted to each of the three profiles of the C43 ribose carbon (C1') and base carbon (C6) separately or together using the Bloch−McConnell 7x7 matrix[352]. Excellent fits were obtained with the two-state model, as demonstrated by the red, black, and blue curves that match well the experimental data (Figure 5.9). An exchange rate constant (k_{ex} =

$k_{AB} + k_{BA}$) of 36 ± 3 s^{-1} and the population of the "invisible" state B (P_B) ≅ 9.5 ± 0.5% were obtained from a global fit of both the C1' and C6 CEST data (Table 5.1) indicating that the C43 residue as a whole experiences similar chemical exchanges rather than undergoing local independent breathings. The lifetime of the state B (τ_B) is 31.0 ± 0.3 ms, which is comparable to that reported in a recent study[386]. The extracted chemical shift differences ($\Delta\omega = \omega_A - \omega_B$) of 1.6 and -2.0 ppm between A and B for C43 ribose C1' and base C6 atoms, respectively, provided important local structural information on the "invisible" state. Note that because of selective labeling technologies, the large $^1J_{CC}$ (~45-60 Hz) couplings are absent with no need to account for these couplings in the data analyses as was necessary in previous studies that used uniformly labeled samples[379,390,393,394].

Table 5.1. ^{13}C CEST fits for SAM-II without Mg^{2+}. Values of the rate of exchange (k_{ex}) and population of the transient state (p_B) obtained from fits (of all CEST profiles for residue C43 of the SAM-II riboswitch) to individual or a global two-site exchange processes. Data were collected on RNA in NMR buffer (10 mM sodium phosphate, pH 6.2, 100 mM NaCl, 0.2 mM EDTA, 50 μM DSS, 0.1% sodium azide, 8% D_2O) at 298 K, and 14.1 T

	Individual fitting		**Global fitting**	
	k_{ex} (s^{-1})	P_B (%)	k_{ex} (s^{-1})	P_B (%)
C1'	36 ± 3	9.6 ± 0.5	37 ± 2	9.2 ± 0.4
C6	36 ± 4	9.3 ± 0.6		

To obtain an added insight into the ^{1}H chemical environment of the minor state, the proton chemical shift was determined using a straightforward

modification ^{13}C CEST relaxation dispersion experiments with weak off-resonance ^{1}H decoupling (Figure 5.10)[380,395].

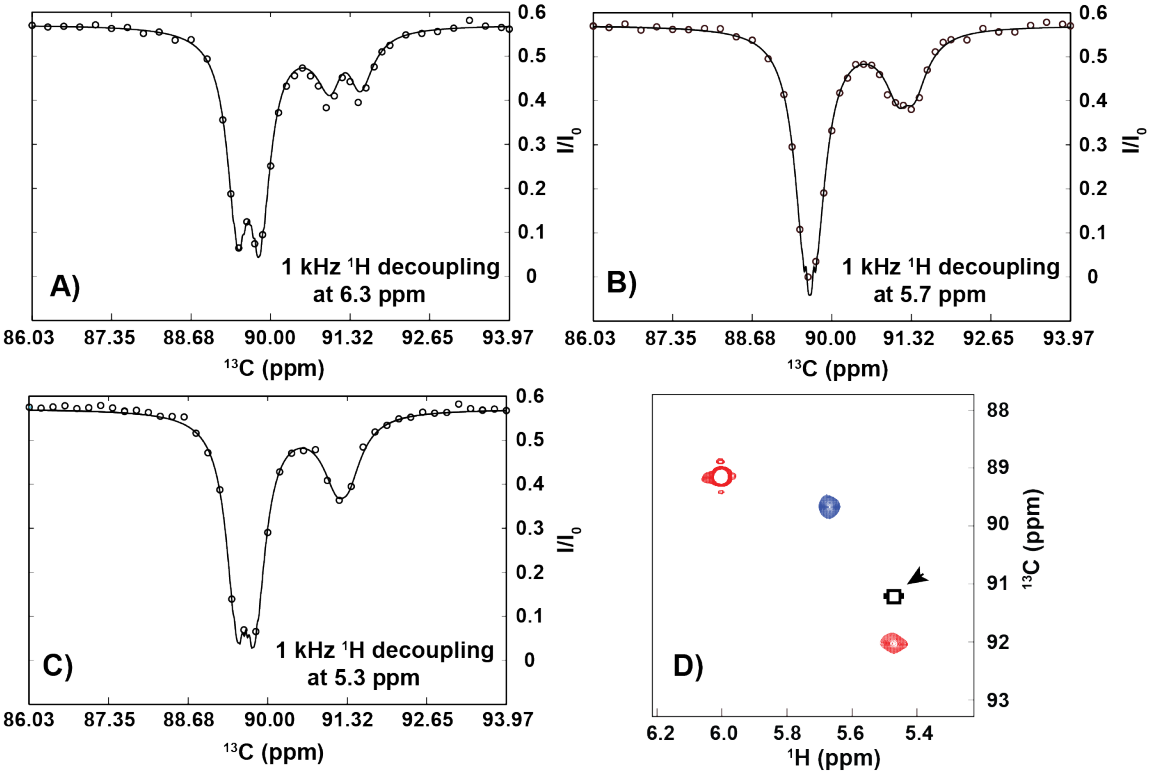

Figure 5.10. ^{1}H offset decoupled CEST profiles. Measurement of ^{1}H chemical shift of the C43-C1' resonance of the RNA in "invisible" minor state through off-resonance decoupling effects. Three experiments with weak (1000 Hz) proton decoupling were recorded at 6.3 (A), 5.7 (B), and 5.3 (C) ppm, respectively. (D) 2D ^{1}H/^{13}C HSQC spectrum of the site-specific labeled SAM-II riboswitch showing the crosspeaks of C43-C1' atom in the absence (blue) or presence (red) of the metabolite. An open box symbol indicated the position of the predicted crosspeak of the C43-C1' and C43-H1' resonances in B state with bars at the left or right of the box indicating the possible range of ^{1}H chemical shift. These were recorded under the same buffer conditions shown in Figure 7.9 legend.

Even though direct ^{1}H CEST experiments are hindered by nuclear Overhauser effects (NOEs) that give rise to artefactual dips in the ^{1}H CEST profile, indirect detection of ^{1}H chemical shifts measured through off-resonance decoupling reveals J_{CH} splitting pattern for the dips of the ^{13}C CEST profile of the

ribose C43-C1' resonance (Figure 5.9A and Figure 5.10). The J_{CH} couplings of the major and minor peaks are easily identified in the CEST profile and are dependent on the B_1 field and offset of the ^{1}H decoupler from the major and minor resonances (Figure 5.10).

The experimental data were analyzed using a two-state exchange Bloch-McConnell 31x31 matrix that describes the time-dependent evolution of magnetization during the CEST period[296,380]. Three experiments with weak (1000 Hz) proton decoupling were run at three ^{1}H offsets (6.3, 5.7, and 5.3 ppm). As expected, the splitting of the major peak is completely decoupled when the ^{1}H decoupler is set at 5.7 ppm. The minor state peak is only decoupled at 5.3 ppm and not at 6.3 or 5.7 ppm indicating that the minor state ^{1}H chemical shift is upfield of the major state ^{1}H chemical shift. An additional CEST experiment with no ^{1}H decoupling during T_{EX} was also recorded, showing complete splitting of both major and minor state peaks which overlap to form three distinct dips in the CEST profile (Figure A.13). All four experiments were globally analyzed using the Bloch-McConnell matrix with errors estimated from 200 Monte-Carlo simulations.

Using the same set of assumptions previously reported (the relaxation parameters of the sparsely populated state are equal to those of the most populated state)[380], the fitted values for the ^{1}H offset decoupled CEST experiment are the transverse relaxation rate of the proton, the transverse multiple quantum relaxation rate, the transverse and longitudinal cross-correlated

relaxation rates of carbon, and the ^{1}H chemical shift offset of the minor state (R_{Ixy}, $R_{2IxySxy}$, η_{Sxy}, η_{Sz}, and δ_H) respectively (Table 5.2).

Table 5.2. ^{1}H offset decoupled CEST fitting results. Extracted rate parameters and chemical shift change of C43 H1' proton for data collected under same buffer conditions specified in Table 7.1

Parameter	Value	Error
R_{Ixy}	18	1
δ_H (ppm)	-0.23	0.03
$R_{2IxySxy}$	18	1
η_{Sxy}	2	1
η_{Sz}	0.8	0.1
$RChi^2$	1.4	-

Smaller deviations from experimental data for the minor state dip indicate that the minor state probably has different relaxation properties from the major state. The assumption of similar relaxation properties between minor and major state will have little effect on the ^{1}H chemical shift value extracted since it is dependent on the splitting of the peak rather than the depth of the dip.

Again the NMR-invisible proton state is not only shifted into a proton frequency region expected for residues in a Watson-Crick base-pair configuration, but also the resulting ^{1}H chemical shift offset matches quite well the chemical shift of ligand bound SAM-II. Thus, the free SAM-2 riboswitch appears to sample a conformation similar to the bound state (Figure 5.10D). Collectively, these results suggest the riboswitch adopts at least two conformations: the apo SAM-II exists in a dynamic equilibrium between an open

highly populated (A,~90.5%) and a partially closed, sparsely populated (B,~9.5%) state (Figure 5.11).

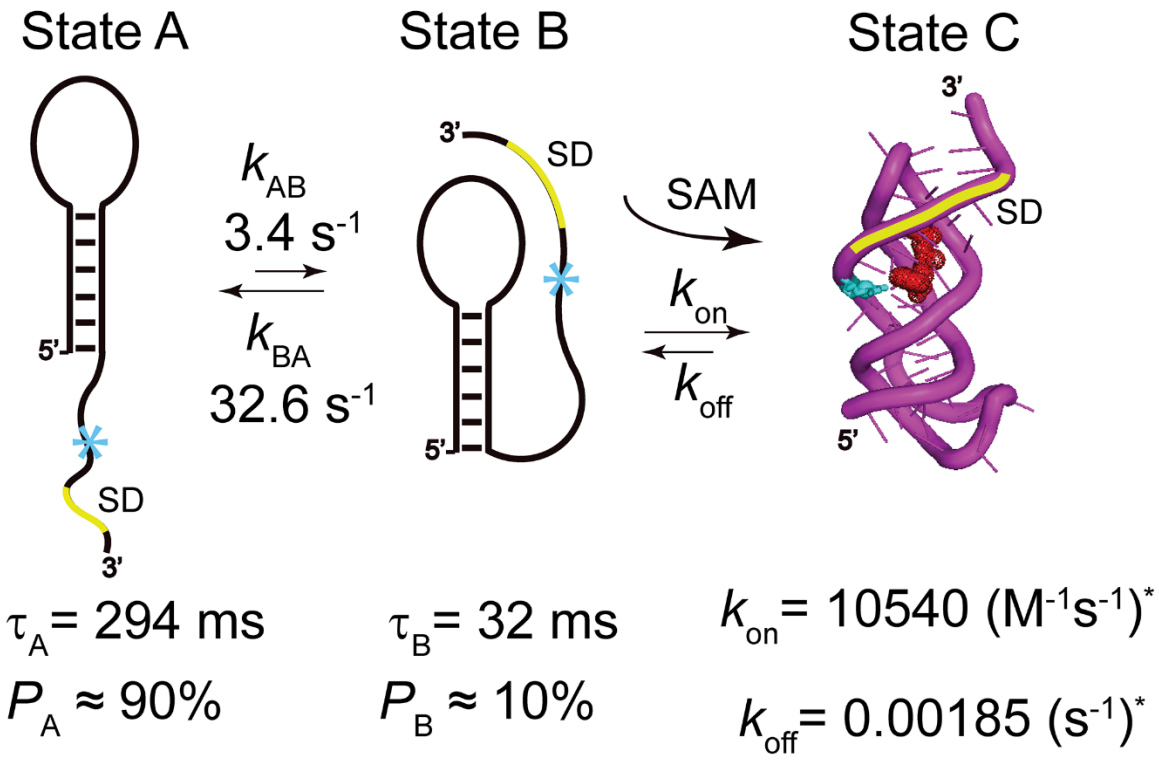

Figure 5.11. Multi-step ligand recognition model for SAM-II. Model illustrating the conformational exchange between states A, B, and C (the X-ray crystal structure of SAM-II riboswitch with a PDB ID: 2qwy)[384]. *Rates for k_{on} and k_{off} are the average of three previously reported pre-steady state fluorescence measurements at 10:1 (ligand:riboswitch) concentrations at 20°C by Haller and co-workers[386].

This apo state exists with a lifetime of ~31 ms. Addition of Mg^{2+} ions increases the sparsely populated B state (from ~11% at 0.25 mM to ~22% at 2 mM) and slightly decreases k_{ex} (28 s^{-1} at 0.25 mM to ~21 s-1 at 2 mM), thereby stabilizing the sparsely populated state by 0.6 kcal mol-1 (Figure 5.12 and Table 5.3). These results also nicely complement earlier MD simulations[385] and smFRET experiments[386] with similar predictions of conformational exchange and

match results we found for the pH stabilization of a minor population of bacterial A-site RNA.

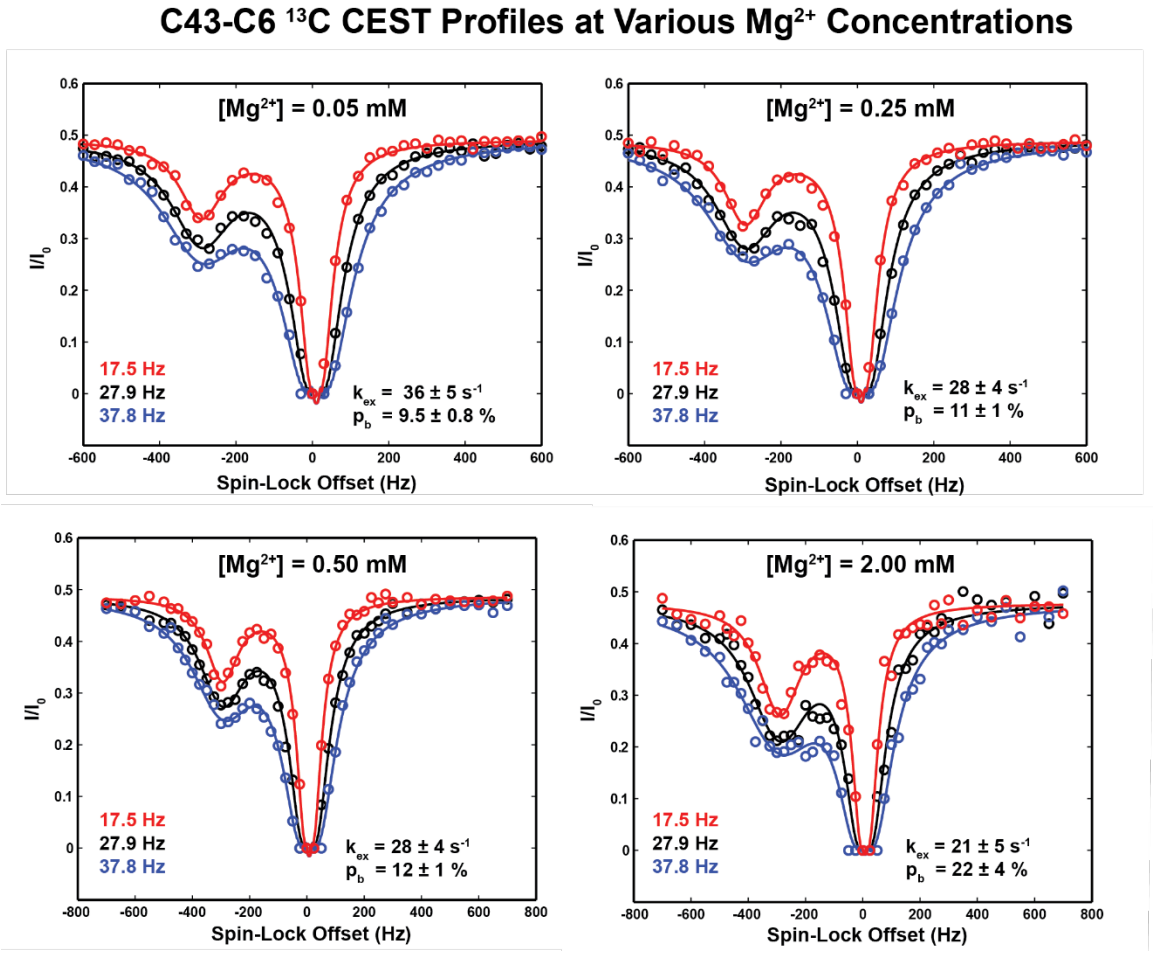

Figure 5.12. ^{13}C CEST profiles at various Mg^{2+} titrations. The ^{13}C CEST profiles for C43-C6 of SAM-II in the presence of different concentrations of Mg^{2+} ions. For each Mg^{2+} titration ^{13}C CEST profile was measured at three different B_1 spin-lock fields (17.5, 27.9, and 37.8 Hz). The fitted k_{ex} and p_b values are given for each Mg^{2+} concentration within the plotted regions. Note that the 2mM data is noisier than the rest because of poorer signal to noise-ratio.

Table 5.3. Thermodynamics and kinetics of Mg^{2+} induced conformational exchange. Extracted rate and free energy parameters for C43-C6 chemical exchange in the presence of Mg^{2+} where k_1 is the rate constant for the transition of the major state to the minor state and k_{-1} is the reverse rate constant. The free energy difference between the minor and major state is given as ΔG, the free energy difference between the transition state and the major state is $\Delta G^{\ddagger}_1$, and the free energy difference between the transition state and minor state is $\Delta G^{\ddagger}_{-1}$. Data were collected under the same buffer conditions specified in supplementary table 1 at different Mg^{2+} concentrations.

$[Mg^{2+}]$ (mM)	p_B (%)	k_1 (s^{-1})	k_{-1} (s^{-1})	ΔG (kcal mol^{-1})	$\Delta G^{\ddagger}_1$ (kcal mol^{-1})	$\Delta G^{\ddagger}_{-1}$ (kcal mol^{-1})
0	9.3 ± 0.6	3.3 ± 0.4	32 ± 4	1.4 ± 0.1	3.1 ± 0.4	1.8 ± 0.2
0.05	9.5 ± 0.8	3.4 ± 0.5	32 ± 5	1.3 ± 0.2	3.1 ± 0.5	1.8 ± 0.3
0.25	11 ± 1	3.1 ± 0.6	25 ± 5	1.2 ± 0.2	3.1 ± 0.6	1.9 ± 0.4
0.50	11 ± 1	3.3 ± 0.6	25 ± 5	1.2 ± 0.2	3.1 ± 0.6	1.9 ± 0.4
2.00	22 ± 4	4 ± 1	17 ± 5	0.8 ± 0.2	2.9 ± 0.9	2.2 ± 0.6

The previous smFRET study probed and observed only the formation of the terminal end (C16-G50) of the pseudo-knot in the absence and presence Mg^{2+} ions[386]. The binding pocket was not monitored directly[386]. However by focusing on the C43 residue that forms a base-triple with G22 and A9 within the SAM ligand binding pocket (Figure 5.1 and Figure 5.7), our results indicate that in the absence of metabolite increasing Mg^{2+} ions (0 to 2mM) increase a minor state (from 9.5% to 22%) that resembles the bound state[386,396]. While our observations do not rule out the possibility of alternative folding pathways, these results nonetheless suggest that this "bound-like" minor state likely follows the initial formation of the terminal end of the pseudoknot observed by Haller and co-workers. SAM-binding, therefore, likely captures this transient, lowly populated state[386] with minimum expenditure of energy[396] (Figure 5.11).

5.6 Conclusion

As anticipated earlier[75] and from this work and recent findings on various RNAs, the model that begins to emerge for RNA ligand binding is conformational selection coupled with induced fit (Figure 5.11)[82,251,274,284,288,397–399]. We anticipate that a number of regulatory RNA processes use a similar dynamic mechanism.

Future work will focus on how this minor state population is tunable by temperature, pH, ion and metabolite concentration as seen with many other RNA systems[205,274,399,400]. Of great interest, the SAM-II pseudoknot structure is able to stimulate -1 frameshifting[401]. Our results predict that the population of the Mg^{2+} induced minor conformation will likely correlate with frameshifting efficiency as a function of the SAM metabolite. These are currently being pursued by other graduate students in our laboratory. We also recently published a TROSY ^{13}C CEST experiment that could allow us to measure the RDCs of the minor state conformation and begin to build 3D models of the “invisible” SAM-II pre-bound conformation[191].

We have also investigated the roles of minor populations in another frame-shifting RNA element within the human coronavirus (SARS). We have added our preliminary research results applying CEST to this system in the Appendix (Figure A.12). The SARS frame-shifting element responds to pH to regulate the ratio of protein products coded within the same block of mRNA[402,403]. By constraining the pseudoknot RNA structure, mechanical forces lead to a slippage of -1 base-pair in the translating codon sequence resulting in the production of an

entirely different protein[397]. This finely tuned system plays a key role in the pathology of the SARS viral lifecycle[403]. Clearly, there is potential for the pseudoknot structure, which folds into a well-formed tertiary structure, to be targeted for anti-viral therapeutics. In fact, recent research shows a pseudoknot stabilizing compound may be an effective anti-viral[404]. Understanding how these pseudoknot structures bind ligands will only increase our ability to target them with anti-viral or antibiotic therapeutics.

6 Conclusion and future directions

6.1 Summary of work

Ribonucleic acid (RNA) is an under-utilized target in the search for new therapeutics. Yet, RNA plays a critical role in essentially all aspects of biology including signaling, gene regulation, catalysis, and retroviral infection[29–35]. Regulatory, structural, and catalytic RNAs unique to bacteria and virus also present an opportunity as highly specific anti-infective drug targets that could curb the rising spread of antibiotic resistance[4,11,18]. In order to better understand how to target these RNAs and how they may function, detailed high-resolution structures are needed. Currently, there is limited information on different RNA structural motifs as compared with proteins.

NMR is a powerful tool for studying RNA structure and dynamics, yet we have seen a stagnation in published RNA structures (70/year). In order to promote continued research and further progress in RNA structure discovery, we endeavored to minimize the cost associated with NMR studies of RNA. We have presented several methods for preparing RNA samples with cost savings upwards of 300-fold compared to commercially available alternatives as well as providing new isotopic labeling patterns that open the door for unique NMR experiments that simplify data collection and analysis.

In particular, we introduced a new NMR resonance assignment strategy for RNA that reduces both the cost and time associated with sample preparation and analysis. In terms of cost savings, this new assignment method could save

upwards of 3 weeks of instrumentation time, which is billed anywhere from $10-40/hr, for a total savings of $5,000-20,000. When coupled with the savings we obtained from our chemo-enzymatic isotopic labelling strategies, we think the door to RNA research by NMR is wide open to new scientists and institutions.

Our unique labeling scheme also provides isolated ^{13}C-^{1}H spin systems probes, and these are ideal for measuring relaxation dispersion, RDC, and RCSA by removing ^{13}C-^{13}C coupling and reducing spectral crowding. We were able to quickly build representative 3D models with minimal NMR data by utilizing cutting edge tools for RNA structure determination. We also developed faster methods for classify molecular motions and chemical exchange, which appears to be ubiquitous within RNA macromolecules. Identifying the structure of the minor populations sampled by RNA has been particularly difficult due to the limited experiments available for measuring chemical exchange in RNA. Before our work, there had only been three publications that successfully implemented CPMG type relaxation dispersion experiments for measuring exchange in RNA[165,333,334]. We have shown that with slight pulse program modifications and proper sample preparation, these experiments are amenable to RNA and will allow us to observe the chemical environment and therefore the structure of transient conformational states of RNA.

We applied our new technologies to uncloaking the “invisible” transient conformational state of the SAM-II riboswitch. Through adaptation of chemical exchange saturation transfer (CEST) pulse programs developed by Lewis Kay[379,380], we were able to identify a minor “invisible” state sampled by the major

conformational on the millisecond timescale. The chemical environmental of this minor state matched very closely to that of the fully-bound SAM-II riboswitch. We purposed that the minor state may represent a pre-bound conformation that is induced into the final bound fold through ligand binding. The model that begins to emerge for RNA ligand binding is conformational selection coupled with induced fit[82,251,274,284,288,397–399]. We anticipate that a number of regulatory RNA processes use a similar dynamic mechanism (Figure A.14).

Understanding how riboswitches recognize their cognate ligand will be crucial for designing new therapeutics. These riboswitches have evolved to recognize specific ligands by adapting complex three-dimensional folds. If we can learn how RNA structure determines its fold and more importantly how to target particular RNA motifs, the opportunity to target RNAs within disease pathologies will be a reality. It is with this goal that we have begun work on screening viral RNAs as potential drug targets. Using our new technologies we will investigate small molecule binding to a series of high impact RNA targets. Our goal is to not only develop new lead compounds in the RNA drug discovery pipeline, but to also provide detailed structural data on RNA-drug interactions.

6.2 Future directions

Building on the work within this thesis, there is still room for continued improvements. As a list of bullet points, here are some of the future opportunities to expand on our results:

- Adapt ribose synthesis methods[405] to obtain 3',4',5',5''-detuerated ribose sugars so that the fast assignment strategy in Chapter 3 can be applied to larger RNA systems. Deuteration of the ribose moiety along with the H5 position in pyrimidines will also reduce interference contributions to MQ and ^{1}H CPMG experiments.
- Assign the HBVε RNA element using our assignment strategy, which we have already began as an ongoing project with the National Cancer Institute. Initial results are extremely promising and we have already assigned the crucial adenine H2 resonances needed for the unambiguous start of the NOESY “walk”.
- Given our initial chemical exchange information on both the bacterial A-site RNA and iron-responsive element RNA, it would be of great interest to measure the residual dipolar couplings (RDCs) of the minor states in order to provide structural restraints for modeling our predicted minor states. We have already developed the NMR tools to conduct these experiments and the next step is to repeat these experiments with alignment media such as Pf1 phage to obtain RDCs.
- For the SAM-II riboswitch, we are already finding strong collaborations with specialists in the field of molecular dynamics based on our initial published results[406]. We will continue to study this system under varying conditions to better understand the flux between ligand recognition models. This system also presents an

opportunity to obtain RDCs of the minor state. Using the TROSY-selected CEST experiment we wrote for our collaboration paper on phosphoramidite solid-phase synthesis of RNA[205].

- We also have a continued collaboration with Victoria D'Souza on the SARS frame-shifting element. We have collected substantial dynamics data using a multitude of new pulse programs which is helping us understand how changes in flexibility and structure correlate with frame-shifting efficiency. As our models and resonance assignment improves, we hope this will contribute to a high impact publication on how dynamics is integral to frame-shifting.

Much of our work depended on the development on new tools and technologies and it is my utmost hope that the effort and time spent refining and developing these tools will allow future graduate students to dive deeper into understanding the structure and dynamics unique to RNA function. We have many promising new projects within our lab with the potential to provide insight and impact in the field of RNA biology. Let this be the groundwork to springboard the next generation.

Appendices

Figure and tables referenced from the text appear first followed by experimental protocols for the SAM-II riboswitch preparation and ligation.

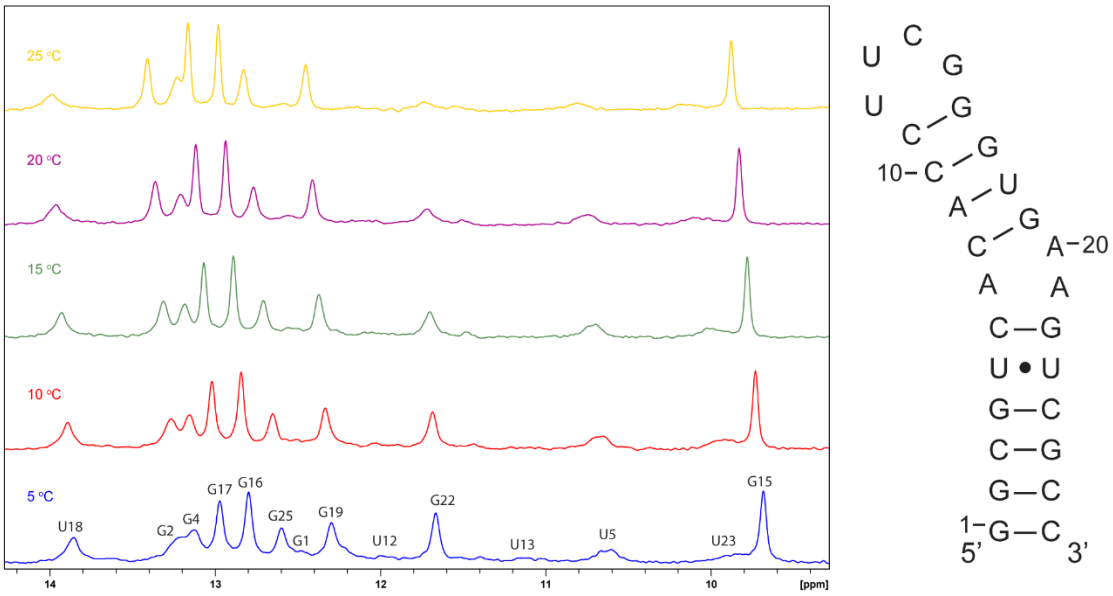

Figure A.1 1D imino spectrum of unlabeled bacterial A-site RNA at various temperatures for optimization of imino spectra.

Table A.1. All NMR experiments used in this study are given in the table along with relevant parameters. The Bruker pulse sequence used or modified for RNA is given under pulse program. The italicized experiments were run in 90/10% H_2O/D_2O. All other experiments were run in 99.98% D_2O. Two samples were used in this study: a 2',8-^{13}C-A; 1',6-^{13}C-1,3-^{15}N-C; 1',8-^{13}C-G; 2',6-^{13}C-1,3-^{15}N-U labeled bacterial Asite (sel-Asite) and a uniform ^{13}C/^{15}N-AUGC bacterial A-site (u-Asite) sample. The carrier, sweep width (SW), and complex points (TD) is given for each dimension of the 2D experiments. All 3D experiments were run as 2D by setting the unobserved dimension to 1 TD. The total number of scans (NS) and recovery delay (d1) is provided for each experiment. The total experimental time (Expt Time) is reported in days (d), hours (h), and minutes (m). All J-coupling transfers were optimized with 1D or short 2D experiments to account for relaxation. The mixing time for the non-exchangeable proton NOESYs was set to 300 ms and 150 ms for the imino NOESY.

			F2 Dimension			F1 Dimension					
Sample	**Experiment**	**Bruker Pulse Program**	**Carrier (ppm)**	**SW (ppm)**	**TD**	**Carrier (ppm)**	**SW (ppm)**	**TD**	**NS**	**d1**	**Expt Time**
sel-Asite	F1eF2e NOESY	noesyhsqcetgpsi3d	4.7	8	1024	4.7	8	256	32	5 s	12h22m
sel-Asite	F1eF2f NOESY	hsqcgpnowgx33d	4.7	8	1024	4.7	8	256	32	1.5 s	4h32m
sel-Asite	F1fF2e NOESY	noesyhsqcgpwgx13d	6.5	7	1024	6.5	8	256	32	1.5 s	4h34m
sel-Asite	^{13}C HSQC	hsqcetgpsi2	4.7	8	1024	115	90	1024	4	1.5 s	1h52m
u-Asite	*1D imino*	*kgzgwg*	*4.7*	*NA*	*NA*	*117*	*22*	*2048*	*16*	*1.5 s*	*2.5m*
u-Asite	*Imino NOESY*	*noesyesfbgpph*	*4.72*	*22*	*2048*	*7*	*15*	*512*	*128*	*2 s*	*1d17h20m*
u-Asite	*^{15}N HSQC*	*hsqcetf3gpsidec*	*4.7*	*22*	*256*	*153*	*28*	*1024*	*32*	*1.2 s*	*3h41m*
u-Asite	*2 bond ^{15}N HSQC*	*hsqcetf3gpsidec*	*4.69*	*9*	*256*	*200*	*84*	*1024*	*120*	*1.5 s*	*13h37m*
u-Asite	TROSY HCN	na_trhcnetsi3d	4.7	10	128	148	12	2048	16	1.5 s	56m
u-Asite	^{13}C Ribose HSQC	hsqcetgpdec	4.7	10	256	79	54	1024	64	1.5 s	7h20n
u-Asite	^{13}C Base HSQC	hsqcetgpdec	4.7	13.3	256	143	24	1024	32	1.5 s	3h37m
u-Asite	HCCH COSY	hcchcogp3d	6.0	5.68	128	6.0	3	2048	8	1 s	38m
u-Asite	HNN-COSY	na_hnncosygpphwg	4.69	22	256	188	100	2048	120	1.5 s	14h20m

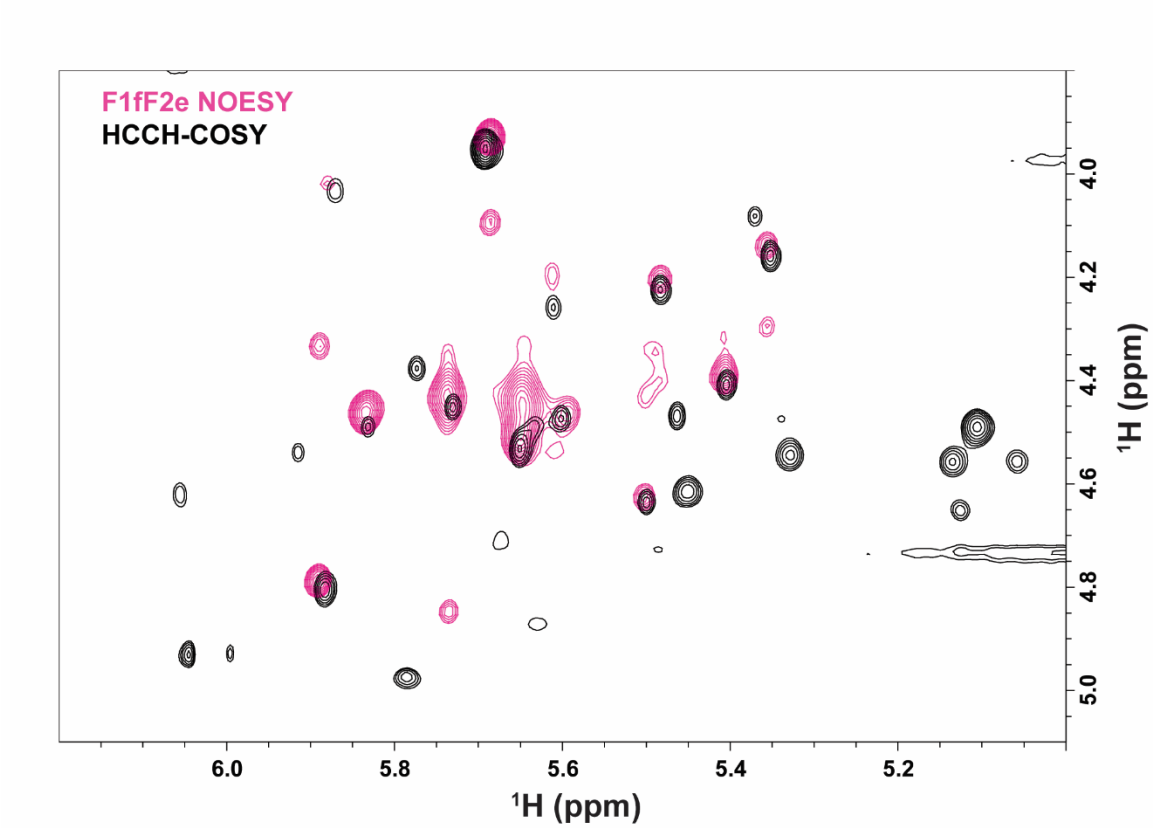

Figure A.2. The yellow region of the F1eF2f NOESY (magenta peaks) is overlaid on the HCCH-COSY (black peaks) of the uniform bacterial A-site RNA sample. The strong cross-peaks of the F1eF2f NOESY correspond to the through bond correlations in the HCCH-COSY indicating intra-nucleotide H1' and H2' cross-peaks.

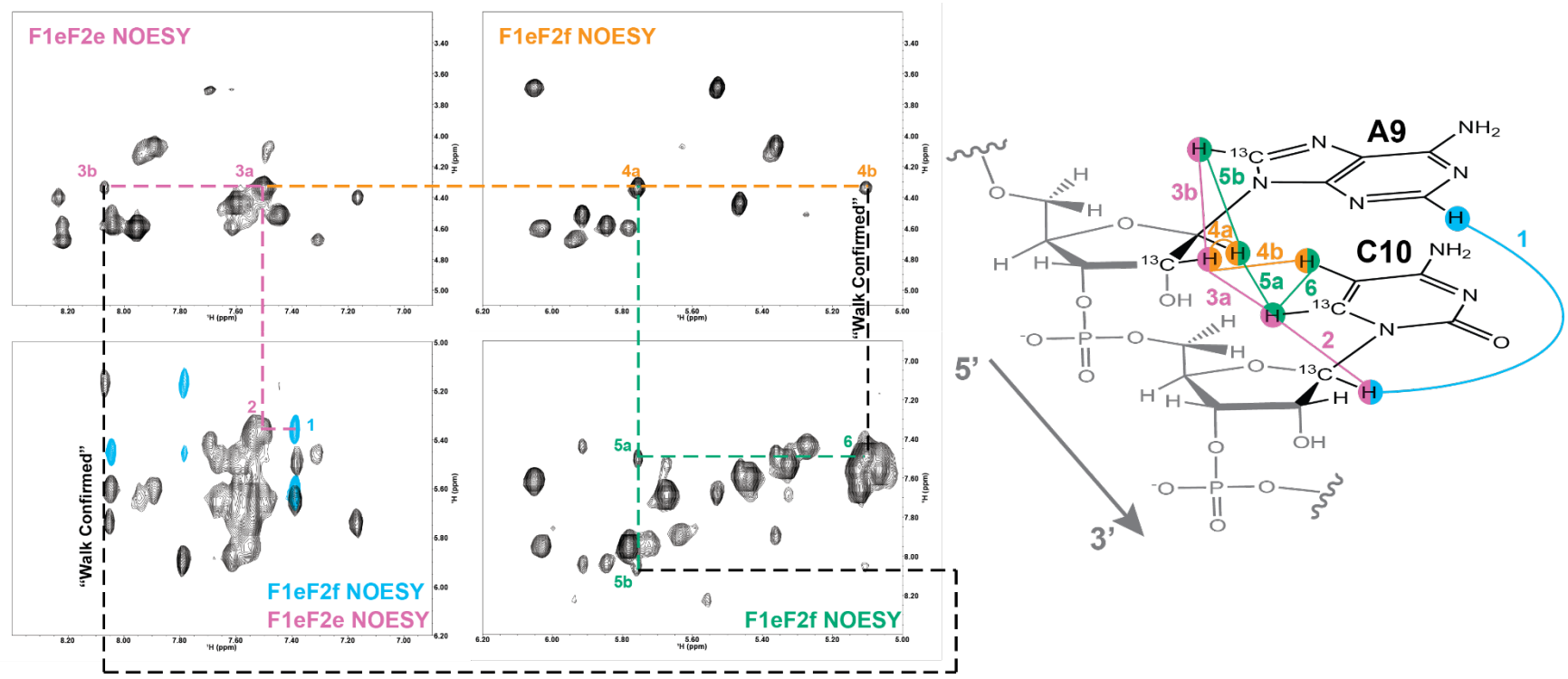

Figure A.3. The full NOESY "walk" scheme is shown with steps connected between spectra. Black dotted lines connected peaks that are self-confirming of the assignment.

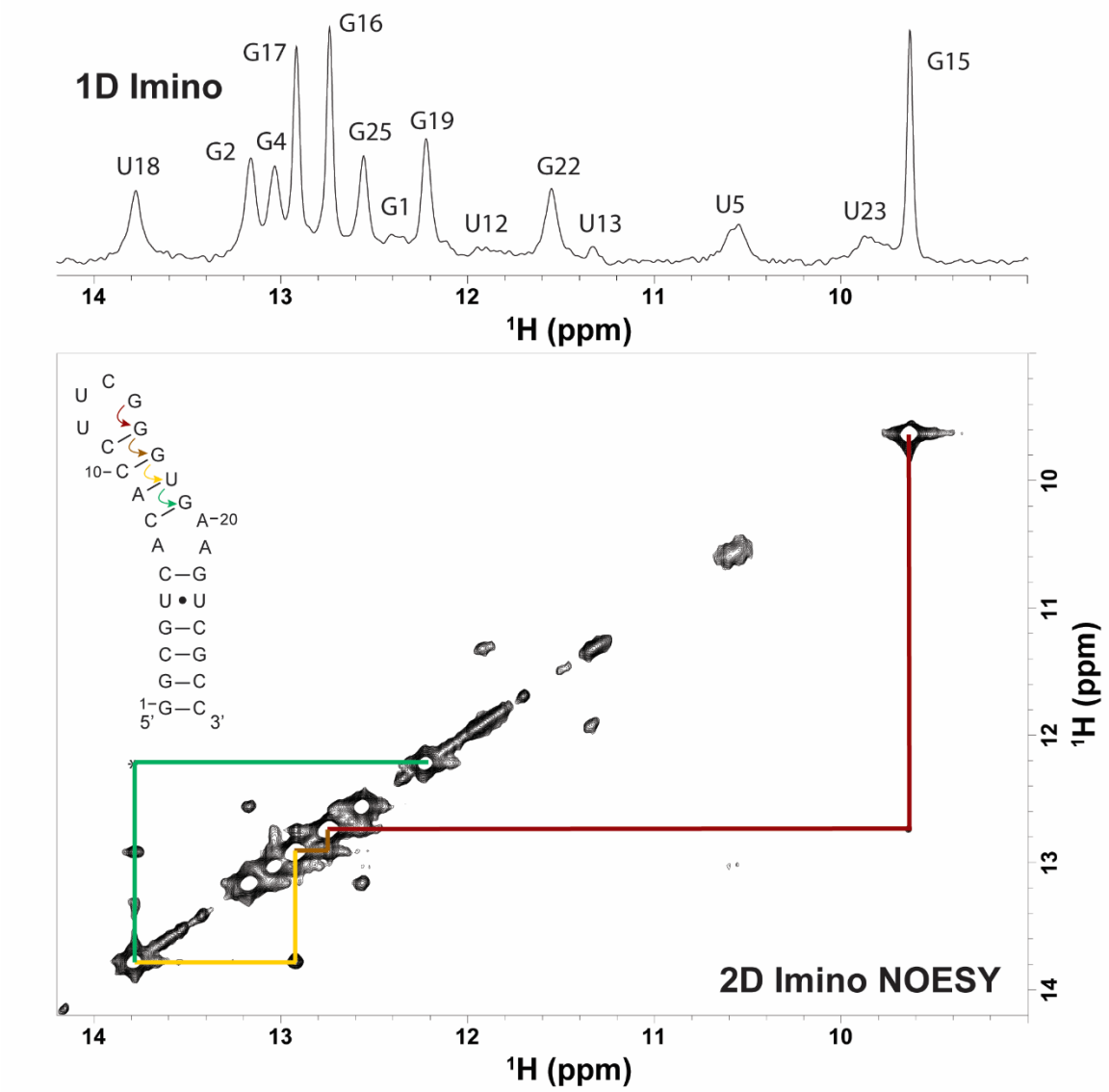

Figure A.4. A 1D imino spectra is shown projected above a 2D ^{1}H-^{1}H imino NOESY spectra with the sequential walk for the upper stem color coded for the secondary structure of bacterial A-site and the cross-peaks of the imino NOESY spectra.

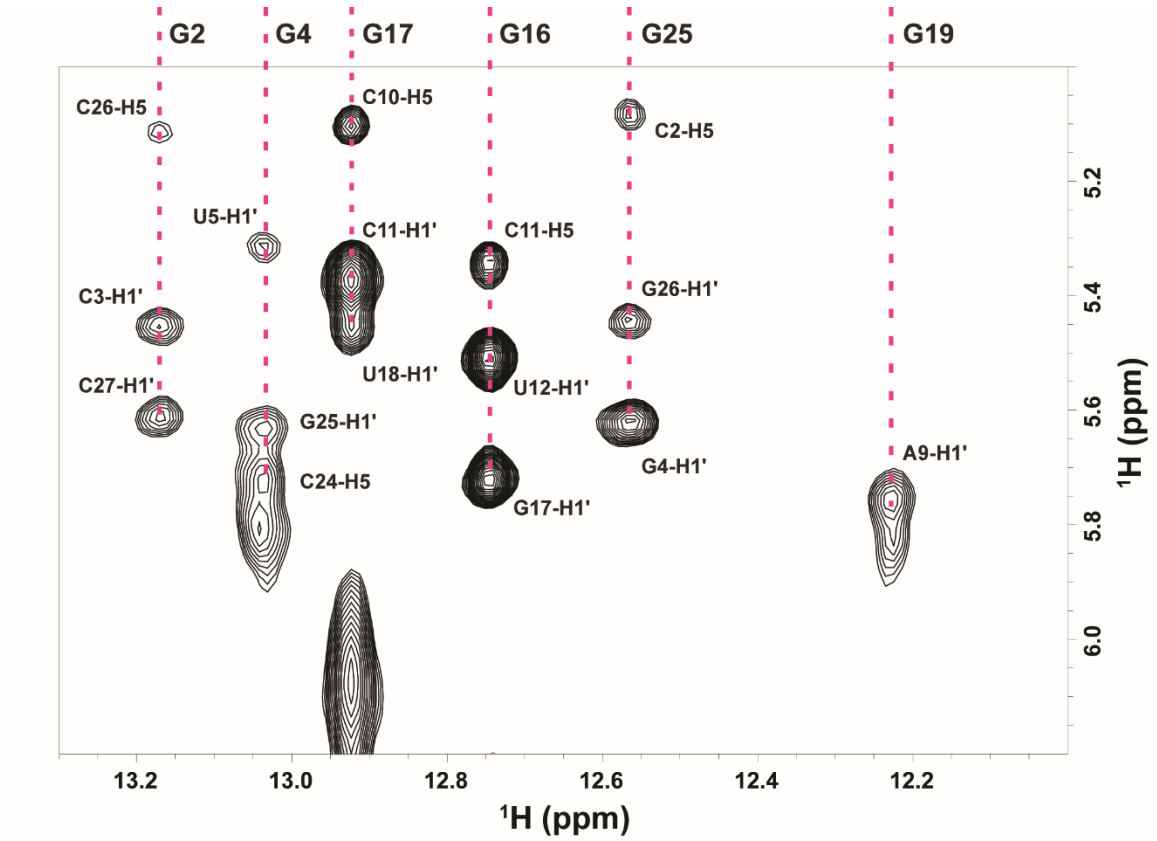

Figure A.5. The zoomed in G-H1 2D ^{1}H-^{1}H NOESY cross-peaks with H1' and H5 through spin-diffusion are labeled within the spectrum. The G-residue corresponding to the H1 (imino) resonance is labeled above the spectra for each set of cross-peaks next to the dashed magenta lines

Table A.2 NOE restraints for MC-SYM model selection of A-site RNA

```
assign (residue 1 and name H1' ) (residue 1 and name H8 ) 6.0 3.0 0.0
assign (residue 1 and name H8 ) (residue 1 and name H1' ) 6.0 3.0 0.0
assign (residue 1 and name H1' ) (residue 2 and name H8 ) 5.0 3.0 0.0
assign (residue 2 and name H8 ) (residue 1 and name H1' ) 6.0 3.0 0.0
assign (residue 1 and name H8 ) (residue 2 and name H8 ) 6.0 3.0 0.0
assign (residue 2 and name H8 ) (residue 1 and name H8 ) 6.0 3.0 0.0
assign (residue 2 and name H1' ) (residue 2 and name H8 ) 5.0 3.0 0.0
assign (residue 2 and name H8 ) (residue 2 and name H1' ) 5.0 3.0 0.0
assign (residue 3 and name H6 ) (residue 2 and name H1' ) 6.0 3.0 0.0
assign (residue 3 and name H6 ) (residue 3 and name H1' ) 6.0 3.0 0.0
assign (residue 3 and name H1' ) (residue 4 and name H8 ) 6.0 3.0 0.0
assign (residue 4 and name H8 ) (residue 3 and name H1' ) 6.0 3.0 0.0
assign (residue 4 and name H1' ) (residue 4 and name H8 ) 6.0 3.0 0.0
assign (residue 4 and name H8 ) (residue 4 and name H1' ) 6.0 3.0 0.0
assign (residue 4 and name H1' ) (residue 5 and name H6 ) 6.0 3.0 0.0
assign (residue 5 and name H6 ) (residue 4 and name H1' ) 6.0 3.0 0.0
assign (residue 5 and name H2' ) (residue 5 and name H6 ) 6.0 3.0 0.0
assign (residue 5 and name H6 ) (residue 5 and name H2' ) 6.0 3.0 0.0
assign (residue 5 and name H2' ) (residue 6 and name H1' ) 6.0 3.0 0.0
assign (residue 6 and name H1' ) (residue 5 and name H2' ) 6.0 3.0 0.0
assign (residue 5 and name H2' ) (residue 6 and name H6 ) 3.5 1.5 0.0
assign (residue 6 and name H6 ) (residue 5 and name H2' ) 3.5 1.5 0.0
assign (residue 5 and name H6 ) (residue 6 and name H6 ) 6.0 3.0 0.0
assign (residue 6 and name H1' ) (residue 6 and name H6 ) 6.0 3.0 0.0
assign (residue 6 and name H6 ) (residue 6 and name H1' ) 6.0 3.0 0.0
assign (residue 6 and name H1' ) (residue 7 and name H8 ) 6.0 3.0 0.0
assign (residue 7 and name H8 ) (residue 6 and name H1' ) 6.0 3.0 0.0
assign (residue 7 and name H2' ) (residue 7 and name H8 ) 6.0 3.0 0.0
assign (residue 7 and name H8 ) (residue 7 and name H2' ) 6.0 3.0 0.0
assign (residue 7 and name H2' ) (residue 8 and name H6 ) 3.5 1.5 0.0
assign (residue 8 and name H6 ) (residue 7 and name H2' ) 3.5 1.5 0.0
assign (residue 7 and name H8 ) (residue 8 and name H6 ) 6.0 3.0 0.0
assign (residue 8 and name H6 ) (residue 7 and name H8 ) 6.0 3.0 0.0
assign (residue 8 and name H1' ) (residue 8 and name H6 ) 6.0 3.0 0.0
assign (residue 8 and name H6 ) (residue 8 and name H1' ) 6.0 3.0 0.0
assign (residue 8 and name H1' ) (residue 9 and name H8 ) 6.0 3.0 0.0
assign (residue 9 and name H8 ) (residue 8 and name H1' ) 6.0 3.0 0.0
assign (residue 8 and name H6 ) (residue 9 and name H8 ) 6.0 3.0 0.0
assign (residue 9 and name H8 ) (residue 8 and name H6 ) 6.0 3.0 0.0
assign (residue 9 and name H2' ) (residue 9 and name H8 ) 6.0 3.0 0.0
assign (residue 9 and name H8 ) (residue 9 and name H2' ) 6.0 3.0 0.0
assign (residue 9 and name H2' ) (residue 10 and name H1' ) 6.0 3.0 0.0
assign (residue 10 and name H1' ) (residue 9 and name H2' ) 6.0 3.0 0.0
assign (residue 9 and name H2' ) (residue 10 and name H6 ) 3.5 1.5 0.0
```

```
assign (residue 10 and name H6 ) (residue 9 and name H2' ) 3.5 1.5 0.0
assign (residue 9 and name H8 ) (residue 10 and name H6 ) 6.0 3.0 0.0
assign (residue 10 and name H6 ) (residue 9 and name H8 ) 6.0 3.0 0.0
assign (residue 10 and name H1' ) (residue 10 and name H6 ) 5.0 3.0 0.0
assign (residue 10 and name H6 ) (residue 10 and name H1' ) 5.0 3.0 0.0
assign (residue 10 and name H1' ) (residue 11 and name H6 ) 5.0 3.0 0.0
assign (residue 11 and name H6 ) (residue 10 and name H1' ) 6.0 3.0 0.0
assign (residue 11 and name H1' ) (residue 11 and name H6 ) 5.0 3.0 0.0
assign (residue 11 and name H6 ) (residue 11 and name H1' ) 5.0 3.0 0.0
assign (residue 11 and name H1' ) (residue 12 and name H6 ) 6.0 3.0 0.0
assign (residue 12 and name H6 ) (residue 11 and name H1' ) 6.0 3.0 0.0
assign (residue 12 and name H6 ) (residue 12 and name H2' ) 6.0 3.0 0.0
assign (residue 13 and name H6 ) (residue 12 and name H2' ) 6.0 3.0 0.0
assign (residue 14 and name H6 ) (residue 12 and name H2' ) 6.0 3.0 0.0
assign (residue 13 and name H2' ) (residue 13 and name H6 ) 3.5 1.5 0.0
assign (residue 13 and name H6 ) (residue 13 and name H2' ) 3.5 1.5 0.0
assign (residue 14 and name H6 ) (residue 13 and name H2' ) 6.0 3.0 0.0
assign (residue 14 and name H6 ) (residue 13 and name H6 ) 6.0 3.0 0.0
assign (residue 14 and name H1' ) (residue 14 and name H6 ) 5.0 3.0 0.0
assign (residue 14 and name H6 ) (residue 14 and name H1' ) 6.0 3.0 0.0
assign (residue 15 and name H1' ) (residue 15 and name H8 ) 5.0 3.0 0.0
assign (residue 15 and name H8 ) (residue 15 and name H1' ) 5.0 3.0 0.0
assign (residue 16 and name H1' ) (residue 15 and name H1' ) 6.0 3.0 0.0
assign (residue 16 and name H1' ) (residue 16 and name H8 ) 6.0 3.0 0.0
assign (residue 16 and name H8 ) (residue 16 and name H1' ) 6.0 3.0 0.0
assign (residue 16 and name H1' ) (residue 17 and name H1' ) 6.0 3.0 0.0
assign (residue 17 and name H1' ) (residue 16 and name H1' ) 6.0 3.0 0.0
assign (residue 16 and name H1' ) (residue 17 and name H8 ) 6.0 3.0 0.0
assign (residue 17 and name H8 ) (residue 16 and name H1' ) 6.0 3.0 0.0
assign (residue 16 and name H8 ) (residue 17 and name H8 ) 6.0 3.0 0.0
assign (residue 17 and name H8 ) (residue 16 and name H8 ) 6.0 3.0 0.0
assign (residue 17 and name H1' ) (residue 17 and name H8 ) 6.0 3.0 0.0
assign (residue 17 and name H8 ) (residue 17 and name H1' ) 5.0 3.0 0.0
assign (residue 17 and name H1' ) (residue 18 and name H6 ) 6.0 3.0 0.0
assign (residue 18 and name H6 ) (residue 17 and name H1' ) 6.0 3.0 0.0
assign (residue 17 and name H8 ) (residue 18 and name H6 ) 6.0 3.0 0.0
assign (residue 18 and name H6 ) (residue 17 and name H8 ) 6.0 3.0 0.0
assign (residue 18 and name H2' ) (residue 18 and name H6 ) 3.5 1.5 0.0
assign (residue 18 and name H6 ) (residue 18 and name H2' ) 3.5 1.5 0.0
assign (residue 18 and name H2' ) (residue 19 and name H1' ) 6.0 3.0 0.0
assign (residue 19 and name H8 ) (residue 18 and name H2' ) 3.5 1.5 0.0
assign (residue 19 and name H1' ) (residue 19 and name H8 ) 5.0 3.0 0.0
assign (residue 19 and name H8 ) (residue 19 and name H1' ) 5.0 3.0 0.0
assign (residue 20 and name H8 ) (residue 19 and name H1' ) 6.0 3.0 0.0
assign (residue 19 and name H8 ) (residue 20 and name H8 ) 6.0 3.0 0.0
assign (residue 20 and name H8 ) (residue 19 and name H8 ) 6.0 3.0 0.0
assign (residue 20 and name H2' ) (residue 20 and name H8 ) 6.0 3.0 0.0
assign (residue 20 and name H8 ) (residue 20 and name H2' ) 5.0 3.0 0.0
```

```
assign (residue 20 and name H2' ) (residue 21 and name H2' ) 3.5 1.5 0.0
assign (residue 21 and name H2' ) (residue 20 and name H2' ) 3.5 1.5 0.0
assign (residue 20 and name H2' ) (residue 21 and name H8 ) 6.0 3.0 0.0
assign (residue 21 and name H8 ) (residue 20 and name H2' ) 5.0 3.0 0.0
assign (residue 21 and name H2' ) (residue 21 and name H8 ) 6.0 3.0 0.0
assign (residue 21 and name H8 ) (residue 21 and name H2' ) 6.0 3.0 0.0
assign (residue 22 and name H8 ) (residue 21 and name H2' ) 6.0 3.0 0.0
assign (residue 22 and name H1' ) (residue 22 and name H8 ) 6.0 3.0 0.0
assign (residue 22 and name H8 ) (residue 22 and name H1' ) 6.0 3.0 0.0
assign (residue 23 and name H6 ) (residue 22 and name H1' ) 6.0 3.0 0.0
assign (residue 23 and name H2' ) (residue 23 and name H6 ) 6.0 3.0 0.0
assign (residue 23 and name H6 ) (residue 23 and name H2' ) 6.0 3.0 0.0
assign (residue 23 and name H2' ) (residue 24 and name H1' ) 6.0 3.0 0.0
assign (residue 23 and name H2' ) (residue 24 and name H6 ) 5.0 3.0 0.0
assign (residue 24 and name H6 ) (residue 23 and name H2' ) 3.5 1.5 0.0
assign (residue 23 and name H6 ) (residue 24 and name H6 ) 6.0 3.0 0.0
assign (residue 24 and name H1' ) (residue 24 and name H6 ) 6.0 3.0 0.0
assign (residue 24 and name H6 ) (residue 24 and name H1' ) 6.0 3.0 0.0
assign (residue 24 and name H1' ) (residue 25 and name H8 ) 5.0 3.0 0.0
assign (residue 25 and name H8 ) (residue 24 and name H1' ) 6.0 3.0 0.0
assign (residue 25 and name H8 ) (residue 24 and name H6 ) 6.0 3.0 0.0
assign (residue 25 and name H8 ) (residue 25 and name H1' ) 5.0 3.0 0.0
assign (residue 26 and name H6 ) (residue 25 and name H1' ) 6.0 3.0 0.0
assign (residue 25 and name H8 ) (residue 26 and name H6 ) 5.0 3.0 0.0
assign (residue 26 and name H6 ) (residue 26 and name H1' ) 6.0 3.0 0.0
assign (residue 27 and name H6 ) (residue 26 and name H1' ) 6.0 3.0 0.0
assign (residue 27 and name H6 ) (residue 27 and name H1' ) 6.0 3.0 0.0
assign (residue 7 and name H2 ) (residue 22 and name H1' ) 6.0 3.0 0.0
assign (residue 7 and name H2 ) (residue 8 and name H1' ) 5.0 3.0 0.0
assign (residue 21 and name H2 ) (residue 22 and name H1' ) 5.0 3.0 0.0
assign (residue 9 and name H2 ) (residue 19 and name H1' ) 5.0 3.0 0.0
assign (residue 9 and name H2 ) (residue 10 and name H1' ) 5.0 3.0 0.0
```

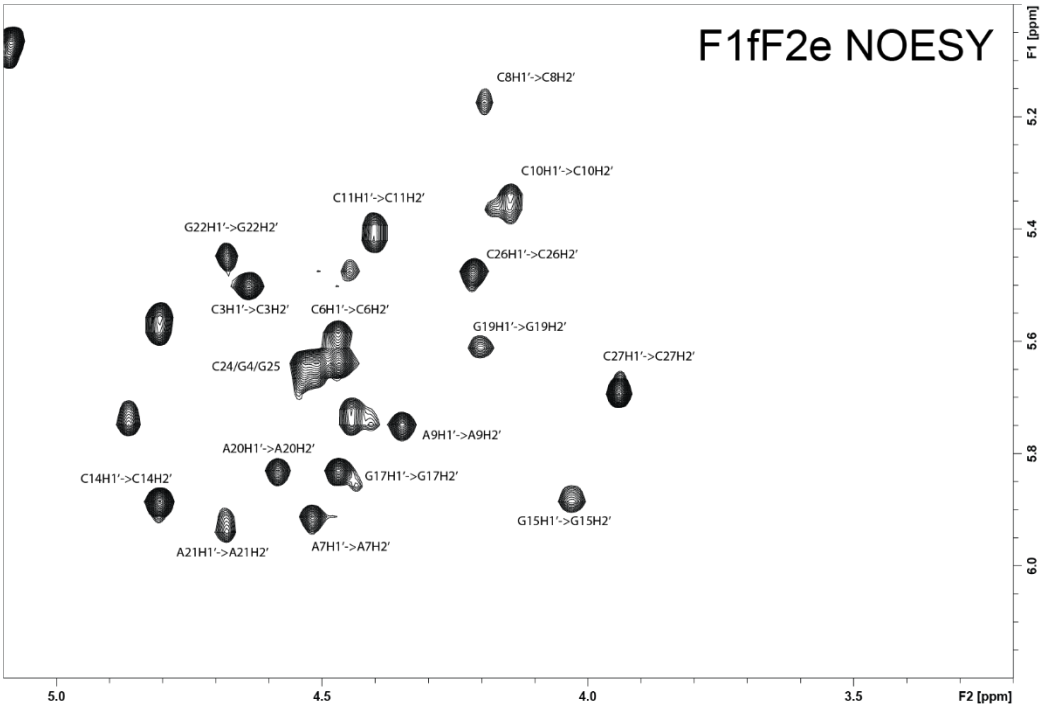

Figure A.6. Assignment confirmation with orthogonal sample of bacterial A-site RNA ladder labeling scheme. U is C1' labeled, AGC is C2' labeled.

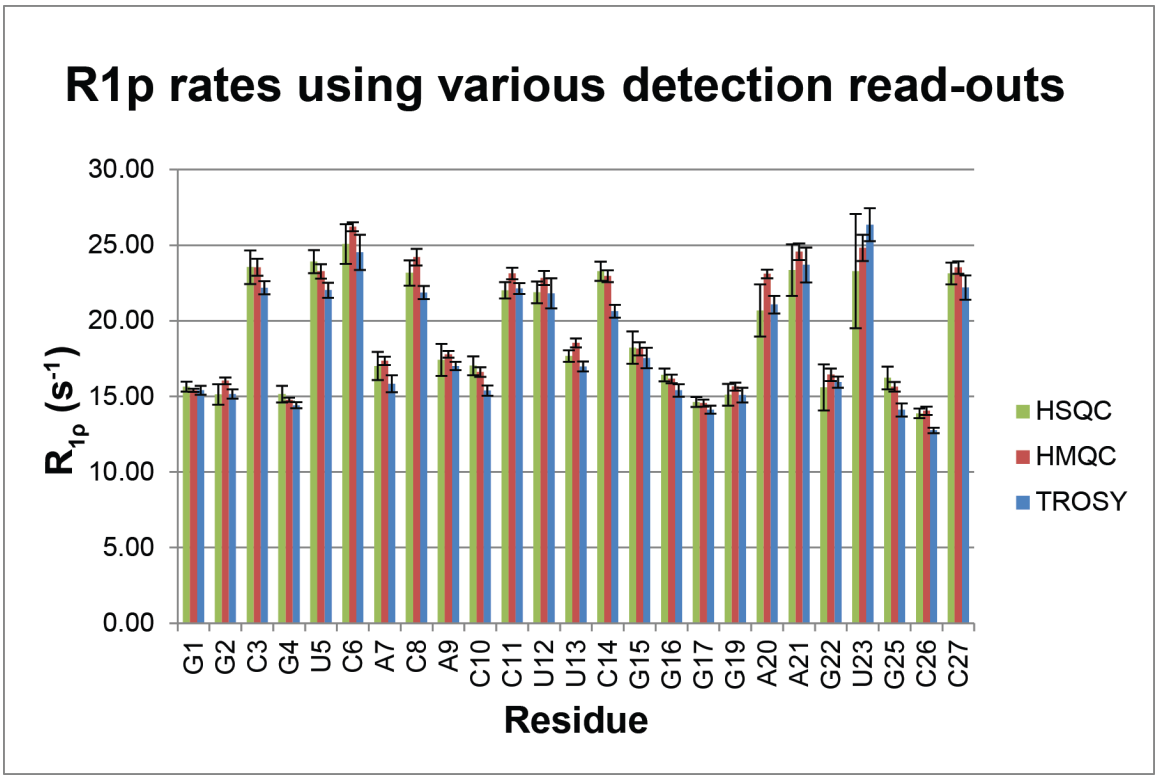

Figure A.7. Comparison of measured $R_{1\rho}$ rates on in-phase ^{13}C magnetization for bacterial A-site RNA using HSQC, HMQC, or TROSY read-outs.

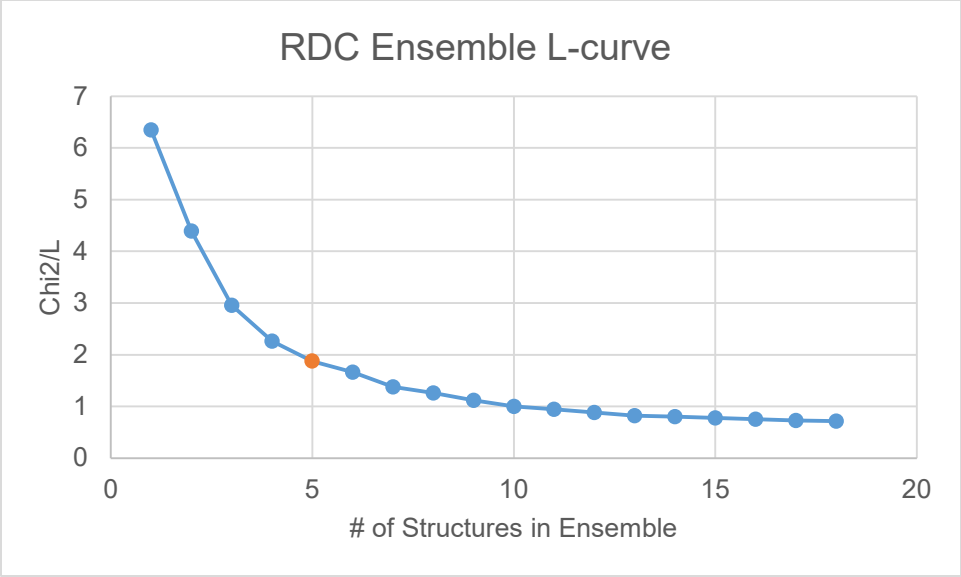

Figure A.8. L-curve analysis of ensemble fit for bacterial A-site RNA based on RDC data and using the program PATI[407] to perform sparse ensemble selection SES[408]**.** Analysis kindly provided by Andrew Longhini.

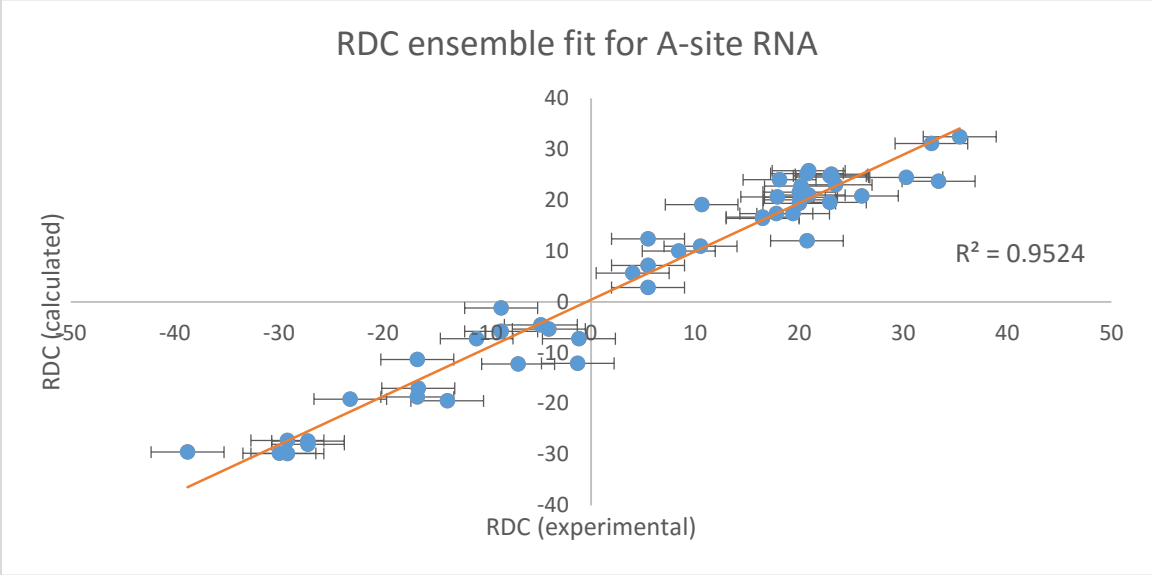

Figure A.9. The experimental RDCs compared to calculated RDCs for the bacterial A-site RNA ensemble selected using SES[408].

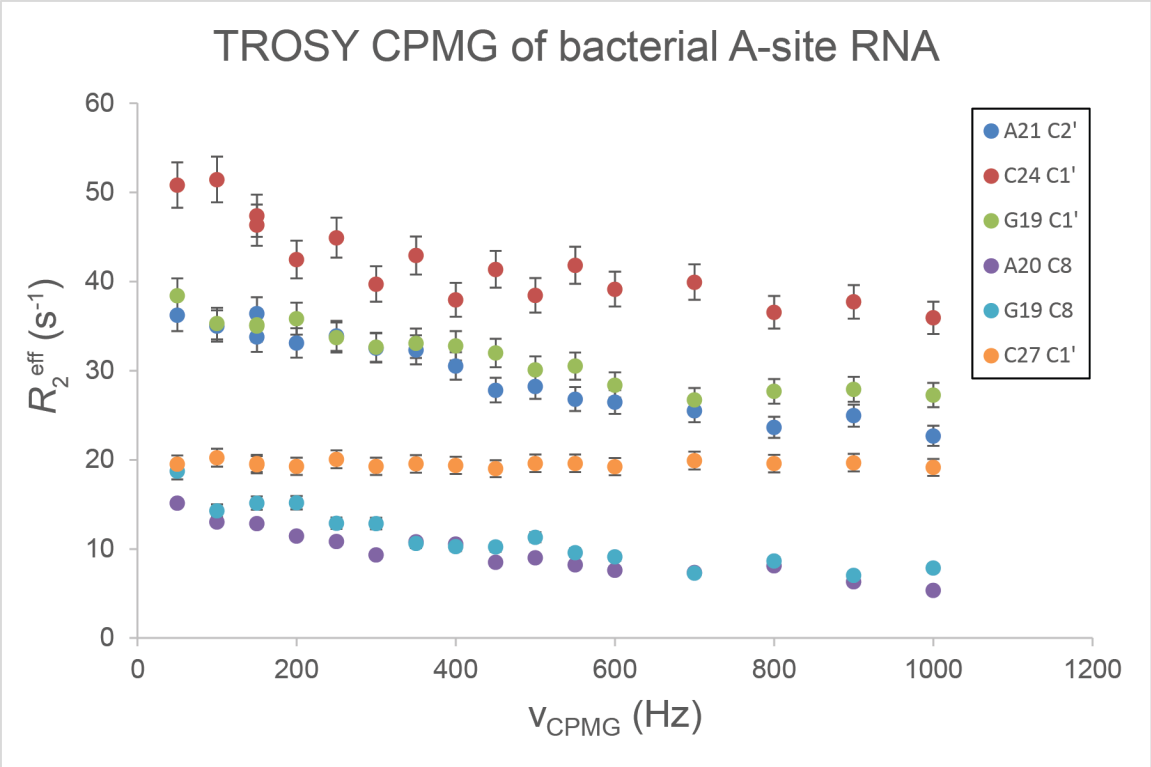

Figure A.10. The TROSY CPMG data for bacterial A-site RNA from the text including a typical flat dispersion profile as shown for C27 C1' (orange) for residues with no exchange on the CPMG timescale.

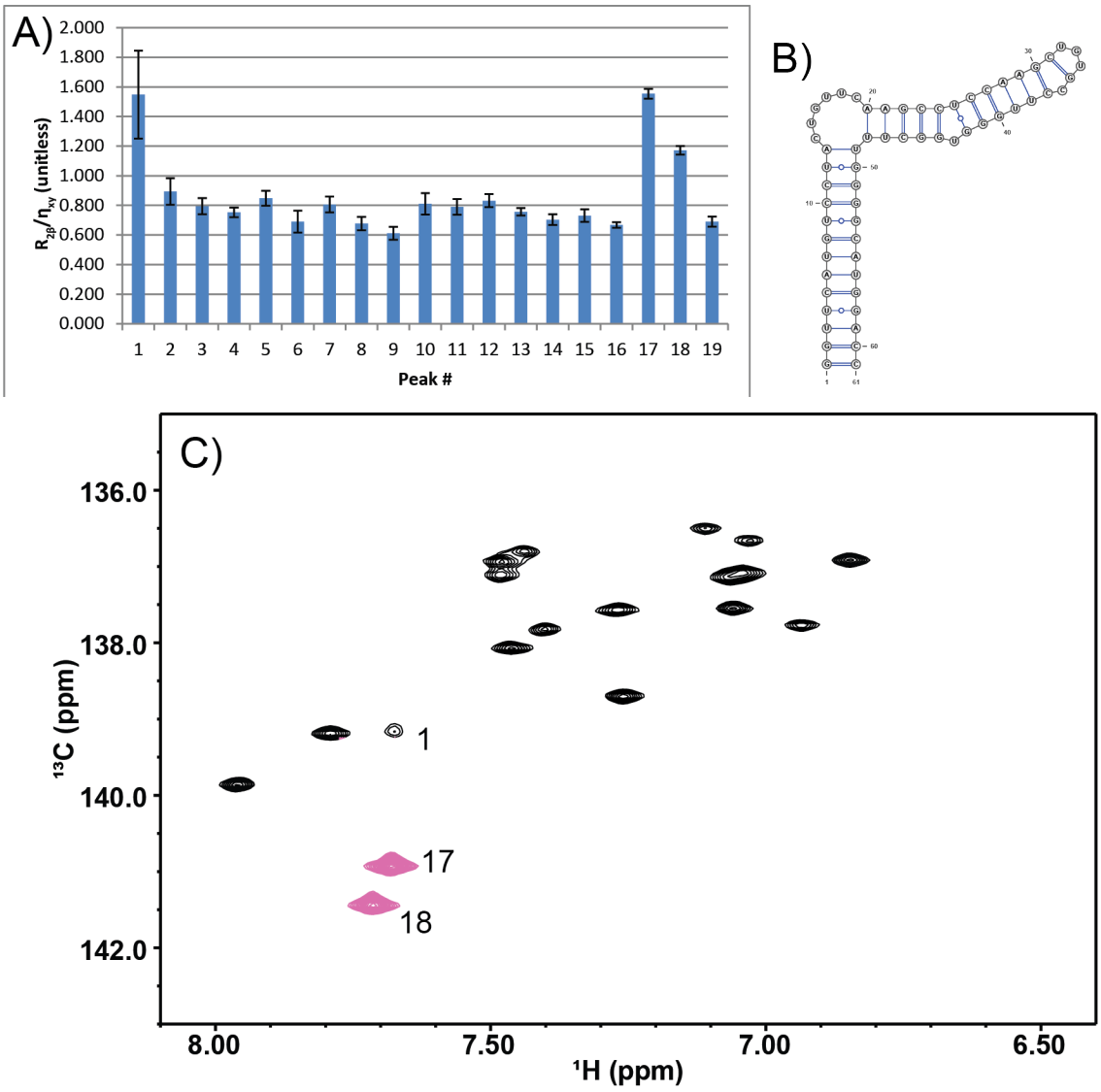

Figure A.11. A) The $R_{2\beta}/\eta_{xy}$ qualitative measure of exchange for HBVe G-C8 labeled sample. Peaks 1, 17, and 18 all have possible exchange. B) The secondary structure of HBVe. Peaks 17 and 18 are most likely G16 and G33 in the bulge and tri-loop, respectively. C) Peak experiencing exchange are labeled in the TROSY-HSQC spectrum with flexible residues colored purple.

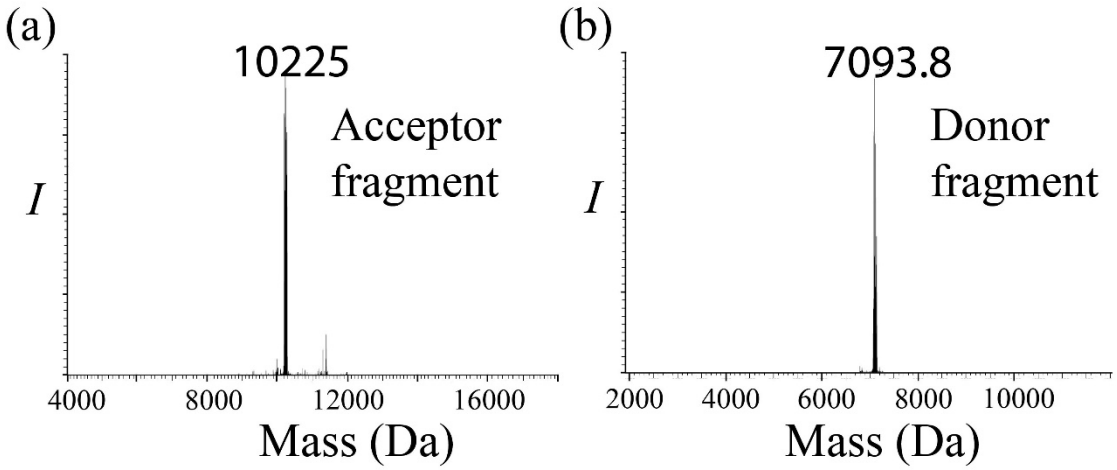

Figure A.12. Mass spectra of the acceptor (a) and donor (b) fragments of SAM-II riboswitch before ligation.

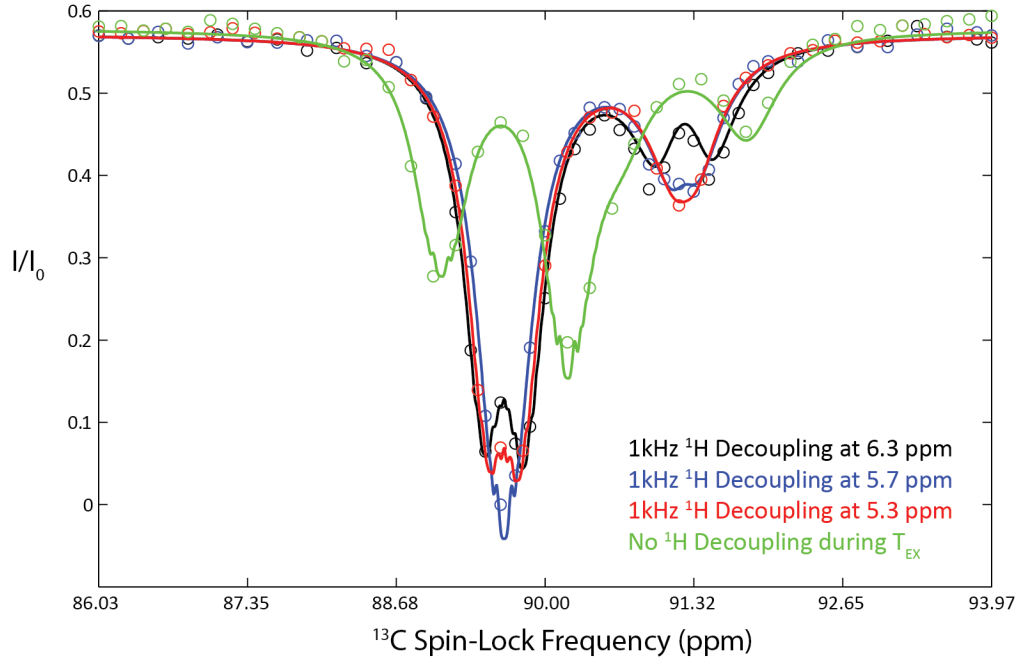

Figure A.13. All ^{13}C CEST profiles with weak ^{1}H decoupling are overlaid with their global fits. No ^{1}H decoupled CEST profile shows a triplet due to overlapping peaks from the major and minor states.

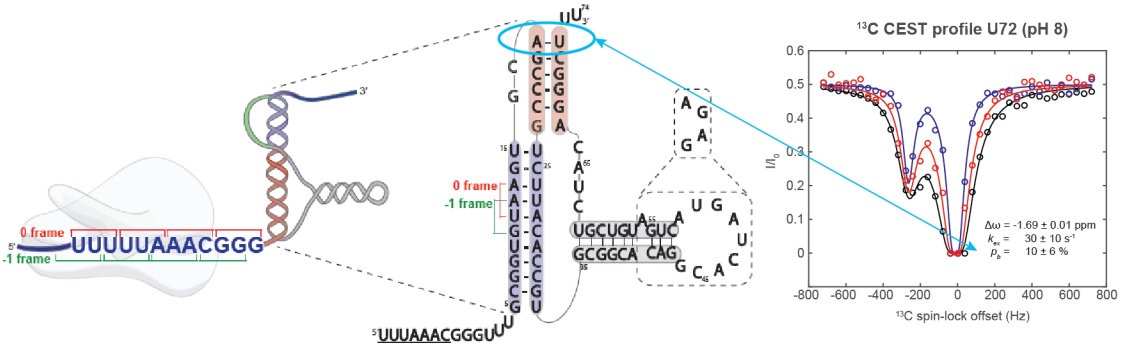

Figure A.14. The secondary structure for the SARS frame-shifting element is shown along with a CEST profile for U72 at pH 8. The observed minor population matches closely with *in vitro* frameshifting data (unpublished) indicating A-U base-opening may occur during compaction of the pseudoknot structure.

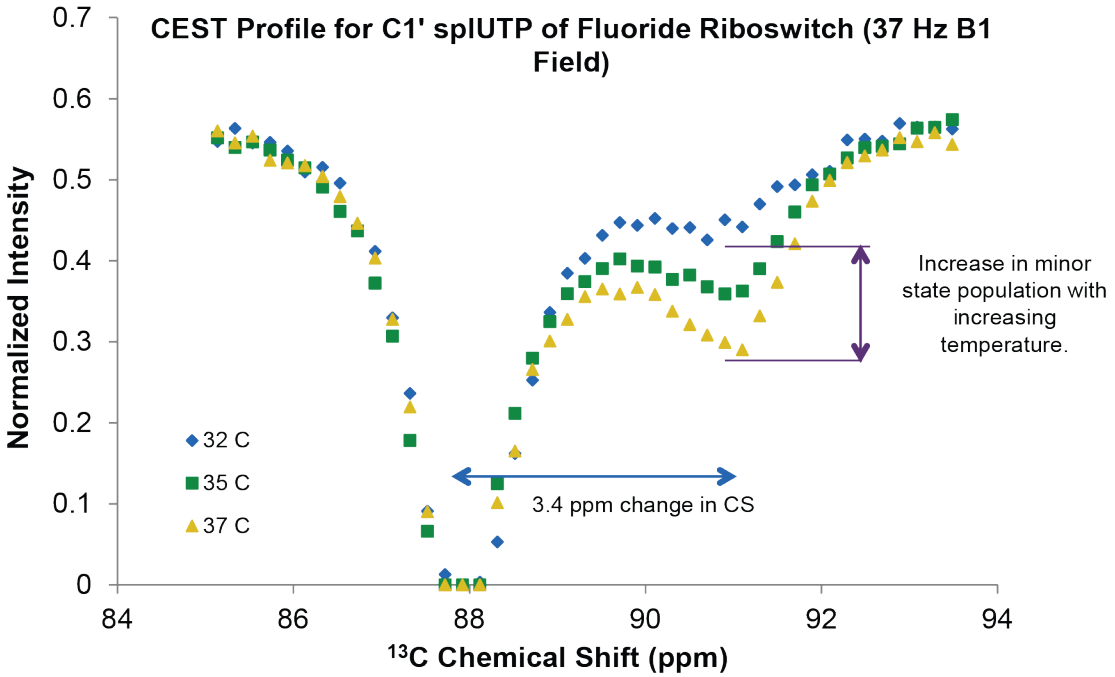

Figure A.15. CEST profile for the *Bacillus anthracis* fluoride riboswitch indicating a shift from flexible to stacked upon increase in temperature. This riboswitch is also proposed to form a pseudoknot as a transient pre-bound state.

Experimental Protocol for SAM-II

Chemicals and materials

S-adenosylmethionine was obtained from Sigma and *Taq* DNA polymerase from New England Biolabs, Inc. Oligonucleotides for polymerase chain reaction (PCR), synthetic DNA template of the acceptor fragment RNA 5′-A_mC_mGC GCT TGC AAT ACG GTT AAA TCA GCG CGC TAT AGT GAG TCG TAT TAG -3', the promoter sequence 5′-CTA ATA CGA CTC ACT ATA G-3′ (CTOPG) of T7 RNA polymerase, and a 40-mer DNA splint 5'-CCTTT TTAGC TACAT TTATC ACGCG CTTGC AATAC GGTTA-3' were purchased from Integrated DNA Technologies, Inc. Ribose C2′-methyoxyl modifications of the 5′ two nucleotides of the template strands were introduced to reduce the amount of N+1 and higher add-on transcripts[409]. Isotopically C6 and C1′-labeled CTP was prepared in house[192]. The T7 RNA polymerase[410] used for *in vitro* transcription was expressed in *Escherichia coli* BL21 (DE3) and purified on a Ni-chelating Sepharose column (Pharmacia) in house. A plasmid containing the donor fragment sequence of the ligated SAM-II RNA along with a self-cleaving trans-acting HδV ribozyme at the 3' end in a pUC57 vector was purchased from GenScript.

RNA preparation

The acceptor fragment RNA was synthesized by *in vitro* transcription with T7 RNA polymerase from the purified synthetic DNA template and the CTOPG [411]. The transcription solution, optimized by a systematic sparse matrix variation of Mg^{2+}, NTP, and enzyme concentrations, contains 10 mM total unlabeled NTPs (2.5 mM each of ATP, CTP, GTP, and UTP), 12 mM Mg^{2+}, 40 mM Tris-HCl, pH 8.0, 1 mM spermidine, 10 mM dithiothreitol (DTT), 0.01% Triton X-100, 300 nM DNA template and CTOPG, 2 u/mL inorganic pyrophosphatase, and 0.16 mg/mL T7 polymerase [411]. After 3h of incubation at 37°C, RNA products were extracted and purified as described previously [412]. The protocol for preparing donor

fragment by *in vitro* transcription is similar except that the DNA template and CTOPG were replaced by a template plasmid containing T7 promotor sequence and hepatitis delta virus (HDV) ribozyme sequence at the 5'- and 3'-ends of the donor fragment RNA plasmid DNA [413], and the NTP mixtures included unlabeled ATP, GTP, UTP and isotopically C6 and C1′-labeled CTP. The purified RNAs was lyophilized, re-suspended in water, and stored at -20°C until use.

RNA ligation

The unlabeled acceptor fragment with a protective 5'-triphosphate group and a reactive 3'-OH group was synthesized without any further processing. The donor fragment was treated with RNA 5′ polyphosphatase at 37°C for 1.5 h to obtain the desired monophosphorylated group at the 5'-terminal prior to a phenol-chloroform extraction and ethanol precipitation. Spontaneous ribozyme cleavage during *in vitro* transcription resulted in a protective cyclic 2'-3'- phosphodiester moiety at the 3'-tail that allowed us to sidestep an oxidative reaction step to block the 3'- hydroxyl group. Then the RNA was pelleted by spinning at maximum speed on a benchtop mini centrifuge at 4°C for 30 min. The RNA pellet was washed with cold 70% ethanol, dried at room temperature for 1 h, and re-suspend into H_2O. 1.5 Molar equivalents of acceptor fragment and DNA splint were added into the tube, respectively. The RNA mixture was mixed well prior to heat denaturation at 90°C for 3 min and cooled down to room temperature for 15 min. Finally, 10X reaction buffer, 25% PEG 8000, and T4 DNA ligase at a ratio of 100 units per picomole of donor fragment RNA in a 1.5 ml final reaction volume at 37°C for 3 h were added into the RNA mixture. The ligation products were separated from unligated species and the DNA splint on a 15% denaturing PAGE. The ligated SAM-II RNA was recovered from the excised small gel pieces using an Elutrap system as described [412].

Sample preparation and NMR experiments

The ligated and purified RNA was lyophilized and dissolved in water, and the corresponding amount of buffer stock solutions was added. The RNA sample solutions were subject to similar refolding process as described [98]. Then 8-10% D_2O and various concentrations of Mg^{2+} and SAM metabolite were added and the sample volume was adjusted to 260 μl with ddH_2O in Shigemi NMR tubes. The final NMR buffer contains 10 mM sodium phosphate, pH 6.2, and 100 mM NaCl, 0.2 mM EDTA, pH 6.2. The sample concentration was ~0.3 mM for NMR dynamic measurements as judged by the magnitude of UV absorbance at 260 nm and the calculated extinction coefficient (754,800 M^{-1} cm^{-1}) of the RNA's base composition. All NMR experiments were performed at a temperature of 298 K on Bruker Avance 600 MHz spectrometer equipped with actively shielded z-axis gradient triple resonance probes unless otherwise specified.

Bibliography

1. World Health Organization. *Causes of death: projections for 2015-2030*. (2015).
2. Centers for Disease Control and Prevention (CDC). *Antibiotic resistance threats in the United States*. Atlanta: CDC (2013).
3. Golkar, Z., Bagasra, O. & Pace, D. G. Bacteriophage therapy: a potential solution for the antibiotic resistance crisis. *J. Infect. Dev. Ctries.* **8,** 129–36 (2014).
4. Gross, M. Antibiotics in crisis. *Curr. Biol.* **23,** R1063–5 (2013).
5. O'Neil, J. *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations*. (2014).
6. Matteelli, A., Carvalho, A. C., Dooley, K. E. & Kritski, A. TMC207: the first compound of a new class of potent anti-tuberculosis drugs. *Future Microbiol.* **5,** 849–58 (2010).
7. Mahajan, R. Bedaquiline: First FDA-approved tuberculosis drug in 40 years. *Int. J. Appl. basic Med. Res.* **3,** 1–2 (2013).
8. LoBue, P. Extensively drug-resistant tuberculosis. *Curr. Opin. Infect. Dis.* **22,** 167–73 (2009).
9. Michael, C. A., Dominey-Howes, D. & Labbate, M. The antimicrobial resistance crisis: causes, consequences, and management. *Front. public Heal.* **2,** 145 (2014).
10. Gupta, S. K. & Nayak, R. P. Dry antibiotic pipeline: Regulatory bottlenecks and regulatory reforms. *J. Pharmacol. Pharmacother.* **5,** 4–7 (2014).
11. Bartlett, J. G., Gilbert, D. N. & Spellberg, B. Seven ways to preserve the miracle of antibiotics. *Clin. Infect. Dis.* **56,** 1445–50 (2013).
12. Gould, I. M. & Bal, A. M. New antibiotic agents in the pipeline and how they can help overcome microbial resistance. *Virulence* **4,** 185–91 (2013).
13. Talbot, G. H. *et al.* Bad bugs need drugs: an update on the development pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America. *Clin. Infect. Dis.* **42,** 657–68 (2006).
14. Bozdogan, B. & Appelbaum, P. C. Oxazolidinones: activity, mode of action, and mechanism of resistance. *Int. J. Antimicrob. Agents* **23,** 113–9 (2004).
15. Schneider, T., Müller, A., Miess, H. & Gross, H. Cyclic lipopeptides as antibacterial agents - potent antibiotic activity mediated by intriguing mode of actions. *Int. J. Med. Microbiol.* **304,** 37–43 (2014).

16. Hochlowski, J. E. *et al.* Tiacumicins, a novel complex of 18-membered macrolides. II. Isolation and structure determination. *J. Antibiot. (Tokyo).* **40,** 575–88 (1987).

17. Butler, M. S., Blaskovich, M. A. & Cooper, M. A. Antibiotics in the clinical pipeline in 2013. *J. Antibiot. (Tokyo).* **66,** 571–91 (2013).

18. Spellberg, B. & Gilbert, D. N. The future of antibiotics and resistance: a tribute to a career of leadership by John Bartlett. *Clin. Infect. Dis.* **59 Suppl 2,** S71–5 (2014).

19. Price, L. B. *et al.* Staphylococcus aureus CC398: host adaptation and emergence of methicillin resistance in livestock. *MBio* **3,** e00305–11– (2012).

20. Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *British journal of experimental pathology* **10,** 226 (1929).

21. Sengupta, S., Chattopadhyay, M. K. & Grossart, H.-P. The multifaceted roles of antibiotics and antibiotic resistance in nature. *Front. Microbiol.* **4,** 47 (2013).

22. Nicolaou, K. C. *et al.* Total synthesis of taxol. *Nature* **367,** 630–4 (1994).

23. Wasko, M. J., Pellegrene, K. A., Madura, J. D. & Surratt, C. K. A Role for Fragment-Based Drug Design in Developing Novel Lead Compounds for Central Nervous System Targets. *Front. Neurol.* **6,** 197 (2015).

24. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9,** 203–14 (2010).

25. Schmidt, M. F. Drug target miRNAs: Chances and challenges. *Trends Biotechnol.* **32,** 578–585 (2014).

26. Nicolaou, K. C. Advancing the drug discovery and development process. *Angew. Chem. Int. Ed. Engl.* **53,** 9128–40 (2014).

27. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1,** 727–30 (2002).

28. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5,** 993–6 (2006).

29. Breaker, R. R. Prospects for riboswitch discovery and analysis. *Mol. Cell* **43,** 867–79 (2011).

30. Mattick, J. S. A new paradigm for developmental biology. *J. Exp. Biol.* **210,** 1526–47 (2007).

31. Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152,** 17–24 (2013).

32. Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **9,** 242–53 (2008).

33. Ramakrishnan, V. The ribosome emerges from a black box. *Cell* **159,** 979–84 (2014).

34. Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. *Science* **349,** 1191–8 (2015).

35. Newman, A. J. & Nagai, K. Structural studies of the spliceosome: blind men and an elephant. *Curr. Opin. Struct. Biol.* **20,** 82–9 (2010).

36. Carter, A. P. *et al.* Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407,** 340–8 (2000).

37. Brodersen, D. E. *et al.* The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* **103,** 1143–54 (2000).

38. Yassin, A., Fredrick, K. & Mankin, A. S. Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 16620–5 (2005).

39. Barton-Davis, E. R., Cordier, L., Shoturma, D. I., Leland, S. E. & Sweeney, H. L. Aminoglycoside antibiotics restore dystrophin function to skeletal muscles of mdx mice. *J. Clin. Invest.* **104,** 375–381 (1999).

40. Howard, M., Frizzell, R. A. & Bedwell, D. M. Aminoglycoside antibiotics restore CFTR function by overcoming premature stop mutations. *Nat. Med.* **2,** 467–9 (1996).

41. Bedwell, D. M. *et al.* Suppression of a CFTR premature stop mutation in a bronchial epithelial cell line. *Nat. Med.* **3,** 1280–1284 (1997).

42. Montange, R. K. & Batey, R. T. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441,** 1172–5 (2006).

43. Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat. Biotechnol.* **24,** 1558–64 (2006).

44. Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. & Polz, M. F. Divergence and Redundancy of 16S rRNA Sequences in Genomes with Multiple rrn Operons. *J. Bacteriol.* **186,** 2629–2635 (2004).

45. Lee, A. S. Y., Kranzusch, P. J. & Cate, J. H. D. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **522,** 111–114 (2015).

46. Ling, H., Fabbri, M. & Calin, G. A. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat. Rev. Drug Discov.* **12,** 847–65 (2013).

47. St George-Hyslop, P. & Haass, C. Regulatory RNA goes awry in Alzheimer's disease. *Nat. Med.* **14,** 711–2 (2008).

48. Faghihi, M., Modarresi, F. & Khalil, A. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nat. Med.* **14,** 723–730 (2008).

49. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12,** 861–74 (2011).

50. Latronico, M. V. G. & Condorelli, G. MicroRNAs and cardiac pathology. *Nat. Rev. Cardiol.* **6,** 419–29 (2009).

51. Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* **2,** 461–72 (2004).

52. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. & Dumas, P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.* **8,** 1064–8 (2001).

53. Stelzer, A. C. *et al.* Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* **7,** 553–9 (2011).

54. Hamy, F. *et al.* An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 3548–53 (1997).

55. Daelemans, D. *et al.* A second target for the peptoid Tat/transactivation response element inhibitor CGP64222: inhibition of human immunodeficiency virus replication by blocking CXC-chemokine receptor 4-mediated virus entry. *Mol. Pharmacol.* **57,** 116–24 (2000).

56. Hyun, S., Na, J., Lee, S. J., Park, S. & Yu, J. RNA grooves can accommodate disulfide-bridged bundles of alpha-helical peptides. *Chembiochem* **11,** 767–70 (2010).

57. Lee, S. J., Hyun, S., Kieft, J. S. & Yu, J. An approach to the construction of tailor-made amphiphilic peptides that strongly and selectively bind to hairpin RNA targets. *J. Am. Chem. Soc.* **131,** 2224–30 (2009).

58. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–10 (1990).

59. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–22 (1990).

60. Lorenz, C. *et al.* Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res.* **38,** 3794–808 (2010).

61. Urak, K. T. *et al.* In vitro RNA SELEX for the generation of chemically-optimized therapeutic RNA drugs. *Methods* (2016). doi:10.1016/j.ymeth.2016.03.003

62. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **333,** 642–6 (2011).

63. Blount, K. F., Wang, J. X., Lim, J., Sudarsan, N. & Breaker, R. R. Antibacterial lysine analogs that target lysine riboswitches. *Nat. Chem. Biol.* **3,** 44–9 (2007).

64. Chen, L., Cressina, E., Leeper, F. J., Smith, A. G. & Abell, C. A fragment-based approach to identifying ligands for riboswitches. *ACS Chem. Biol.* **5,** 355–8 (2010).

65. Johnson, E. C., Feher, V. A., Peng, J. W., Moore, J. M. & Williamson, J. R. Application of NMR SHAPES screening to an RNA target. *J. Am. Chem. Soc.* **125,** 15724–5 (2003).

66. Mayer, M. & James, T. L. Detecting ligand binding to a small RNA target via saturation transfer difference NMR experiments in D(2)O and H(2)O. *J. Am. Chem. Soc.* **124,** 13376–7 (2002).

67. Mayer, G. & Famulok, M. High-throughput-compatible assay for glmS riboswitch metabolite dependence. *Chembiochem* **7,** 602–4 (2006).

68. Schüller, A. *et al.* The concept of template-based de novo design from drug-derived molecular fragments and its application to TAR RNA. *J. Comput. Aided. Mol. Des.* **22,** 59–68 (2008).

69. Durrant, J. D., Friedman, A. J. & McCammon, J. A. CrystalDock: a novel approach to fragment-based drug design. *J. Chem. Inf. Model.* **51,** 2573–80 (2011).

70. Velagapudi, S. P., Gallo, S. M. & Disney, M. D. Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nat. Chem. Biol.* **10,** 291–7 (2014).

71. Moumné, R., Catala, M., Larue, V., Micouin, L. & Tisné, C. Fragment-based design of small RNA binders: promising developments and contribution of NMR. *Biochimie* **94,** 1607–19 (2012).

72. Bodoor, K. *et al.* Design and implementation of an ribonucleic acid (RNA) directed fragment library. *J. Med. Chem.* **52,** 3753–61 (2009).

73. Kim, J. N. *et al.* Design and antimicrobial action of purine analogues that bind Guanine riboswitches. *ACS Chem. Biol.* **4,** 915–27 (2009).

74. Mulhbacher, J. *et al.* Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS Pathog.* **6,** e1000865 (2010).

75. Vicens, Q., Mondragón, E. & Batey, R. T. Molecular sensing by the

aptamer domain of the FMN riboswitch: a general model for ligand binding by conformational selection. *Nucleic Acids Res.* **39,** 8586–98 (2011).

76. Wilson, R. C. *et al.* Tuning riboswitch regulation through conformational selection. *J. Mol. Biol.* **405,** 926–38 (2011).

77. Haller, A., Soulière, M. F. & Micura, R. The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc. Chem. Res.* **44,** 1339–48 (2011).

78. Liberman, J. A. & Wedekind, J. E. Riboswitch structure in the ligand-free state. *Wiley Interdiscip. Rev. RNA* **3,** 369–84 (2012).

79. Ottink, O. M. *et al.* Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined induced fit mechanism. *RNA* **13,** 2202–12 (2007).

80. Noeske, J. *et al.* Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res.* **35,** 572–83 (2007).

81. Suddala, K. C., Wang, J., Hou, Q. & Walter, N. G. Mg(2+) shifts ligand-mediated folding of a riboswitch from induced-fit to conformational selection. *J. Am. Chem. Soc.* **137,** 14075–83 (2015).

82. Rinnenthal, J. *et al.* Mapping the landscape of RNA dynamics with NMR spectroscopy. *Acc. Chem. Res.* **44,** 1292–301 (2011).

83. Hann, M. M., Leach, A. R. & Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **41,** 856–64

84. Jeppsson, F. *et al.* Discovery of AZD3839, a potent and selective BACE1 inhibitor clinical candidate for the treatment of Alzheimer disease. *J. Biol. Chem.* **287,** 41245–57 (2012).

85. Cheng, Y. *et al.* From fragment screening to in vivo efficacy: optimization of a series of 2-aminoquinolines as potent inhibitors of beta-site amyloid precursor protein cleaving enzyme 1 (BACE1). *J. Med. Chem.* **54,** 5836–57 (2011).

86. Stamford, A. W. *et al.* Discovery of an Orally Available, Brain Penetrant BACE1 Inhibitor that Affords Robust CNS Aβ Reduction. *ACS Med. Chem. Lett.* **3,** 897–902 (2012).

87. Efremov, I. V *et al.* Discovery and optimization of a novel spiropyrrolidine inhibitor of β-secretase (BACE1) through fragment-based drug design. *J. Med. Chem.* **55,** 9069–88 (2012).

88. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res.* **31,** 6748–57 (2003).

89. Batey, R. T., Gilbert, S. D. & Montange, R. K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432,** 411–5 (2004).

90. Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11,** 1729–41 (2004).

91. Suddala, K. C. & Walter, N. G. Riboswitch structure and dynamics by smFRET microscopy. *Methods Enzymol.* **549,** 343–73 (2014).

92. Shebl, B., Norman, Z. & Cornish, P. V. Ribosome structure and dynamics by smFRET microscopy. *Methods Enzymol.* **549,** 375–406 (2014).

93. Rodgers, M. L. *et al.* Conformational dynamics of stem II of the U2 snRNA. *RNA* **22,** 225–36 (2016).

94. Stephenson, J. D., Kenyon, J. C., Symmons, M. F. & Lever, A. M. L. Characterizing 3D RNA structure by single molecule FRET. *Methods* (2016). doi:10.1016/j.ymeth.2016.02.004

95. Bercy, M. & Bockelmann, U. Hairpins under tension: RNA versus DNA. *Nucleic Acids Res.* **43,** 9928–36 (2015).

96. Wu, Y.-J., Wu, C.-H., Yeh, A. Y.-C. & Wen, J.-D. Folding a stable RNA pseudoknot through rearrangement of two hairpin structures. *Nucleic Acids Res.* **42,** 4505–15 (2014).

97. Savinov, A., Perez, C. F. & Block, S. M. Single-molecule studies of riboswitch folding. *Biochim. Biophys. Acta* **1839,** 1030–1045 (2014).

98. Chen, B., Zuo, X., Wang, Y.-X. & Dayie, T. K. Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res.* **40,** 3117–30 (2012).

99. Grishaev, A., Ying, J., Canny, M. D., Pardi, A. & Bax, A. Solution structure of tRNAVal from refinement of homology model against residual dipolar coupling and SAXS data. *J. Biomol. NMR* **42,** 99–109 (2008).

100. Baird, N. J. & Ferré-D'Amaré, A. R. Analysis of riboswitch structure and ligand binding using small-angle X-ray scattering (SAXS). *Methods Mol. Biol.* **1103,** 211–25 (2014).

101. Kazantsev, A. V *et al.* Solution structure of RNase P RNA. *RNA* **17,** 1159–71 (2011).

102. Brown, A., Shao, S., Murray, J., Hegde, R. S. & Ramakrishnan, V. Structural basis for stop codon recognition in eukaryotes. *Nature* **524,** 493–6 (2015).

103. Yamamoto, H. *et al.* Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA. *EMBO J.* **34,** 3042–58

(2015).

104. Agirrezabala, X. *et al.* Structural insights into cognate versus near-cognate discrimination during decoding. *EMBO J.* **30,** 1497–507 (2011).

105. Nguyen, T. H. D. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530,** 298–302 (2016).

106. Keane, S. C. *et al.* RNA structure. Structure of the HIV-1 RNA packaging signal. *Science* **348,** 917–21 (2015).

107. D'Souza, V. & Summers, M. F. Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* **431,** 586–90 (2004).

108. D'Souza, V., Dey, A., Habib, D. & Summers, M. F. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J. Mol. Biol.* **337,** 427–42 (2004).

109. Lukavsky, P. J., Kim, I., Otto, G. A. & Puglisi, J. D. Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol.* **10,** 1033–8 (2003).

110. Holden, S. J. *et al.* Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophys. J.* **99,** 3102–11 (2010).

111. Neuman, K. C. & Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **5,** 491–505 (2008).

112. Gajda, M. J., Martinez Zapien, D., Uchikawa, E. & Dock-Bregeon, A.-C. Modeling the structure of RNA molecules with small-angle X-ray scattering data. *PLoS One* **8,** e78007 (2013).

113. Yang, S., Parisien, M., Major, F. & Roux, B. RNA structure determination using SAXS data. *J. Phys. Chem. B* **114,** 10039–48 (2010).

114. Cochrane, J. C., Lipchock, S. V & Strobel, S. A. Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. *Chem. Biol.* **14,** 97–105 (2007).

115. Ke, A., Zhou, K., Ding, F., Cate, J. H. D. & Doudna, J. A. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429,** 201–5 (2004).

116. Adams, P. L. *et al.* Crystal structure of a group I intron splicing intermediate. *RNA* **10,** 1867–87 (2004).

117. Klein, D. J. & Ferré-D'Amaré, A. R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **313,** 1752–6 (2006).

118. Garst, A. D., Héroux, A., Rambo, R. P. & Batey, R. T. Crystal structure of the lysine riboswitch regulatory mRNA element. *J. Biol. Chem.* **283,** 22347–51 (2008).

119. Montange, R. K. & Batey, R. T. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441,** 1172–5 (2006).

120. Spitale, R. C., Torelli, A. T., Krucinska, J., Bandarian, V. & Wedekind, J. E. The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain. *J. Biol. Chem.* **284,** 11012–6 (2009).

121. Ban, N. *et al.* Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit. *Nature* **400,** 841–7 (1999).

122. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313,** 1935–42 (2006).

123. Savchenko, A. *et al.* Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. *Proteins* **50,** 392–9 (2003).

124. Kimber, M. S. *et al.* Data mining crystallization databases: knowledge-based approaches to optimize protein crystal screens. *Proteins* **51,** 562–8 (2003).

125. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28,** 171 (2009).

126. Lu, P. *et al.* Three-dimensional structure of human γ-secretase. *Nature* **512,** 166–70 (2014).

127. Quade, N., Boehringer, D., Leibundgut, M., van den Heuvel, J. & Ban, N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nat. Commun.* **6,** 7646 (2015).

128. Carlomagno, T. Present and future of NMR for RNA-protein complexes: a perspective of integrated structural biology. *J. Magn. Reson.* **241,** 126–36 (2014).

129. Lukavsky, P. J. & Puglisi, J. D. Structure determination of large biological RNAs. *Methods Enzymol.* **394,** 399–416 (2005).

130. Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *ChemBioChem* **4,** 936–962 (2003).

131. Dayie, K. T. Resolution enhanced homonuclear carbon decoupled triple resonance experiments for unambiguous RNA structural characterization. *J. Biomol. NMR* **32,** 129–139 (2005).

132. Ramakrishnan, V. The ribosome emerges from a black box. *Cell* **159,** 979–84 (2014).

133. Dayie, K. T. Key labeling technologies to tackle sizeable problems in RNA

structural biology. *Int. J. Mol. Sci.* **9,** 1214–40 (2008).

134. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–42 (2000).

135. Farjon, J. *et al.* Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. *J. Am. Chem. Soc.* **131,** 8571–7 (2009).

136. Sathyamoorthy, B., Lee, J., Kimsey, I., Ganser, L. R. & Al-Hashimi, H. Development and application of aromatic [(13)C, (1)H] SOFAST-HMQC NMR experiment for nucleic acids. *J. Biomol. NMR* **60,** 77–83 (2014).

137. Dallmann, A. & Sattler, M. Detection of hydrogen bonds in dynamic regions of RNA by NMR spectroscopy. *Curr. Protoc. Nucleic Acid Chem.* **59,** 7.22.1–7.22.19 (2014).

138. Stanek, J., Podbevšek, P., Koźmiński, W., Plavec, J. & Cevec, M. 4D Non-uniformly sampled C,C-NOESY experiment for sequential assignment of 13C, 15N-labeled RNAs. *J. Biomol. NMR* **57,** 1–9 (2013).

139. Bonneau, E. & Legault, P. Nuclear magnetic resonance structure of the III-IV-V three-way junction from the Varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement. *Biochemistry* **53,** 6264–75 (2014).

140. Otting, G. & Wüthrich, K. Heteronuclear filters in two-dimensional [1H,1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Q. Rev. Biophys.* **23,** 39–96 (1990).

141. Zhang, N. *et al.* Dimeric DNA quadruplex containing major groove-aligned A-T-A-T and G-C-G-C tetrads stabilized by inter-subunit Watson-Crick A-T and G-C pairs. *J. Mol. Biol.* **312,** 1073–88 (2001).

142. Frank, A. T., Bae, S.-H. & Stelzer, A. C. Prediction of RNA 1H and 13C chemical shifts: a structure based approach. *J. Phys. Chem. B* **117,** 13497–506 (2013).

143. Krähenbühl, B., Lukavsky, P. & Wider, G. Strategy for automated NMR resonance assignment of RNA: application to 48-nucleotide K10. *J. Biomol. NMR* **59,** 231–40 (2014).

144. Bermejo, G. A., Clore, G. M. & Schwieters, C. D. Improving NMR Structures of RNA. *Structure* (2016). doi:10.1016/j.str.2016.03.007

145. Salmon, L., Yang, S. & Al-Hashimi, H. M. Advances in the determination of nucleic acid conformational ensembles. *Annu. Rev. Phys. Chem.* **65,** 293–316 (2014).

146. Berlin, K. *et al.* Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *J. Am. Chem. Soc.* **135,** 16595–609 (2013).

147. Varani, G., Aboul-Ela, F. & Allain, F. H.-T. NMR investigation of RNA structure. *Prog. Nucl. Magn. Reson. Spectrosc.* **29,** 51–127 (1996).

148. Bothe, J. R. *et al.* Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. **8,** (2011).

149. Pardi, A. Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA. *Methods Enzymol.* **261,** 350–80 (1995).

150. Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D. & Williamson, J. R. Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res.* **20,** 4515–23 (1992).

151. Thakur, C. S., Sama, J. N., Jackson, M. E., Chen, B. & Dayie, T. K. Selective 13C labeling of nucleotides for large RNA NMR spectroscopy using an E. coli strain disabled in the TCA cycle. *J. Biomol. NMR* **48,** 179–92 (2010).

152. Martino, L. & Conte, M. R. Biosynthetic preparation of 13C/15N-labeled rNTPs for high-resolution NMR studies of RNAs. *Methods Mol. Biol.* **941,** 227–45 (2012).

153. Schultheisz, H. L., Szymczyna, B. R., Scott, L. G. & Williamson, J. R. Enzymatic de novo pyrimidine nucleotide synthesis. *J. Am. Chem. Soc.* **133,** 297–304 (2011).

154. Schultheisz, H. L., Szymczyna, B. R., Scott, L. G. & Williamson, J. R. Pathway engineered enzymatic de novo purine nucleotide synthesis. *ACS Chem. Biol.* **3,** 499–511 (2008).

155. Alvarado, L. J. *et al.* Chemo-enzymatic synthesis of selectively $^{13}C/^{15}N$-labeled RNA for NMR structural and dynamics studies. *Methods Enzymol.* **549,** 133–62 (2014).

156. Longhini, A. P., Leblanc, R. M. & Dayie, T. K. Chemo-enzymatic Labeling for Rapid Assignment of RNA Molecules. *Methods* **Submitted,** (2015).

157. Batey, R. T., Battiste, J. L. & Williamson, J. R. Preparation of isotopically enriched RNAs for heteronuclear NMR. *Methods Enzymol.* **261,** 300–22 (1995).

158. Johnson, J. E., Julien, K. R. & Hoogstraten, C. G. Alternate-site isotopic labeling of ribonucleotides for NMR studies of ribose conformational dynamics in RNA. *J. Biomol. NMR* **35,** 261–74 (2006).

159. Nikonowicz, E. P. *et al.* Preparation of 13C and 15N labelled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Res.* **20,** 4507–13 (1992).

160. Soupene, E. *et al.* Physiological studies of Escherichia coli strain MG1655: growth defects and apparent cross-regulation of gene expression. *J.*

Bacteriol. **185,** 5611–26 (2003).

161. Fraenkel, D. G. Selection of Escherichia coli mutants lacking glucose-6-phosphate dehydrogenase or gluconate-6-phosphate dehydrogenase. *J. Bacteriol.* **95,** 1267–71 (1968).

162. Lemaster, D. M. & Kushlan, D. M. Dynamical Mapping of E . coli Thioredoxin via 13 C NMR Relaxation Analysis. **7863,** 9255–9264 (2001).

163. Hoffman, D. W. & Holland, J. A. Preparation of carbon-13 labeled ribonucleotides using acetate as an isotope source. *Nucleic Acids Res.* **23,** 3361–2 (1995).

164. Nikonowicz, E. P. Preparation and use of 2H-labeled RNA oligonucleotides in nuclear magnetic resonance studies. *Methods Enzymol.* **338,** 320–41 (2001).

165. 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.* **130,** 16757–69 (2008).

166. Thakur, C. S., Brown, M. E., Sama, J. N., Jackson, M. E. & Dayie, T. K. Growth of wildtype and mutant E. coli strains in minimal media for optimal production of nucleic acids for preparing labeled nucleotides. *Appl. Microbiol. Biotechnol.* **88,** 771–9 (2010).

167. Dayie, T. K. & Thakur, C. S. Site-specific labeling of nucleotides for making RNA for high resolution NMR studies using an E. coli strain disabled in the oxidative pentose phosphate pathway. *J. Biomol. NMR* **47,** 19–31 (2010).

168. Thakur, C. S., Luo, Y., Chen, B., Eldho, N. V & Dayie, T. K. Biomass production of site selective 13C/15N nucleotides using wild type and a transketolase E. coli mutant for labeling RNA for high resolution NMR. *J. Biomol. NMR* **52,** 103–14 (2012).

169. Thakur, C. S. & Dayie, T. K. Asymmetry of 13C labeled 3-pyruvate affords improved site specific labeling of RNA for NMR spectroscopy. *J. Biomol. NMR* **52,** 65–77 (2012).

170. Lu, K., Miyazaki, Y. & Summers, M. F. Isotope labeling strategies for NMR studies of RNA. *J. Biomol. NMR* **46,** 113–25 (2010).

171. Nyholm, T. *et al.* A method for production of 13C/15N double labelled RNA in E. coli, and subsequent in vitro synthesis of ribonucleotide 5' triphosphates. *J. Biochem. Biophys. Methods* **30,** 59–68 (1995).

172. Kratovich, M. & Roe, B. A. Improved separation of nucleosides, nucleotides, and aminoacyl tRNA a strong anion-exchange resin. *J. Chromatogr.* **155,** 407–13 (1978).

173. Lilly, M. D. Extraction and separation of nucleotides fromEscherichia coli. *Biotechnol. Bioeng.* **7,** 335–342 (1965).

174. LeMaster, D. M. & Cronan, J. E. Biosynthetic production of 13C-labeled amino acids with site-specific enrichment. *J. Biol. Chem.* **257,** 1224–30 (1982).

175. Feist, A. M. *et al.* A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3,** 121 (2007).

176. Cohn, W. E. The Anion-Exchange Separation of Ribonucleotides. *J. Am. Chem. Soc.* **72,** 1471–1478 (1950).

177. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays* **29,** 288–299 (2007).

178. Boumendjel, A., Blanc, M., Williamson, G. & Barron, D. Efficient synthesis of flavanone glucuronides. *J. Agric. Food Chem.* **57,** 7264–7 (2009).

179. Ahmad, A. L., Oh, P. C. & Abd Shukor, S. R. Sustainable biocatalytic synthesis of L-homophenylalanine as pharmaceutical drug precursor. *Biotechnol. Adv.* **27,** 286–96

180. Nuijens, T. *et al.* Enzymatic synthesis of C-terminal arylamides of amino acids and peptides. *J. Org. Chem.* **74,** 5145–50 (2009).

181. Rowan, A. S. & Hamilton, C. J. Recent developments in preparative enzymatic syntheses of carbohydrates. *Nat. Prod. Rep.* **23,** 412–43 (2006).

182. Melnick, J. S., Sprinz, K. I., Reddick, J. J., Kinsland, C. & Begley, T. P. An efficient enzymatic synthesis of thiamin pyrophosphate. *Bioorg. Med. Chem. Lett.* **13,** 4139–41 (2003).

183. Rocha-Uribe, A. & Hernandez, E. Solvent-free enzymatic synthesis of structured lipids containing CLA from coconut oil and tricaprylin. *J. Am. Oil Chem. Soc.* **81,** 685–689 (2004).

184. Zhu, D., Mukherjee, C. & Hua, L. 'Green' synthesis of important pharmaceutical building blocks: enzymatic access to enantiomerically pure α-chloroalcohols. *Tetrahedron: Asymmetry* **16,** 3275–3278 (2005).

185. Bendich, A., Getler, H. & Brown, G. B. A SYNTHESIS OF ISOTOPIC CYTOSINE AND A STUDY OF ITS METABOLISM IN THE RAT. *J. Biol. Chem.* **177 ,** 565–570 (1949).

186. SantaLucia, J., Shen, L. X., Cai, Z., Lewis, H. & Tinoco, I. Synthesis and NMR of RNA with selective isotopic enrichment in the bases. *Nucleic Acids Res.* **23,** 4913–21 (1995).

187. Lagoja, I. M. *et al.* A short path synthesis of [13C/15N] multilabeled pyrimidine nucleosides starting from glucopyranose nucleosides. *J. Org. Chem.* **68,** 1867–71 (2003).

188. Lafrancois, C. J., Fujimoto, J. & Sowers, L. C. Synthesis and Utilization of 13 C(8)-Enriched Purines. *Nucleosides and Nucleotides* **18,** 23–37 (1999).

189. Barrio, M. del C., Scopes, D. I., Holtwick, J. B. & Leonard, N. J. Syntheses of all singly labeled [N]adenines: Mass spectral fragmentation of adenine. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 3986–8 (1981).

190. Abad, J.-L., Gaffney, B. L. & Jones, R. A. 15N-Multilabeled Adenine and Guanine Nucleosides. Syntheses of [1,3,NH(2)-(15)N(3)]- and [2-(13)C-1,3,NH(2)-(15)N(3)]-Labeled Adenosine, Guanosine, 2'-Deoxyadenosine, and 2'-Deoxyguanosine. *J. Org. Chem.* **64,** 6575–6582 (1999).

191. Longhini, A. P. *et al.* Chemo-enzymatic synthesis of site-specific isotopically labeled nucleotides for use in NMR resonance assignment, dynamics and structural characterizations. *Nucleic Acids Res.* **44,** e52 (2015).

192. Alvarado, L. J. *et al.* Regio-selective chemical-enzymatic synthesis of pyrimidine nucleotides facilitates RNA structure and dynamics studies. *Chembiochem* **15,** 1573–7 (2014).

193. Jiang, F., Patel, D. J., Zhang, X., Zhao, H. & Jones, R. A. Specific labeling approaches to guanine and adenine imino and amino proton assignments in the AMP–RNA aptamer complex. *J. Biomol. NMR* **9,** 55–62

194. Abad, J.-L., Shallop, A. J., Gaffney, B. L. & Jones, R. A. Use of 13C tags with specifically 15N-labeled DNA and RNA. *Biopolymers* **48,** 57–63 (1998).

195. Longhini, A. P., LeBlanc, R. & Dayie, T. K. Chemo-enzymatic Labeling for Rapid Assignment of RNA Molecules. *Methods* **Submitted,** (2015).

196. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15,** 8783–98 (1987).

197. Brakmann, S. & Grzeszik, S. An error-prone T7 RNA polymerase mutant generated by directed evolution. *Chembiochem* **2,** 212–9 (2001).

198. Milligan, J. F. & Uhlenbeck, O. C. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180,** 51–62 (1989).

199. Ponchon, L. & Dardel, F. Recombinant RNA technology: the tRNA scaffold. *Nat. Methods* **4,** 571–6 (2007).

200. Ponchon, L. & Dardel, F. Large scale expression and purification of recombinant RNA in Escherichia coli. *Methods* **54,** 267–73 (2011).

201. Nelissen, F. H. T. *et al.* Fast production of homogeneous recombinant RNA--towards large-scale production of RNA. *Nucleic Acids Res.* **40,** e102 (2012).

202. Liu, Y. *et al.* DNAzyme-mediated recovery of small recombinant RNAs from a 5S rRNA-derived chimera expressed in Escherichia coli. *BMC Biotechnol.* **10,** 85 (2010).

203. Le, M. T., Brown, R. E., Simon, A. E. & Dayie, T. K. In vivo, large-scale preparation of uniformly (15)N- and site-specifically (13)C-labeled homogeneous, recombinant RNA for NMR studies. *Methods Enzymol.* **565,** 495–535 (2015).

204. Liu, Y. *et al.* Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* **522,** 368–72 (2015).

205. Wunderlich, C. H. *et al.* Stable Isotope-Labeled RNA Phosphoramidites to Facilitate Dynamics by NMR. *Methods Enzymol.* **565,** 461–94 (2015).

206. Sumita, M. *et al.* Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA* **11,** 1420–9 (2005).

207. Vorbrüggen, H., Lagoja, I. M. & Herdewijn, P. Synthesis of ribonucleosides by condensation using trimethylsilyl triflate. *Curr. Protoc. Nucleic Acid Chem.* **Chapter 1,** Unit 1.13 (2007).

208. Micura, R. Small interfering RNAs and their chemical synthesis. *Angew. Chem. Int. Ed. Engl.* **41,** 2265–9 (2002).

209. Phan, A. T. & Patel, D. J. A site-specific low-enrichment (15)N,(13)C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.* **124,** 1160–1 (2002).

210. 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. Chem. Int. Ed. Engl.* **Submitted,** (2015).

211. Duss, O., Lukavsky, P. J. & Allain, F. H.-T. Isotope labeling and segmental labeling of larger RNAs for NMR structural studies. *Adv. Exp. Med. Biol.* **992,** 121–44 (2012).

212. Nelissen, F. H. T., Girard, F. C., Tessari, M., Heus, H. A. & Wijmenga, S. S. Preparation of selective and segmentally labeled single-stranded DNA for NMR by self-primed PCR and asymmetrical endonuclease double digestion. *Nucleic Acids Res.* **37,** e114 (2009).

213. Nelissen, F. H. T. *et al.* Multiple segmental and selective isotope labeling of large RNA for NMR structural studies. *Nucleic Acids Res.* **36,** e89 (2008).

214. Lang, K. & Micura, R. The preparation of site-specifically modified riboswitch domains as an example for enzymatic ligation of chemically synthesized RNA fragments. *Nat. Protoc.* **3,** 1457–66 (2008).

215. Kurschat, W. C., Müller, J., Wombacher, R. & Helm, M. Optimizing splinted ligation of highly structured small RNAs. *RNA* **11,** 1909–14 (2005).

216. Nozinovic, S., Fürtig, B., Jonker, H. R. a, Richter, C. & Schwalbe, H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* **38,** 683–94 (2010).

217. Butcher, S. E., Dieckmann, T. & Feigon, J. Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.* **16,** 7490–9 (1997).

218. Peterson, R. D., Theimer, C. A., Wu, H. & Feigon, J. New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR* **28,** 59–67 (2004).

219. Dieckmann, T. & Feigon, J. Assignment methodology for larger RNA oligonucleotides: application to an ATP-binding RNA aptamer. *J. Biomol. NMR* **9,** 259–72 (1997).

220. Pardi, A. Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA. *Methods Enzymol.* **261,** 350–80 (1995).

221. Cromsigt, J., van Buuren, B., Schleucher, J. & Wijmenga, S. Resonance assignment and structure determination for RNA. *Methods Enzymol.* **338,** 371–99 (2001).

222. Lukavsky, P. J. & Puglisi, J. D. RNAPack: an integrated NMR approach to RNA structure determination. *Methods* **25,** 316–32 (2001).

223. Wijmenga, S. S. & van Buuren, B. N. M. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **32,** 287–387 (1998).

224. Bahrami, A., Clos, L. J., Markley, J. L., Butcher, S. E. & Eghbalnia, H. R. RNA-PAIRS: RNA probabilistic assignment of imino resonance shifts. *J. Biomol. NMR* **52,** 289–302 (2012).

225. Aeschbacher, T. *et al.* Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.* **41,** e172 (2013).

226. Krähenbühl, B., Lukavsky, P. & Wider, G. Strategy for automated NMR resonance assignment of RNA: application to 48-nucleotide K10. *J. Biomol. NMR* **59,** 231–40 (2014).

227. Krähenbühl, B., El Bakkali, I., Schmidt, E., Güntert, P. & Wider, G. Automated NMR resonance assignment strategy for RNA via the phosphodiester backbone based on high-dimensional through-bond APSY experiments. *J. Biomol. NMR* **59,** 87–93 (2014).

228. Krähenbühl, B., El Bakkali, I., Schmidt, E., Güntert, P. & Wider, G. Automated NMR resonance assignment strategy for RNA via the phosphodiester backbone based on high-dimensional through-bond APSY experiments. *J. Biomol. NMR* **59,** 87–93 (2014).

229. Wuthrich, K. *NMR of Proteins and Nucleic Acids*. (Wiley Interscience, 1986).

230. Wijmenga, S. S. & van Buuren, B. N. M. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **32,** 287–387 (1998).

231. Alvarado, L. J. *et al.* Chemo-enzymatic synthesis of selectively $^{13}C/^{15}N$-labeled RNA for NMR structural and dynamics studies. *Methods Enzymol.* **549,** 133–62 (2014).

232. Breeze, A. L. Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. Nucl. Magn. Reson. Spectrosc.* **36,** 323–372 (2000).

233. Duszczyk, M. M., Zanier, K. & Sattler, M. A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. *Nucleic Acids Res.* **36,** 7068–77 (2008).

234. Webba da Silva, M. NMR methods for studying quadruplex nucleic acids. *Methods* **43,** 264–277 (2007).

235. Fourmy, D., Recht, M. I., Blanchard, S. C. & Puglisi, J. D. Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **274,** 1367–71 (1996).

236. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15,** 8783–98 (1987).

237. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4,** 603–14 (1994).

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

239. Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson.* **93,** 151–170 (1991).

240. Kay, L., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114,** 10663–10665 (1992).

241. Schleucher, J. *et al.* A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR* **4,** 301–6 (1994).

242. Dingley, A. J. & Grzesiek, S. Direct Observation of Hydrogen Bonds in Nucleic Acid Base Pairs by Internucleotide 2 J NN Couplings. *J. Am. Chem. Soc.* **120,** 8293–8297 (1998).

243. Grzesiek, S. & Bax, A. The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115,** 12593–12594 (1993).

244. Piotto, M., Saudek, V. & Sklenář, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2,** 661–665 (1992).

245. Brutscher, B. & Simorre, J. P. Transverse relaxation optimized HCN experiment for nucleic acids: combining the advantages of TROSY and MQ spin evolution. *J. Biomol. NMR* **21,** 367–72 (2001).

246. Sklenár, 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* **3,** 721–7 (1993).

247. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R. & Formankay, J. D. A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations in H2O Samples of Proteins. *J. Magn. Reson. Ser. B* **101,** 333–337 (1993).

248. Aeschbacher, T., Schubert, M. & Allain, F. H.-T. A procedure to validate and correct the 13C chemical shift calibration of RNA datasets. *J. Biomol. NMR* **52,** 179–90 (2012).

249. Aeschbacher, T., Schubert, M. & Allain, F. H.-T. A procedure to validate and correct the 13C chemical shift calibration of RNA datasets. *J. Biomol. NMR* **52,** 179–90 (2012).

250. Lynch, S. R. & Puglisi, J. D. Structure of a eukaryotic decoding region A-site RNA. *J. Mol. Biol.* **306,** 1023–35 (2001).

251. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* **491,** 724–8 (2012).

252. Hoffman, D. W. Resolution of the 1H-1H NOE spectrum of RNA into three dimensions using 15N-1H two-bond couplings. *J. Biomol. NMR* **16,** 165–9 (2000).

253. Simon, B., Zanier, K. & Sattler, M. A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly 13C-labeled RNA molecules. *J. Biomol. NMR* **20,** 173–6 (2001).

254. Fürtig, B., Richter, C., Bermel, W. & Schwalbe, H. New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop. *J. Biomol. NMR* **28,** 69–79 (2004).

255. Scott, L. G. & Hennig, M. RNA structure determination by NMR. *Methods Mol. Biol.* **452,** 29–61 (2008).

256. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452,** 51–5 (2008).

257. Ren, P. & Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **107,** 5933–5947 (2003).

258. Ampt, K. A. M. *et al.* 1H, 13C and 15N NMR assignments of Duck HBV apical stem loop of the epsilon encapsidation signal. *Biomol. NMR Assign.* **2,** 159–62 (2008).

259. van der Werf, R. M., Girard, F. C., Nelissen, F., Tessari, M. & Wijmenga, S. S. 1H, 13C and 15N NMR assignments of Duck HBV primer loop of the encapsidation signal epsilon. *Biomol. NMR Assign.* **2,** 143–5 (2008).

260. Flodell, S., Cromsigt, J., Schleucher, J., Kidd-Ljunggren, K. & Wijmenga, S. Structure Elucidation of the Hepatitis B Virus Encapsidation Signal by NMR on Selectively Labeled RNAs. *J. Biomol. Struct. Dyn.* **19,** 627–636 (2002).

261. Flodell, S. *et al.* The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.* **30,** 4803–11 (2002).

262. Frank, A. T., Horowitz, S., Andricioaei, I. & Al-Hashimi, H. M. Utility of 1H NMR chemical shifts in determining RNA structure and dynamics. *J. Phys. Chem. B* **117,** 2045–52 (2013).

263. Frank, A. T., Law, S. M. & Brooks, C. L. A Simple and Fast Approach for Predicting 1 H and 13 C Chemical Shifts: Toward Chemical Shift-Guided Simulations of RNA. *J. Phys. Chem. B* **118,** 12168–12175 (2014).

264. Duss, O., Diarra Dit Konté, N. & Allain, F. H.-T. Cut and Paste RNA for Nuclear Magnetic Resonance, Paramagnetic Resonance Enhancement, and Electron Paramagnetic Resonance Structural Studies. *Methods Enzymol.* **565,** 537–62 (2015).

265. Nelissen, F. H. T. *et al.* Multiple segmental and selective isotope labeling of large RNA for NMR structural studies. *Nucleic Acids Res.* **36,** e89 (2008).

266. Duss, O., Maris, C., von Schroetter, C. & Allain, F. H.-T. A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. *Nucleic Acids Res.* **38,** e188 (2010).

267. Tzakos, A. G., Easton, L. E. & Lukavsky, P. J. Preparation of large RNA oligonucleotides with complementary isotope-labeled segments for NMR structural studies. *Nat. Protoc.* **2,** 2139–2147 (2007).

268. Kim, I., Lukavsky, P. J. & Puglisi, J. D. NMR study of 100 kDa HCV IRES RNA using segmental isotope labeling. *J. Am. Chem. Soc.* **124,** 9338–9 (2002).

269. Scott, L. G., Tolbert, T. J. & Williamson, J. R. Preparation of specifically 2H- and 13C-labeled ribonucleotides. *Methods Enzymol.* **317,** 18–38 (2000).

270. Miyazaki, Y. *et al.* Structure of a conserved retroviral RNA packaging element by NMR spectroscopy and cryo-electron tomography. *J. Mol. Biol.* **404,** 751–72 (2010).

271. Kolberg, M., Strand, K. R., Graff, P. & Andersson, K. K. Structure, function, and mechanism of ribonucleotide reductases. *Biochim. Biophys. Acta* **1699,** 1–34 (2004).

272. Hansen, A. L. & Al-hashimi, H. M. Dynamics of large elongated RNA by NMR carbon relaxation. *J. Am. Chem. Soc.* **129,** 16072–82 (2007).

273. Berlin, K., Longhini, A., Dayie, T. K. & Fushman, D. Deriving quantitative dynamics information for proteins and RNAs using ROTDIF with a graphical user interface. *J. Biomol. NMR* **57,** 333–52 (2013).

274. Zhao, B. & Zhang, Q. Characterizing excited conformational states of RNA by NMR spectroscopy. *Curr. Opin. Struct. Biol.* **30,** 134–46 (2015).

275. Stelzer, A. C., Frank, A. T., Bailor, M. H., Andricioaei, I. & Al-Hashimi, H. M. Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. *Methods* **49,** 167–73 (2009).

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

277. Dayie, T. K. *Encyclopedia of Magnetic Resonance*. (John Wiley & Sons, Ltd, 2007).

278. Kang, M., Eichhorn, C. D. & Feigon, J. Structural determinants for ligand capture by a class II preQ1 riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E663–71 (2014).

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

280. Dayie, T. K. Nucleic Acids: Dynamics Studies by Solution NMR. *eMagRes* (2011).

281. Trott, O. & Palmer, A. G. R1rho relaxation outside of the fast-exchange limit. *J. Magn. Reson.* **154,** 157–60 (2002).

282. Mulder, F., de Graaf RA, Kaptein, R. & Boelens, R. An Off-resonance Rotating Frame Relaxation Experiment for the Investigation of Macromolecular Dynamics Using Adiabatic Rotations. *J. Magn. Reson.* **131,** 351–7 (1998).

283. 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(1rho) NMR spectroscopy. *J. Am. Chem. Soc.* **131,** 3818–9 (2009).

284. Xue, Y. *et al.* Characterizing RNA Excited States Using NMR Relaxation Dispersion. *Methods Enzymol.* **558,** 39–73 (2015).

285. Yamazaki, T., Muhandiram, R. & Kay, L. E. NMR Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly 13C,15N-Labeled Proteins: Application to 13C.alpha. Carbons. *J. Am. Chem. Soc.* **116,** 8266–8278 (1994).

286. Vallurupalli, P., Bouvignies, G. & Kay, L. E. A computational study of the effects of (13) C-(13) C scalar couplings on (13) C CEST NMR spectra: towards studies on a uniformly (13) C-labeled protein. *Chembiochem* **14,** 1709–13 (2013).

287. Zhou, Y. & Yang, D. Effects of J couplings and unobservable minor states on kinetics parameters extracted from CEST data. *J. Magn. Reson.* **249,** 118–125 (2014).

288. Zhao, B., Hansen, A. L. & Zhang, Q. Characterizing Slow Chemical Exchange in Nucleic Acids by Carbon CEST and Low Spin-Lock Field R1ρ NMR Spectroscopy. *J.Am.Chem.Soc.* 20–23 (2014).

289. Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *Chembiochem* **4,** 936–62 (2003).

290. Shajani, Z. & Varani, G. 13C NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J. Mol. Biol.* **349,** 699–715 (2005).

291. Ferner, J. *et al.* NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res.* **36,** 1928–40 (2008).

292. Ying, J., Grishaev, A., Bryce, D. L. & Bax, A. Chemical shift tensors of protonated base carbons in helical RNA and DNA from NMR relaxation and liquid crystal measurements. *J. Am. Chem. Soc.* **128,** 11443–54 (2006).

293. Ying, J., Grishaev, A. & Bax, A. Carbon-13 chemical shift anisotropy in DNA bases from field dependence of solution NMR relaxation rates. *Magn. Reson. Chem.* **44,** 302–10 (2006).

294. Grishaev, A., Yao, L., Ying, J., Pardi, A. & Bax, A. Chemical shift anisotropy of imino 15N nuclei in Watson-Crick base pairs from magic angle spinning liquid crystal NMR and nuclear spin relaxation. *J. Am. Chem. Soc.* **131,** 9490–1 (2009).

295. Allard, P., Helgstrand, M. & Hard, T. The complete homogeneous master

equation for a heteronuclear two-spin system in the basis of cartesian product operators. *J. Magn. Reson.* **134,** 7–16 (1998).

296. Helgstrand, M., Härd, T. & Allard, P. Simulations of NMR pulse sequences during equilibrium and non-equilibrium chemical exchange. *J. Biomol. NMR* **18,** 49–63 (2000).

297. Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Rance, M. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*. (Academic Press, 2007).

298. Pervushin, K. Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q. Rev. Biophys.* **33,** 161–97 (2000).

299. Fernández, C. & Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr. Opin. Struct. Biol.* **13,** 570–80 (2003).

300. Carlomagno, T. Present and future of NMR for RNA-protein complexes: a perspective of integrated structural biology. *J. Magn. Reson.* **241,** 126–36 (2014).

301. Simorre, J.-P. & Marion, D. Acquisition schemes and quadrature artifacts in phase-sensitive two-dimensional NMR. *J. Magn. Reson.* **89,** 191–197 (1990).

302. Kong, X. M., Sze, K. H. & Zhu, G. Gradient and sensitivity enhanced multiple-quantum coherence in heteronuclear multidimensional NMR experiments. *J. Biomol. NMR* **14,** 133–40 (1999).

303. Geen, H. & Freeman, R. Band-selective radiofrequency pulses. *J. Magn. Reson.* **93,** 93–141 (1991).

304. Tannús, A. & Garwood, M. Adiabatic pulses. *NMR Biomed.* **10,** 423–34 (1997).

305. Lakomek, N.-A., Ying, J. & Bax, A. Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR* **53,** 209–21 (2012).

306. Weininger, U., Respondek, M. & Akke, M. Conformational exchange of aromatic side chains characterized by L-optimized TROSY-selected ^{13}C CPMG relaxation dispersion. *J. Biomol. NMR* **54,** 9–14 (2012).

307. Weininger, U., Brath, U., Modig, K., Teilum, K. & Akke, M. Off-resonance rotating-frame relaxation dispersion experiment for 13C in aromatic side chains using L-optimized TROSY-selection. *J. Biomol. NMR* **59,** 23–9 (2014).

308. Pervushin, K., Vögeli, B. & Eletsky, A. Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy. *J. Am. Chem. Soc.* **124,**

12898–902 (2002).

309. Akke, M., Fiala, R., Jiang, F., Patel, D. & Palmer, A. G. Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: correlations with structure and stability. *RNA* **3,** 702–9 (1997).

310. Kaul, M., Barbieri, C. M. & Pilch, D. S. Fluorescence-based approach for detecting and characterizing antibiotic-induced conformational changes in ribosomal RNA: comparing aminoglycoside binding to prokaryotic and eukaryotic ribosomal RNA sequences. *J. Am. Chem. Soc.* **126,** 3447–53 (2004).

311. Kaul, M., Barbieri, C. M. & Pilch, D. S. Aminoglycoside-induced reduction in nucleotide mobility at the ribosomal RNA A-site as a potentially key determinant of antibacterial activity. *J. Am. Chem. Soc.* **128,** 1261–71 (2006).

312. Fourmy, D., Recht, M. I., Blanchard, S. C. & Puglisi, J. D. Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **274,** 1367–71 (1996).

313. Kadeřávek, P. *et al.* Spectral density mapping at multiple magnetic fields suitable for (13)C NMR relaxation studies. *J. Magn. Reson.* **266,** 23–40 (2016).

314. Schwalbe, H. *et al.* Cross-correlated relaxation for measurement of angles between tensorial interactions. *Methods Enzymol.* **338,** 35–81 (2001).

315. Ravindranathan, S., Kim, C.-H. & Bodenhausen, G. Cross correlations between 13C-1H dipolar interactions and 15N chemical shift anisotropy in nucleic acids. *J. Biomol. NMR* **27,** 365–75 (2003).

316. Duchardt, E. *et al.* Determination of the glycosidic bond angle chi in RNA from cross-correlated relaxation of CH dipolar coupling and N chemical shift anisotropy. *J. Am. Chem. Soc.* **126,** 1962–70 (2004).

317. Richter, C. *et al.* Determination of sugar conformation in large RNA oligonucleotides from analysis of dipole-dipole cross correlated relaxation by solution NMR spectroscopy. *J. Biomol. NMR* **15,** 241–50 (1999).

318. Chiarparin, E., Rüdisser, S. & Bodenhausen, G. Hydrogen bonds in RNA base pairs investigated by cross-correlated relaxation of multiple-quantum coherence in NMR. *Chemphyschem* **2,** 41–5 (2001).

319. Hennig, M., Williamson, J. R., Brodsky, A. S. & Battiste, J. L. Recent advances in RNA structure determination by NMR. *Curr. Protoc. Nucleic Acid Chem.* **Chapter 7,** Unit 7.7 (2001).

320. Lakomek, N.-A. *et al.* Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales. *Angew. Chem. Int. Ed. Engl.* **52,** 3911–5 (2013).

321. Goldman, M. Interference effects in the relaxation of a pair of unlike nuclei. *J. Magn. Reson.* **60,** 437–452 (1984).

322. Tjandra, N., Szabo, A. & Bax, A. Protein Backbone Dynamics and 15 N Chemical Shift Anisotropy from Quantitative Measurement of Relaxation Interference Effects. *J. Am. Chem. Soc.* **118,** 6986–6991 (1996).

323. Palmer, A. G., Kroenke, C. D. & Loria, J. P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339,** 204–38 (2001).

324. Fushman, D., Tjandra, N. & Cowburn, D. Direct Measurement of 15 N Chemical Shift Anisotropy in Solution. *J. Am. Chem. Soc.* **120,** 10947–10952 (1998).

325. Cai, Z. & Tinoco, I. Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry* **35,** 6026–36 (1996).

326. Bayfield, M. A., Thompson, J. & Dahlberg, A. E. The A2453-C2499 wobble base pair in Escherichia coli 23S ribosomal RNA is responsible for pH sensitivity of the peptidyltransferase active site conformation. *Nucleic Acids Res.* **32,** 5512–8 (2004).

327. Cromsigt, J. A., Hilbers, C. W. & Wijmenga, S. S. Prediction of proton chemical shifts in RNA. Their use in structure refinement and validation. *J. Biomol. NMR* **21,** 11–29 (2001).

328. Dejaegere, A., Bryce, R. A. & Case, D. A. *Modeling NMR Chemical Shifts*. **732,** (American Chemical Society, 1999).

329. VanLoock, M. S. *et al.* Movement of the decoding region of the 16 S ribosomal RNA accompanies tRNA translocation. *J. Mol. Biol.* **304,** 507–15 (2000).

330. Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406,** 318–22 (2000).

331. Carr, H. & Purcell, E. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **94,** 630–638 (1954).

332. Meiboom, S. & Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **29,** 688 (1958).

333. 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.* **39,** 4340–51 (2011).

334. Wunderlich, C. H. *et al.* Synthesis of (6- 13 C) Pyrimidine Nucleotides as Spin-Labels for RNA Dynamics. (2012).

335. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches

for measuring protein dynamics. *Biochim. Biophys. Acta* **1814,** 942–68 (2011).

336. Palmer, A. G., Grey, M. J. & Wang, C. Solution NMR spin relaxation methods for characterizing chemical exchange in high-molecular-weight systems. *Methods Enzymol.* **394,** 430–65 (2005).

337. Lisi, G. P. & Patrick Loria, J. Using NMR spectroscopy to elucidate the role of molecular motions in enzyme function. *Prog. Nucl. Magn. Reson. Spectrosc.* **92-93,** 1–17 (2016).

338. Hansen, D. F., Vallurupalli, P. & Kay, L. E. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J. Biomol. NMR* **41,** 113–20 (2008).

339. Bouvignies, G., Vallurupalli, P. & Kay, L. E. Visualizing side chains of invisible protein conformers by solution NMR. *J. Mol. Biol.* **426,** 763–74 (2014).

340. Hansen, A. L. & Kay, L. E. Quantifying millisecond time-scale exchange in proteins by CPMG relaxation dispersion NMR spectroscopy of side-chain carbonyl groups. *J. Biomol. NMR* **50,** 347–55 (2011).

341. Hansen, A. L., Lundström, P., Velyvis, A. & Kay, L. E. Quantifying millisecond exchange dynamics in proteins by CPMG relaxation dispersion NMR using side-chain 1H probes. *J. Am. Chem. Soc.* **134,** 3178–89 (2012).

342. Baldwin, A. J., Religa, T. L., Hansen, D. F., Bouvignies, G. & Kay, L. E. 13CHD2 methyl group probes of millisecond time scale exchange in proteins by 1H relaxation dispersion: an application to proteasome gating residue dynamics. *J. Am. Chem. Soc.* **132,** 10992–5 (2010).

343. Hansen, D. F., Neudecker, P., Vallurupalli, P., Mulder, F. A. A. & Kay, L. E. Determination of Leu side-chain conformations in excited protein states by NMR relaxation dispersion. *J. Am. Chem. Soc.* **132,** 42–3 (2010).

344. Vallurupalli, P., Hansen, D. F., Lundström, P. & Kay, L. E. CPMG relaxation dispersion NMR experiments measuring glycine 1H alpha and 13C alpha chemical shifts in the ‘invisible’ excited states of proteins. *J. Biomol. NMR* **45,** 45–55 (2009).

345. 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 (13)C labeled samples. *J. Biomol. NMR* **42,** 35–47 (2008).

346. Vallurupalli, P., Bouvignies, G. & Kay, L. E. Increasing the exchange time-scale that can be probed by CPMG relaxation dispersion NMR. *J. Phys. Chem. B* **115,** 14891–900 (2011).

347. Korzhnev, D. M., Neudecker, P., Mittermaier, A., Orekhov, V. Y. & Kay, L.

E. Multiple-site exchange in proteins studied with a suite of six NMR relaxation dispersion experiments: an application to the folding of a Fyn SH3 domain mutant. *J. Am. Chem. Soc.* **127,** 15602–11 (2005).

348. Orekhov, V. Y., Korzhnev, D. M. & Kay, L. E. Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins. *J. Am. Chem. Soc.* **126,** 1886–91 (2004).

349. Loria, J. P., Rance, M. & Palmer, A. G. A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J. Biomol. NMR* **15,** 151–5 (1999).

350. Carver, J. . & Richards, R. . A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. *J. Magn. Reson.* **6,** 89–105 (1972).

351. Luz, Z. & Meiboom, S. Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution—Order of the Reaction with Respect to Solvent. *J. Chem. Phys.* **39,** 366 (1963).

352. McConnell, H. M. Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.* **28,** 430 (1958).

353. Weininger, U., Respondek, M. & Akke, M. Conformational exchange of aromatic side chains characterized by L-optimized TROSY-selected 13C CPMG relaxation dispersion. *J. Biomol. NMR* **54,** 9–14 (2012).

354. Vallurupalli, P., Hansen, D. F. & Kay, L. E. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11766–71 (2008).

355. Otten, R., Villali, J., Kern, D. & Mulder, F. A. A. Probing microsecond time scale dynamics in proteins by methyl (1)H Carr-Purcell-Meiboom-Gill relaxation dispersion NMR measurements. Application to activation of the signaling protein NtrC(r). *J. Am. Chem. Soc.* **132,** 17004–14 (2010).

356. Korzhnev, D. M., Kloiber, K. & Kay, L. E. Multiple-quantum relaxation dispersion NMR spectroscopy probing millisecond time-scale dynamics in proteins: theory and application. *J. Am. Chem. Soc.* **126,** 7320–9 (2004).

357. Kleckner, I. R. & Foster, M. P. GUARDD: user-friendly MATLAB software for rigorous analysis of CPMG RD NMR data. *J. Biomol. NMR* **52,** 11–22 (2012).

358. Owen, D. & Kühn, L. C. Noncoding 3’ sequences of the transferrin receptor gene are required for mRNA regulation by iron. *EMBO J.* **6,** 1287–1293 (1987).

359. Hentze, M. W. *et al.* Identification of the iron responsive element for the translational regulation of human ferritin mRNA. *Science (80-.).* **238,** 1570–1573 (1987).

360. Aziz, N. & Munro, H. N. Iron regulates ferritin mRNA translation through a segment of its 5' untranslated region. *Proc. Natl. Acad. Sci.* **84,** 8478–8482 (1987).

361. Kaptain, S. *et al.* A regulated RNA binding protein also possesses aconitase activity. *Proc. Natl. Acad. Sci.* **88,** 10109–10113 (1991).

362. Guo, B., Yu, Y. & Leibold, E. A. Iron Regulates Cytoplasmic Levels of a Novel Iron-responsive Element-binding Protein without Aconitase Activity. *J. Biol. Chem.* **269,** 24252–24260 (1994).

363. McCallum, S. A. & Pardi, A. Refined Solution Structure of the Iron-responsive Element RNA Using Residual Dipolar Couplings. *J. Mol. Biol.* **326,** 1037–1050 (2003).

364. Walden, W. E. *et al.* Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science* **314,** 1903–8 (2006).

365. Laing, L. G. & Hall, K. B. A model of the iron responsive element RNA hairpin loop structure determined from NMR and thermodynamic data. *Biochemistry* **35,** 13586–96 (1996).

366. Hansen, D. F., Vallurupalli, P. & Kay, L. E. An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J. Phys. Chem. B* **112,** 5898–904 (2008).

367. Hall, K. B. & Tang, C. 13C relaxation and dynamics of the purine bases in the iron responsive element RNA hairpin. *Biochemistry* **37,** 9323–32 (1998).

368. Miclet, E. *et al.* Relaxation-optimized NMR spectroscopy of methylene groups in proteins and nucleic acids. *J. Am. Chem. Soc.* **126,** 10560–70 (2004).

369. Latham, M. P. & Pardi, A. Measurement of imino 1H-1H residual dipolar couplings in RNA. *J. Biomol. NMR* **43,** 121–9 (2009).

370. Addess, K. J., Basilion, J. P., Klausner, R. D., Rouault, T. A. & Pardi, A. Structure and dynamics of the iron responsive element RNA: implications for binding of the RNA by iron regulatory binding proteins. *J. Mol. Biol.* **274,** 72–83 (1997).

371. De Gregorio, E., Preiss, T. & Hentze, M. W. Translation driven by an eIF4G core domain in vivo. *EMBO J.* **18,** 4865–74 (1999).

372. Wrabl, J. O., Shortle, D. & Woolf, T. B. Correlation between changes in nuclear magnetic resonance order parameters and conformational entropy: molecular dynamics simulations of native and denatured staphylococcal nuclease. *Proteins* **38,** 123–33 (2000).

373. Li, Z., Raychaudhuri, S. & Wand, A. J. Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Sci.* **5,** 2647–50

(1996).

374. Prabhu, N. V, Lee, A. L., Wand, A. J. & Sharp, K. A. Dynamics and entropy of a calmodulin-peptide complex studied by NMR and molecular dynamics. *Biochemistry* **42,** 562–70 (2003).

375. Frederick, K. K., Sharp, K. A., Warischalk, N. & Wand, A. J. Re-evaluation of the model-free analysis of fast internal motion in proteins using NMR relaxation. *J. Phys. Chem. B* **112,** 12095–103 (2008).

376. Kasinath, V., Sharp, K. A. & Wand, A. J. Microscopic insights into the NMR relaxation-based protein conformational entropy meter. *J. Am. Chem. Soc.* **135,** 15092–100 (2013).

377. Allnér, O., Foloppe, N. & Nilsson, L. Motions and entropies in proteins as seen in NMR relaxation experiments and molecular dynamics simulations. *J. Phys. Chem. B* **119,** 1114–28 (2015).

378. Frank, A. T., Law, S. M. & Brooks, C. L. A simple and fast approach for predicting (1)H and (13)C chemical shifts: toward chemical shift-guided simulations of RNA. *J. Phys. Chem. B* **118,** 12168–75 (2014).

379. Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying 'invisible' excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134,** 8148–61 (2012).

380. 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* **116,** 14311–7 (2012).

381. Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V. & Wedekind, J. E. Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.* **286,** 24626–37 (2011).

382. Copeland, R. A. Conformational adaptation in drug-target interactions and residence time. *Future Med. Chem.* **3,** 1491–501 (2011).

383. Corbino, K. A. *et al.* Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol.* **6,** R70 (2005).

384. Gilbert, S. D., Rambo, R. P., Van Tyne, D. & Batey, R. T. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.* **15,** 177–82 (2008).

385. Kelley, J. M. & Hamelberg, D. Atomistic basis for the on-off signaling mechanism in SAM-II riboswitch. *Nucleic Acids Res.* **38,** 1392–400 (2010).

386. Haller, A., Rieder, U., Aigner, M., Blanchard, S. C. & Micura, R. Conformational capture of the SAM-II riboswitch. *Nat. Chem. Biol.* **7,** 393–400 (2011).

387. Sekhar, A. & Kay, L. E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 12867–74 (2013).

388. Anthis, N. J. & Clore, G. M. Visualizing transient dark states by NMR spectroscopy. *Q. Rev. Biophys.* **48,** 35–116 (2015).

389. 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* **480,** 268–72 (2011).

390. 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.* **136,** 20–3 (2014).

391. McConnell, H. M. Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.* **28,** 430–1 (1958).

392. Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. (1999).

393. Hansen, A. L., Bouvignies, G. & Kay, L. E. Probing slowly exchanging protein systems via ^{13}Cα-CEST: monitoring folding of the Im7 protein. *J. Biomol. NMR* **55,** 279–89 (2013).

394. Zhou, Y. & Yang, D. 13Cα CEST experiment on uniformly 13C-labeled proteins. *J. Biomol. NMR* **61,** 89–94 (2015).

395. Grace, C. R. R. & Riek, R. Pseudomultidimensional NMR by spin-state selective off-resonance decoupling. *J. Am. Chem. Soc.* **125,** 16104–13 (2003).

396. Savinov, A., Perez, C. F. & Block, S. M. Single-molecule studies of riboswitch folding. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839,** 1030–1045 (2014).

397. Houck-Loomis, B. *et al.* An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* **480,** 561–4 (2011).

398. Santner, T., Rieder, U., Kreutz, C. & Micura, R. Pseudoknot preorganization of the preQ1 class I riboswitch. *J. Am. Chem. Soc.* **134,** 11928–31 (2012).

399. Lee, J., Dethoff, E. A. & Al-Hashimi, H. M. Invisible RNA state dynamically couples distant motifs. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9485–90 (2014).

400. Fürtig, B., Nozinovic, S., Reining, A. & Schwalbe, H. Multiple conformational states of riboswitches fine-tune gene regulation. *Curr. Opin. Struct. Biol.* **30,** 112–24 (2015).

401. Yu, C.-H. & Olsthoorn, R. C. L. Monitoring ribosomal frameshifting as a platform to screen anti-riboswitch drug candidates. *Methods Enzymol.* **550,**

385–93 (2015).

402. Plant, E. P., Sims, A. C., Baric, R. S., Dinman, J. D. & Taylor, D. R. Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production. *Viruses* **5,** 279–94 (2013).

403. Plant, E. P. & Dinman, J. D. The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front. Biosci.* **13,** 4873–81 (2008).

404. Park, S.-J., Kim, Y.-G. & Park, H.-J. Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J. Am. Chem. Soc.* **133,** 10094–100 (2011).

405. Serianni, A. S. & Bondo, P. B. 13C-Labeled D-Ribose: Chemi-Enzymatic Synthesis of Various Isotopomers. *J. Biomol. Struct. Dyn.* **11,** 1133–1148 (1994).

406. 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.* **55,** 2724–2727 (2016).

407. Berlin, K., O'Leary, D. P. & Fushman, D. Improvement and analysis of computational methods for prediction of residual dipolar couplings. *J. Magn. Reson.* **201,** 25–33 (2009).

408. Berlin, K. *et al.* Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *J. Am. Chem. Soc.* **135,** 16595–609 (2013).

409. Kao, C., Zheng, M. & Rudisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5,** 1268–1272 (1999).

410. Guillerez, J., Lopez, P. J., Proux, F., Launay, H. & Dreyfus, M. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 5958–5963 (2005).

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

412. Gumbs, O. H., Padgett, R. A. & Dayie, K. T. Fluorescence and solution NMR study of the active site of a 160-kDa group II intron ribozyme. *RNA* **12,** 1693–1707 (2006).

413. Kieft, J. S. & Batey, R. T. A general method for rapid and nondenaturing purification of RNAs ABSTRACT. *RNA* **10,** 988–995 (2004).