ABSTRACT

Title of Dissertation: CHARACTERIZATION OF NON-CODING RNAS VIA NMR SPECTROSCOPY: ANALYSIS OF STRUCTURE, THERMAL STABILITY, AND DYNAMICS

Hyeyeon Nam, Doctor of Philosophy, 2019

Dissertation directed by: Theodore K. Dayie, Associate Professor, Biochemistry

Non-coding RNAs are involved in various cellular processes and characterization of these RNAs may provide better insights into their functional roles. NMR spectroscopy is a powerful biophysical tool that can provide residue-specific information. Herein we examine an RNA triple helix at the 3' end of the lncRNA MALAT1, which may be a potential therapeutic target for cancer treatment. We investigate the local stability of the MALAT1 triple helix by analyzing the individual base-pair stability via NMR spectroscopy. In addition, we screened small molecules to identify the compounds that can selectively target the MALAT1 triple helix. In the second part, we studied the effect of dipolar couplings on the relaxation measurements of various non-coding RNAs using both computational and experimental measurements. The results suggest an increasing contribution of the dipolar coupling effect with the increasing size of the RNA.

CHARACTERIZATION OF NON-CODING RNAS VIA NMR SPECTROSCOPY: ANALYSIS OF STRUCTURE, THERMAL STABILITY, AND DYNAMICS

by

Hyeyeon Nam

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
2019

Advisory Committee:

Professor Theodore K. Dayie, Chair
Professor Nicole LaRonde
Professor Paul Paukstelis
Professor Wade C. Winkler
Professor Jonathan D. Dinman

Dedication

To my beloved family for their endless love and support.

Acknowledgements

Throughout my graduate career, I have been very fortunate to receive great help from many people. I would like to acknowledge all those who have supported me in this challenging journey towards a Ph.D.

I would like to first thank my advisor, Dr. T. Kwaku Dayie, for guiding me through all the work to become an independent scientist. I have not only gained the scientific knowledge but also learned how to be a good researcher with critical perspectives. I greatly appreciate his patience, wisdom, and support.

I appreciate all my committee members for their patience and thoughtful advice throughout my Ph.D. research. I would also like to thank our collaborator, Dr. Stuart F.J. Le Grice for giving me an opportunity to work on the MALAT1 project. He has supported my research in many ways and has always been willing to share his thoughts on the project.

I appreciate all the members of our lab. To Regan, Andrew, My Le, and Luigi, our former lab members, who have been always supportive and kindly answered all my questions. Owen, who has shared most of my journey in graduate school. Thanks for all the chats and discussions we had. Lukasz and Mary, for reading through my thesis and giving all the thoughtful comments. Daniel, for making all those samples for my project. He has made most of the samples required for the second part of this thesis. To Amr and Sydney, for all the hard work and making our lab a cheerful space.

I would also like to thank all my friends. Diana and Yeji, for making my graduate school life enjoyable. I missed all the chats we had after you left. Thomas, for always encouraging me and reading through my thesis even though your major is far from science. Very special thanks to my family. To my sister, Hye Woong, for all the cheerful chats we had over the phone. Your jokes always made me laugh and it was a huge encouragement for my life away from home. To my parents for your unconditional love and support. None of this would have been possible without you. Dad, thank you for always understanding me and all the advice you gave me whenever I felt lost. Mom, thank you for always believing in me and being there to encourage me. Finally, I would like to thank my grandparents for always trusting me and praying for me.

TABLE OF CONTENTS

List of Tables

Part 1. Characterization of MALAT1 Triple Helix and Identification of Small Molecule Selectively Targeting MALAT1

Part 2. Deleterious Effects of Carbon-Carbon Coupling on RNA NMR Dynamics

List of Figures

Part 1. Characterization of MALAT1 Triple Helix and Identification of Small Molecules Selectively Targeting MALAT1

Part 2. Deleterious Effects of Carbon-Carbon Coupling on RNA NMR Dynamics

Abbreviation

ASO: Antisense Oligonucleotide

COSY: Correlation Spectroscopy

DSS: 3-(Trimethylsilyl)propane-1-sulfonate

EMT: Epithelial-to-Mesenchymal Transition

ENE: Expression and Nuclear Retention Element

GTP: Guanosine Triphosphate

HDV: Hepatitis Delta Virus

HSQC: Heteronuclear Single Quantum Coherence

KSHV: Kaposis's Sarcoma-Associated Herpesvirus

ITC: Isothermal Titration Calorimetry

lncRNA: Long Noncoding RNA

MALAT1: Metastasis – Associated Lung Adenocarcinoma Transcript 1

mascRNA: MALAT1-Associated Small Cytoplasmic RNA

MENβ: Multiple Endocrine Neoplasia-β

METLL16: Methyltransferase-like protein 16

MMTV: Mouse Mammary Tumor Virus

NEAT2: Nuclear Enriched Abundant Transcripts 2

NMR: Nuclear Magnetic Resonance

NOESY: Nuclear Overhauser Effect Spectroscopy

NSCLC: Non-Small Cell Lung Cancer

nt: nucleotides

rNTP: Ribonucleoside Triphosphate

PAGE: Poly Acrylamide Gel Electrophoresis

PAN: Polyadenylated Nuclear

PCR: Polymerase Chain Reaction

PDB: Protein Databank

PyMT: Polyoma Middle Tumor-Antigen

REBURP: Refocusing Band-Selective Uniform Response Pure-Phase

RF: Radiofrequency

RNA: Ribonucleic Acid

SMM: Small Molecules Microarray

STD: Saturation Transfer Difference

tRNA: Transfer RNA

UTP: Uridine Triphosphate

Part 1

Characterization of MALAT1 Triple Helix and Identification of Small Molecules Selectively Targeting MALAT1

1. Introduction

1.1 Background and history of lncRNA MALAT1

Long noncoding RNAs (lncRNAs) are defined as transcripts > 200 nucleotides (nt) in length and are known to be involved in many biological functions, such as transcription, translation, cellular differentiation, gene expression regulation, cell cycle regulation, and chromatin modification.[1, 2] In addition, increasing studies have shown the involvement of lncRNAs in various diseases including cancer.[3, 4] In particular, numerous studies have focused on MALAT1 (Metastasis-Associated Lung Adenocarcinoma Transcript 1) to characterize its role in various human cancers.[5, 6]

MALAT1 is an 8.5 kb lncRNA, located on chromosome 11q13, and was identified in 2003 by Ji *et al*. from a study of metastasizing non-small cell lung cancer (NSCLC).[7] NSCLC represents ~ 80 % of all lung carcinoma and has an overall 5-year survival rate of 15 %. In this study, the authors showed a greater accumulation level of MALAT1 by three-fold in metastasizing NSCLCs compared to non-metastasizing tumors.[7] Since the metastatic spread of tumor cells is the main cause of cancer-related death,[8] metastasis-associated genes, such as MALAT1, can potentially predict patient survival. Indeed, only 9 % (2/22) of patients with low MALAT1 accumulation levels died, while more than 40 % (12/28) of patients with high accumulation levels died during a 5-year follow-up period.[7] Further analysis of the MALAT1 sequence in detailed database searches led to the finding of homologs for *Pan troglodytes* (chimpanzee), *Macaca mulatta* (rhesus monkey) and *Mus musculus* (mouse). The mouse MALAT1 (mMALAT1) sequence,

in particular, was 69.3 % identical to the human MALAT1 (hMALAT1) sequence, indicating the conservation of the MALAT1 gene.[7]

Following this previous work, numerous ongoing research efforts have been engaged in characterizing the function of MALAT1 in various cancers (Table 1.1).[6] These studies demonstrated a substantial contribution of MALAT1 in cancer development and progression by influencing cell proliferation, migration, invasion, and tumor growth.[5, 6] Cancer metastasis, in particular, was shown to be modulated by MALAT1 in various cancer types through regulation of epithelial-to-mesenchymal transition (EMT).[25, 26, 29] Epithelial cells are organized and lack

Table 1. 1. Role of MALAT1 in human cancer

Cancer type	Role of MALAT1	Reference
NSCLC	Associated with poor prognosis	[7, 9]
	Involved in cell proliferation, migration and invasion	[9-11]
	Regulates expression of metastasis-associated genes	[12]
Breast cancer	Involved in cell proliferation an cell cycle progression	[13, 14]
	Influence the tumor growth and metastasis	[15, 16]
Liver cancer	Up-regulated in murine and human tumor samples	[17, 18]
	Increase risk of tumor recurrence after liver transplantation	[18]
	Maintain the proliferation potential	[19]
Pancreatic cancer	Diagnostic biomarker	[20, 21]
	Influence cell proliferation, migration and invasion ability	[22, 23]
Renal cell carcinoma	Involved in cell proliferation, migration and invasion	[24]
Bladder cancer	Affect cell migration by regulating the expression of EMT associated gene level	[25]
	Promote metastasis by TGFβ induced EMT	[26]
Cervical cancer	Associated with cell proliferation, invasion and apoptosis	[27, 28]
	Influence cell invasion and metastasis by regulating EMT	[29]
Colorectal carcinoma	Involved in cell proliferation, migration and invasion	[30]
	Prognostic parameter for survival and promotes chemoresistance	[31]

NSCLC: non-small cell lung cancer, EMT: epithelial-to-mesenchymal transition, TGFβ: transforming growth factor β

mobility, in contrast to mesenchymal cells, which have a greater ability of migration. Therefore, regulation of EMT affects the migration of tumor cells and thus, the metastasis of cancer. Although the molecular mechanisms of how MALAT1 functions are not fully understood, these studies suggest that MALAT1 influences cancer cells in diverse ways by directly or indirectly interacting with other biomolecules including proteins, RNAs and DNAs.[32, 33]

In addition, many studies have shown the effect of MALAT1 knockdown in different tumor cells. In human lung cancer cells, depletion of MALAT1 using zinc-finger nucleases showed a significant reduction of both cell migration and formation of tumor nodules in a mouse xenograft.[12] Furthermore, inhibition of MALAT1 using antisense oligonucleotides (ASOs) prevented formation of metastasis after tumor implantation.[12] Studies conducted by Spector and co-workers showed slower tumor growth and loss of migration in the mouse mammary tumor virus (MMTV) – polyoma middle tumor antigen (PyMT) mouse mammary tumor model when MALAT1 was knocked down with ASOs or gene knockdown strategies.[15] These previous reports show a critical function of MALAT1 in the development of various cancers and suggest a potential role of MALAT1 as a therapeutic target for cancer treatment.

1.2 Expression and 3' end processing of MALAT1

MALAT1 shows broad expression in normal human and mouse tissues and is also named Nuclear Enriched Abundant Transcripts 2 (NEAT2) due to its

localization in nuclear speckles, which are small subnuclear domains for storage and modification of pre-mRNA splicing factors.[7, 34-36] Analysis of the protein composition of nuclear speckles indicates the presence of RNA polymerase II, suggesting the involvement of speckles in mechanisms of gene expression.[35, 37] Thus, localization of MALAT1 in nuclear speckles may implicate the involvement of MALAT1 in the organization and/or regulation of gene expression.[5]

Sequence analysis of the MALAT1 transcripts shows short blocks of highly conserved sequences, especially in the 3' end of the RNA.[38] Spector and co-workers identified mascRNA (MALAT1-Associated Small Cytoplasmic RNA) which is a highly conserved 61 nt small RNA at the 3' end of an ~160 nt region.[38] Immediately upstream of mascRNA, there are two highly conserved U-rich motifs, separated by a stem-loop, followed by a conserved poly A-rich tract. The authors showed RNase P cleavage of nascent MALAT1 transcripts at the 3' end of poly-A rich tract to generate abundant long MALAT1 transcripts and mascRNA (Fig. 1.1).[38] The mascRNA forms a tRNA-like structure and is exported to the cytoplasm after further processing of its 3' end by RNase Z cleavage and CCA addition.[38] In contrast, the resulting 5'-cleavage product of RNase P localizes to nuclear speckles (Fig. 1.1). Interestingly, despite having a short poly(A) tail-like moiety, MALAT1 is a stable noncoding RNA with a half-life of 9-12 hours.[5, 38]

Rather than being added on post-transcriptionally, like canonical poly(A) tails, the MALAT1 poly(A) tail-like moiety is encoded in the genome and is, therefore, part of the nascent MALAT1 transcript. This A-rich motif and the two

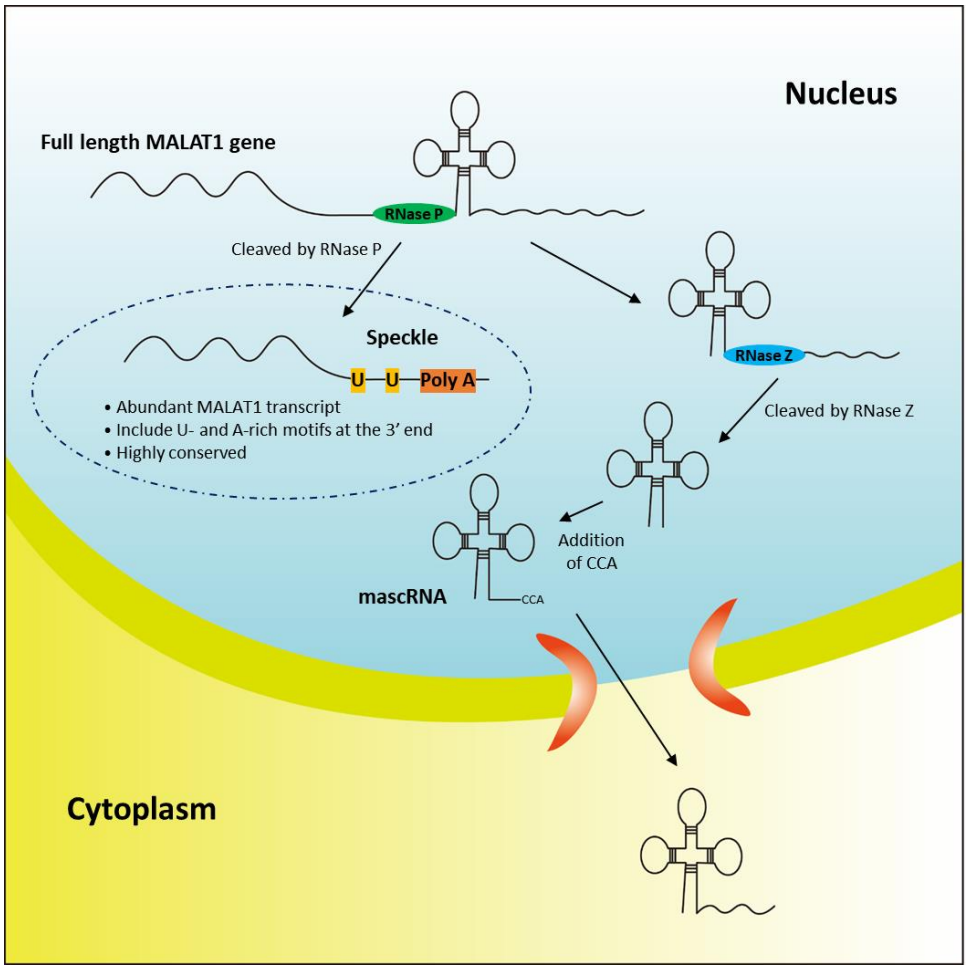


Figure 1. 3' end processing of nascent MALAT1 transcripts. The MALAT1 transcript is cleaved by RNase P immediately after the poly A-rich tract to generate abundant long MALAT1 transcripts and mascRNA. The mascRNA is exported to the cytoplasm after its 3' end is further processed by RNase Z cleavage and CCA addition. The abundant long MALAT1 transcript, with two U-rich motifs and poly(A) tail-like moiety at the 3' end, localizes in nuclear speckle.

upstream U-rich motifs are highly conserved in multiple species suggesting that these sequences may represent functional elements. Mutation of these sequences had no effect on mascRNA biogenesis but the long MALAT1 transcripts were efficiently degraded, suggesting that these motifs are required to stabilize the 3' end of MALAT1.[39] In fact, Wilusz *et al*. showed that the U-rich motifs are important for stabilization of MALAT1 by inhibiting uridylation and degradation by 3'-5' exonucleases.[39] Moreover, the authors proposed a structural model for the 3' end

of MALAT1 in which the two U-rich motifs sequester the downstream poly(A) tail by forming an RNA triple helix (see section 1.3).[39] There is likely a significant free energy cost to unwind the triple helix and thus, the 3' end of MALAT1 is well protected from degradation. These results suggest that the MALAT1 triple helix at the 3' end may have critical regulatory roles by determining the stability of MALAT1. The stability of the MALAT1 transcripts affects the abundance of the MALAT1, which is involved in cancer metastasis as indicated in section 1.1. Taken together, the unique triple helix structure at the 3' end of MALAT1 may serve as a potential therapeutic target.

1.3 Triple helical structure at the 3' end of MALAT1

Due to the unique structure and contribution in stabilizing the MALAT1 transcripts, significant effort has been made to determine the tertiary structure of the triple helix at the 3' end of MALAT1.[39-41] These structural studies have focused on characterizing the highly conserved ~ 90-nt region at the 3' end which includes the two U-rich motifs, separated by a stem-loop, and the downstream poly(A) tail. To solve the X-ray crystal structure of this region, Steitz and co-workers designed a 76-nt modified MALAT1 triple helix core (Fig. 1.2A,B) that has (i) an additional G at the 5' end, (ii) a 6-nt deletion in the linker, and (iii) a shorter stem-loop with a GAAA tetraloop.[41] The accumulation level of the MALAT1 transcripts with the 76-nt triple helix core sequence was comparable to that of the wildtype MALAT1 transcript, indicating the full functionality of the modified construct.[41]

The triple helical structure of MALAT1 is assembled by the poly(A) tail (Fig. 1.2A,B, magenta) at the 3' end of the sequence and the internal U-rich loop (Fig. 1.2A,B, cyan) forming nine U•A-U, one C^+•G-C base triples, and a C-G doublet. One characteristic feature of the MALAT1 triple helix is the C^+•G-C triple and the C-G doublet interrupting the U•A-U triple-helical stack in the middle to generate a bipartite triple helix composed of two independent stacks of base triples (denoted as triplex I and triplex II, Fig. 1.2A).[41] The intervening G-C doublet and C^+•G-C triple contribute to the accumulation level of the transcripts by forming a precisely-

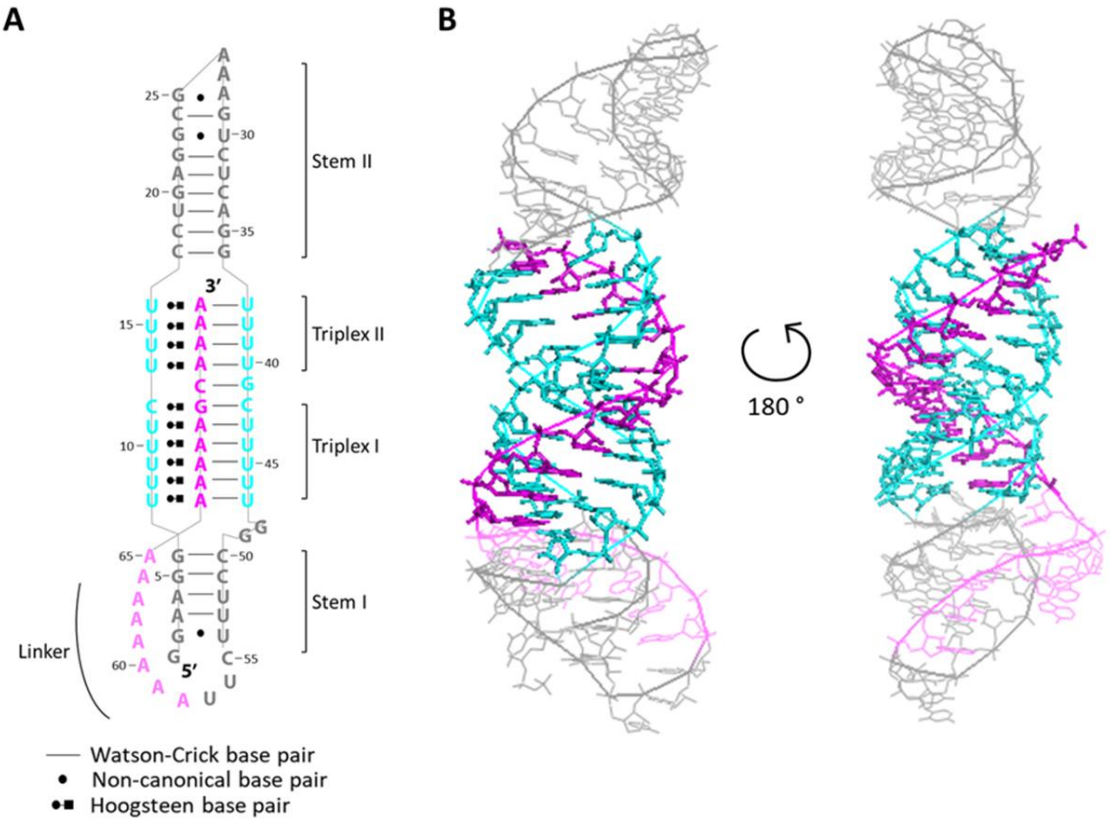


Figure 1. 2. Structure of the MALAT1 triple helix core. (A) Schematic diagram of the 76-nt MALAT1 triple helix core. The poly(A) tail (magenta) at the 3' end and the internal U-rich loop (cyan) form nine U•A-U, one C^+•G-C base triples, and a C-G doublet. The C^+•G-C triple and the C-G doublet interrupt the U•A-U triple-helical stack to generate a bipartite triple helix composed of two independent stacks of base triples (denoted as triplex I and triplex II).[41] **(B)** X-ray crystal structure of MALAT1 triple helix core with 3.1 Å resolution (PDB ID 4plx[41]).

aligned triple helix with a blunt end.[40, 41] In their absence, overhanging A nucleotides at the 3' end may lead to degradation by 3'-5' exonucleases.[40]

Furthermore, a 1-nt gap in the Hoogsteen strand at the C-G doublet allows reorientation of ~ 15° between the helical axes of triplex I and triplex II. Brown *et al*. described this reorientation as a 'reset' of the helical axis.[41] The reset of the helical axis enables the formation of a relatively long triple helix (11 consecutive Watson-Crick base pairs and 10 Hoogsteen base pairs) without destabilization. In fact, most RNA triple helices that have been characterized by structural studies have three to six consecutive base triples.[41-46] Indeed, a computational analysis showed that the successive RNA base triples of more than four to six can be a destabilizing factor depending on the specific sequence due to steric hindrance.[41] It is also important to note that the N3 position of C12 needs to be protonated to form C^+•G-C triple which suggests that the stability of the MALAT1 triple helix may depend on pH. Melting temperature analysis of the MALAT1 triple helix at pH 5 and pH 7 showed the existence of Hoogsteen face interactions at both pHs. While the melting temperature of the Watson-Crick face was pH-independent, that of the Hoogsteen face increased at pH 5, indicating greater stability at lower pH.[40]

1.4 lncRNAs with MALAT1-like structures

In addition to MALAT1, there are a few known lncRNAs with a MALAT1-like triple helical structure at their 3' ends, characterized by an internal U-rich loop that sequesters downstream poly(A)-rich tract (Fig. 1.3).[42, 47, 48] For instance, Kaposi's

sarcoma-associated herpesvirus (KSHV) produces a 1.1-kb lncRNA called polyadenylated nuclear (PAN) RNA.[42, 47, 49] PAN RNA contains the expression and nuclear retention element (ENE), a 79-nt RNA with two internal U-rich loops separated by a stem-loop. The ENE sequesters the downstream poly(A) tail to form a triple helix with five U•A-U triples (Fig. 1.3, right) .[42] In contrast to the MALAT1 poly(A) tail (Fig. 1.3, left), which is located immediately downstream of

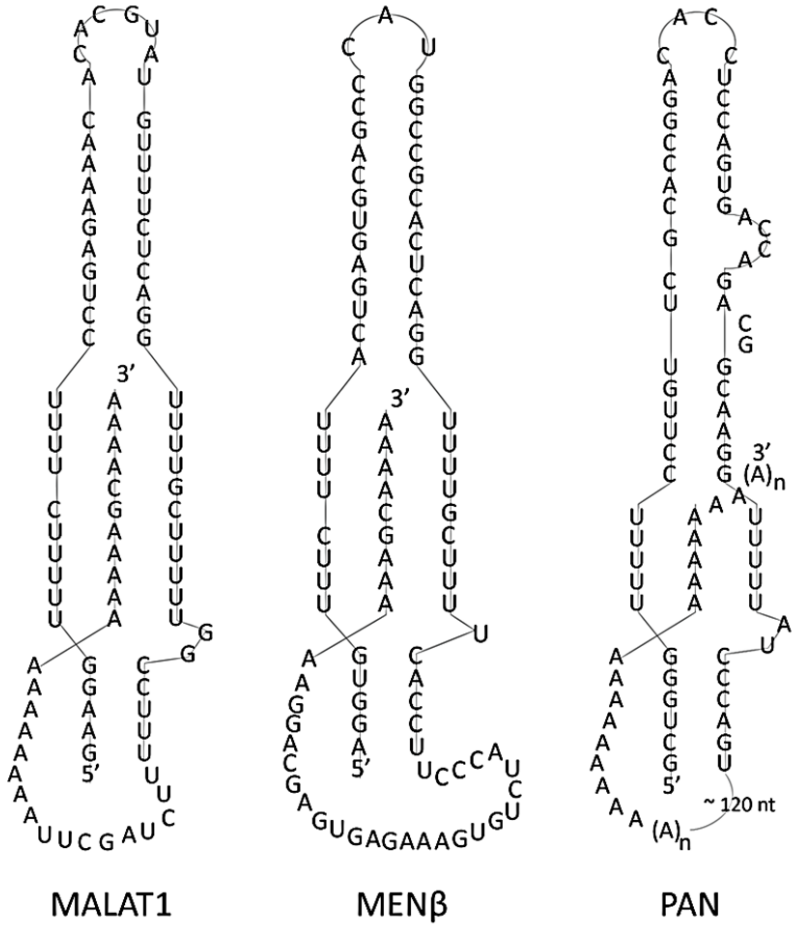


Figure 1. 3. Schematic diagram of triple helical structures from different lncRNAs. Triple helices, responsible for the accumulation of the RNA, are located at the 3' end of each lncRNA The ENE of PAN RNA sequester the downstream poly(A) tail to form a triple helix with five U•A-U triples. In contrast to the MALAT1 poly(A) tail (left), which is located immediately downstream of the U-rich motifs, poly(A) tail in the PAN RNA is ~ 120 nt downstream of the ENE (left). For MENβ, the poly(A) tail is located immediately downstream of two U-rich motifs like MALAT1. Compared to MALAT1 (left), MENβ has one less U•A-U triple base at the bottom of the triplex but has the same C^+•G-C triplet and C-G reset doublet in the middle of the triple helix (middle). The major difference between the triple helices of MALAT1 and MENβ is the bulge sequence at the bottom of the triplex; MALAT1 has a GG dinucleotide bulge (left), while MENβ has a mononucleotide U-bulge (middle).

the U-rich motifs, the PAN RNA poly(A) tail is ~120 nt downstream of the ENE (Fig. 1.3, right). In addition, PAN poly(A) tail is added by post-transcriptional processing, whereas the MALAT1 poly(A)-tail moiety is genomically encoded.[42, 49]

Another example of a lncRNA with a MALAT1-like triple helix is Multiple Endocrine Neoplasia-β (MENβ), which is highly abundant and localizes to the nucleus (Fig. 1.3, middle).[40, 48, 50] MENβ and MALAT1 are both located on chromosome 11q13 and are in close genomic proximity.[48] While MENβ RNA is longer than MALAT1, it is not as well conserved.[48] Like MALAT1, the poly(A) tail of MENβ is genomically encoded and forms triple bases with two U-rich motifs located immediately upstream of poly(A)-rich tract (Fig. 1.3, middle).[40] Compared to MALAT1, MENβ has one less U•A-U triple at the bottom of triplex but has the same C^+•G-C triplet and C-G reset doublet in the middle (Fig. 1.3, left and middle). As described in section 1.3, the C-G reset base pair is the most important region of the triplex to maintain the stability of the molecule when the triple helix has more than 4-5 consecutive triple bases.[40, 41] The major difference between the triple helices of MALAT1 and MENβ is the bulge sequence at the bottom of the triplex. MALAT1 has a GG dinucleotide bulge (Fig. 1.3, left), while MENβ has a mononucleotide U-bulge (Fig. 1.3, middle).[40, 41] Interestingly, all of these triple helical structures are located at the 3' end of the lncRNAs and are responsible for the accumulation of the corresponding RNAs by protecting them from degradation.[40, 42]

1.5 What is next for study of the MALAT1 triple helix?

Studies on MALAT1 have revealed important structural features of the MALAT1 triple helix via mutational and computational analyses and X-ray crystallography.[39-41, 51] It is important to note that not many RNA triple helices have been characterized, especially triple helices with more than 4-6 consecutive base triples.[41-46, 49] Additional studies on the MALAT1 triple helix using nuclear magnetic resonance (NMR) spectroscopy can provide residue-specific information, and further insight into MALAT1 structure and function and broaden our knowledge of RNA triple helices. Here, we focus on the imino proton resonance assignments as an initial step of NMR studies. Then, using the information from the assignments, we investigated the local stability of the MALAT1 triple helix by determining the thermodynamic parameters of individual base pairs.

In addition to NMR characterization of the MALAT triple helix, we have focused on identifying small molecules that selectively target the MALAT1 triple helix. As described in section 1.3, the triple helix contributes to the accumulation level of MALAT1, suggesting its involvement in regulatory roles. Since the accumulation level of MALAT1 transcripts is involved in cancer development, the unique triple helical structure of MALAT1 may serve as a potential drug target. Therefore, we have focused on identifying small molecules that can selectively target the MALAT1 triple helix using various biochemical and biophysical methods.

2. NMR assignments

2.1 MALAT1 constructs for NMR studies

Although NMR studies can provide atom-specific information on biomolecules, they can be limited for RNA due to the fact that RNA encompasses only four bases leading to spectral crowding. The spectral crowding problem worsens as the size of the RNA increases, which hinders accurate NMR assignments. To reduce the spectral crowding, we prepared two MALAT1 constructs for NMR studies: MALAT1 core *cis* (Fig. 1.4A, left) and MALAT1 core *trans* (Fig. 1.4A, right). A 76-nt MALAT1 core *cis* sequence was adapted from the crystal structure solved by Steitz and co-workers.[41] As indicated in section 1.3, this sequence only includes the core part of the triple helix and is shorter than the wildtype sequence.

To further reduce the size of the RNA sample, we also prepared a MALAT1 core *trans* construct (Fig. 1.4A, right) composed of two RNAs: a 55-nt stem-loop (SL) sequence which includes the U-rich loop between stem I and stem II, and a 15-nt poly(A) tail. NMR assignments for stem I and stem II would be easier with the 55-nt SL construct since less spectral crowding is expected due to the smaller size. Prior to starting the NMR studies, native polyacrylamide gel electrophoresis (PAGE) analysis confirmed that the poly(A) tail binds the SL in the *trans* construct to form the complex (Fig. 1.4B). The poly(A) tail, concentration ranging between 0-1000 μM, was mixed with 100 μM of SL sample. The SL band on the native PAGE showed binding with 10 μM of poly(A) tail and was almost completely shifted with 100 μM poly(A) tail. This result suggests a 1:1 ratio of SL:poly(A) binding to form the complex.

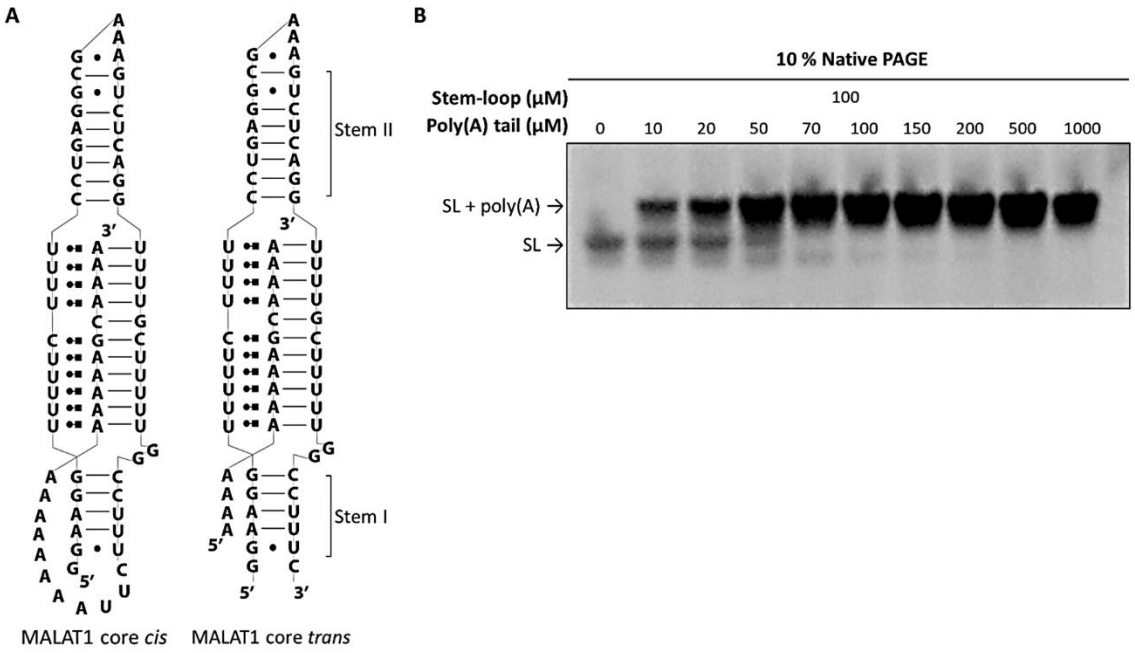


Figure 1. 4. MALAT1 constructs for NMR studies. (A) Schematic diagram of MALAT1 core *cis* (left) and MALAT1 core *trans* (right). MALAT1 core *trans* construct consists of a 55-nt stem-loop (SL) and a 15-nt poly(A) tail. **(B)** 10% native PAGE for the MALAT1 core *trans* construct. The 55-nt SL and the 15-nt poly(A) tail forms a complex in 1:1 ratio.

Although the native PAGE confirmed the formation of the SL:poly(A) complex, it is not clear that they are binding in the orientation needed to form the same triple helix structure as the *cis* construct. To demonstrate the orientation of binding, we used 1D ^{1}H NMR experiments. In a 1D ^{1}H spectrum, the imino protons of guanosines and uridines resonate between 10 – 15 ppm providing useful information on base pairing in the RNA molecule.[52] The peaks for the imino protons are observable only when they are involved in base pairing and therefore protected from the solvent exchange. We collected proton spectra for the MALAT1 core *cis* (Fig. 1.5, Blue) and the *trans* (SL+poly(A) tail) constructs (Fig. 1.5, Red) at 15 °C. For the *trans* construct, the ratio of SL:poly(A) was 1:1.1 to ensure complete formation of the complex for the SL. The observed result showed almost identical spectra for both of the constructs with the same number of peaks and chemical

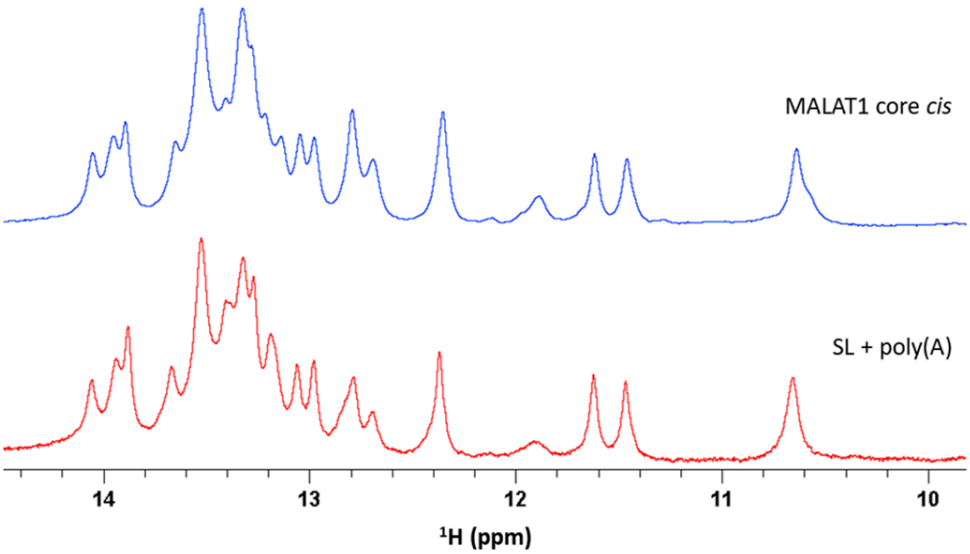


Figure 1. 5. 1D ^{1}H spectra for MALAT1 core *cis* and *trans* constructs. The proton spectra for the MALAT1 core *cis* (Blue) and the *trans* (SL+poly(A) tail) construct (Red) showed almost identical results with the same number of peaks and chemical shifts. This suggests that the *trans* construct has the same base-pairing pattern as the MALAT1 core *cis*, which indicates that it is forming the same triple helical structure.

shifts, suggesting that the *trans* construct has the same base-pairing pattern as the *cis* construct and forms the same triple helical structure. Moreover, the result reconfirmed our observation from native PAGE, which showed a 1:1 ratio of SL:poly(A) tail is sufficient for full conversion of the SL to the complex (Fig. 1.4B). Taken together, our data suggest that the MALAT1 core *trans* construct can be used for NMR assignments and will provide information for further studies for the MALAT1 triple helix.

2.2 Optimal pH for stabilization of C^{+}•G-C triple

As noted in section 1.3, the MALAT1 triple helix has a C^{+}•G-C triple in the middle of the triplex (Fig. 1.2A). The N3 position of cytosine on the Hoogsteen face

needs to be protonated in order to form the hydrogen bond (Fig. 1.6A). Thus, acidic pHs (< 7) are required to stabilize the C^+•G-C triple. Although there is only one C^+•G-C, it is close to the C-G reset base pair and thus, its stability may affect the triple helix. Based on earlier work, we chose the initial buffer conditions as 10 mM potassium phosphate (pH 6.5), 50 mM potassium chloride and 1 mM magnesium chloride.[41] Previous NMR studies on RNA triple helices have reported that imino proton peaks from protonated cytosines resonate around 15 – 16 ppm.[53, 54] We collected 1D proton spectrum (Fig. 1.6B, Blue) at 25 °C in order to check the presence of C^+•G-C triple at pH 6.5. A C^+ peak near 15 ppm (marked with an asterisk in Fig. 1.6B) has a weak signal even after 2048 scans. To test whether lowering pH stabilizes the C^+•G-C triple and give a better signal, we repeated the

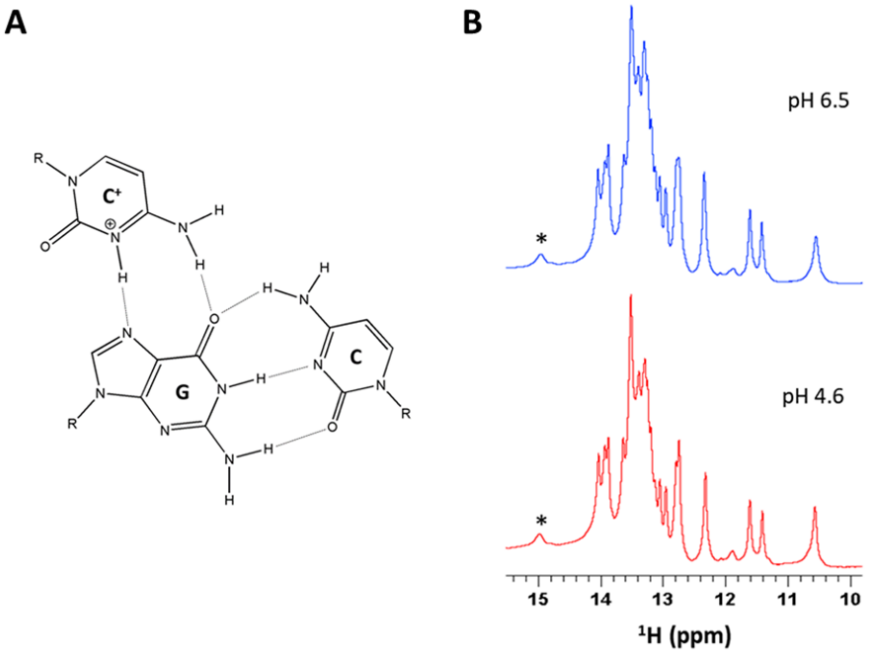


Figure 1. 6. C^+•G-C triple at acidic pH. (A) Schematic diagram of C^+•G-C triple. **(B)** 1D ^{1}H spectra of MALAT1 core cis at pH 6.5 (Blue) and pH 4.6 (Red). The imino proton of protonated cytosine in Hoogsteen face is marked with an asterisk. As can be seen from the spectra, lowering pH did not change the intensity of the signal from C^+ indicating no further stabilization of C^+•G-C triple at pH 4.6.

experiment at pH 4.6 (Fig. 1.6B, Red). We observed no dramatic change in the signal intensity, indicating no further stabilization of C^+•G-C in more acidic pH. Therefore, pH 6.5 is sufficient to stabilize the C^+•G-C triple and was used for further studies.

2.3 Imino assignments of stem-loop

For the assignment of the MALAT1 triple helix, we started from assigning stem I and stem II using the 55-nt SL of the *trans* construct (Fig. 1.7C). To achieve this, two RNA samples were prepared by *in vitro* transcription: one using ribonucleoside triphosphates (rNTPs) that are enriched with ^{15}N and another using unlabeled rNTPs. The first sample was used for both imino Heteronuclear Single Quantum Coherence (HSQC) and ^{15}N-edited 3D nuclear Overhauser effect spectroscopy (NOESY)-HSQC experiments. The unlabeled sample was used for 2D ^{1}H-^{1}H NOESY experiments. The ^{1}H-^{1}H NOESY experiments were recorded with a mixing time of 100 ms, which was determined to be optimal after trial and error. We also performed a 3D NOESY-HSQC at 15 °C with the same mixing time to resolve the peaks that were overlapped in the ^{1}H dimension. Imino assignments were performed mainly using ^{1}H-^{1}H NOESY (Fig. 1.7B) and ^{15}N-^{1}H imino HSQC (Fig. 1.7A) recorded at 15 °C. 3D NOESY-HSQC data was used to resolve the cross-peaks for: i) a G peak at 13.52 ppm and a U peak at 13.54 ppm, ii) a G peak at 11.47 ppm and a U peak at 11.49 ppm, and iii) a G peak at 10.69 ppm and a U peak at 10.70 ppm. (Fig. 1.7A). The imino HSQC spectrum showed peaks corresponding to 4 A-U base pairs, 7 G-C base pairs, 2 Gs and 9 Us involved in

non-canonical base pairs (Fig. 1.7A). Compared to the secondary structure (Fig. 1.7C), we do not observe two G-C base pairs, which would belong to G1-C55 and C17-G36 at the bottom of each stem. Typically, the base pairs at the beginning of the stem are frayed and not observed in the imino HSQC spectrum. Interestingly,

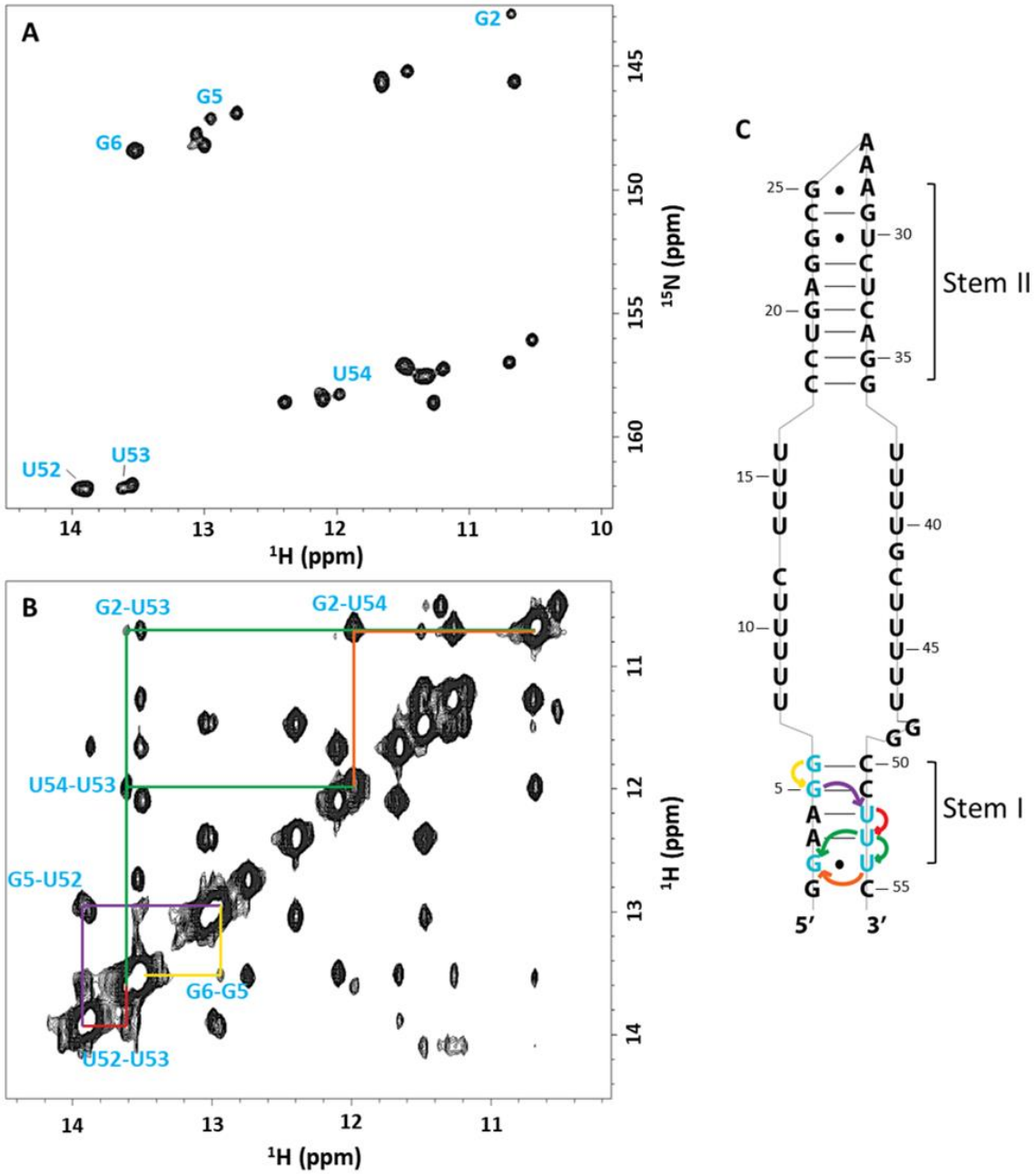


Figure 1. 7. ^{1}H-^{1}H NOESY walk for stem I of the MALAT1 SL. (A) ^{15}N-^{1}H imino HSQC recorded at 15 °C. **(B)** ^{1}H-^{1}H NOESY collected at 15 °C. **(C)** Schematic diagram of MALAT1 SL. The color of arrows corresponds to the color of NOESY walk in (B).

we also observed more U peaks in the non-canonical base pair region than what we expected from the secondary structure. This suggests that the Us in the U-rich loops are forming U•U base pairs, which will be discussed in more detail.

For the assignment process, we first began with stem I of SL (Fig. 1.7). The starting point for the assignment was A3-U53 and A4-U52 base pairs (Fig. 1.7B,C). The SL construct has four A-U base pairs (Fig. 1.7C) and stem I has two consecutive A-U base pairs, A3-U53 and A4-U52, which are not present in stem II. Therefore, a cross-peak between two A-U peaks will correspond to the cross-peak between A3-U53 and A4-U52. Fig. 1.7B shows a NOESY walk, from G6 through U54, with the color of each correlation corresponding to the colored arrows in Fig. 1.7C. G6 at 13.52 ppm has a cross-peak with G5 at 12.95 ppm (yellow in Fig. 1.7B,C) and G5 has a cross-peak with U52 at 13.93 ppm (purple in Fig. 1.7B,C). U52 then correlates to U53 at 13.61 ppm (red in Fig 1.7B,C), which has two other cross-peaks with G2 and U54 at 10.69 ppm and 11.98 ppm, respectively (green in Fig. 1.7B,C). Since G2 and U54 are forming a base pair, there is a cross-peak between these two residues colored in orange in Fig. 1.7B,C.

For the assignment of stem II, we began with the G20-C33 base pair (Fig. 1.8C) between the two A-U base pairs. A G peak with two cross-peaks with the A-U base pairs correspond to G20 as an unambiguous starting point. Fig. 1.8B is displaying a NOESY walk for stem II, starting from G35 to G29, with colors for each walk correlating to the colored arrows in Fig. 1.8C. G35 at 12.75 ppm correlates to U19 at 13.54 ppm (green in Fig. 1.8B,C) and U19 has a cross-peak with G20 at 11.65 ppm (yellow in Fig. 1.8B,C). G20 has a cross-peak with U32 at 13.88 ppm

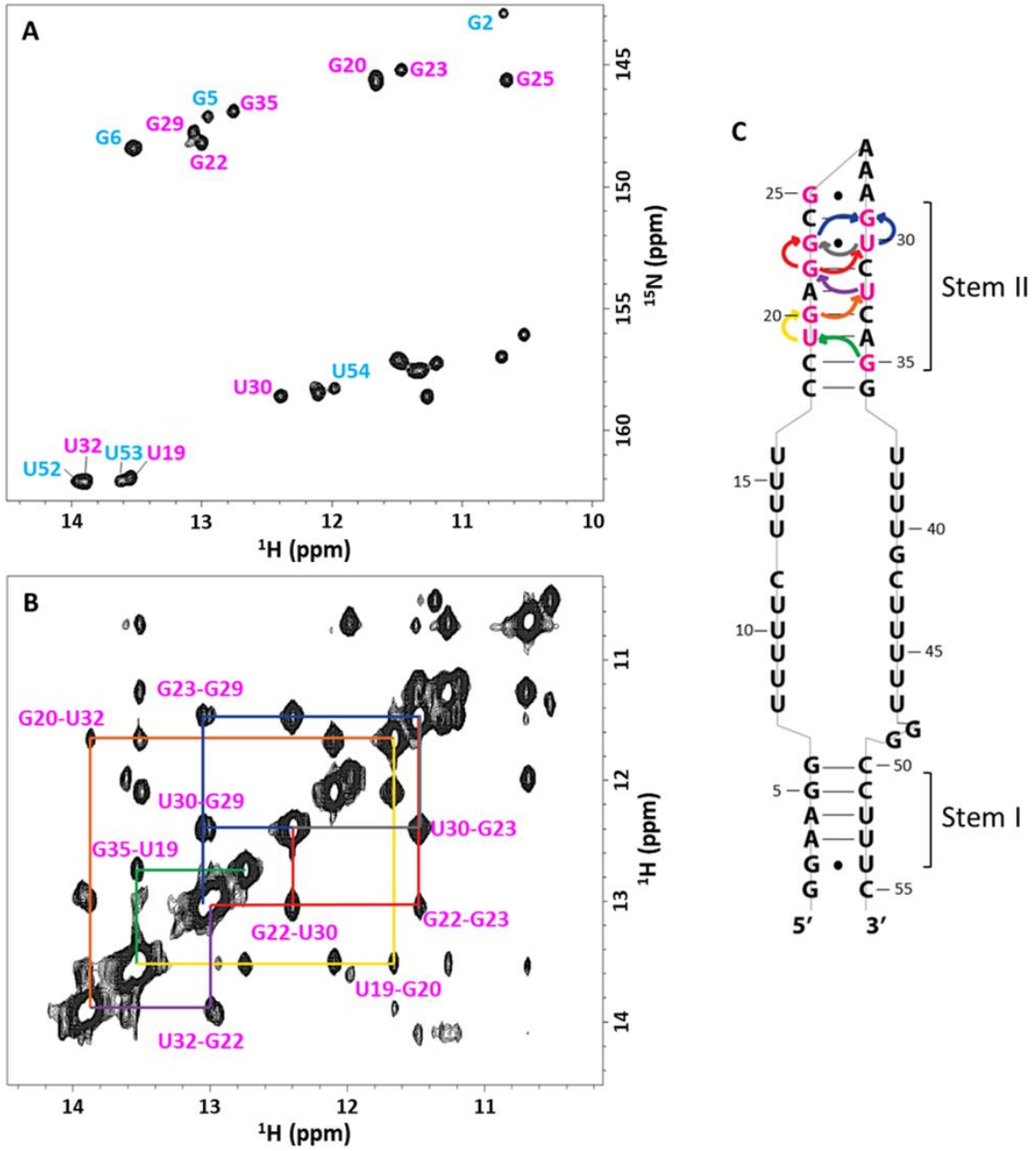


Figure 1. 8. ^{1}H-^{1}H NOESY walk for stem II of the MALAT1 SL. (A) ^{15}N-^{1}H imino HSQC recorded at 15 °C. **(B)** ^{1}H-^{1}H NOESY collected at 15 °C. **(C)** Schematic diagram of MALAT1 SL. The color of arrows corresponds to the color of NOESY walk in (B).

(orange in Fig. 1.8B,C), which then correlates to G22 at 13.00 ppm (purple in Fig. 1.8B,C). G22 has two cross-peaks with G23 at 11.47 ppm and U30 at 12.40 ppm (red in Fig. 1.8B,C). G23 and U30 form a G•U base pair and thus, there is a cross-

peak between these two residues (gray in Fig. 1.8B,C). Both G23 and U30 correlate to G29 at 13.06 ppm (blue in Fig. 1.8B, C). A cross-peak between G29 and G25 was not observed. However, there was only one unassigned G peak left with a chemical shift of 10.65 ppm. This chemical shift is indicative of a non-canonical base pair and thus, we were able to assign this peak as G25.

Fig. 1.8A shows the imino HSQC with the full assignments of the stem I and stem II. As seen, there are several unassigned U peaks in the non-canonical base pair region. These peaks suggest the presence of U•U base pairs in the U-rich loop of SL. Interestingly, we observe cross-peaks between G6 and these U peaks (Fig. 1.9A). G6 has cross-peaks with U at 10.69 and 11.27 ppm (red in Fig. 1.9A), and these two U peaks are also correlated to each other. U at 11.27 ppm also had a cross-peak with U at 11.48 ppm (yellow in Fig. 1.9A). These cross-peaks suggest at least two U•U base pairs form right above G6-C50 (Fig. 1.9B, left). In addition,

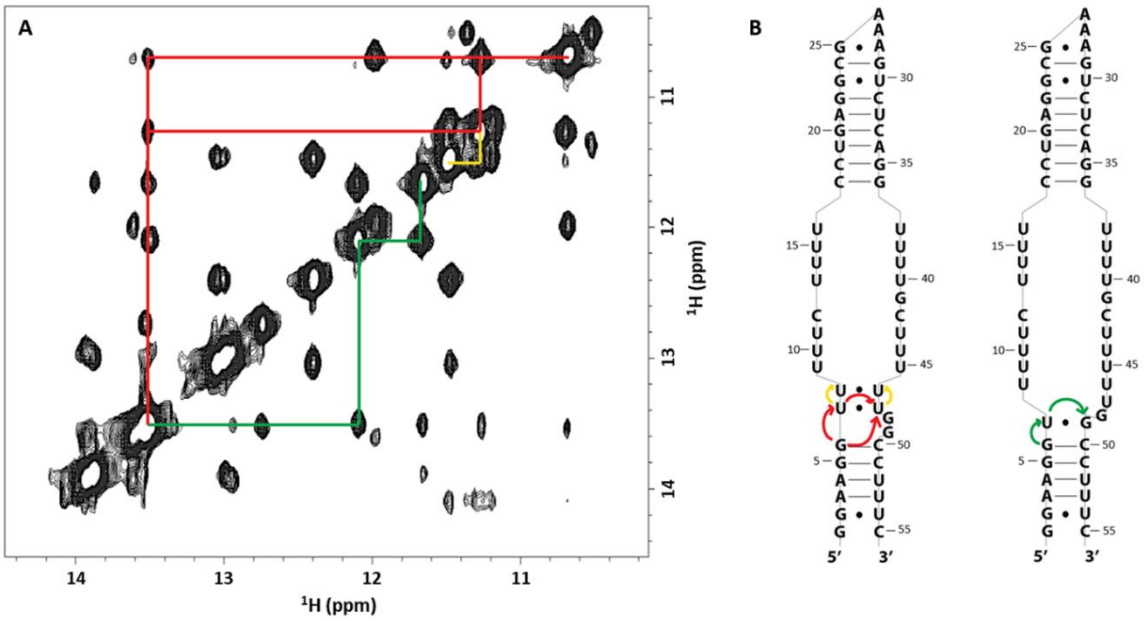


Figure 1. 9. U•U and G•U base pairs forming in U-rich loop. (A) A ^{1}H-^{1}H NOESY walk showing G6 correlating to Us in the non-canonical base pair region. The color of each walk corresponds to the colored arrows in 9B. **(B)** Schematic diagram of U•U and G•U base pairs forming in U-rich loop.

we also observe G6 correlating to a U at 12.10 ppm, which also correlates to a G at 11.65 ppm (green in Fig. 1.9A). These cross-peaks are indicative of a G•U base pair forming next to G6-C50 (Fig. 1.9B). These results suggest that there are at least two different species with different base pair patterns in the U-rich loop (Fig. 1.9B). At higher temperature (30 °C), we still observed the G•U base pair forming but did not observe the U•U base pairs (data not shown), indicating that these U•U base pairs are weaker than other base pairs in stem I and II.

2.4 Imino assignments of SL:poly(A) complex

To monitor the structural change of SL in the triple helix, we mixed SL and poly(A) tail in a 1:1.1 ratio. The SL sample was prepared with uniformly labeled ^{15}N rNTPs, whereas the poly(A) tail was unlabeled. Therefore, all observed imino HSQC peaks would only belong to the SL residues. Fig. 1.10A shows the overlay of ^{15}N-^{1}H imino HSQC spectra of SL (black spectrum) and SL:poly(A) complex (red spectrum) recorded at 15 °C. The first thing we noticed was the absence of U peaks in the non-canonical base pair region (Fig.1.10A), which indicates disruption of the G•U and U•U base pairs in the U-rich loop. This was expected as these U residues are interacting with the poly (A) tail to form the base triples (Fig. 1.10B). We also observed crowded U peaks around 13 – 14 ppm (Fig. 1.10A), which belong to the residues involved in U•A-U triples. It is also important to note that there is one additional G peak for SL:poly(A) complex due to G41 forming a C-G doublet.

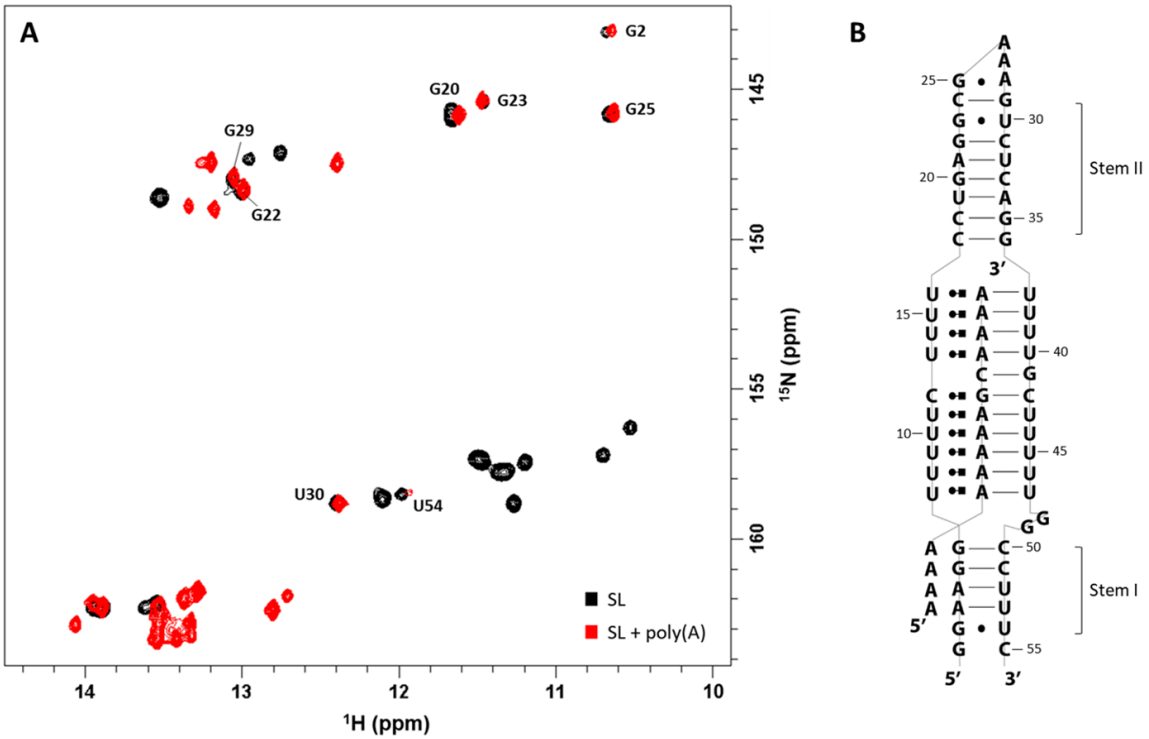


Figure 1. 10. ^{15}N-^{1}H imino HSQC for the SL:poly(A) complex. (A) Overlay of imino HSQC spectra for SL (black) and SL:poly(A) tail complex (red). Both of the spectra were recorded at 15 °C. **(B)** Schematic diagram of MALAT1 core *trans* (SL:poly(A) complex) construct

As poly(A) tail is not directly binding to stem I and stem II, we did not observe dramatic shifts for the majority of stem residues. For those peaks, we were able to assign them from the assignments of SL (Fig. 1.8A). For instance, G20, G22, G23, G25, G29, and U30 are the stem II residues and G2 and U54 are the stem I residues. Peaks of these residues overlaid nicely and therefore were unambiguously assigned (Fig. 1.10A). Although G5, G6, U19, and G35 are also the stem residues, their peaks significantly shifted (Fig. 1.10A), as they are closer to the triple helix and will be perturbed by the addition of poly(A) tail. Assignments of U52 and U53 were also ambiguous as new U peaks were resonating in close proximity. Therefore, assignments for these residues required more careful analysis using ^{1}H-^{1}H NOESY.

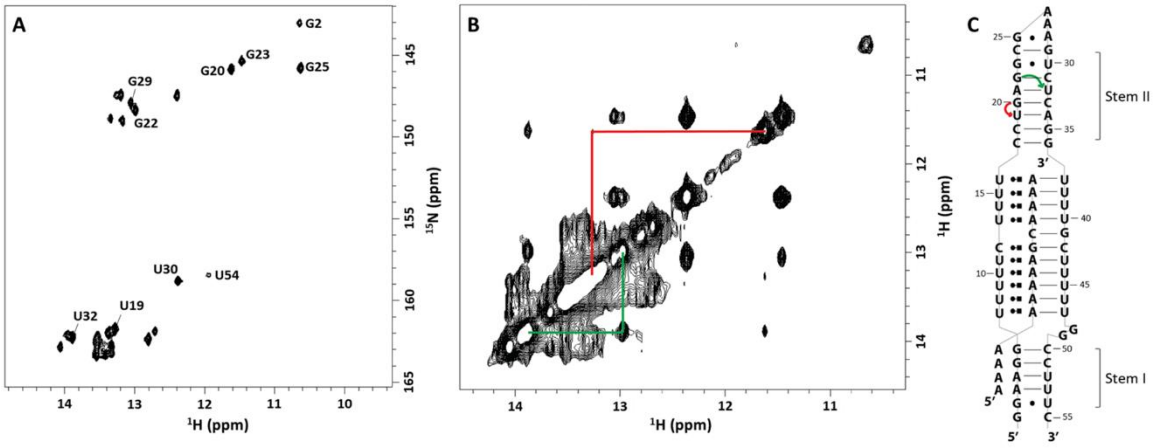


Figure 1. 11. Imino assignments for the SL:poly(A) complex. (A) ^{15}N-^{1}H imino HSQC spectra recorded at 15 °C for SL:poly(A) complex. Assigned peaks are labeled with their residues. **(B)** ^{1}H-^{1}H NOESY spectrum recorded at 15 °C for SL:poly(A) complex. Due to the spectral crowding, many of the cross-peaks were overlapped. **(C)** Schematic diagram of SL:poly(A) complex (MALAT1 core trans). Colored arrows correlated the color of NOESY walk on (B).

Unfortunately, the ^{1}H-^{1}H NOESY spectrum of SL:poly(A) complex was very crowded with overlapping peaks that hindered full assignments (Fig. 1.11B). We observed a few well-resolved cross-peaks and were able to confirm the assignments for two additional peaks. G20 at 11.61 ppm had a cross-peak with U19 at 13.28 ppm (red in Fig. 1.11B,C) and G22 at 12.98 ppm correlated to U32 at 13.88 ppm (green in Fig. 11B,C). However, this NOESY spectrum (Fig. 1.11B) was unable to assign the triple helix residues, due to the overlapped cross-peaks between 13 - 14 ppm.

2.5 Imino assignments of MALAT1 core *cis*

For further assignments of the residues in the triple helix, we recorded a ^{15}N-^{1}H imino HSQC spectrum for the MALAT1 core *cis* construct (Fig. 1.12A, red spectrum) and overlaid it with the HSQC spectrum for SL:poly(A) complex (Fig.

1.12A, black spectrum). For the MALAT1 core *trans* construct, the SL was uniformly ^{15}N-labeled (colored black in Fig. 1.12B, left) and the poly(A) tail was unlabeled (colored gray in Fig. 1.12B, left). The MALAT1 core *cis* construct was also uniformly ^{15}N-labeled (Fig. 1.12B, right). As seen in the overlaid spectra (Fig. 1.12A), there is one extra peak at 13.67 ppm for the *cis* construct. Since the poly(A) tail in the *trans* construct was unlabeled, this peak necessarily belongs to G71 (highlighted with a red circle in Fig. 1.12B, right), which is ^{15}N labeled in the *cis* construct but not in the *trans* construct. This peak was used as the starting point for the NOESY walk to assign the triple helix residues.

For the imino assignments of the triple helix residues, we mainly used spectra recorded in a buffer condition without $MgCl_2$, which showed more

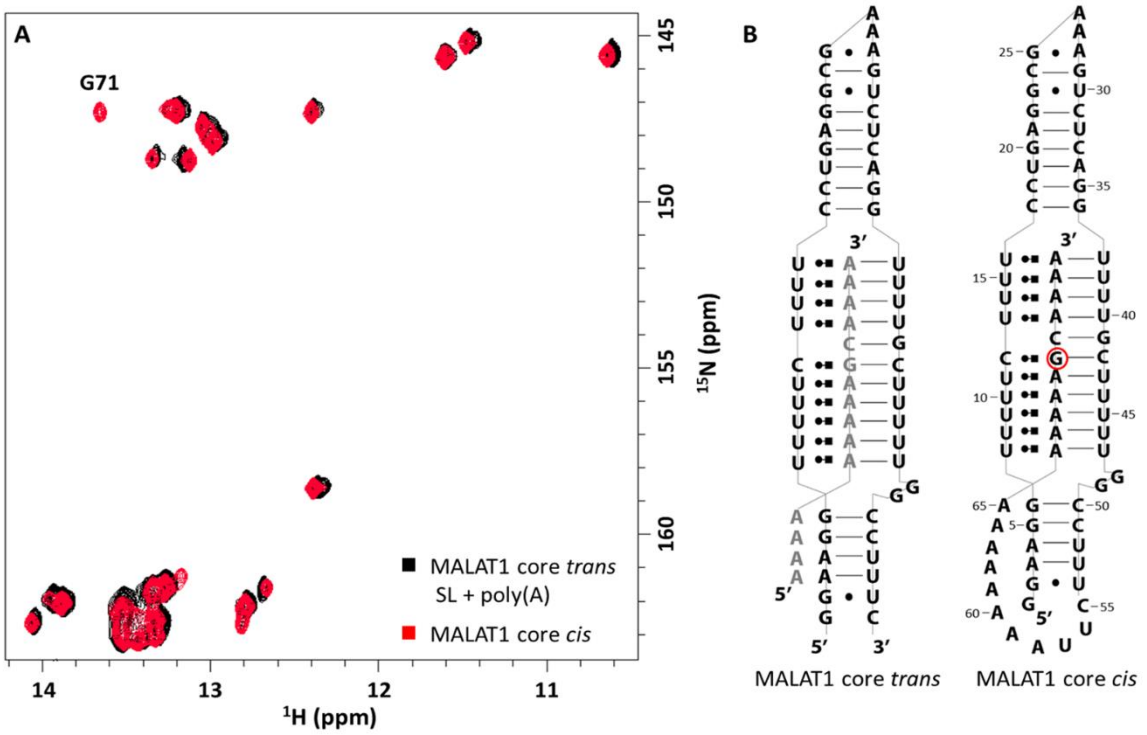


Figure 1. 12. ^{15}N-^{1}H imino HSQC for the MALAT1 core *cis*. (A) Overlay of imino HSQC spectra for MALAT1 core *trans* (black) and MALAT1 core *cis* (red). Both spectra were recorded at 15 °C. **(B)** Schematic diagram of MALAT1 core *trans* (SL:poly(A) complex) construct (left) and MALAT1 core *cis* (right).^{15}N labeled residues are colored black and unlabeled residues are colored gray.

dispersed peaks that simplified the assignment process. We have examined the effect of Mg^{2+} on the secondary structure of MALAT1 triple helix. The role of Mg^{2+} in RNA stability is well established and has been studied for various systems.[55, 56] Many studies have reported the effect of Mg^{2+} on the formation of stable tertiary structure for various RNAs such as riboswitches[57, 58], tRNAs[55], ribozymes[59] and ribosomal RNAs[60]. Although Mg^{2+} plays a crucial role in inducing conformational changes for some RNAs, other studies have shown that Mg^{2+} is not necessarily required for native folding - but simply contributes to structural stabilization, especially in the presence of monovalent ions.[55] In addition, Russu and co-workers have shown that Mg^{2+} has a minor effect on the conformation of DNA triple helices but contributes to stabilizing triple-helical base pairs.[61]

In order to check whether Mg^{2+} is required for the formation of the MALAT1 triple helix, we performed Mg^{2+} titrations and recorded the ^{15}N-^{1}H HSQC spectrum at 15 °C and 25 °C. A 0.4 mM of uniformly ^{15}N-labeled sample in 150 ul buffer (10 mM potassium phosphate, pH 6.5, 50 mM KCl, 0.1 mM EDTA, 10 % D_2O) was used and $MgCl_2$ was titrated to a final concentration of 0.2 – 10 mM. By gradually increasing the concentration of Mg^{2+}, we were able to follow the shifts of each peak. Although there were slight chemical shift perturbations for some peaks, most peaks were observable both in the presence and absence of Mg^{2+} at 15 °C (Fig. 1.13). Comparable results were observed at 25 °C (data not shown). This result indicates that all the base pairs in the triple helix are already formed in a condition with just monovalent ions and the presence of Mg^{2+} causes no dramatic change in the secondary structure. However, as previously reported, Mg^{2+} may have a

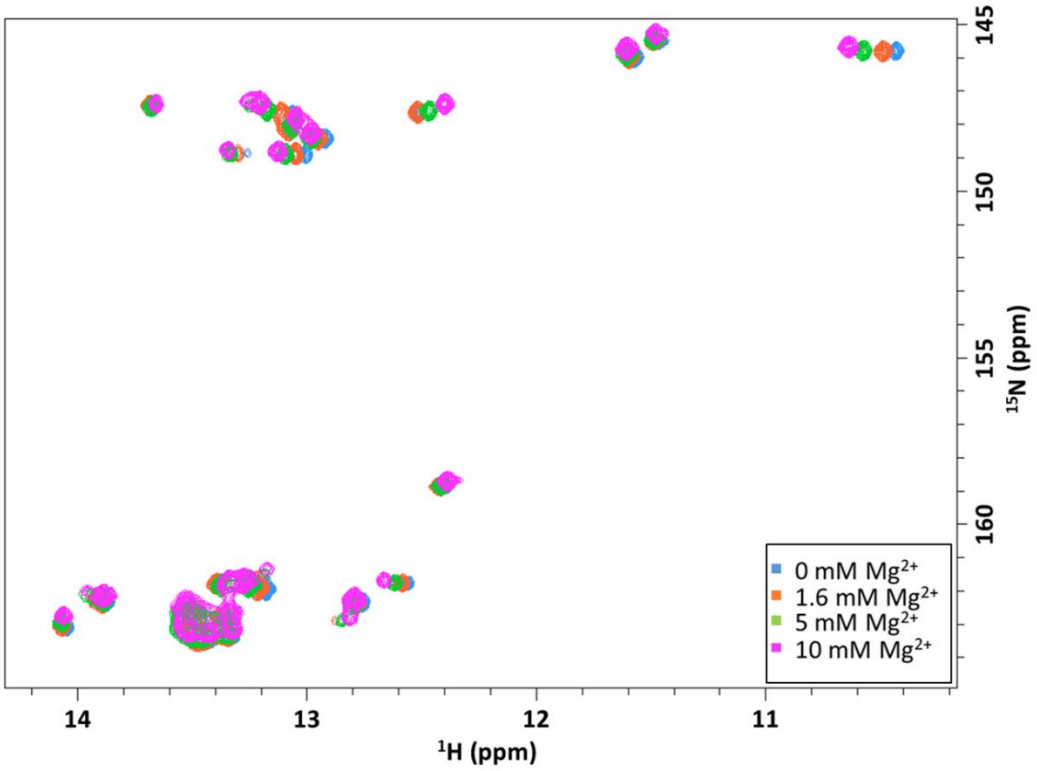


Figure 1. 13. Mg^{2+} titration for the MALAT1 triple helix at 15 °C. A 0.4 mM of RNA sample was used and $MgCl_2$ was titrated to a final concentration of 0.2 – 10 mM. There are slight chemical shift perturbations but most of the peaks are observable both in the absence and presence of Mg^{2+}, indicating a minor contribution of Mg^{2+} to the secondary structure formation.

contribution to stabilizing the MALAT1 triple helix.[55, 61] Based on the titration results, we concluded that Mg^{2+} has a minimum effect on the secondary structure of MALAT1 triple helix. Therefore, we mainly used the spectra collected in the absence of Mg^{2+} (which showed more dispersed peaks) for the assignment process.

For the imino assignment, we recorded ^{1}H-^{1}H NOESY, HNN COSY, $^{2}J_{HN}$ – ^{15}N-^{1}H HSQC, and 3D NOESY-HSQC spectra for the *cis* construct at 15 °C. An unlabeled sample was used for the NOESY and a ^{15}N – A/U labeled sample was used for the HNN COSY, $^{2}J_{HN}$ – ^{15}N-^{1}H HSQC, and the 3D NOESY-HSQC experiments. The ^{1}H dimensions for these spectra were referenced with 3-

(trimethylsilyl)propane-1-sulfonate (DSS) and the ^{15}N dimensions were indirectly referenced using the chemical shift ratios.[62] The NOESY spectrum was collected with a mixing time of 200 ms. For the 3D NOESY-HSQC, the ^{15}N dimension carrier was set to 161.75 ppm with a spectral width of 3.5 ppm to resolve the U peaks that are crowded between 13.1 – 13.6 ppm (Fig. 1.12A).

To aid the assignment process, we first narrowed down the U peaks that are involved in the Hoogsteen base pairs using HNN COSY and $^{2}J_{HN}$ – ^{15}N-^{1}H HSQC spectra. Hydrogen bonds have a partial covalent character that allows the J_{NN} scalar coupling to be observed via NMR.[63] Specifically, the HNN COSY experiment utilizes this coupling to obtain direct correlations of nitrogen atoms in the hydrogen bond donor and acceptor.[52, 63] For instance, in Watson – Crick A – U base pairs, the HNN COSY experiment correlates uridine H3 to N3, which then correlates to adenosine N1. In the case of U•A Hoogsteen base pairs, correlations between uridine N3 and adenosine N7 are observed instead (Fig. 1.14A, left). Since the chemical shifts of adenosine N1 and N7 resonate between 220 – 226 ppm and 224 – 232 ppm, respectively[64], we can distinguish Watson – Crick and Hoogsteen base pairs by monitoring the chemical shifts of adenosine cross-peaks. In the HNN COSY spectrum (Fig. 1.14B) of the MALAT1 triple helix, we observed five adenosine peaks that were well separated (225 – 229 ppm) from other adenosine peaks involved in Watson – Crick base pairs (black spectrum in Fig. 1.14B).

We also performed $^{2}J_{HN}$ – ^{15}N-^{1}H HSQC experiment to confirm that these five identified peaks indeed correspond to adenosine N7. In this experiment, the

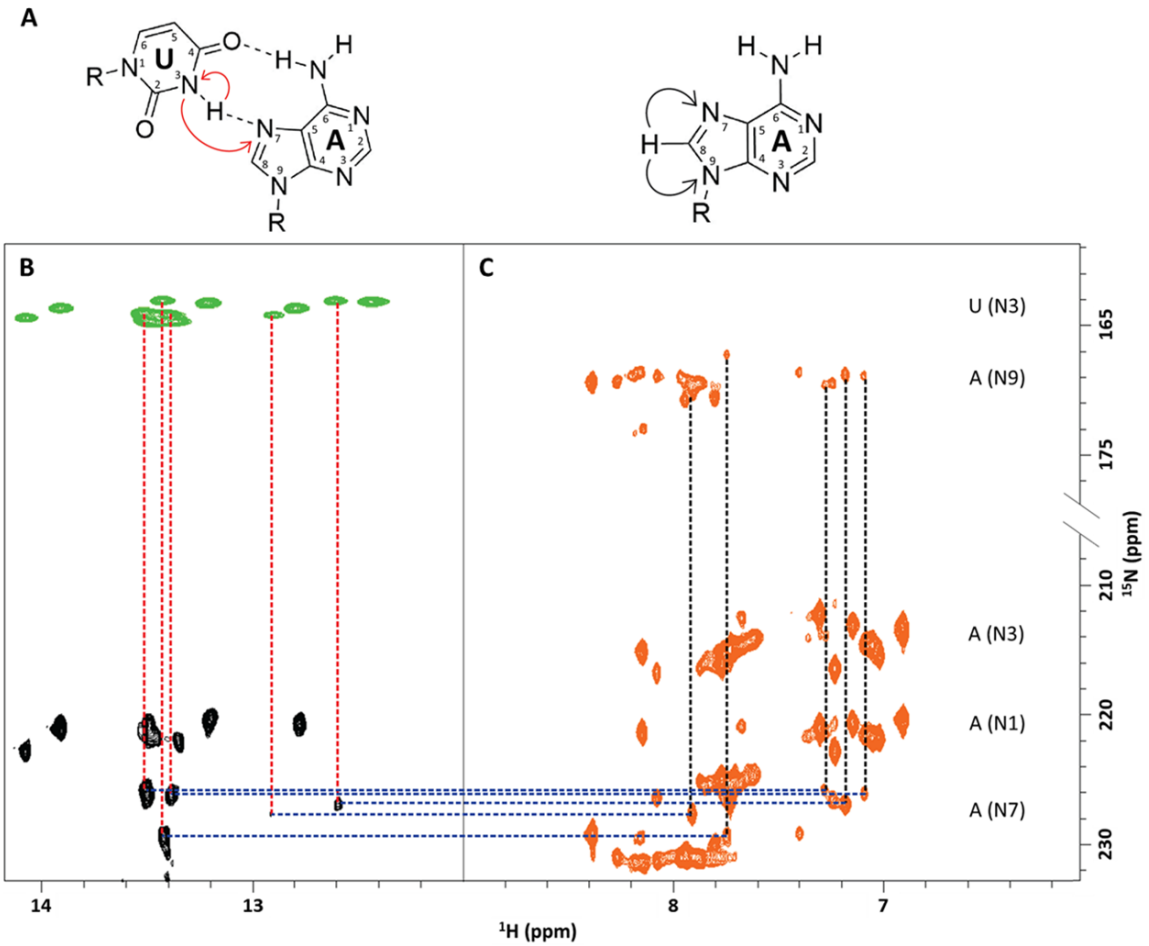


Figure 1. 14. Identification of U peaks involved in U•A Hoogsteen base pairs. (A) The geometry of U•A Hoogsteen base pair (Left). Schematic of the correlation between H8 and N7/N9 in adenosine (Right). Arrows indicate scalar couplings. **(B)** HNN COSY spectrum recorded at 15 °C. Green and black peaks correspond to N3 of uridine and N1/N7 of adenosine, respectively. Dotted red lines indicate cross-correlations between uridine N3 and adenosine N7. **(C)** $^{2}J_{HN}$ – ^{15}N-^{1}H HSQC spectrum recorded at 15 °C. Dotted black lines indicate H8 of adenosine correlating with N7/N9.

delay of magnetization transfer from the proton to nitrogen is optimized for the $^{2}J_{HN}$ coupling.[52] Thus, this experiment correlates adenosine H8 and H2 protons with adenosine N7/N9 and N1/N3, respectively. Since the chemical shifts of N1, N3, N7, and N9 resonances are well separated, H8 – N7/N9 and H2 – N1/N3 correlations can be easily distinguished. The five adenosine peaks identified from the HNN COSY experiment showed cross-peaks with N9 resonances in the $^{2}J_{HN}$ – ^{15}N-^{1}H HSQC spectrum (Fig. 1.14C). This result confirmed that these peaks

correspond to adenosine N7. Thus, uridine N3 resonances that correlate with these adenosine N7 resonances are necessarily involved in U•A Hoogsteen base pairs. The proton chemical shifts of five U peaks on the Hoogsteen side were 13.49, 13.38, 13.36, 12.87, and 12.56 ppm.

We then utilized ^{1}H-^{1}H NOESY, ^{15}N-^{1}H HSQC, and 3D NOESY-HSQC experiments of the *cis* construct to assign the triple helix residues. G71 identified from Fig. 1.12A was the starting point for a NOESY walk. As before, Fig. 1.15B shows the NOESY walk with the color for each walk corresponding to the colored arrows in Fig. 1.15A. At 13.67 ppm, G71 has two cross-peaks with G41 at 12.51 ppm (red in Fig. 1.15A,B) and U43 at 14.04 ppm (purple in Fig. 1.15A,B). G41 has two additional cross-peaks with U40 and U13 at 12.76 and 13.38 ppm, respectively (yellow in Fig. 1.15A,B). Both U40 and U13 have a cross-peak with U14 at 12.55 ppm (green in Fig. 1.15A,B), which correlates to U39 and U15 at 13.47 and 13.36 ppm, respectively (orange and blue in Fig.1.15A,B). We also observed a NOESY walk from U13 to U40 and from U40 to U39. Finally, U15 has a cross-peak with U16 at 12.87 ppm (magenta in Fig. 1.15A,B). For the lower part of the triple helix, U43 has two additional cross-peaks with U11 and U44 at 13.48 and 13.33 ppm, respectively (dark blue in Fig. 1.15A,B). Interestingly, the Us on the Watson-Crick side of the triple helix show cross-peaks to two Us on the Hoogsteen side: one of these Us is on the opposite side of the U•A-U base triple and the other is 3' of it. The distances between these residues in the crystal structure (PDB ID 4plx[41]) ranged between 6.3 – 6.8 Å, which are at the edge of observable NOE contacts. Fig. 1.15C shows an HSQC spectrum with the assign-

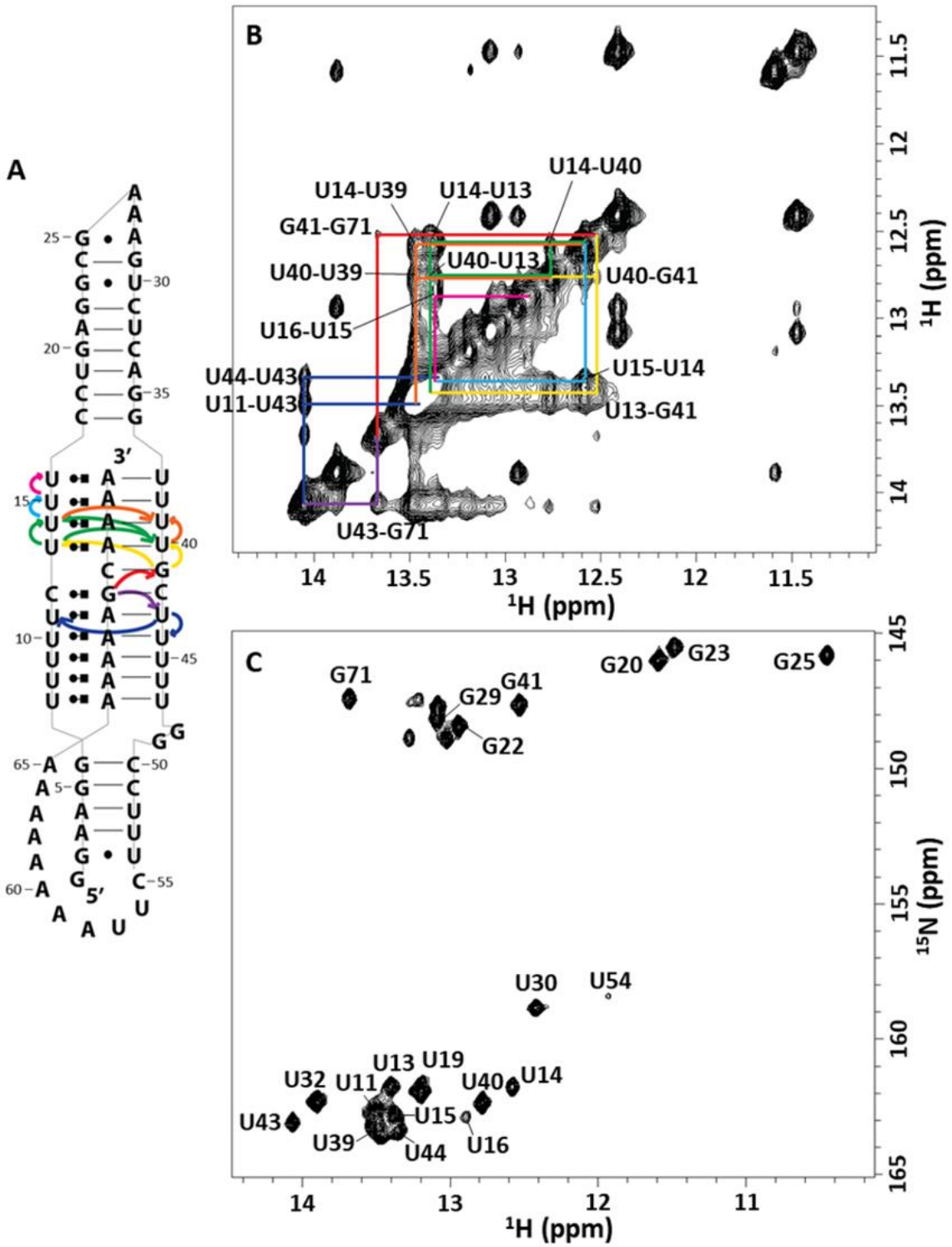


Figure 1. 15. Imino assignments of the triple helix residues. (A) Schematic diagram of MALAT1 cis construct. The color of arrows correlates the color of the NOESY walk in (B). **(B)** ^{1}H-^{1}H NOESY collected at 15 °C. **(C)** ^{15}N-^{1}H imino HSQC recorded at 15 °C.

ments of identified residues.

We observed the same NOE walk in the presence of Mg^{2+} and confirmed the assignments (Fig. 1.16, NOESY walk not shown). With Mg^{2+}, we observed an additional cross-peak between U16 and G36 at 13.13 ppm. We also observe few more U peaks with low signal intensities that may correspond to the unassigned U residues in the lower part of the triple helix. Unfortunately, the signal intensities were too weak for accurate assignment. The weak intensities of these peaks may

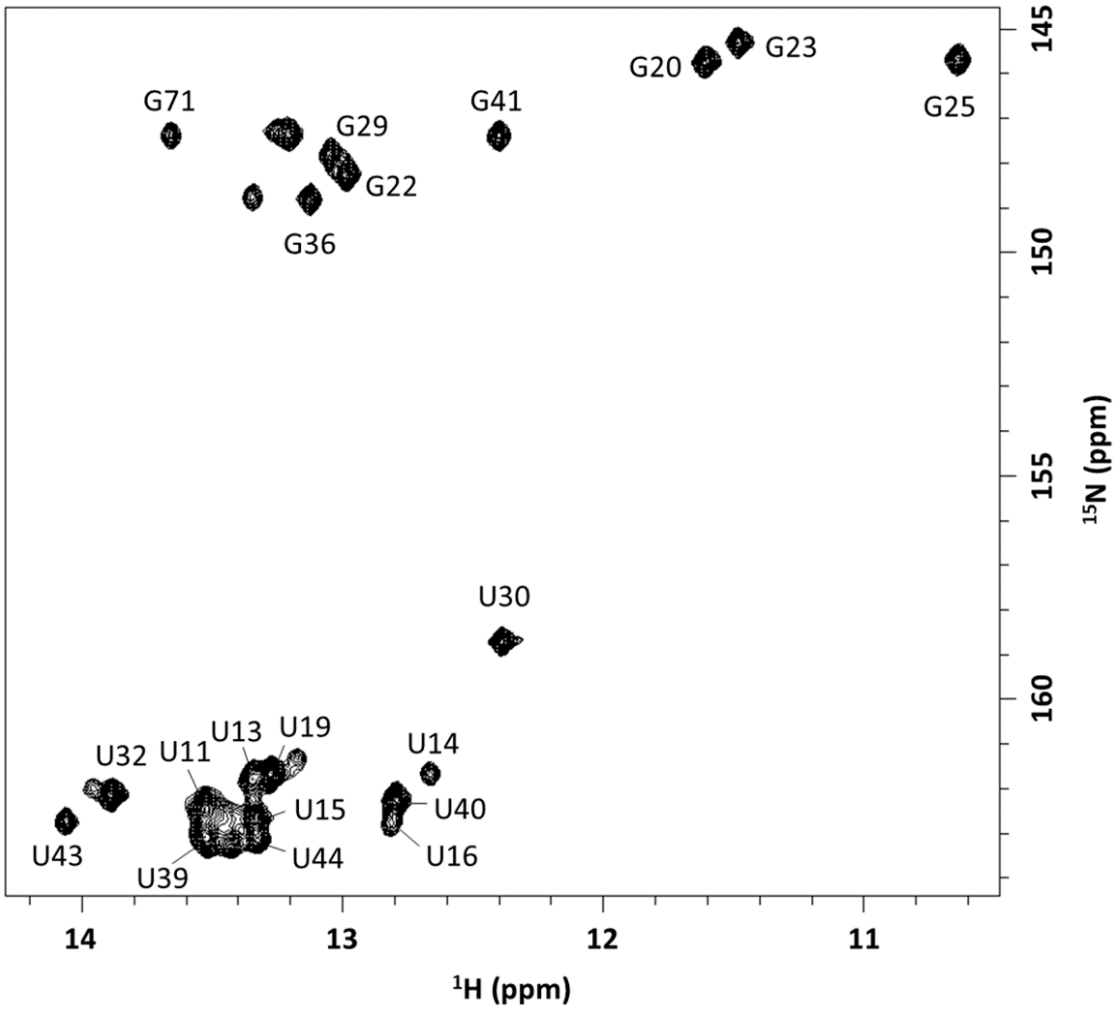


Figure 1. 16. Imino assignments of MALAT1 core *cis* in the presence of Mg^{2+}. We observed the same NOESY walk in the presence of Mg^{2+}. Additionally, we observed a cross-peak between U16 and G36 and were able to assign G36.

be due to the fast solvent exchange of imino protons. We observe three more G peaks as well, belonging to the unassigned G5, G6, and G35. The cross-peaks for theses residues were weak and overlapped requiring further evidence for unambiguous assignment.

2.6 Conclusion

For the imino assignments of the MALAT1 triple helix, we used ^{1}H-^{1}H NOESY, ^{15}N-^{1}H HSQC, and 3D NOESY-HSQC spectra recorded for different constructs. Stem I and II residues were assigned mainly using the MALAT1 core *trans* construct. For the triple helix residues, we were able to identify the U residues on the Hoogsteen face of U•A-U base triples by performing HNN COSY and $^{2}J_{HN}$ – ^{15}N-^{1}H HSQC experiments. The triple helix residues were assigned by a NOESY walk using the MALAT1 core *cis* construct. As a result, we were able to assign the majority of the upper triplex residues and a few lower triplex residues. We use these assignments for further studies on characterizing the MALAT1 triple helix.

3. Investigation of local stability of MALAT1 triple helix

Because MALAT1 triple helix prevents its mRNA transcripts from degradation by sequestering the poly (A) tail, the stability of the triple helix can also impact the accumulation level of MALAT1 transcripts. Multiple studies recently reported the global stability of the MALAT1 triple helix in varying concentrations of monovalent and divalent cations using FRET efficiency and UV melting profiles.[65, 66] However, less is known about the local melting and stability of the MALAT1 triple helix. In nucleic acids, local fluctuations involving base-pair opening and closing occur frequently, and studies have shown the importance of this local melting for various biological functions.[67-69] Therefore, our studies have focused on understanding the local stability of the MALAT1 triple helix via NMR spectroscopy.

With NMR spectroscopy, the local stability of nucleic acids can be examined at an individual base-pair level by monitoring the imino protons in guanosine and uridine, which are solvent exchangeable.[61, 70] Even base-paired imino protons undergo some degree of solvent exchange due to the frequent base-pair opening and closing events. When the base pairs open, the imino protons are exposed to the solvent and can undergo exchange. The less stable base pairs will open and close more frequently resulting in a higher exchange rate of their imino protons.[70, 71] Therefore, analyzing imino proton exchange rates provide information on the stability of individual base-pairs.

3.1 Inversion recovery NMR experiment

One common method to measure imino proton exchange rates is an inversion recovery experiment, which has been applied for various DNA and RNA systems.[61, 71-73] This experiment consists of a 180° soft pulse that selectively inverts the water proton resonance followed by variable mixing delays.[71-73] When the base-pairs open during the mixing delays and undergo solvent exchange, there is a change in the imino proton signals due to the selectively inverted water proton resonances. Therefore, monitoring the signal change provides imino proton exchange rates according to the following equation[71, 74]:

$$\frac{I_n(t)}{I_n(0)} - 1 = \left(\frac{I_w(0)}{\langle I_n(0)\rangle} - 1\right) k_{ex} \frac{[\exp(-R_{1n} \times t) - \exp(-R_{1w} \times t)]}{(-R_{1n} + R_{1w})} \qquad (1.1)$$

where $I_n(t)$ is the intensity of the imino protons at a delay point t, $I_n(0)$ is the intensity of the imino proton at an initial delay point $t = 0$, $I_w(0)$ is the intensity of the water signal normalized to the number of water protons within the sample and $\langle I_n(0)\rangle$ is the intensity of an individual imino proton species at delay point $t = 0$ normalized to the number of protons of that species. R_{1w} is the R_1 relaxation rate of the water signal, R_{1n} is the R_1 relaxation rate of the imino protons, and k_{ex} is the exchange rate of imino protons. Assuming that the selective 180 ° pulse completely inverts water magnetization to the –z orientation and leaves imino proton magnetization unperturbed along the z orientation, $I_w(0)/\langle I_n(0)\rangle$ can be set to -1.[71-73] R_{1w}, R_{1n}, and k_{ex} are allowed to adjust freely when fitting the data by eq 1.1. For the measurement of k_{ex}, the spectra were recorded with 16 – 21 different delay points and the data were fitted to eq 1.1 to extract k_{ex} (Fig. 1.17A). The

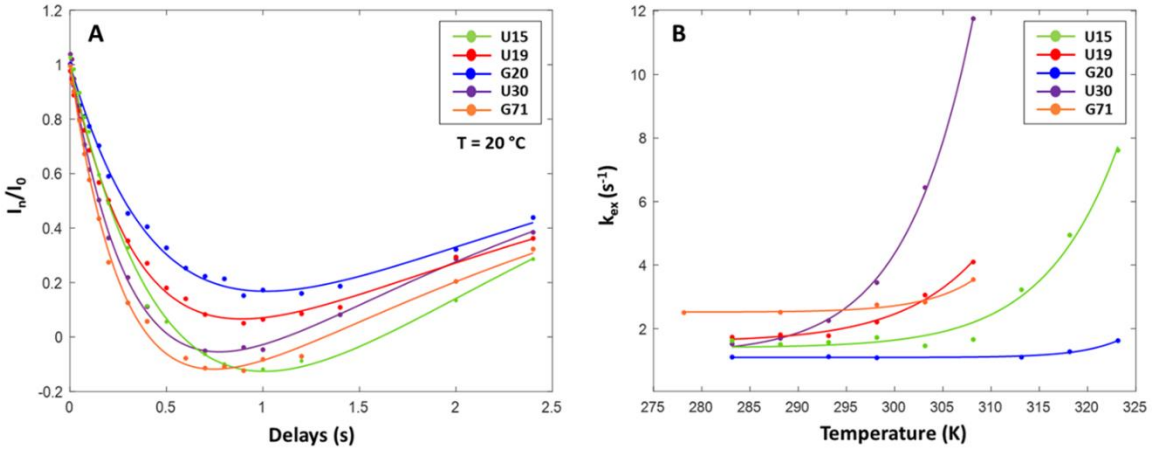


Figure 1. 17. Inversion recovery experiment on the MALAT1 triple helix. (A) Dependence of the imino proton signals on mixing delays (s) at 20 °C. **(B)** The inversion recovery experiment was performed at various temperatures between 5 – 50 °C to extract thermodynamic parameters.

experiments were performed at temperatures between 5 – 50 °C to extract thermodynamic parameters (Fig.1. 17B).

3.2 Analysis of imino proton exchange rate (k_{ex})

There are two steps involved in the imino proton exchange reaction: i) opening of the base-pair followed by ii) proton transfer to solvent.[70, 71] The latter step is catalyzed by small basic molecules, such as OH^-, NH_3, and $HPO4^{2-}$, all of which are in the buffer. For the analysis of the imino proton exchange rates, two different conditions can be considered. In the first condition (EX1), the proton transfer in the open state (k_{tr}) is so fast that every time the base-pair opens a proton transfer occurs.

$$\text{Closed} \overset{k_0}{\rightarrow} \text{Open} \overset{k_{tr}}{\rightarrow} \text{Water} \qquad (1.2)$$

In this case, the base-pair opening reaction is the rate liming step and k_{ex} is simply k_0. This type of kinetic process occurs in the presence of high catalyst (buffer) concentration.

The second condition (EX2) is when the proton transfer is sufficiently slow so that base-pairs open and close many times before the proton transfer occurs. In this case, there is an equilibrium between the open and the closed state.

$$\text{Closed} \overset{K_{diss}}{\rightleftarrows} \text{Open} \overset{k_{tr}}{\rightarrow} \text{Water} \tag{1.3}$$

For our experiments, we have adjusted the catalyst (HPO_4^{2-}) concentration based on earlier works[72, 74], so that the imino proton exchange is in the EX2 condition. The exchange rate in the EX2 condition is dependent on the k_{tr} and the fraction of the open state and can be expressed as follows[70, 71]:

$$k_{ex} = \frac{k_{tr}}{1 + (1/K_{diss})} \tag{1.4}$$

K_{diss} can also be described by:

$$K_{diss} = \exp\left(-\frac{\Delta G_{diss}}{RT}\right) = \exp\left(-\frac{\Delta H_{diss} - T\Delta S_{diss}}{RT}\right) \tag{1.5}$$

where ΔG_{diss} in eq 1.5 is substituted with enthalpy and entropy terms according to the Gibbs-Helmholtz equation. Substituting K_{diss} in eq 1.4 with eq 1.5 results in the following equation:

$$k_{ex} = \frac{k_{tr}}{(1 + \exp((\Delta H_{diss} - T\Delta S_{diss})/(RT)))} \tag{1.6}$$

Furthermore, the dipolar cross-relaxation of magnetization may lead to an overestimation of k_{ex} and thus, it needs to be considered in the analysis. To account for the NOE contributions in the measurement, a d term is added to the eq 1.6 as follows:[71, 72]

$$k_{ex} = \frac{k_{tr}}{(1 + \exp((\Delta H_{diss} - T\Delta S_{diss})/(RT)))} + d \qquad (1.7)$$

where k_{tr} is estimated from the control experiments which will be discussed more in detail in the following section. ΔH_{diss}, ΔS_{diss}, and d are freely adjustable parameters when fitting the data. Using eq 1.7, we can extract ΔH_{diss} and ΔS_{diss} parameters of individual base-pairs and determine their stability.

3.3 Measurement of proton transfer rate (k_{tr})

To extract the enthalpy and entropy parameters from eq 1.7, we have to estimate the proton transfer rate (k_{tr}) with additional measurements. The imino proton transfer can be catalyzed through both external and internal pathways.[72, 73, 75] The externally catalyzed pathway is mediated by small basic molecules in the buffer such as NH_3, OH^- and HPO_4^{2-}, as stated above. On the other hand, the internal pathway is catalyzed by nucleophilic functional groups of neighboring nucleobase moieties such as carbonyl O, N3 of cytidine, and N1 of adenosine.[73, 75] Therefore, k_{tr} can be expressed as[71, 72]:

$$k_{tr} = k_{tr,ext} + k_{tr,int} \qquad (1.8)$$

where $k_{tr,ext}$ and $k_{tr,int}$ correspond to the rates of the external and internal transfer pathways, respectively. According to the Eyring formula, $k_{tr,ext}$ and $k_{tr,int}$ can be also described as[71, 72]:

$$k_{tr,ext} = \frac{k_B T}{h} \exp\left(-\frac{\Delta H_{tr,ext} - T\Delta S_{tr,ext}}{RT}\right) \quad (1.9)$$

$$k_{tr,int} = \frac{k_B T}{h} \exp\left(-\frac{\Delta H_{tr,int} - T\Delta S_{tr,int}}{RT}\right) \quad (1.10)$$

where k_B is Boltzmann constant, h is Planck constant, R is universal gas constant, and T is the temperature (in Kelvin). $\Delta H_{tr,ext(int)}$ and $\Delta S_{tr,ext(int)}$ are enthalpy and entropy of the external(internal) proton transfer pathway, respectively. Equations 1.9 and 1.10 can be rearranged as follows:

$$ln\left(\frac{k_{tr,ext}}{T}\right) = -\frac{\Delta H_{tr,ext}}{R}\frac{1}{T} + \frac{\Delta S_{tr,ext}}{R} - ln\left(\frac{h}{k_B}\right) \quad (1.11)$$

$$ln\left(\frac{k_{tr,int}}{T}\right) = -\frac{\Delta H_{tr,int}}{R}\frac{1}{T} + \frac{\Delta S_{tr,int}}{R} - ln\left(\frac{h}{k_B}\right) \quad (1.12)$$

Substituting $k_{tr,ext}$ and $k_{tr,int}$ in eq 1.8 with eq 1.9 and 1.10 results:

$$k_{tr} = \frac{k_B T}{h}\left(\exp\left(-\frac{\Delta H_{tr,ext} - T\Delta S_{tr,ext}}{RT}\right) + \exp\left(-\frac{\Delta H_{tr,int} - T\Delta S_{tr,int}}{RT}\right)\right) \quad (1.13)$$

In order to estimate $k_{tr,ext}$, it is assumed that the accessibility of the catalysts to the imino protons in the base-pair open state is the same as that of mononucleotides.[71-73] Therefore, the imino proton transfer rates of free guanosine triphosphate (GTP) and uridine triphosphate (UTP) ($k_{tr,GTP(UTP)}$) are measured in a temperature-dependent manner under the same NMR buffer conditions.

$k_{tr,GTP(UTP)}$ was calculated from the line-widths of the imino signals in ^{1}H 1D spectra using the following equation[71, 72]:

$$k_{tr,GTP(UTP)} = \Delta\nu \cdot \pi - R_2 - R_{2(B0)} \quad (1.14)$$

where $\Delta\nu$ is the line-widths of the imino signals, R_2 is the rate constant of spin-spin relaxation of the imino resonance, and $R_{2(B0)}$ is the line-broadening caused by B_0 field inhomogeneity. $\Delta\nu$ is obtained by deconvolution of the imino signals using the software in Bruker Topspin 3.2. The contribution of R_2 and $R_{2(B0)}$ are assumed to be negligible compared to the contribution of $\Delta\nu$ to the $k_{tr,GTP(UTP)}$.[71, 72]

Furthermore, since the proton transfer by external catalysis is controlled by diffusion, $k_{tr,GTP(UTP)}$ has to be corrected to account for the difference in diffusion coefficient between the mononucleotides and the nucleobases within the RNA according to[71, 72]:

$$k_{tr,ext} = k_{tr,GTP(UTP)} \frac{1 + \sqrt[3]{M_{HPO_4{}^{2-}}/M_{RNA}}}{1 + \sqrt[3]{M_{HPO_4{}^{2-}}/M_{GTP(UTP)}}} \quad (1.15)$$

where $k_{tr,ext}$ is the external proton transfer rate of guanosine (uridine) within the RNA. $M_{HPO_4{}^{2-}}$, $M_{GTP(UTP)}$, and M_{RNA} are the molecular weight of the catalyst (HPO_4^{2-}), free GTP(UTP), and RNA molecule, respectively.

To estimate the contribution of the external catalysis ($k_{tr,ext}$), we measured the $k_{tr,GTP(UTP)}$ in the same NMR buffer condition. Fig. 1.18A shows the change in line-width of the imino proton signal of GTP in the ^{1}H 1D spectrum with increasing

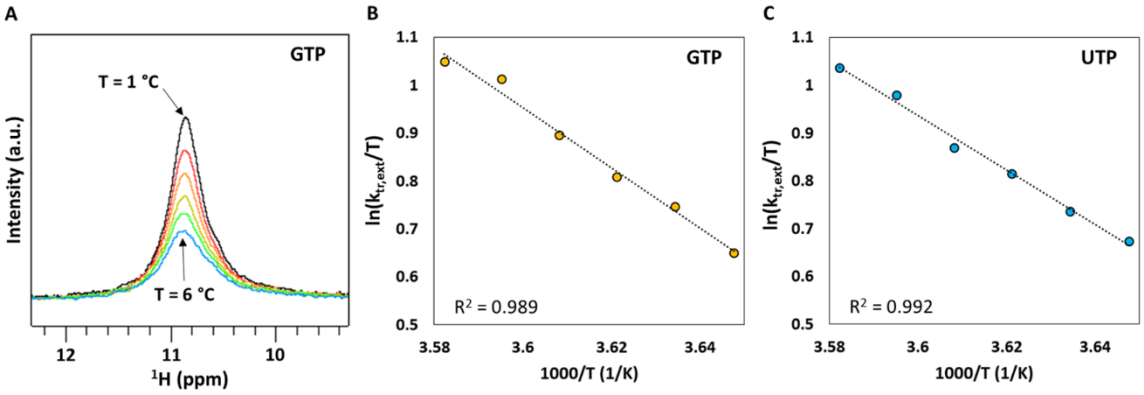


Figure 1. 18. Measurement of $k_{tr,ext}$. (A) The change in line-width of the imino proton signal of GTP in the ^{1}H 1D spectrum with increasing temperature. Eyring plots describing the temperature dependence of the external proton transfer rate for **(B)** GTP and **(C)** UTP.

	$\Delta H_{tr,ext}$ (kJ/mol)	$\Delta\Delta H_{tr,ext}$ (kJ/mol)	$\Delta S_{tr,ext}$ (J/mol·K)	$\Delta\Delta S_{tr,ext}$ (J/mol·K)
GTP	52.3	2.8	-1.43	0.02
UTP	47.3	2.1	-19.3	0.3

Table 1. 2. Thermodynamic parameters of imino proton transfer by external catalysis. $\Delta\Delta H_{tr,ext}$ and $\Delta\Delta S_{tr,ext}$ represent fit errors.

temperature. The $k_{tr,ext}$ was calculated according to eq 1.14 and 1.15 at each temperature and $ln(k_{tr,ext}/T)$ was plotted against $1/T$ (Fig. 1.18B,C). According to eq 1.11, $\Delta H_{tr,ext}$ and $\Delta S_{tr,ext}$ for GTP and UTP can be calculated from the slope and intercept of the plot, respectively (Fig. 1.18 B,C). The calculated values are summarized in Table 1.2. For GTP, the experimentally determined $\Delta H_{tr,ext}$ and $\Delta S_{tr,ext}$ were 52.3 ± 2.8 kJ/mol and -1.43 ± 0.02 J/mol·K, respectively. For UTP, the determined values were $\Delta H_{tr,ext}$ = 47.3 ± 2.1 kJ/mol and $\Delta S_{tr,ext}$ = -19.3 ± 0.3 J/mol·K. These values are comparable to the reported values in an earlier study.[71]

The measurements of the internal catalysis ($k_{tr,int}$) contribution were performed as described in earlier works.[71-73] The temperature dependence of k_{ex}

at different catalyst (HPO_4^{2-}) concentrations were measured. Based on previous reports, the buffer concentrations were adjusted for the EX2 regime, in which the k_{ex} is proportional to the concentration of catalyst.[70, 72, 74] From these measurements, the dependence of k_{ex} on catalyst concentration can be calculated. While $k_{tr,ext}$ increases with increasing catalyst concentration, $k_{tr,int}$ is independent of the buffer concentration. Therefore, the change in k_{ex} with different catalyst concentrations is due to the contribution of $k_{tr,ext}$. This allows us to dissect the contribution of external catalysis from the contribution of internal catalysis at different temperatures.

For the estimation of $k_{tr,int}$, the buffer conditions with low catalyst concentration (10 mM potassium phosphate) and high catalyst concentration (50 mM potassium phosphate) were used. The concentration of KCl was adjusted in such a way that the concentration of monovalent cation (K^+) is constant in both buffer conditions. The concentration of $MgCl_2$ was also kept constant in both conditions to minimize the structural change due to the monovalent and divalent cations. Using calculated $k_{tr,int}$ values, $ln(k_{tr,int}/T)$ was plotted against $1/T$ (Fig. A.1) and $\Delta H_{tr,int}$ and $\Delta S_{tr,int}$ values were extracted for each residue according to eq 1.12 (Table 1.3). Using the calculated values in Table 1.2 and 1.3, we are now able to estimate the k_{tr} at different temperatures using eq 1.13.

Residue	$\Delta H_{tr,int}$ (kJ/mol)	$\Delta\Delta H_{tr,int}$ (kJ/mol)	$\Delta S_{tr,int}$ (J/mol·K)	$\Delta\Delta S_{tr,int}$ (J/mol·K)	$k_{tr,int}$ (Hz) (293 K)	$\Delta k_{tr,int}$ (Hz) (293K)
U11	10	3	-55	14	11986	5041
U13	13	6	-15	6	9457	5636
U14	9	1	-61	5	9014	1124
U15	7	3	-94	31	8484	4762
U16	12	3	-32	6	7745	2301
U19	17	2	20	2	4944	676
G20	16	6	20	6	13224	6381
G22	17	6	31	9	12191	5264
G23	15	3	0.9	0.2	10118	2960
G29	25	7	127	31	9564	3533
U30	12	1	-30	2	6603	639
U32	10	6	-46	20	14027	10017
U39	7	3	-95	35	9126	5851
U40	15	1	3.1	0.2	9830	1196
G41	27	16	151	80	18275	14452
U43	10	1	-45	4	17251	2810
U44	10	2	-44	1	15140	5095
G71	23	3	108	10	24843	3589

Table 1. 3. Thermodynamic parameters of imino proton transfer by internal catalysis. $\Delta\Delta H_{tr,ext}$ and $\Delta\Delta S_{tr,ext}$ represent fit errors.

3.4 Thermodynamic stability of the individual base-pairs

The change in the imino proton signal was monitored with increasing delay time to obtain the imino proton exchange rate (k_{ex}) of individual base-pairs (Fig. 1.17A). For each residue, the temperature dependence of the k_{ex} was measured (Fig 1.17B) and analyzed using eq 1.7 and k_{tr} was estimated as described in section 3.3. Table 1.4 contains ΔH_{diss}, ΔS_{diss}, ΔG_{diss}, and $T\Delta S_{diss}$ values determined for the base-pair opening of individual residues. ΔH_{diss} and ΔS_{diss}values ranged between 58 – 210 kJ/mol and 16 – 438 J/mol·K, respectively. These values are in the same range as what has been reported for other RNAs.[71-

Residue	ΔH_{diss} (kJ/mol)	$\Delta\Delta H_{diss}$ (kJ/mol)	ΔS_{diss} (J/mol·K)	$\Delta\Delta S_{diss}$ (J/mol·K)	ΔG_{diss} (kJ/mol) (293 K)	$\Delta\Delta G_{diss}$ (kJ/mol) (293 K)	$T\Delta S_{diss}$ (kJ/mol) (293 K)	$T\Delta\Delta S_{diss}$ (kJ/mol) (293 K)
U11	77	8	89	25	50.4	0.6	26	7
U13	58	2	17	7	52.7	0.4	5	2
U14	170	5	393	14	55.0	0.6	115	4
U15	107	1	214	6	44.3	0.2	63	2
U16	90	13	127	40	53.2	0.9	37	12
U19	80	2	57	6	63.4	0.2	17	2
G20	169	3	301	9	80.5	0.3	88	3
G22	213	17	438	54	84.4	1.7	128	16
G23	85	4	90	12	58.6	0.4	26	3
G29	138	12	158	40	91.2	0.9	46	12
U30	98	3	161	9	50.5	0.1	47	3
U32	99	1	168	3	50.0	0.1	49	1
U39	87	5	156	13	41.6	0.9	46	4
U40	59	2	16	5	54.5	0.1	5	2
G41	132	9	146	34	89.4	1.0	43	10
U43	68	6	61	18	50.2	0.3	18	5
U44	61	7	37	19	49.9	0.3	11	5
G71	134	15	160	50	87.3	0.2	47	15

Table 1. 4. Thermodynamic parameters for the base-pair opening of individual residues within the MALAT1 triple helix. Multiple measurements were performed to estimate the experimental errors.

[73] The ΔG_{diss} values were between 41.6 – 91.2 kJ/mol, which are on the same scale, but greater than the values determined for other RNAs.[71-73] The difference in the range of ΔG_{diss} values may be due to the difference in the RNA systems. The earlier studies were performed on simple RNA hairpins with a size of less than 40 nt.[71-73] Compared to these RNAs, MALAT1 triple helix has a more compact structure with a greater number of base pairs, which may lead to increased stability of individual base pairs.

Fig. 1.19 is a summary of determined thermodynamic parameters for the MALAT1 triple helix. The relative stabilities of individual residues are represented

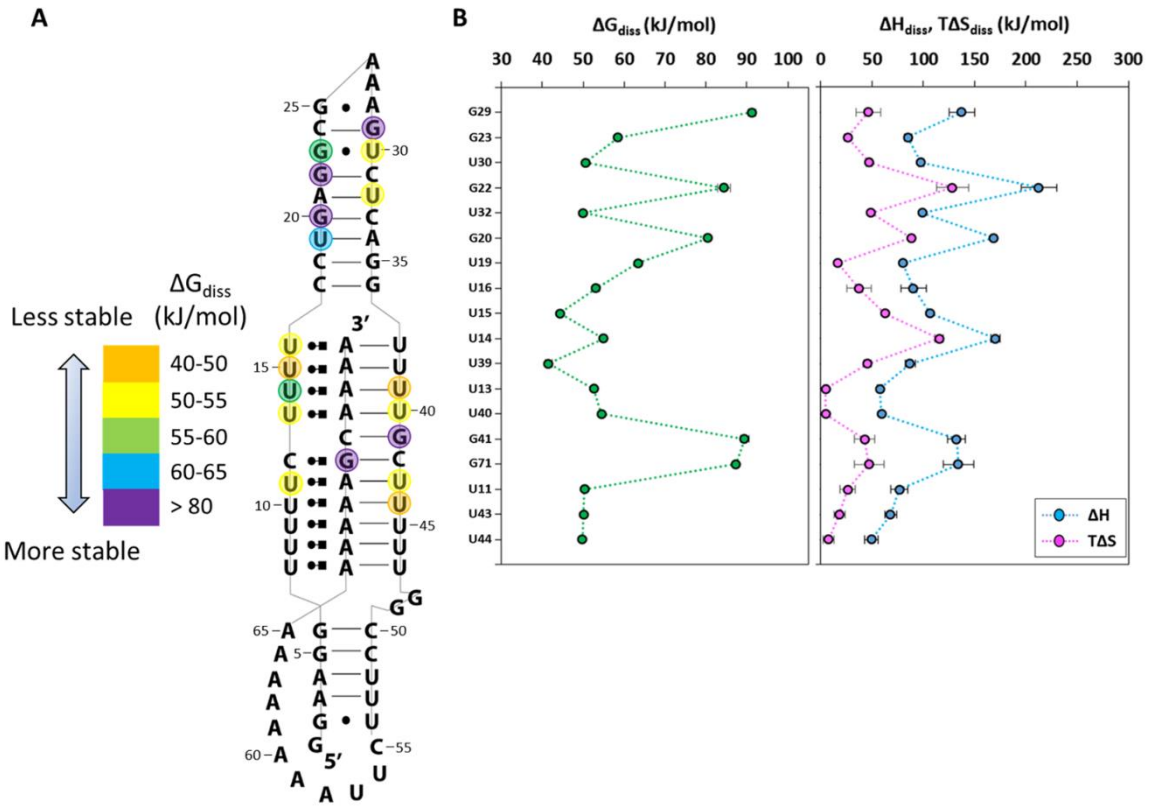


Figure 1. 19. Individual base-pair stability of MALAT1 triple helix (A) ΔG_{diss} for individual base-pair. Relative stability is represented in different colors. **(B)** Diagram of thermodynamic parameters determined for MALAT1 triple helix residues.

with different colors (Fig. 1.19A). Greater ΔG_{diss} values indicate a less favorable base-pair opening reaction and thus, a more stable base-pair. The diagram of thermodynamic parameters shows a strong correlation between ΔH_{diss} and $T\Delta S_{diss}$ (Fig. 1.19B, right). This has also been observed for other nucleic acids and leads to a smaller spread of ΔG_{diss} values.[71-73, 76] The correlation between ΔH_{diss} and $T\Delta S_{diss}$ is discussed further in detail in section 3.5.

A comparison of ΔG_{diss} values of the stem II residues shows that G-C base pairs are more stable than A-U or G•U base pairs. Interestingly, within the same G•U base pair, the imino proton of G23 had slightly greater stability compared to

the imino proton of U30. For the U residues in the triple helix, ΔG_{diss} values ranged between 45 – 55 kJ/mol and had similar base-pair stabilities. Melting curve analyses of synthetic RNA and DNA triple helices in earlier studies have shown that the strand on the Hoogsteen face is less stable than that of the Watson-Crick face.[77, 78] However, our analysis shows similar stabilities between the U residues on the Hoogsteen and the Watson-Crick face. Furthermore, ΔG_{diss} values of G41 and G71 were significantly greater than that of other residues within the triple helix. This result suggests that C72 – G41 and G71 – C42 base pairs have a greater contribution to stabilizing the triple helix compared to other U•A-U base triples.

The contribution of C72 – G41 and G71 – C42 base pairs for the stabilization of the triple helix is also supported by previously reported mutational studies.[40, 41] Mutation of these residues dramatically reduced the accumulation level of the reporter gene.[40] While the C12U mutation only reduced ~ 40 % of accumulation level, mutations that perturb C72 – G41 and G71 – C42 base-pairs completely abolished the accumulation.[40] In addition, C72 – G41 and G71 – C42 base-pairs allow the precise alignment of poly(A) tail in the triple helix. Steitz and co-workers have shown the importance of the blunt end at the 3' end of the sequence in the overall stability of the triple helix.[40] The presence of C72 – G41 and G71 – C42 base-pairs in the middle prevents the poly(A) tail sliding within the triple helix. Taken together, these results suggest that these base pairs are important for the triple helix due to their specific sequences and greater stability.

3.5 Enthalpy – entropy correlations and global unfolding of the triple helix

As seen in Fig. 1.20A, enthalpy and entropy values are linearly correlated. This has been frequently observed for other nucleic acids and understood as a solvent-solute interaction.[71, 72, 76] When base-pairs open, the water molecules surrounding the base-pairs are released. Therefore, the loss in enthalpy is compensated with the increase in entropy. The correlations between ΔH_{diss} and ΔS_{diss} can be described with the following equation[71, 72]:

$$\Delta H_{diss} = m\Delta S_{diss} + y_0 \tag{1.16}$$

where m and y_0 are the slope and y-intercept of the linear plot, respectively. Substitution of ΔH_{diss} in eq 1.16 with the Gibbs-Helmholtz equation results the following equation[71, 72]:

$$\Delta G_{diss} = (m - T)\Delta S_{diss} + y_0 \tag{1.17}$$

Eq 1.17 shows that at the temperature of $m = T$, ΔG_{diss} of all base pairs is equal to y_0. This temperature is also called compensation temperature (T_C), and at T_C, all residues have relatively similar stabilities. Interestingly, studies on RNA hairpins

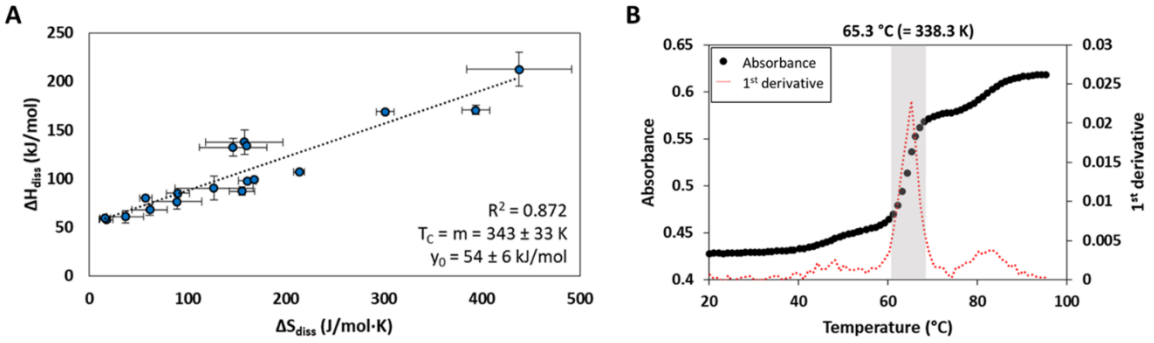


Figure 1. 20. Enthalpy – entropy correlation (A) Linear correlation between ΔH_{diss} and ΔS_{diss}. T_C determined from the plot is 343 ± 33 K and y_o is 54 ± 6 kJ/mol. **(B)** UV melting curve of MALAT1 triple helix. The first transition is observed at 65.3 °C (=338.3 K).

have reported that T_C is close to the melting temperature (T_m).[71-73] The interpretation of these observations was that global unfolding occurs at the temperature where the individual base-pairs have similar stabilities.[71-73]

We observed similar results for the MALAT1 triple helix. The observed T_C in Fig 1.20A was 343 ± 33 K (≈ 70 °C), which is close to the temperature for the major transition in the UV melting curve, 65.3 °C. At 70 °C, the average ΔG_{diss} is 53.8 kJ/mol. Interestingly, ΔG_{diss} values for triple helix residues were below the average and ΔG_{diss} values for the stem II residues were above the average. Furthermore, plotting ΔH_{diss} against ΔS_{diss} values for the stem II residues resulted in T_C = 362 ± 48 K (≈ 89 °C) (Fig. 1.21A). This temperature is close to the second transition temperature in the melting curve, 85.3 °C (Fig. 1.21B). These observations suggest that the first transition at 65.3 °C is the unfolding of the triple helix region and the second transition at 85.3 °C is the unfolding of the stem II. Moreover, the enthalpy – entropy correlation of the residues in the triple helix region showed T_C = 311± 49 K (≈ 38 °C) (Fig. 1.22). This T_C is close to the recently

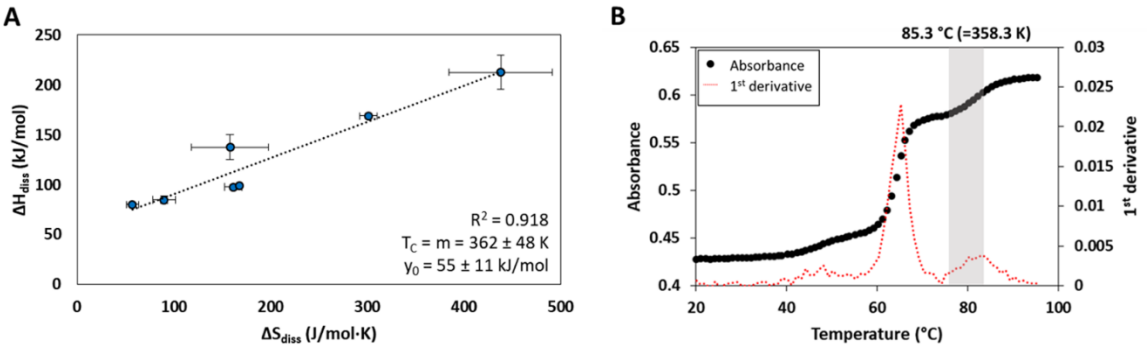


Figure 1. 21 Enthalpy – entropy correlation for the residues in the stem II (A) Linear correlation between ΔH_{diss} and ΔS_{diss} values for the residues in the stem II. T_C determined from the plot is 362 ± 48 K and y_o is 55 ± 11 kJ/mol. **(B)** UV melting curve of MALAT1 triple helix. The second transition is observed at 85.3 °C (= 358.3 K).

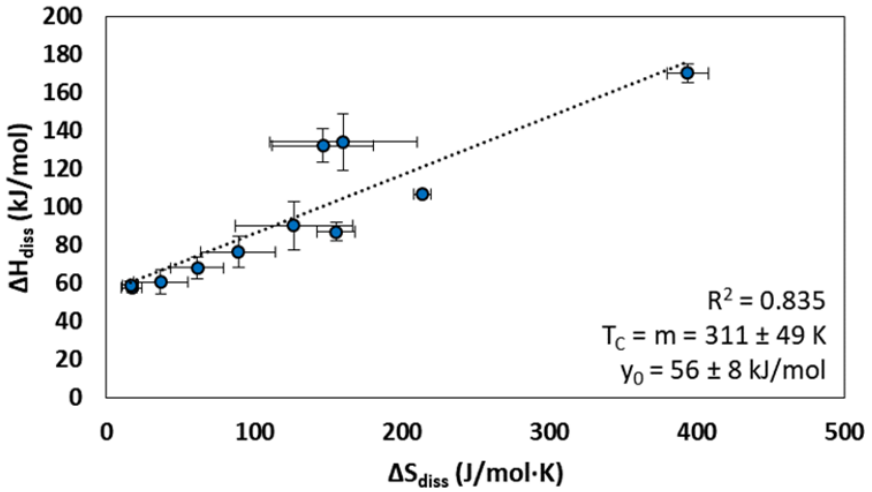


Figure 1. 22 Enthalpy – entropy correlation for the residues in the triple helix region. T_C determined from the plot is 311 ± 49 K and y_o is 56 ± 8 kJ/mol.

reported melting temperature (45 °C) of the MALAT1 construct that only includes the triple helix sequence.[66] This observation suggests that the presence of stem II further stabilizes the triple helix region.

3.6 Effect of Mg^{2+} ions on the MALAT1 triple helix stability

We further examined the effect of Mg^{2+} ions on the stability of the MALAT1 triple helix. Many studies have shown that Mg^{2+} stabilizes the structure of various RNAs and DNAs.[55, 61] To analyze the effect of Mg^{2+}, we carried out inversion recovery experiments in the presence and absence of Mg^{2+} and compared the imino proton exchange rates at 20 °C. Assuming that Mg^{2+} has a negligible effect on the proton transfer rate (k_{tr}), the change in ΔG_{diss} in the presence of Mg^{2+} ($\delta\Delta G^{Mg}$) can be described as follows[61, 79]:

$$\delta\Delta G^{Mg} = \Delta G_{diss}^{Mg} - \Delta G_{diss}^{NoMg} = -RTln\frac{k_{ex}^{Mg}}{k_{ex}^{NoMg}} \quad (1.18)$$

where ΔG_{diss}^{Mg} and ΔG_{diss}^{NoMg} are Gibb's free energy of base-pair opening in the presence and absence of Mg^{2+}, respectively. k_{ex}^{Mg} and k_{ex}^{NoMg} are the imino proton exchange rate in the presence and absence of Mg^{2+}, respectively.

Table 1.5 contains the calculated $\delta\Delta G^{Mg}$ values for individual base-pairs according to eq 1.18. The changes in free energy ranged between -0.6 – 2.4 kJ/mol. Most of the base-pairs had positive $\delta\Delta G^{Mg}$ values indicating greater ΔG_{diss} values in the presence of Mg^{2+} ions. Interestingly, there were few residues with negative $\delta\Delta G^{Mg}$ values, which suggest less stable base-pairs in the presence of Mg^{2+}. Similar results were observed in the previous study on a DNA triple helix.[61] This result suggests that not all base-pairs are stabilized by Mg^{2+}. There may be a few base-pairs that are less stable with Mg^{2+}; however, a greater number of base- pairs

Residue	$\delta\Delta G^{Mg}$ (kJ/mol)	Residue	$\delta\Delta G^{Mg}$ (kJ/mol)
U11	-0.59 ± 0.09	U30	0.11 ± 0.01
U13	-0.40 ± 0.05	U32	0.5 ± 0.1
U14	0.21 ± 0.02	U39	-0.031 ± 0.002
U15	0.24 ± 0.03	U40	0.31 ± 0.06
U16	0.6 ± 0.1	G41	0.5 ± 0.1
U19	2.4 ± 0.3	U43	0.38 ± 0.08
G20	0.13 ± 0.01	U44	0.21 ± 0.02
G22	-0.12 ± 0.01	G71	0.48 ± 0.05
G23	-0.25 ± 0.02		

Table 1. 5. Effect of Mg^{2+} ions on the free energy of base-pair opening reaction. Error are estimated by propagating the errors in imino proton exchange rates (k_{ex}).

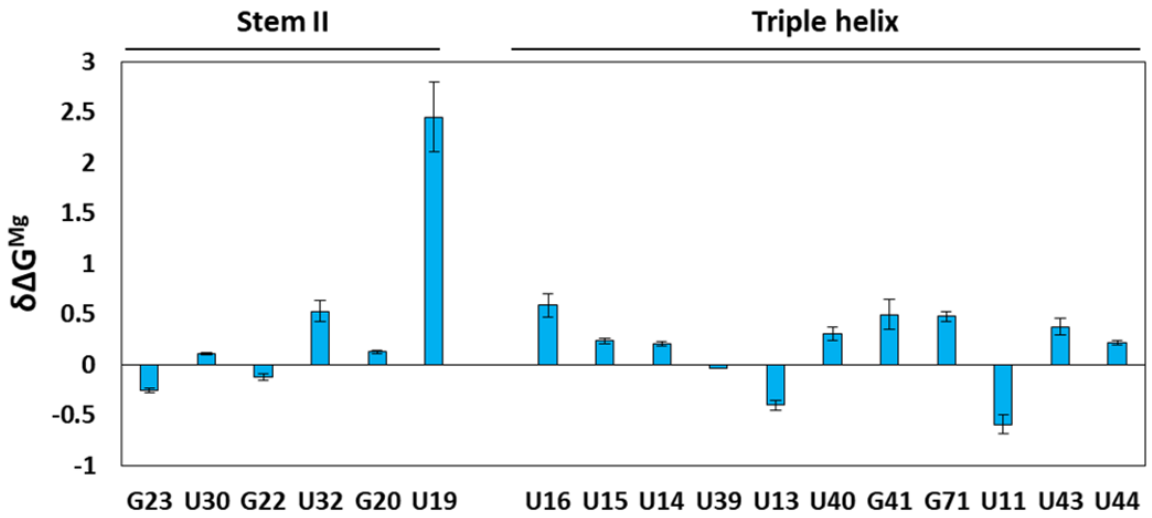


Figure 1. 23. Diagram of $\delta\Delta G^{Mg}$ for individual base-pairs. U19 has the highest $\delta\Delta G^{Mg}$ value among all the residues. Within the triple helix, U16, G41, and G71 have relatively high $\delta\Delta G^{Mg}$ values.

are stabilized by Mg^{2+}, which leads to increased overall stability. Overall, the ΔG_{diss} values increased by ~ 4.7 kJ/mol.

Fig. 1.23 shows a diagram of calculated $\delta\Delta G^{Mg}$ for individual base-pairs. As seen, U19 in the stem II has the highest $\delta\Delta G^{Mg}$, indicating the greatest stabilization effect by Mg^{2+} compared to the other base-pairs. Within the triple helix, U16, G41, and G71 have relatively high $\delta\Delta G^{Mg}$ values, suggesting that these residues are most stabilized by Mg^{2+}. Interestingly, two residues on the Hoogsteen side base-pairs, one below (U11) and above (U13) C-G and C•G-C base pairs, had negative $\delta\Delta G^{Mg}$. Overall, $\delta\Delta G^{Mg}$ values of the Watson-Crick base-pairs were greater than that of the Hoogsteen base-pairs. A similar result was observed for the DNA triple helix, which showed greater stabilization of the Watson-Crick base pairs compared to the Hoogsteen base pairs in the presence of Mg^{2+}.[61]

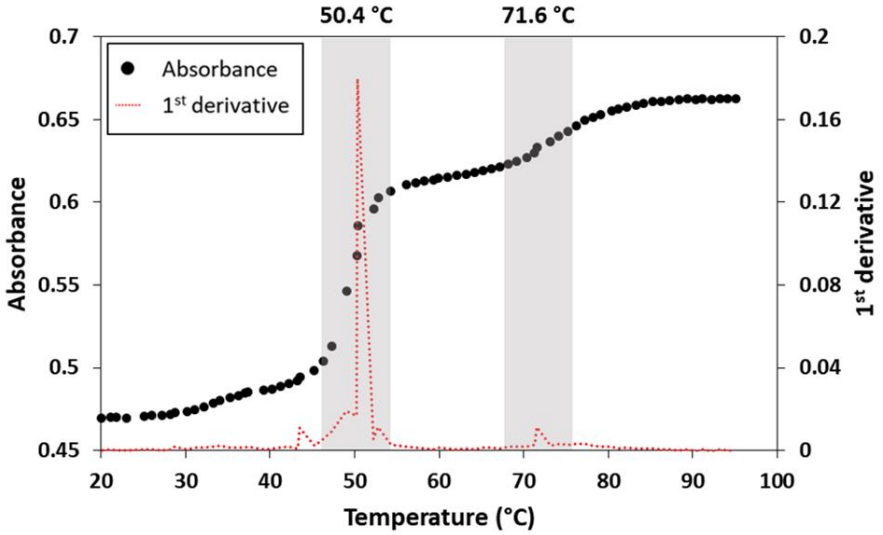


Figure 1. 24 UV melting profile in the absence of Mg^{2+} ions. The first transition temperature is at 50.4 °C and the second transition temperature is at 71.6 °C.

The UV melting profile also suggests a decrease in global stability in the absence of Mg^{2+}. Without Mg^{2+}, the first melting transition was observed at 50.4 °C and the second transition was observed at 71.6 °C. These temperatures are ~ 15 °C lower than the melting temperatures observed in the presence of Mg^{2+}, which indicates the global stabilization effect of Mg^{2+}. As indicated in section 2.5, Mg^{2+} has no dramatic effect on the formation of the secondary structure of the MALAT1 triple helix. This is also observed in the melting profile showing two transitions, which is indicative of triple helix formation. Although Mg^{2+} has a minor effect on the secondary structure formation, it contributes to the overall stabilization of the MALAT1 triple helix.

3.7 Discussion

Considering that the MALAT1 triple helix is a potential therapeutic target for cancer, it is important to understand its stability in greater detail. In addition to the previous reports on the global stability of the MALAT1 triple helix,[65, 66] we examined its local stability via NMR spectroscopy. Individual base-pair stabilities were analyzed by measuring the imino proton exchange rate (k_{ex}) with an inversion recovery experiment. The advantage of this approach is that it provides residue-specific information and better insights on which region of the molecule is essential for maintaining the overall stability. As a therapeutic target, this information is important as it will provide guidance on which part of the RNA needs to be targeted.

Our data indicate that the C72 – G41 and G71 – C42 base pairs are the most stable base pairs within the triple helix (Fig. 1.19 and Table 1.4), suggesting the importance of these base pairs for the overall stability of the triple helix. This result is also supported by the previous reports of the mutational analyses on the triple helix region.[39-41] These studies have shown that the mutations on the C72 – G41 and G71 – C42 base pairs significantly reduce the accumulation levels of the reporter genes *in vivo*. In addition, a recent *in vitro* study has shown that the mutations on G71 and C72 lead to complete degradation of the triple helix by 3' – 5' exonuclease (RNase R).[66] Taken together, our data suggest that the presence of stable C72 – G41 and G71 – C42 base pairs is important due to their high contribution to the triple helix stability, which affects the RNA accumulation level by protecting from the 3' degradation.

The ΔG_{diss} values for the U residues in the triple helix were between 45 – 55 kJ/mol, showing relatively similar stabilities (Fig. 1.19 and Table 1.4). This was unexpected as earlier studies on the synthetic DNA or RNA triple helices have shown that the Hoogsteen side is less stable than the Watson-Crick side.[54, 78] The thermal melting profiles in these studies showed two melting transitions: the first transition corresponded to the unfolding of the Hoogsteen side, while the second transition corresponded to the unfolding of the Watson-Crick side.[54, 78] However, our data suggest relatively similar stabilities between the Hoogsteen and the Watson-Crick sides.

In fact, the average ΔG_{diss} for the Hoogsteen side U residues was 51.1 kJ/mol, whereas that of the Watson-Crick side U residues were 49.1 kJ/mol. This shows that the Hoogsteen side base pairs are slightly more stable by ~ 2 kJ/mol. Indeed, a mutational study on the U•A-U base triples showed the significance of the Hoogsteen positions for the overall stability of the MALAT1 triple helix.[51] This study showed that the mutations on the Hoogsteen base-pairs result in more dramatic reduction in the RNA accumulation levels compared to the mutations on the Watson-Crick side. Taken together, these results suggest the importance of the Hoogsteen U•A base pairs for the stability of the triple helix. This observation is not in agreement with earlier studies on synthetic triple helices. This may be due to the presence of the stem I and stem II in the MALAT1 triple helix. The synthetic constructs in earlier studies only included the triple-helical region.[54, 78] On the other hand, the MALAT1 triple helix has stem residues on both sides of the triple helix, which may contribute to the stabilization of the Hoogsteen strand.

Furthermore, we observed a linear correlation between the ΔH_{diss} and ΔS_{diss} values. Correlation of ΔH_{diss} and ΔS_{diss} values have been frequently observed for various nucleic acid systems.[71-73, 76] This is often understood as the effect of the solute-solvent interaction. When base-pairs open, the water molecules surrounding the base-pairs are released. Therefore, the loss in enthalpy is compensated with the increase in entropy.[76] Moreover, the formation of the base-pairs is an enthalpically favorable interaction, which is counterbalanced by entropic costs of degrees of freedom of translation, rotation, and vibration. Therefore, the higher exothermicity results in the higher entropic penalty and vice versa, which accounts for the linear correlation between the ΔH_{diss} and ΔS_{diss} values.[76]

The slope of the linear correlation plot is also known as the compensation temperature (T_C). Theoretically, the ΔG_{diss} values for all residues are approximately the same as the y-intercept at T_C. Interestingly, earlier studies on the RNA hairpins showed that the melting temperature is close to T_C, which suggest that the global unfolding occurs when all residues have relatively similar stabilities.[71-73] The y-intercept of the linear correlation plot (Fig. 1.20A) was 54 ± 6 kJ/mol and the observed T_C was ≈ 70 °C. This temperature is close to the first transition temperature (= 65.3 °C) in the thermal melting profile (Fig. 1.20). At 70 °C, the ΔG_{diss} values for most residues were within the error of the y-intercept. However, we also observed a few residues with the ΔG_{diss} values far off from the 54 kJ/mol. The data points for these residues were off from the linear trend line on the plot (Fig. 1.20A).

At 70 °C, the average ΔG_{diss} for the triple helix and stem II residues were 51.7 and 58.0 kJ/mol, respectively. Interestingly, the average ΔG_{diss} of triple helix residues was below 54 kJ/mol (y-intercept) and that of stem II residues were above 54 kJ/mol. Moreover, when we plotted ΔH_{diss} against ΔS_{diss} values for the stem II residues, the observed T_C was ≈ 89 °C (Fig. 1.21A). This temperature is close to the second transition temperature (= 85.3 °C) in the thermal melting profile (Fig. 1.21). These observations suggest that the relatively less stable triple helix region unfolds first at 65.3 °C (first transition in the melting curve). Then the remaining stem II residues unfold at 85.3 °C (second transition in the melting curve).

Lastly, we examined the effect of Mg^{2+} ions on the MALAT1 triple helix. The results indicate that Mg^{2+} has no dramatic effect on the formation of the secondary structure (Fig. 1.13). However, the analysis of $\delta\Delta G^{Mg}$ (Fig 1.23) and the comparison of thermal melting profiles (Fig. 24) suggest that the Mg^{2+} contributes to the stabilization. Interestingly, we observed negative $\delta\Delta G^{Mg}$ values for some residues, suggesting that the stabilization effect of Mg^{2+} depends on the specific sequence and position of residues within the RNA structure. However, we observe a greater number of base-pairs stabilized with the addition of Mg^{2+}, which leads to increased overall stability. In addition, residues on the Watson-Crick side of the base triples had greater $\delta\Delta G^{Mg}$ values compared to the residues on the Hoogsteen side. This result suggests that Mg^{2+} has a greater contribution to stabilizing the Watson-Crick base pairs than the Hoogsteen base pairs in the triple helix.

4. Identification of small molecules selectively targeting MALAT1 triple helix

4.1 Small molecule microarray (SMM) screening of chemical compounds

The MALAT1 triple helix contributes to the stability of its mRNA transcripts, which affects the accumulation level of MALAT1. As indicated in section 1.1, the accumulation level of MALAT1 has been correlated with the development of various cancers, suggesting the potential role of the MALAT1 triple helix as a therapeutic target.[5] Small molecules provide an opportunity to target structured RNA motifs, such as hairpins, bulges, pseudoknots and three-helix junctions to modulate their biological functions.[80, 81] Therefore, targeting the MALAT1 triple helix with small molecules may offer the possibility to therapeutically modulate numerous cellular processes that affect cancer growth.

To identify the molecules that bind the MALAT1 triple helix, a ~ 26,000 compound library was screened using a small molecule microarray (SMM) approach by Le Grice and Schneekloth group (Fig. 1.25).[65, 82] This method utilizes a glass slide with small molecules immobilized on the surface by covalent links through their functional group. In order to aid visualization, RNA molecules of interest are fluorescently tagged at the 5' end and are flowed over the glass slides to monitor small molecule binding (Fig. 1.25A). A control experiment was also performed with the buffer in the absence of RNA. Small molecules with positive hits show a bright fluorescent signal (bottom panel in Fig. 1.25B) compared to the control (upper panel in Fig. 1.25B). SMM screening was performed with various RNAs, including hairpins, riboswitches, PAN triple helix, and MALAT triple helix to evaluate the selectivity of binding.[65]

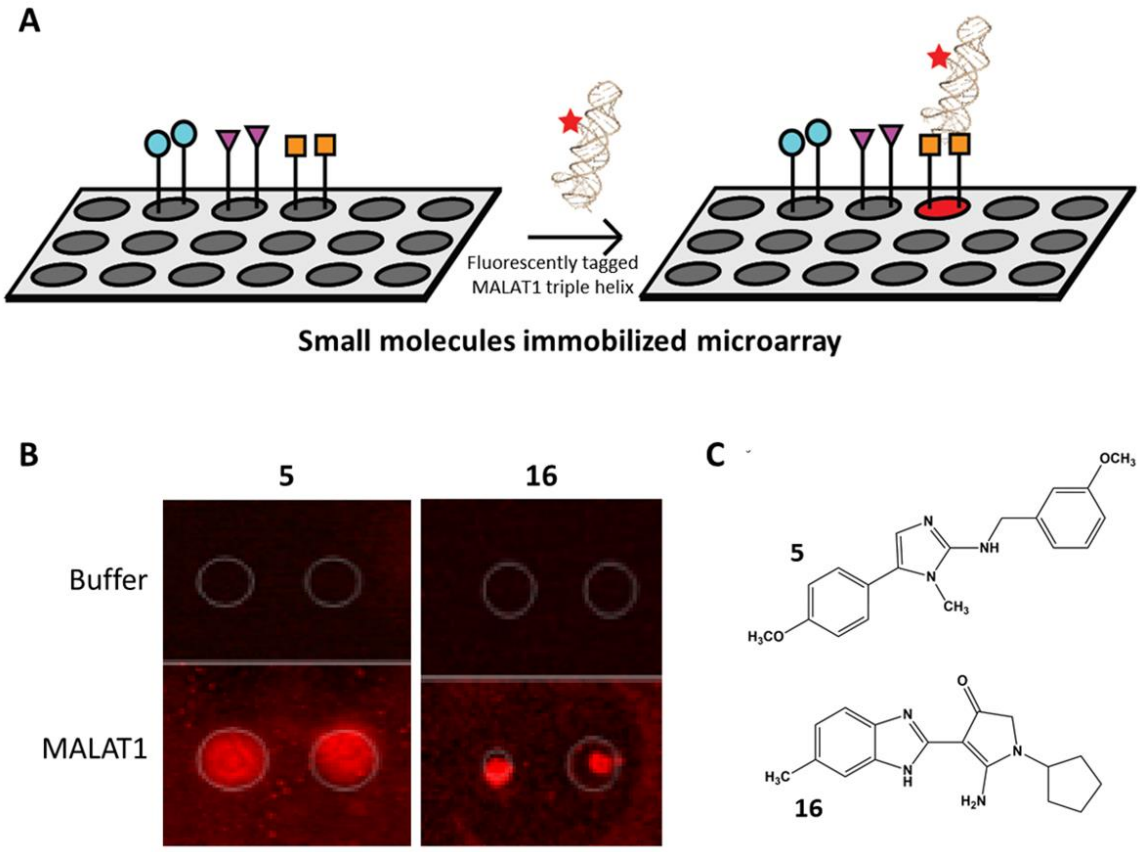


Figure 1. 25 SMM screening of small molecules for MALAT1 triple helix. (A) Schematic of SMM screening. Small molecules are covalently attached to the surface of a glass slide. Fluorescently tagged MALAT1 triple helix is flowed over the glass slides and the fluorescent signal is observed for the molecules that bind to the RNA. **(B)** Small molecules with positive hits show a bright fluorescent signals (bottom panel) compared to the control experiment with buffer (upper panel). **(C)** Structures of compounds 5 and 16.

The SMM screen identified 188 molecules as initial hits, representing a hit rate of 0.7 %. These molecules were further narrowed to 28 compounds by examination of their selectivity, commercial availability, and pharmacophore properties.[65] For further evaluation, these compounds were tested in an organoid model of mammary cancer derived from MMTV-PyMT luminal B tumors in mice. The reduction of the MALAT1 accumulation level was monitored upon the treatment of each compound. The results showed that compounds 5 and 16 induced the largest

reductions in MALAT1 accumulation, 54 % and 41%, respectively.[65] These results motivated our efforts to further characterize compounds 5 and 16 (Fig. 1.25C).

4.2 Detecting small molecules binding to MALAT1 triple helix via saturation transfer difference (STD) NMR

To confirm the direct interaction between the identified compounds and the MALAT1 triple helix *in vitro*, we performed saturation transfer difference (STD) NMR experiments. STD is one of the most popular ligand-based NMR screening methods and has been applied for detecting small molecules that bind to receptor molecules such as proteins[83-86] and RNAs[87-89]. This experiment is cost-efficient since isotope labeling is not necessary and only requires a relatively small amount of sample. The optimal K_D range for this experiment is $10^{-8} < K_D < 10^{-3}$ M.[90, 91] As determined from isothermal titration calorimetry (ITC) experiments, the K_D of compounds 5 and 16 for the MALAT1 triple helix were 2.9 ± 1.6 μM and 6.1 ± 2.1 μM, respectively.[65] Therefore, these compounds were in the right K_D range for the STD NMR.

In an STD experiment, a collection of off- and on-resonance ^{1}H spectra are recorded. The latter spectra have a train of frequency-selective radiofrequency (RF) pulses inserted in the relaxation delay for selective saturation.[90, 91] In the off-resonance spectrum, the RF pulses are applied at a frequency devoid of the receptor and the compound resonances. On the other hand, in the on-resonance spectrum, the receptor molecules are selectively saturated by irradiating at a

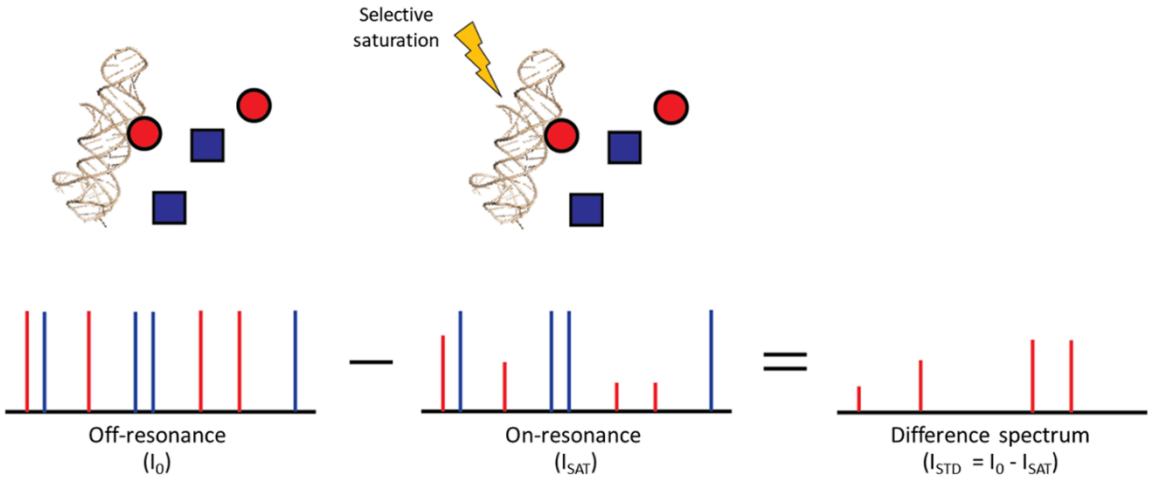


Figure 1. 26. Scheme of the STD NMR experiment. In the on-resonance spectrum, the receptor molecules are selectively saturated by irradiating at a frequency that contains the resonances of the receptor but not the small molecule. When the magnetization is transferred from the receptor to the bound compound, the signal intensity of the bound compound decreases in the on-resonance spectrum (I_{SAT}). The on-resonance spectrum is then subtracted from the off-resonance spectrum and the positive signal in the difference spectrum ($I_{STD} = I_0 - I_{SAT}$) corresponds to the receptor-bound compounds (red spectrum). For the unbound small molecules, there is no magnetization transfer between the receptor and the compound and signal intensity in the on-resonance spectrum remains the same as the off-resonance spectrum. Therefore, no positive signal is observed in the difference spectrum for the unbound compounds (blue spectrum).

frequency that contains the resonances of the receptor but not the small molecule (Fig. 1.26). Typically, the RF pulses are applied at -1 – 0 ppm for proteins to selectively saturate the methyl protons in the protein.[90, 91] For RNAs, however, the irradiation frequency pulses are applied at 4-5 ppm or 6-8 ppm, which saturates ribose or nucleobase protons, respectively.[87-89]

When the receptor molecule is irradiated, the saturation propagates to all protons within the receptor via intramolecular ^{1}H-^{1}H cross-relaxation pathways. In addition, the saturation is transferred to the small molecules that are bound to the receptor via intermolecular ^{1}H-^{1}H cross-relaxation at the ligand-receptor interface. These cross-relaxation pathways are responsible for the NOE and for the large molecules, such as proteins and RNAs, the NOE is negative. When the

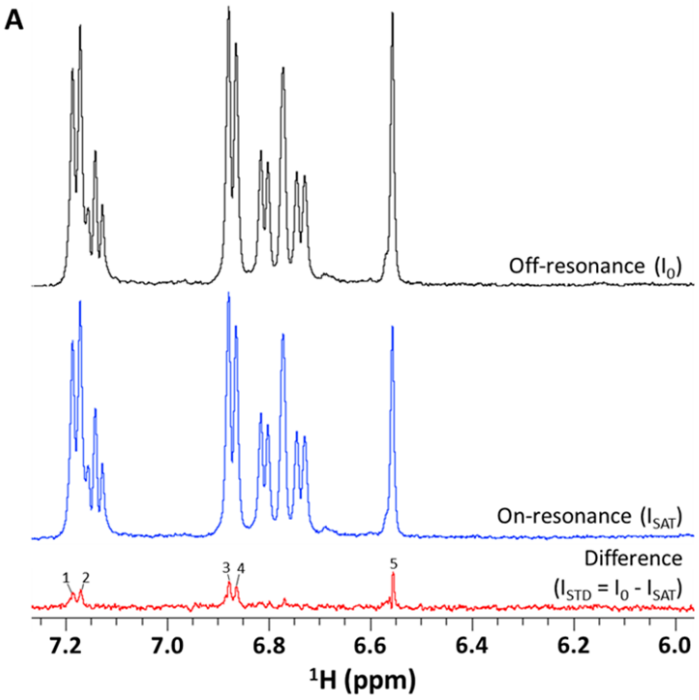
magnetization is transferred from receptor to the bound compound, the compound behaves as part of the receptor and thus, the NOE becomes negative. Therefore, the signal intensities (I_{SAT}) of the bound ligand decreases in the on-resonance spectrum (Fig. 1.26). In the off-resonance spectrum, however, the RF pulses are applied far from the resonances of the compound and the receptor and thus, the signal intensity (I_0) of the compound remains unaffected (Fig 1.26).

The on-resonance spectrum is then subtracted from the off-resonance spectrum and the positive signal in the difference spectrum ($I_{STD} = I_0 - I_{SAT}$) corresponds to the receptor-bound compound (Fig. 1.26). For an unbound small molecule, there is no magnetization transfer between the receptor and the compound and thus, the signal intensity in the on-resonance spectrum remains the same as the off-resonance spectrum. Therefore, no positive signal is observed in the difference spectrum for an unbound compound. Using the STD method, we can monitor direct interaction between the MALAT1 triple helix (receptor) and the small molecules 5 and 16 by observing the signal in the difference spectrum.

Figure 1.27A shows the STD results for compound 5. The red spectrum at the bottom is the difference spectrum between the off-resonance spectrum (black spectrum in Fig. 1.27A) and the on-resonance spectrum (blue spectrum in Fig. 1.27A). The peak intensities in the difference spectrum were weak relative to that of the off-resonance reference spectrum. For each peak, we calculated % STD (Fig. 1.27B), defined as $100*I_{STD}/I_0$, which is a simple and commonly used method to determine the positive hits.[90, 91] Typically, the % STD threshold for positive hit selection is within 5 – 10 % and can also be set as low as 2 %, depending on

certain circumstances.[90] The % STD was calculated for five peaks (labeled 1 – 5 in Fig. 1.27A) and the values ranged from ~ 7 – 13 % (Fig. 1.27B), which were within the threshold for positive hits. In particular, peak 5 at 6.55 ppm showed a higher % STD compared to the other peaks. This suggests a proton corresponding to peak 5 may be the closest region of compound 5 interacting with MALA1 triple helix. Based on these results, we concluded that compound 5 is directly binding to the MALAT1 triple helix.

STD experiments on compound 16 were inconclusive due to solubility issues and weak signal intensities (data not shown). As a result, we focused solely on compound 5 for further characterization.



B

Peak #	^{1}H (ppm)	% STD
1	7.18	7.0
2	7.17	6.8
3	6.88	9.2
4	6.86	8.5
5	6.55	12.9

Figure 1. 27 STD result for compound 5 (A) Off-resonance spectrum (black), on-resonance spectrum (blue) and the difference spectrum (red). The positive signal in the difference spectrum confirms the direct interaction between compound 5 and MALAT1 triple helix. **(B)** % STD (= $100*I_{STD}/I_0$) calculated for each peak. Typically, the % STD threshold for positive hit selection is within 5-10 %.

4.3 Selectivity and interacting moiety of compound 5

To evaluate the selectivity of compound 5, we performed the STD NMR experiment with MENβ and PAN triple helix (Fig. 1.3). As indicated in section 1.4, MENβ and PAN also have similar triple helical structure as MALAT1, with U-rich loop sequestering downstream poly(A) tail.[40, 42] Thus, we can determine whether compound 5 binds to RNA triple helices in a general manner, or is selective for the MALAT1 triple helix. The STD experiments were performed under the same condition as MALAT1 triple helix. The STD results showed positive signals for the MALAT1 triple helix, but not for the MENβ and PAN triple helices (Fig. 1.28B-D). Thus, these data show that compound 5 selectively binds only the MALAT1 triple helix.

In addition to the selectivity of compound 5, we were also able to map the interacting region by assigning the peaks that appeared in the STD spectrum. The positive peaks in the STD spectrum appear when the magnetization is transferred from the MALAT1 triple helix (receptor), which indicates that these peaks belong to compound 5 (ligand) protons that are close to the RNA. A peak at 6.55 ppm (labeled with blue asterisk in Fig. 1.28A,B) and four peaks between 6.85 – 7.20 ppm (labeled with red asterisks in Fig. 1.28A,B) in the MALAT1 STD spectrum are resonances of protons on the methyl-substituted imidazole ring and the *para*-methoxy phenyl ring, respectively. These results suggest that these moieties in compound 5 are interacting with the MALAT1 triple helix. The proton on the imidazole ring, in particular, had the highest % STD (Fig. 1.27B), indicating that the imidazole ring may be the closest to the RNA.

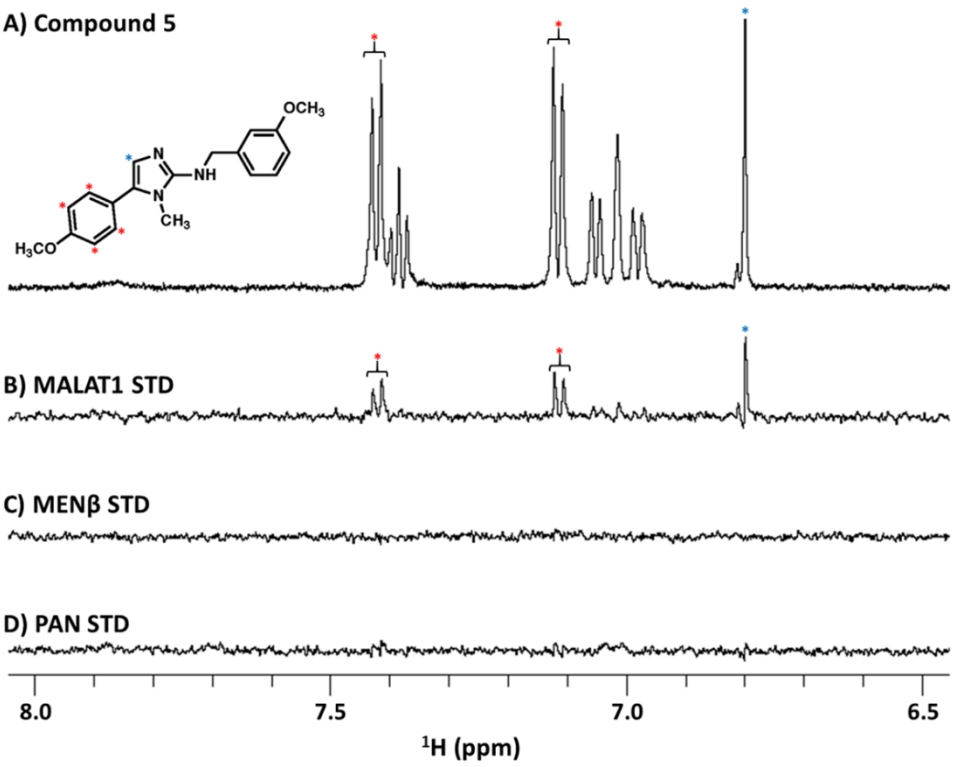


Figure 1. 28 Selectivity of compound 5 [65] **(A)** 1D reference spectrum of compound 5. **(B-D)** STD spectra compound 5 with MALAT1, MENβ, and PAN triple helices, respectively. The STD signals for compound 5 appear with MALAT1 triple helix but not with MENβ and PAN, which highlights the selectivity of compound 5.

We also tested 13 different compound 5 analogs using the STD NMR (Fig. 1.29 and A.2). Nine compounds showed binding with the MALAT1 triple helix and two compounds were weak-binder. Compound 50 and 51 were non-binders showing no STD signal. The binding of these compounds may be perturbed due to the steric hindrance caused by the large functional groups that have been substituted. The STD spectra for all the binding compounds commonly showed the highest signal for the proton on methyl-substituted imidazole ring. Even the weak-binders showed the STD signal for this proton. Taken together, these results suggest that the imidazole ring is closely interacting with the RNA.

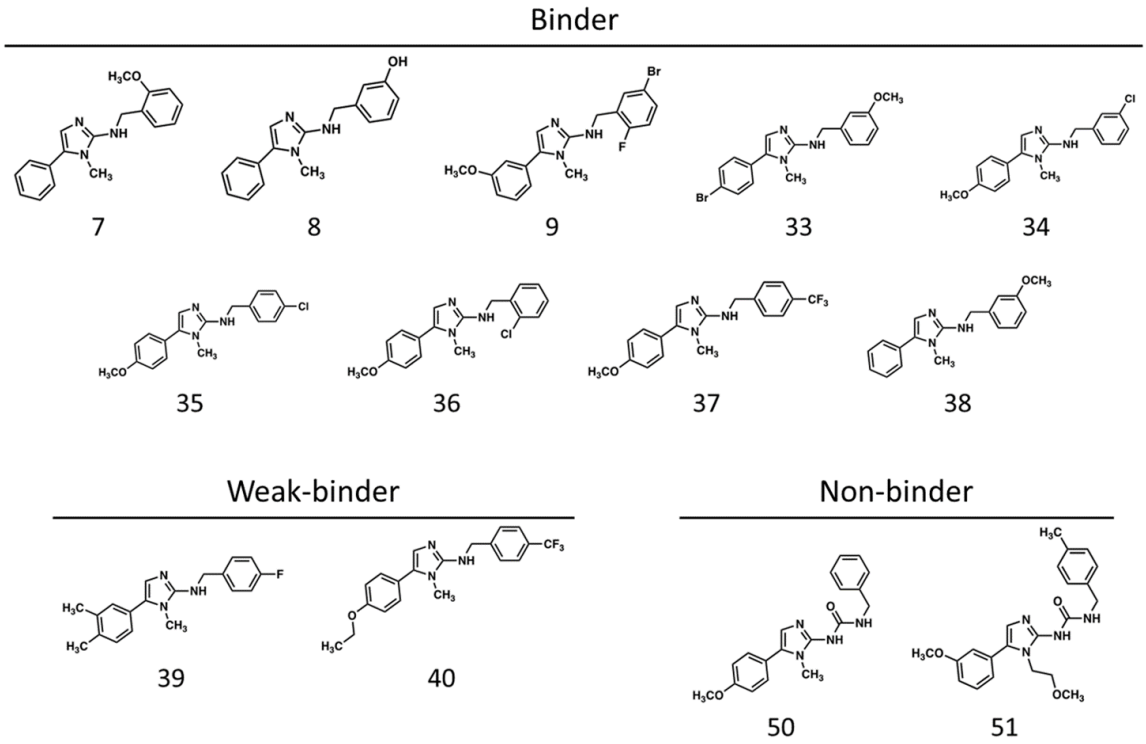


Figure 1. 29 Analogs of compound 5. 13 different analogs were tested. The results showed that 9 compounds are a relatively strong binders. Two compounds are weak-binders and two compounds are non-binders.

4.4 Effect of compound 5 on the MALAT1 triple helix

To examine the structural change of the MALAT1 triple helix upon binding of compound 5, we performed a titration and recorded ^{15}N-^{1}H HSQC spectra at 15 °C. A 0.3 mM of uniformly ^{15}N-labeled RNA sample was titrated with compound 5 at a final concentration between 0 – 1.2 mM. Fig. 1.30 shows an overlay of the titration results: the spectrum without compound 5 in black and the spectrum with 1.2 mM of compound 5 in red. The results show no significant change in the chemical shifts of imino peaks upon binding of compound 5 (Fig. 1.30). Similar results were observed when the compound was titrated into a ^{13}C labeled RNA sample for monitoring C8 and C6 residues (data not shown). These results suggest

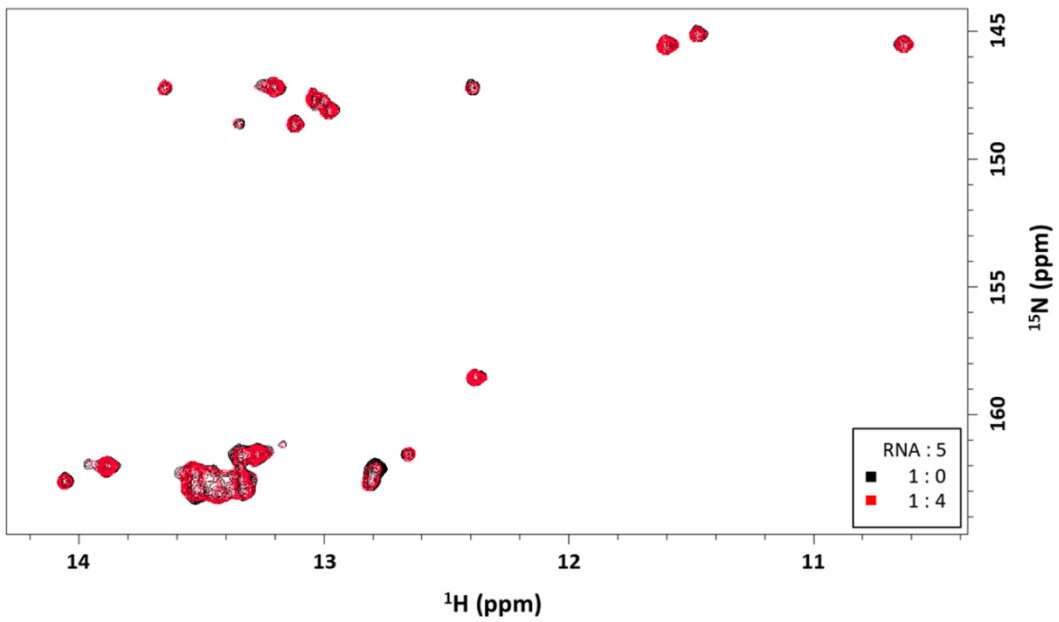


Figure 1. 30 Compound 5 titration at 15 °C. A 0.3 mM RNA sample was titrated with compound 5 in the final concentration ranging between 0 – 1.2 mM. There was no significant change in the chemical shifts upon binding of compound 5, however, we observed line-broadening of few peaks.

that there is no dramatic change in the structure of MALAT1 triple helix upon binding compound 5.

However, we did observe significant changes in signal intensity for a few peaks due to line-broadening. It is commonly understood that both slow tumbling or chemical exchange can cause line-broadening.[92] The effect of slow tumbling is caused by the large molecular weight of the interacting molecule. In this case, the signal intensity of all RNA peaks would decrease in a similar extent.[92] Since compound 5 is a small molecule (323 Da), it is unlikely that the observed line-broadening is due to the slow tumbling. In this case, the broadening may be due to the intermediate or slow exchange between free and bound states of the RNA.[92]

For each peak, we analyzed the relative intensity change described as $(I_0 - I_5)/I_0$, where I_0 and I_5 are the signal intensities in the absence and presence

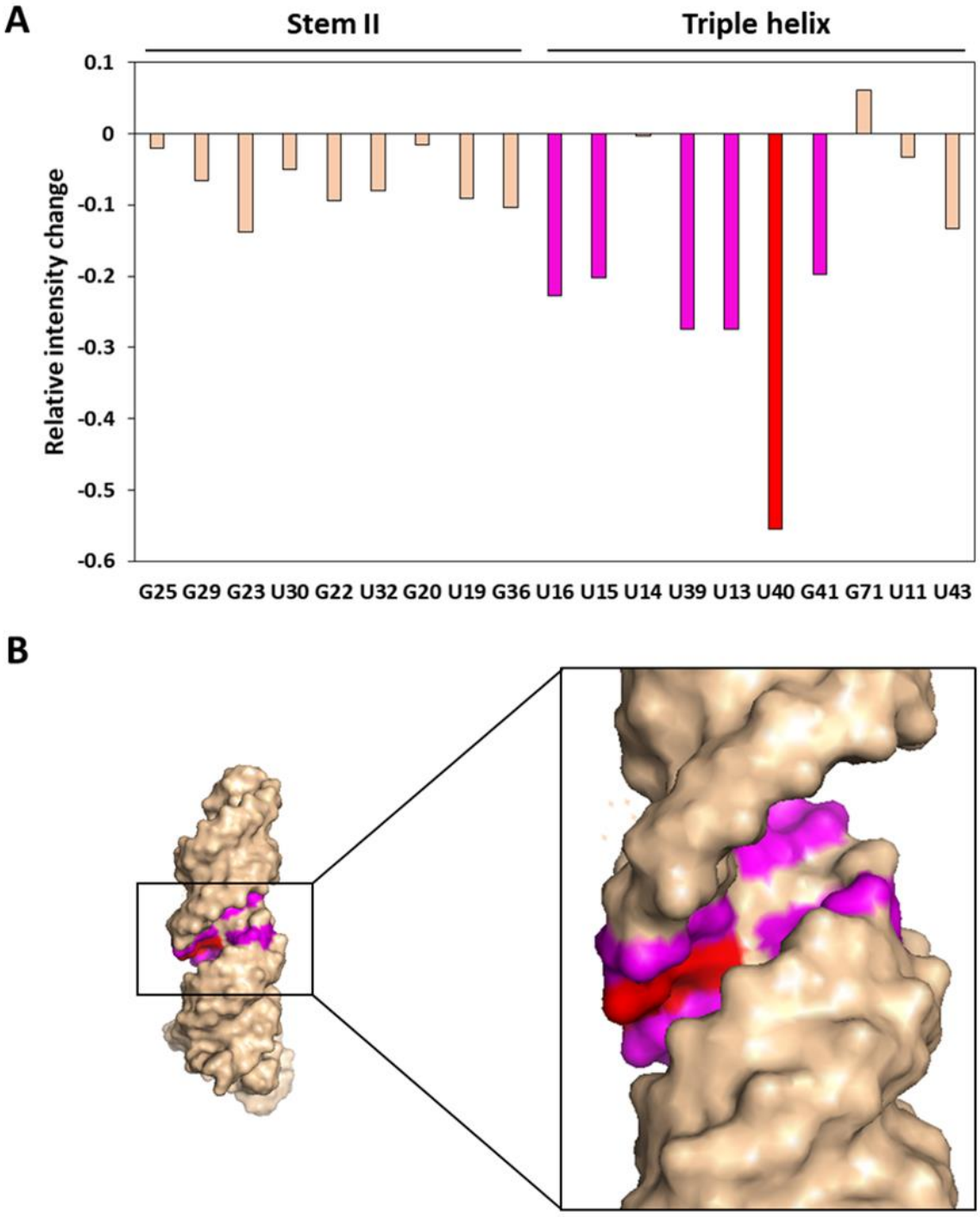


Figure 1. 31 Analysis of relative intensity changes upon binding of compound 5. (A) For each peak, the change in relative intensities upon binding of compound 5 was calculated. The change in relative intensity was calculated according to: $(I_5 - I_0)/I_0$, where I_0 and I_5 are the peak intensity in the absence and presence of compound 5, respectively. **(B)** Surface representation of the MALAT1 triple helix. The upper triple helix residues that showed a significant change in the relative intensity upon binding compound 5 are colored with magenta and red.

of compound 5, respectively (Fig. 1.31A). From the analysis, we identified a few residues that had a greater change in the signal intensity compared to the other residues. Interestingly, the identified residues, U13, U15, U16, U39, U40, and G41, are on the major groove of the upper triple helix region (colored with magenta and red in Fig. 1.31A,B). U40, in particular, had the most dramatic effect upon binding with compound 5 (colored red in Fig.1.31A,B). This result suggests that the major groove of the upper triple helix may be the binding pocket of the compound 5. This is also in agreement with a molecular docking model, which showed compound 5 interacting with the residues on the major groove of upper triple helix.[65]

4.5 Discussion

As a potential therapeutic method, we have identified small molecules that can selectively target the MALAT1 triple helix. Schneekloth and Le Grice group performed an initial high-throughput screening using the SMM assay. Starting from ~26,000 compounds, two compounds were identified that reduced the MALAT1 accumulation level in MMTV-PyMT tumors.[65] Although the SMM approach allows the screening of a large number of compounds, one end of the small molecules is covalently linked to the surface of the glass slides and thus has limited accessibility. Therefore, an alternative method is required to confirm the binding between the flexible small molecules and the target RNA. To confirm the direct binding between the small molecules and the MALAT1 triple helix, we performed the STD experiment. The binding affinity of the identified small molecules was in the

appropriated range for the STD NMR measurements.[65] Our STD data confirmed the selective binding between the MALAT1 triple helix and compound 5.

Spector's group monitored the effect of compound 5 in the MMTV-PyMT tumor organoids. Treatment of the tumor with the compound caused a reduction of MALAT1 accumulation level and branching morphogenesis, which were the same effects as MALAT1 knockout using ASOs.[15, 65] Contrary to our expectation, there was no dramatic structural change of the MALAT1 triple helix upon binding of compound 5 (Fig. 1.30). This suggests that the MALAT1 triple helix is a very stable structure and binding with a small molecule is not enough to perturb or completely disrupt the formation of the triple helix. Although there were no significant chemical shift perturbations, we observed line-broadening for several peaks (Fig. 1.30). The most dramatic changes in the peak intensity were observed for the residues in the major groove of the upper triple helix region. This suggests that the binding pocket for compound 5 is in the upper triple helix region, which is in agreement with a previous docking model.[65]

Furthermore, we analyzed the effect of compound 5 on the stability of the MALAT1 triple helix by measuring the imino proton exchange rate (data not shown). The data indicated reduced stability for some residues but others had increased stability, which resulted in no significant change in overall stability. FRET analysis by Baird group also showed no significant change in the overall stability in the presence of Mg^{2+}. However, binding of compound 5 resulted in increased stability of the MALAT1 triple helix in the absence of Mg^{2+}.[65]

This was unexpected as our initial hypothesis was that the observed reduction of cancer development is mainly due to the destabilization of the triple helix. A possible explanation for these observations is that the binding of compound 5 can affect the interactions between the MALAT1 triple helix and other biomolecules, which may lead to inhibition of the tumor growth. In fact, many of the bioactive small molecules bind in the RNA – protein interfaces.[80] Although binding of compound 5 did not lead to the destabilization or structural changes of the MALAT1 triple helix, it may perturb the interaction between the triple helix and other proteins or nucleic acids.

As a matter of fact, there has been a study, which identified a protein that interacts with the MALAT1 triple helix. This study showed that methyltransferase-like protein 16 (METTL16) binds the MALAT1 triple helix.[93] This protein is predicted to be an RNA methyltransferase that methylates the N^6 position of adenine. The mutational analysis of the triple helix region suggested that the upper triple helix is important for the binding of the METTL16. Although the functional role of this RNA – protein interaction is unknown, the disruption of this interaction could lead to a reduction of cancer development. Since the predicted binding pocket for compound 5 is also in the major groove of the upper triple helix, binding of compound 5 may disrupt this RNA-protein interaction. However, further studies have to be performed in order to test whether compound 5 can actually affect the interaction between METTL16 and the MALAT1 triple helix.

5. Materials and Methods

5.1 Sample preparation

RNA samples were prepared by *in vitro* transcription using T7 RNA polymerase.[94] Uniform ^{15}N- labeled rNTPs (Cambridge Isotope Laboratories, Inc (MA)) or unlabeled rNTPs (Chem-Impex, Inc (IL) were used depending on the sample requirements. For the SL construct, a synthetic DNA template (Integrated DNA Technologies, Inc (CA)) and an RNA polymerase promoter sequences (5′-CTA ATA CGA CTC ACT ATA G-3′) were used as a template. Poly(A) tail for the *trans* construct was purchased from Dharmacon, Inc (IL). For the MALAT1 core *cis* construct, a double-stranded DNA template was prepared by polymerase chain reaction (PCR) reaction. The template for the *cis* construct included RNA polymerase promoter, MALAT1 core *cis*, and Hepatitis delta virus (HDV) sequences.[95] HDV sequence was added to give homogeneity of 3' end in poly(A) tail.

The transcription reaction contained the following components: 40 mM Tris-HCl (pH 8.0), 1 mM spermidine, 10 mM dithiothreitol (DTT), 0.01 % Triton X-100, 300 nM DNA template, 20 % DMSO, 2 units/ml of thermostable inorganic pyrophosphatase (New England BioLabs, MA) and 0.1 mg/ml T7 polymerases for 10 ml reaction volume. For the SL construct, 300 nM promoter sequence and 400 mg/ml PEG were also included in the transcription buffer. The concentration of rNTPs and Mg^{2+} were optimized to obtain maximum yields. Transcribed RNAs were purified by gel extraction followed by electroelution. The eluted sample was solvent exchanged into the distilled H_2O (Fisher Scientific, Inc (NH)) using 3K

MWCO tube (Fisher Scientific, Inc (NH)). The RNA was diluted to the concentration below ~ 0.1 mM then folded by heating up to 95 °C for 3 min followed by adding 10X NMR buffer and snap-cooled on ice for 15 min. The folded sample was then solvent exchanged and concentrated into 1X NMR buffer (10 mM potassium phosphates, 50 mM KCl, 1 mM $MgCl_2$, 0.1 mM EDTA (pH 6.5)) for 2 – 3 times. Finally, NaN_3, DSS, and D_2O were added to a final concentration of 0.02 %, 20 μM, and 10 %, respectively. The final RNA concentrations ranged between 0.5 – 1 mM.

5.2 Inversion recovery experiment

The inversion recovery experiments were performed on a 600 MHz Bruker spectrometer. A modified ^{15}N-^{1}H imino HSQC experiment was used as described in earlier works with a hard Watergate pulse scheme for the water repression.[71-73] For the selective inversion of water protons, 180° refocusing band-selective uniform response pure-phase (REBURP) pulse was used with a pulse length of 8.5 ms.[96] The spectra were recorded with 16 – 21 different delay points. The delay points were between 2 ms and 2.4 s. 32 scans were used for each spectrum and the recycling delay was set to 4 s. Spectra were processed and the peak intensities were analyzed using Topspin (Bruker).

5.3 Estimation of proton transfer rate (k_{tr})

5.3.1 External proton transfer rate ($k_{tr,ext}$) measurements

$k_{tr,ext}$ was measured using free GTP and UTP as described in earlier works.[71, 72] 5 mM of GTP and UTP were dissolved in the NMR buffer (see section 5.1). 1D ^{1}H spectra were collected with 512 scans at the temperatures between 1 – 6 °C. The spectra were processed with Topspin (Bruker) and the line-width was obtained by deconvolution of the imino signals. $k_{tr,ext}$ was calculated using eq 1.14 and 1.15.

5.3.2 Internal proton transfer rate ($k_{tr,int}$) measurements

The estimation of $k_{tr,int}$ was performed as described in earlier works.[72, 73] The inversion recovery experiments were repeated as described in section 5.2 using different catalyst (phosphate ions) concentrations. k_{ex} was measured with low (10 mM potassium phosphates, 50 mM KCl, 1 mM $MgCl_2$, 0.1 mM EDTA (pH 6.5)) and high (50 mM potassium phosphates, 10 mM KCl, 1 mM $MgCl_2$, 0.1 mM EDTA (pH 6.5)) phosphate concentrations. The ratio (q) of externally and intrinsically catalyzed rate and its temperature dependence were derived from the temperature dependence of k_{ex} in different buffer conditions. $k_{tr,int}$ was calculated using q and $k_{tr,ext}$.

5.4 Saturation transfer difference NMR

The STD experiments were performed at 283.15 K on a 600 MHz Bruker spectrometer equipped with a cryoprobe. NMR samples contained 5 μM RNA in NMR buffer. Compound 5 was added as a stock solution in DMSO-d_6 to a final concentration of 150 μM and 4% DMSO-d_6. The final volume of each NMR sample was 300 μl and DSS was used as an internal standard. STD experiments were performed as described previously[86] using a standard Bruker pulse (stddiffesgp.3) with a 2 s saturation time and excitation sculpting for water suppression with 2 ms sinc pulses.[97] The RNA protons were saturated with a train of 50 ms Gaussian-shaped pulses. For selective saturation the nucleobase region of the MALAT1 triple helix and MENβ, the on-resonance irradiation frequency was set to 6.46 ppm. For PAN RNA, the on-resonance irradiation frequency was set to 5.90 ppm. The off-resonance irradiation frequency was set to 40000 Hz (≈ 67 ppm). A total of 128 scans with 16 dummy scans were collected for each. A second control experiment was performed in the absence of RNA to ensure that 5 resonances were not being excited. Spectra were processed in Topspin (Bruker).

6. Conclusion and future direction

6.1 Summary of work

Considering that the MALAT1 triple helix may be a potential therapeutic target for cancer, it is important to understand its stability in greater detail. In addition to the previous reports on the global stability of the MALAT1 triple helix,[65, 66] we examined its local stability via NMR spectroscopy. The imino proton exchange rates for individual base-pairs were measured using an inversion recovery experiment. Analysis of the exchange rates revealed that C72-G41 and G41-C42 base-pairs were the most stable base-pairs within the triple helix with relatively high ΔG_{diss} values. These residues have a large contribution to the stabilization of the triple helix. Moreover, a comparison of T_C, determined from the linear correlation plot of ΔH_{diss} and ΔS_{diss}, and the melting temperature (T_m) suggests that the first melting transition corresponds to the global unfolding of the triple helix region. In addition, the second transition in the melting profile corresponds to the unfolding of the stem II. Further analysis suggests that the presence of stem II contributes to stabilizing the triple helix region. Finally, we have examined the effect of Mg^{2+} on the MALAT1 triple helix. Although Mg^{2+} has no dramatic effect on the formation of the secondary structure, it contributes to stabilizing the MALAT1 triple helix.

We also screened small molecules to identify compounds that selectively target the MALAT1 triple helix and thus might serve as a potential therapeutic for cancer treatment. More than 20,000 compounds were screened with SMM screening and two compounds, compound 5 and 16, were identified by Le Grice

and Schneekloth group.[65] We then performed STD experiments to confirm the direct interaction between these compounds and MALAT1 triple helix. Due to the solubility issues of compound 16, the NMR studies focused on compound 5. The STD results indicate that compound 5 binds to MALAT1 triple helix and the analysis of % STD suggests that the imidazole ring of compound 5 is most closely interacting with the RNA. Moreover, STD experiments on compound 5 analogs further demonstrated the importance of the imidazole ring for binding. Furthermore, we performed the STD experiment on other triple-helical RNAs, PAN and MEN β, and confirmed the selectivity of compound 5 for MALAT1. Monitoring the RNA peaks upon binding of compound 5 showed no significant chemical shift perturbations indicating no dramatic change in the RNA structure. However, we observe large changes in relative peak intensities for several peaks, which suggests a slow-to-intermediate exchange of free and bound states of RNA. Interestingly, the identified residues were on the major groove of the upper triple helix region suggesting that this region may be the binding pocket for compound 5. This agrees with a molecular docking model, which showed compound 5 binding the major groove of the upper triple helix region.

6.2 Future direction

There are more aspects of the MALAT1 triple helix that can be further investigated. For the analysis of the stability, we can mutate some of the key residues within the triple helix and monitor how those mutations affect the stability of other residues by repeating the inversion recovery experiment. *In vivo* and *in*

vitro assays in earlier studies have shown that some of the mutations cause the complete abolition of the RNA accumulation level.[39-41, 66] However, these studies have not shown whether these mutations lead to a partial or complete disruption of the triple helix. It would be interesting to analyze to what extent the triple helix is perturbed due to these mutations and whether partial disruption is enough to cause complete removal of RNA accumulation.

In addition, syntheses of compound 5 analogs that are more soluble and have a greater binding affinity are needed. Due to the limited solubility of this compound, we were not able to test the higher concentration of compound 5. Although we did not observe significant structural changes of the triple helix in the current condition, more noticeable structural changes could occur with a higher concentration of the compound. Moreover, a compound with a better binding affinity could cause a more significant change in structure and stability. As an alternative compound, Hargrove's group recently identified a small molecule that can target MALAT1 triple helix with a greater binding affinity.[98] It would be interesting to test whether this compound has a greater effect on the structure and stability of the triple helix.

Lastly, we can examine whether the binding of compound 5 perturbs the RNA – protein interaction. As indicated in section 4.5, METTL16 binds to the MALAT1 triple helix and the upper triple-helical region is important for this interaction.[93] Since compound 5 is binding to the major groove of the upper triple helix, testing whether this compound affects the RNA – protein interaction will provide better insights on the effect of the compound in the cell.

Part 2

Deleterious Effects of Carbon-Carbon Coupling on RNA NMR Dynamics

Adapted from a manuscript by Nam, H.; Becette, O.; LeBlanc, R.M.; Oh, D.; Case, D.A.; Dayie, T.K.

(H.N designed the experiments, prepared the site-selectively labeled purines, and performed all the relaxation measurements and simulations. O.B prepared the site-selectively labeled pyrimidines. O.B and D.O prepared RNA samples for the dynamics measurements. R.L collected the preliminary data. D.C calculated CSA tensor values. T.K.D simulated the CSA contributions to the relaxation rates. H.N and T.K.D wrote the manuscript with inputs from all co-authors.)

1. Introduction

1.1 Studies on RNA Dynamics

The biological function of many regulatory RNAs depends on both the structure and dynamics of the molecule. Studies continue to show that dynamic exchanges of RNA molecules occur over a wide range of timescales that are important for their biological activity.[99, 100] Riboswitches, for example, undergo dynamic conformational changes in response to ligand binding to regulate gene expression at the level of transcription, splicing, or translation.[101, 102] Similarly, interhelical motions of tRNA can significantly change the relative orientation of helical domains and affect the formation of tRNA-protein, tRNA-ribosome complexes.[103] Furthermore, active sites of ribozymes are conformationally flexible and can undergo dynamic change during catalytic cycles to form reactive structures.[104, 105]

To better understand these dynamic properties of RNA molecules, different biophysical tools such as cryo-EM[106, 107], small molecule FRET[108], and time-resolved X-ray crystallography[109] have been applied. Among these tools, NMR spectroscopy enables the direct monitoring and quantification of dynamic motions at atomic resolution. The two commonly measured dynamic parameters comprise longitudinal (R_1) and transverse (R_2) relaxation rates.[110, 111] The R_1 relaxation rate measures the return of the longitudinal magnetization to thermal equilibrium whereas the R_2 relaxation rate measures the decay of x- and y- magnetization (loss of coherence as individual spins experience slightly different resonance frequencies). The relaxation rates R_1 and R_2 are directly related to spectral density

functions, $J(\omega)$, which describes the probability of finding dynamic motions at a given frequency (ω). In addition to R_1 and R_2, hNOE measures the change in heteronuclear spin magnetization in response to saturating protons spins and can provide meaningful information on rotational diffusion, correlation time, internal motions, and flexibility.[112, 113]

In proteins, dynamic motions are typically measured using ^{15}N nuclei in the amide backbone.[110, 114] For RNAs, ^{15}N nuclei are only present in the bases and due to the water exchange properties of imino protons ^{15}N nuclei are suitable reporters of hydrogen-bonding and dynamics of base-paired guanine and uracil bases[115]; solvent exposed imino regions are usually broadened beyond detection. Nonetheless, ^{15}N chemical shift is very sensitive to RNA secondary structure and makes imino ^{15}N relaxation measurements a useful complement for characterizing the secondary structure of RNA. The limited availability of suitable imino nitrogen probes has necessitated the use of primarily protonated carbons as alternative relaxation probes. These carbon sites are found in all the ribose sugar moiety (C1', C2', C3', C4', and C5') and all 4 bases (C2-Ade, C8-Ade, C8-Gua, C5-Cyt, C6-Cyt, C6-Cyt, C6-Ura). Thus, ^{13}C relaxation measurements can potentially provide a more complete coverage of the structure necessary to obtain further insights into dynamic properties of these RNAs.

1.2 Effect of dipolar coupling on relaxation measurements

Despite the greater number of detectable ^{13}C nuclei in RNA compared to ^{15}N nuclei, there are significant challenges that complicate measurements and analysis of ^{13}C relaxation. First in uniformly labeled RNA, ^{13}C spins do not approximate an isolated 2-spin system; these carbon sites are linked by intricate multi-bond couplings to other ^{15}N, ^{13}C, and ^{1}H nuclei and positioned within 2.5 Å of several nuclei. These extensive dipolar couplings between adjacent carbons complicate analysis of the carbon relaxation rates.[116-118] The relative contribution to the relaxation of a ^{13}C nucleus from covalently attached carbon and proton spins can be determined according to the following approximation:[116]

$$\frac{\rho_{CC}}{\rho_{HC}} \approx \frac{(\omega_C \tau_C)^2}{3}\left(\frac{{\gamma_C}^2}{{\gamma_H}^2}\right)\left(\frac{r_{HC}}{r_{CC}}\right)^6 \qquad (2.1)$$

where ρ_{ic} is the contribution to the selective relaxation rate of the ^{13}C nucleus from an attached proton or carbon, γ_j is the gyromagnetic ratio of spin *j*, r_{ij} is the distance between spins *i* and *j*, ω_C is the Larmor frequency of carbon, and τ_C is the overall molecular tumbling time. As shown in eq 2.1, the contribution of the attached carbon relative to the attached proton on the ^{13}C relaxation rate is proportional to the square of the overall correlation time.

Previous works have shown that uniform isotopic labeling negatively impacts longitudinal relaxation rate measurements in proteins and nucleic acids; discrepant R_1 measurements increase with increased τ_C as captured by eq 1.[116, 118-124] Comparison of uniformly ^{13}C labeled and alternate site ^{13}C labeled adenine ribonucleotide monophosphate (rAMP) showed greater difference in relaxation

rate with larger τ_C, again in agreement with eq 2.1.[118] These network of carbon-carbon couplings hinder accurate measurements and complicate extraction of relaxation rate constants. These prior works suggest that the interpretation of relaxation measurements to determine dynamic models of macromolecules can be erroneously impacted especially for biopolymers of large size (τ_C > 7 ns).

Here we extend those earlier studies and address the significant effect of adjacent ^{13}C on RNA dynamics measurements for not just longitudinal but also transverse and cross-relaxation measurements in RNA. We transcribe various RNAs *in vitro* using uniform or site selective labeled ribonucleoside-5'-triphosphates (rNTPs). The site-selective labeled rNTPs, synthesized in house using chemo-enzymatic methods (Fig. 2.1A), are strategically enriched with ^{13}C-isotopes at the ribose C1', purine C8, and pyrimidine C6 carbon sites (Fig. 2.1B).[125, 126] TROSY-detected pulse sequences[121] are used to measure carbon R_1, on-resonance $R_{1\rho}$ rates and hNOE for RNAs with sizes varying between 14 to 61 nt. We also compare our experimental results with the computational simulation of relaxation rates. Our simulation includes the effect of dipolar coupling of nearby nuclei calculated using X-ray crystal coordinates derived distances and CSA effect of each nuclei using values derived from density functional theory (DFT) calculation. Such an approach allows us to compare theoretical predictions with experimentally measured values.

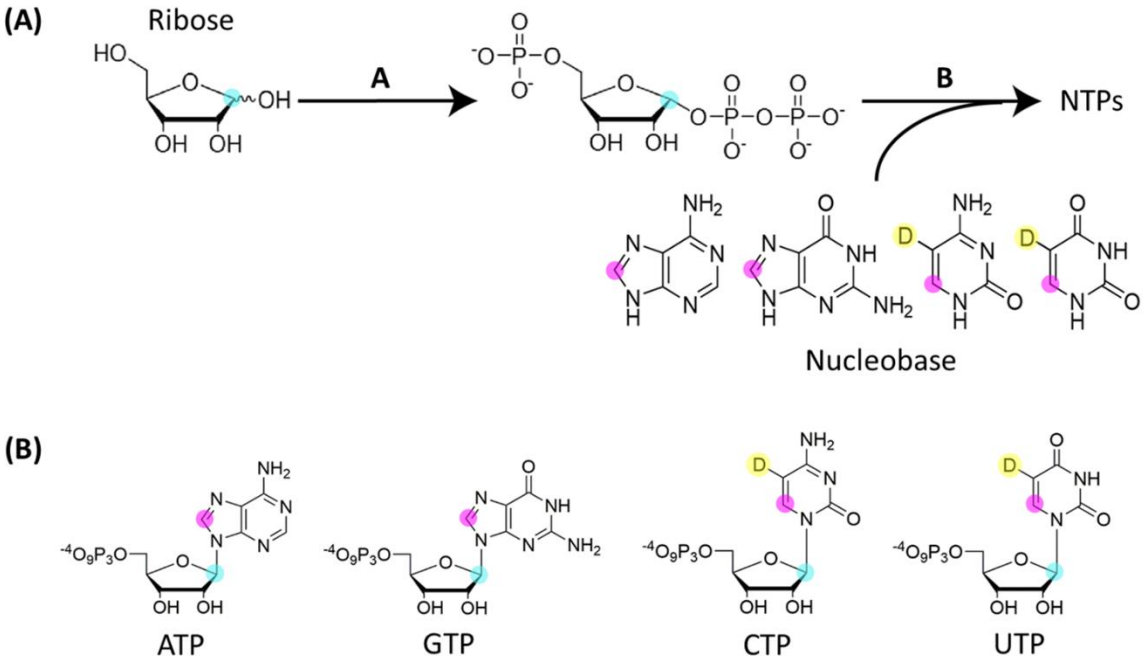


Figure 2. 1. **Synthesis of site-selective label by chemo-enzymatic method. (A)** Outline of NTP synthesis by chemo-enzymatic methods. Enzymes in step A includes RK and PRPPS. Enzymes in Step B includes MK and CK for ATP synthesis, GK and CK for GTP synthesis, UMPK and CK for UTP synthesis. UTP is converted to CTP using CTPS. For all steps MK and CK is added for dATP regeneration. **(B)** Scheme of four nucleotides used for the study. C1' on ribose (cyan) was labeled for all nucleotides. C8 position on base was labeled (magenta) for ATP and GTP. For pyrimidines, C6 position on base was labeled (magenta) and H5 position was deuterated (yellow).

Enzyme abbreviation RK: ribokinase (E.C. 2.7.1.15), PRPPS: ribose-phosphate diphosphokinase (E.C. 2.7.6.1), MK: myokinase (E.C. 2.7.4.3), CK: creatine kinase (E.C. 2.7.3.2), GK: guanylate kinase (E.C. 2.7.4.8), UMPK: UMP kinase (E.C. 2.7.4.22), CTPS: CTP synthetase (E.C. 6.3.4.2)

2. ^{13}C Longitudinal Relaxation (R_1)

2.1 Simulation of ^{13}C longitudinal relaxation (R_1)

Longitudinal relaxation (R_1) is one of the most commonly measured relaxation parameters to obtain dynamics information on biomolecules.[110, 114, 121] The longitudinal relaxation rate of the ^{13}C spin is given by the equation[116, 127, 128]:

$$\frac{dC_z{}^K}{dt} = -R_{1,C_K}\left(C_z{}^K - C_z{}^{K,0}\right) - \sum_i \sigma_{C_K,C_i}(C_z{}^i - C_z{}^{i,0}) - \sum_j \sigma_{C_K,H_j}\left(I_z{}^j - I_z{}^{j,0}\right) \quad (2.2)$$

where $C_z{}^K$ is the longitudinal magnetization of carbon K, $C_z{}^{K,0}$ is the equilibrium longitudinal magnetization, $C_z{}^i$ and $I_z{}^j$ are longitudinal magnetizations of directly attached carbon i and proton j, respectively. R_{1,C_K} is auto-relaxation rates and σ is cross-relaxation rates given by following equations[116, 127]:

$$R_{1,C_K} = \sum_i R_{1,C_K,H_i} + \sum_j R_{1,C_K,C_j} + \sum_j R_{1,C_K,N_j} + R_{1,CSA} \quad (2.3)$$

$$\sigma_{C_K,i} = \frac{1}{10} d_{C_K,i}{}^2\left[6J\left(\omega_{C_K} + \omega_i\right) - J\left(\omega_{C_K} - \omega_i\right)\right] \quad (2.4)$$

in which $R_{1,C_K,i}$, $R_{1,CSA}$, and $d_{C_K,i}$ are[116, 127]:

$$R_{1,C_K,i} = \frac{1}{10} d_{C_K,i}{}^2\left[3J\left(\omega_{C_K}\right) + 6J\left(\omega_{C_K} + \omega_i\right) + J\left(\omega_{C_K} - \omega_i\right)\right] \quad (2.5)$$

$$R_{1,CSA} = \frac{1}{10} C_{C_K}{}^2\left[3J\left(\omega_{C_K}\right)\right] \quad (2.6)$$

$$d_{C_K,i} = \frac{\mu_0 \gamma_{C_K} \gamma_i h}{8\pi^2 r^3} \quad (2.7)$$

Eq 2.2 shows that the longitudinal relaxation rates have contribution from both auto-relaxation rate (R_{1,C_K}) and cross-relaxation rate ($\sigma_{C_K,i}$) between neighboring spins. $\sigma_{C,H}$ can be neglected since it is generally 2 – 3 orders of magnitude smaller than auto-relaxation rate and C – C cross-relaxation is experimentally suppressed as previously described.[116, 121] Therefore, the major contribution to the longitudinal relaxation rates comes from auto-relaxation rates. As shown in eq 2.3, auto-

relaxation rate (R_{1,C_K}) incorporates R_{1,C_K,H_i}, R_{1,C_K,C_j}, and R_{1,C_K,N_j} terms: these represent dipolar interaction of carbon K with its directly attached proton, other ^{13}C nuclei, and ^{15}N nuclei respectively. In addition R_{1,C_K} includes $R_{1,CSA}$, representing CSA relaxation mechanism of carbon K. C_{C_K} is a CSA constant for carbon K, defined as $\omega_{C_K}\Delta\sigma_{C_K}/\sqrt{3}$, where $\Delta\sigma_{C_K} = \sqrt{{\sigma_x}^2 + {\sigma_x}^2 - \sigma_x\sigma_y}$, $\sigma_x = \sigma_{33} - \sigma_{11}$, and $\sigma_y = \sigma_{33} - \sigma_{22}$. σ_{11}, σ_{22}, and σ_{33} are the principal components of the chemical shielding tensor.[129-131] $J(\omega)$ is a spectral density function which is assumed to be a Lorentzian $J(\omega) = \frac{\tau_C}{1+(\omega\tau_C)^2}$, γ_i is the gyromagnetic ratio of spin i, r is the distance between the two spins and h is Plank's constant.

We utilized these equations to analyze and compare the relaxation rates of uniform and site selective labeled NTPs. Unlike the auto-relaxation rate of C6 on selectively labeled pyrimidines (${R_{1,C_6}}^s$), that of C6 on uniformly labeled pyrimidines (${R_{1,C_6}}^u$) includes dipolar contributions from ^{13}C labeled C2, C4 and C5 (R_{1,C_6,C_i}), as well as the dipolar contributions from ^{15}N labeled N1 and N3 (R_{1,C_6,N_i}). Of great interest the $J(\omega_{C_K} - \omega_i)$ term in eq 2.5 is proportional to $J(0)$ when spin i is carbon, which increases with increasing τ_C (Fig. 2.2). Thus, increased dipolar contributions from the neighboring carbons are expected with greater τ_C. Among all the R_{1,C_6,C_i} terms, R_{1,C_6,C_5} is a major contributor to the difference between ${R_{1,C_6}}^s$ and ${R_{1,C_6}}^u$, as C5 is the closest carbon to C6 and, thus, d_{C_6,C_5} is greater than other d_{C_6,C_i}. This predicts C5 having the largest dipolar contributions to the ${R_{1,C_6}}^u$. Similarly, ${R_{1,C_1}{}'}^u$ includes $R_{1,C_1{}',C_2{}'}$ term in contrast to ${R_{1,C_1}{}'}^s$, which leads to a greater deviation

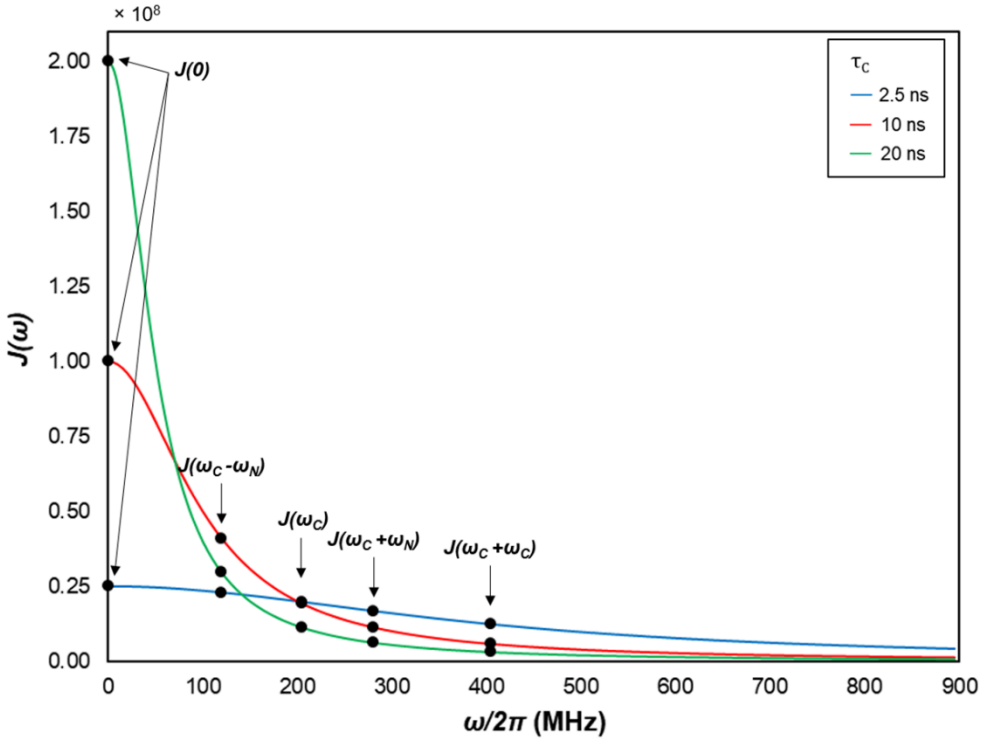


Figure 2. 2. A simulation of spectral density function at 800 MHz. The simulation shows dramatic increase of $J(0)$ term with increasing τ_C. Compared to the change in the $J(0)$ term, there is no significant changes in other terms with different τ_C.

between $R_{1,C_1'}{}^{u}$ and $R_{1,C_1'}{}^{s}$ as τ_C increases. On the other hand, because C8 in purines do not have directly bonded carbon neighbors, dipolar contributions for C8 are usually assumed to be relatively small compared to CH-groups in pyrimidines or ribose moieties. However, the contributions from neighboring 2-bond carbons, such as C4 and C5 carbons, increase for large RNAs (τ_C > 20ns).

Simulation of auto-relaxation rates for uniform and site selective labeled C6 in UTP (Fig. 2.3A) and C8 in ATP (Fig. 2.3B) predict discrepant rates between them. The internuclear distances used for the simulation of dipolar effect were calculated from the structural coordinates. The distances for $^{1}H_6$ – $^{13}C_6$, $^{13}C_6$ – $^{13}C_5$, $^{13}C_6$ – $^{13}C_4$, $^{13}C_6$ – $^{13}C_2$, $^{13}C_6$ – $^{15}N_1$, $^{13}C_6$ – $^{15}N_3$, $^{1}H_8$ – $^{13}C_8$, $^{13}C_8$ – $^{13}C_6$, $^{13}C_8$ – $^{13}C_5$,

$^{13}C_8$ – $^{13}C_4$, $^{13}C_8$ – $^{15}N_7$, and $^{13}C_8$ – $^{15}N_9$ are 1.081 Å, 1.349 Å, 2.419 Å, 2.408 Å, 1.371 Å, 2.658 Å, 1.079 Å, 3.514 Å, 2.118 Å, 2.197 Å, 1.329 Å and 1.372 Å, respectively. CSA values, derived from DFT calculation, were taken into account for the simulation as well (details are described in supplementary materials). For UTP, no notable difference exists between $R_{1,C_6}{}^{u}$ and $R_{1,C_6}{}^{s}$ when τ_C is less than ~ 2 ns, (Fig. 2.3A). However, this difference becomes significant with increasing τ_C due to increased contribution from R_{1,C_6,C_5} (Fig. 2.3A). Similarly, the calculated

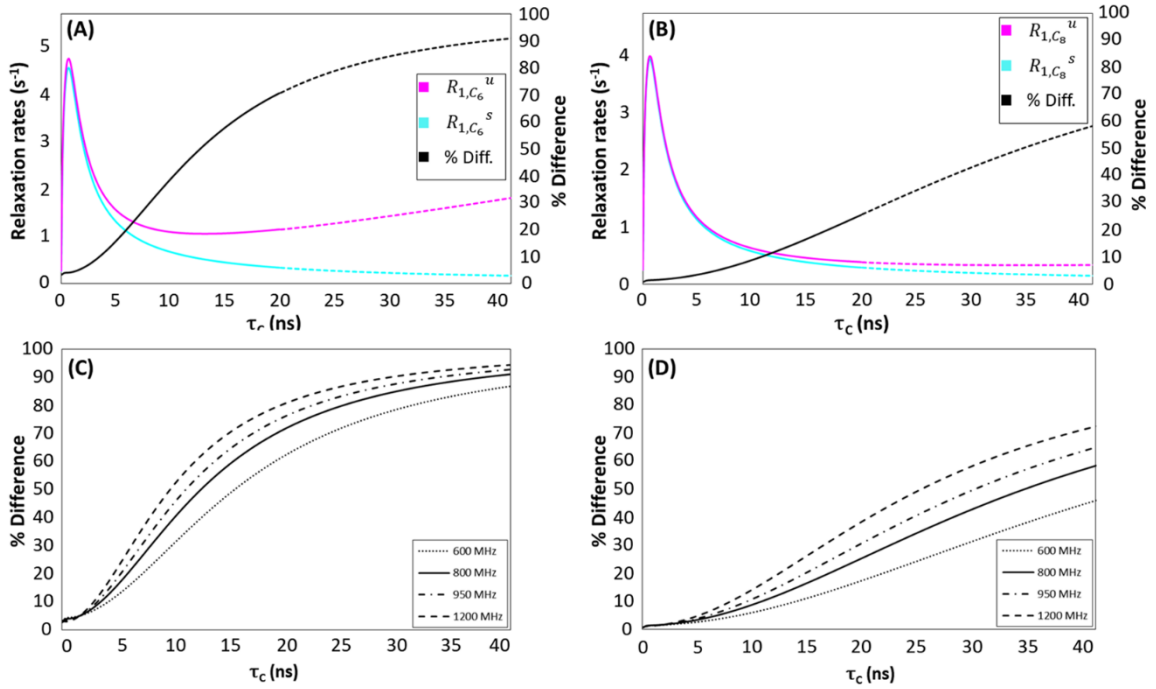


Figure 2. 3. Simulations of auto-relaxation rates (R_{1,C_K}) for uniform- and site-selective labeled RNAs and % differences as a function of correlation time (τ_C). **(A)** Simulation of R_{1,C_6} for U-labeled RNAs at 800 MHz. $R_{1,C_6}{}^{u}$, $R_{1,C_6}{}^{s}$ and % differences correspond to magenta, cyan and black lines, respectively. When τ_C is less than ~ 2 ns, there is no notable difference between $R_{1,C_6}{}^{u}$ and $R_{1,C_6}{}^{s}$. However, this difference becomes significant with increasing τ_C due to increasing contribution from R_{1,C_6,C_5}. **(B)** Simulation of R_{1,C_8} for A-labeled RNAs at 800 MHz. $R_{1,C_8}{}^{u}$, $R_{1,C_8}{}^{s}$ and % differences correspond to magenta, cyan and black lines, respectively. $R_{1,C_8}{}^{u}$ and $R_{1,C_8}{}^{s}$ values are comparable for τ_C < 10 ns with the differences below ~ 10 % (solid lines). When τ_C > 20 ns, the simulation predicts greater differences reaching ~ 60% at 40 ns (dotted lines). **(C)** Simulation of % differences between $R_{1,C_6}{}^{u}$ and $R_{1,C_6}{}^{s}$ for U-labeled RNAs at different frequencies. **(D)** Simulation of % differences between $R_{1,C_8}{}^{u}$ and $R_{1,C_8}{}^{s}$ for A-labeled RNAs at different frequencies. For both R_{1,C_6} and R_{1,C_8}, greater % differences are predicted with higher frequencies.

difference between $R_{1,C_8}{}^{u}$ and $R_{1,C_8}{}^{s}$ values are negligibly small for τ_C < 5 ns with the differences below ~ 10 % (Fig. 2.3B, solid lines). This is expected since C8 has no directly bonded carbon-carbon coupling partners. When τ_C > 20 ns, the simulation predicts even greater differences for both UTP and ATP, reaching ~ 90% and ~ 60% at 40 ns for C8 and C6, respectively (Fig. 2.3A and B, dotted lines). It is noteworthy that purine C8, with no directly attached carbon, also have increased dipolar contributions from neighboring ^{13}C labeled carbons. This indicates the R_{1,C_8,C_i} term, though negligible for τ_C < 5ns, leads to larger difference between $R_{1,C_8}{}^{u}$ and $R_{1,C_8}{}^{s}$ for τ_C > 20ns. In addition, the simulations of R_1 at different magnetic fields showed greater % differences at higher fields for both UTP (Fig. 2.3C) and ATP (Fig. 2.3D). This implies that while NMR experiments at higher frequencies may provide results with better resolution and sensitivity, these higher fields may paradoxically hinder accurate measurements of R_1 for uniform labeled RNA sample.

2.2 Dipolar coupling effects on ^{13}C longitudinal relaxation (R_1)

To see how the contribution of neighboring nuclei on R_1 rates bear out experimentally, uniformly labeled RNAs and site-selectively labeled (Fig. 2.1B) RNA samples of three different sizes (14, 29 and 61 nt) were prepared to examine how the contribution on R_1 rates varies with the size of the macromolecules. The measurements were performed at 25 °C and 800 MHz for all RNAs. In agreement with the simulations, there were no notable differences in R_1 of uniformly labeled ($R1^u$) and site-selectively labeled ($R1^s$) samples for the 14 nt RNA, indicating

insignificant carbon-carbon coupling effects for both base and ribose carbon atoms (Fig. 2.4A). The average difference in R_1 rates were < 5 % (Table 2.1) suggesting little to no dipolar coupling effect on R_1 relaxation for small RNAs. R_1^u and R_1^s for C8 position on purine for 29 nt RNA were also comparable resulting in an average difference of ~2 % (Fig. 2.4B, Table 2.1). The C6 position of pyrimidine and C1' position of ribose, however, have directly bonded carbon coupling partners and this effect can be observed by the significant differences on R_1^u and R_1^s (Fig. 2.4B). The average difference was ~13 % for C6 and ~ 16 % for C1' showing increased contribution of the attached carbon compared to the 14 nt RNA (Table 2.1). For 61 nt RNA, the % differences were even greater (Fig. 2.4C, Table 1) indicating increased dipolar coupling effects. These results show increasing contribution of the adjacent carbon with increasing size of RNA supporting the prediction from eq

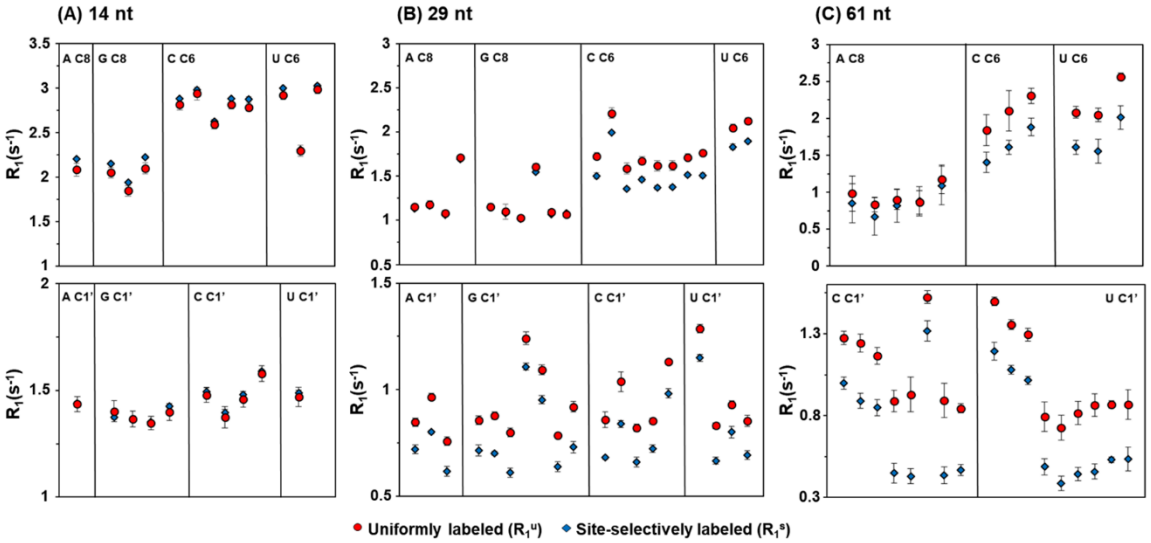


Figure 2. 4. R_1 rates of uniformly labeled sample (R_1^u) and site-selectively labeled sample (R_1^s) at 800 MHz. (A) For 14 nt RNA, there were no notable differences between R_1^u and R_1^s indicating no significant carbon-carbon coupling effects for both base and ribose. **(B)** R_1^u and R_1^s for C8 position on purine for 29 nt RNA were comparable which was expected since C8 is fairly isolated with no adjacent carbon-carbon coupling partners. The C6 position of pyrimidine and C1' position of ribose, however, have adjacent carbon-carbon coupling partners and this effect is reflected in the significant differences on R_1^u and R_1^s. **(C)** For 61 nt RNA, the discrepancies between R_1^u and R_1^s were even greater for C8, C6 and C1' indicating increased coupling effects.

RNA Size (nt)	^{13}C Position	% Difference in R_1		
		Maximum	Minimum	Average
14	C8	7.5	4.9	6.3
	C6	3.6	0.3	2.0
	C1'	3.9	0.1	1.5
29	C8	6.8	0.1	2.2
	C6	15.8	10.7	13.1
	C1'	23.6	5.4	15.7
61	C8	19.6	1.7	10.4
	C6	23.9	18.4	22.3
	C1'	54.0	13.6	37.8

Table 2. 1. % Differences between R_1^u and R_1^s for varying size of RNAs at 800 MHz.

% differences ($= \frac{|{R_1}^u - {R_1}^s|}{{R_1}^u} \times 100$ %) were calculated for multiple peaks and averaged. Maximum and minimum values shows the range of observed values.

2.1 and simulations (Fig. 2.3). The τ_C values range between 0 - 15 ns for the RNA sizes (14, 29, 61 nt) used for the measurements. In this τ_C range, the simulated differences between ${R_{1,C_6}}^u$ and ${R_{1,C_6}}^s$ were within 0 – 25 % (Fig. 2.3A, solid lines). This % difference is in agreement with the experimentally observed % differences (Table 2.1). For the same τ_C range, calculated differences between ${R_{1,C_8}}^u$ and ${R_{1,C_8}}^s$ are between 0 – 15 % (Fig. 2.3B, solid lines). Again, these calculated values are in agreement with the observed results in Figure 2.4 and Table 2.1.

These results are also consistent with the observation that for small nucleic acids (τ_C < 2 ns) the contribution of carbon-carbon dipolar coupling interaction to the relaxation is negligible compared to the proton-carbon interaction (ρ_{CC}/ρ_{CH} < 0. 06) at 800 MHz proton frequency. However, for larger nucleic acids the ^{13}C-^{13}C

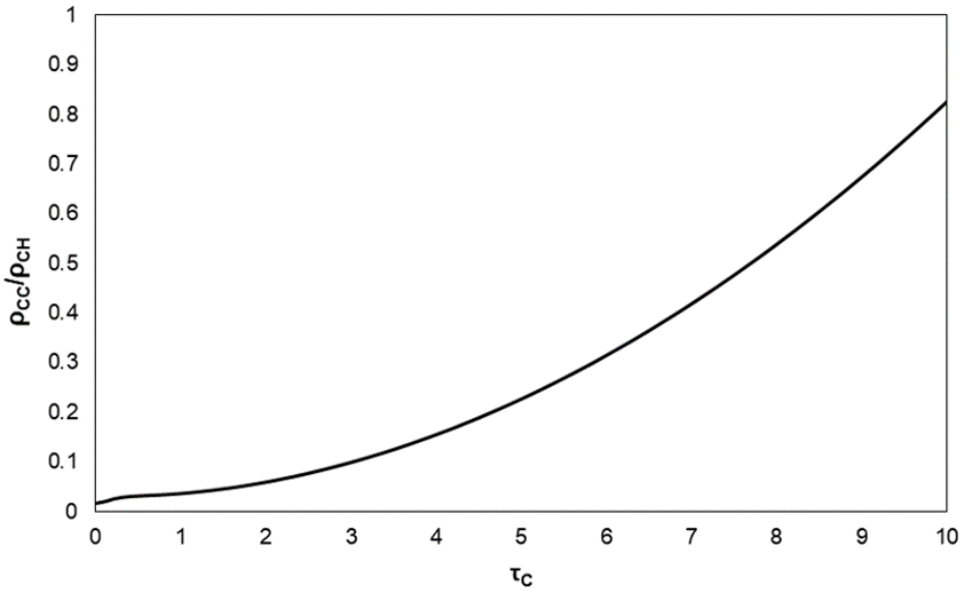


Figure 2. 5. Simulation of ρ_{CC}/ρ_{CH} as a function of correlation time (τ_C) at 800 MHz. The contribution of carbon-carbon dipolar interaction to the relaxation is negligible compared to the proton-carbon interaction (ρ_{CC}/ρ_{CH} < 0. 06) when τ_C < 2 ns. However, the ^{13}C-^{13}C contribution is no longer negligible when τ_C > 2 ns. For instance, ρ_{CC}/ρ_{CH} = 0.83 for τ_C = 10 ns.

contribution is no longer negligible (ρ_{CC}/ρ_{CH} = 0.83 for τ_C = 10 ns) (Fig. 2.5). Overall, the observed results and simulations suggest necessity of more careful analysis of R_1 rates for the uniform labeled RNAs, especially when the size is large and experiments are conducted at very high fields.

3. Steady-State ^{1}H-^{13}C hNOE

3.1 Steady-state ^{1}H-^{13}C hNOE measurement

Accurate measurements of heteronuclear Overhauser effect (hNOE) parameters can also provide meaningful dynamics information on biomolecules when combined with R_1 and R_2.[112] A value of NOE enhancement, η (= NOE-1), is defined as the difference between the peak intensity in the irradiated spectrum

(I_{sat}) and the reference non-irradiated spectrum (I_{ref}) scaled by the signal intensity from the reference experiment ($\eta = (I_{sat} - I_{ref})/I_{ref}$). The steady-state NOE enhancement (η_{SS}) of the spin K is given by[127, 128]:

$$\eta_{SS} = \frac{\gamma_i \sigma_{i,K}}{\gamma_X R_{1,K}} \qquad (2.8)$$

where $\sigma_{i,K}$ is cross-relaxation rate between spin i and K, $R_{1,K}$ is auto-relaxation rate of spin K and $\gamma_{i(H)}$ is gyromagnetic ratio of spin i (K). For this experiment, proton spins were saturated using hard pulses and the NOE enhancement of C1' on ribose is monitored. As shown in eq 2.8, auto-relaxation rate ($R_{1,C_1'}$) of C1' and cross relaxation ($\sigma_{C_1',H_1'}$) between C1' and H1' determines the NOE enhancement (η_{SS}) of the C1'. Table 2.2 shows the results from the steady-state ^{1}H-^{13}C hNOE measurements on ribose C1' of a 29 nt RNA. As shown in the Table 2.2, uniformly labeled RNA (NOE^u) had higher NOE values compared to site-selectively labeled RNA (NOE^s). These results are also in agreement with the earlier protein work of Kay and co-workers[116] which showed increased NOE for uniformly labeled alanine compared to singly labeled alanine when τ_C is greater than ~ 6 ns. Similarly,

Peak	NOE^u	NOE^s
1	1.48 ± 0.08	1.39 ± 0.01
2	1.32 ± 0.01	1.25 ± 0.01
3	1.27 ± 0.02	1.19 ± 0.01
4	1.26 ± 0.02	1.17 ± 0.03
5	1.28 ± 0.01	1.20 ± 0.01
6	1.38 ± 0.02	1.29 ± 0.03

Table 2. 2. Steady-state ^{1}H-^{13}C hNOE measurements on ribose C1' of a 29 nt RNA.

NOE^u and NOE^s represent observed NOE values for uniformly- and site-selectively labeled sample, respectively.

Engelke and Rüterjans showed the contribution of carbon-carbon interaction in protein hNOE measurement and speculated greater effect with larger rotational correlation time.[132]

Unlike site-selectively labeled ribose, in which only the C1' is isotopically enriched, uniformly labeled ribose has both C1' and C2' labeled (Fig. 2.1). Since both C1' and C2' has a proton attached, saturation of proton spins enhances the longitudinal ^{13}C-magnetization of both carbon positions. Enhanced intensity of the C2' can then be transferred to the C1' spin and further increase the intensity of C1' as a result of cross relaxation. This effect becomes more significant with greater τ_C as cross-relaxation rate ($\sigma_{C_1',C_2'}$) for C1' – C2' increases with increasing correlation time. In contrast, C2' in site-selectively labeled ribose is unlabeled resulting in smaller NOE^s compared to NOE^u. These NOE results suggest that the analysis of steady-state NOE values for the uniformly labeled RNA should be done with more care due to the cross-relaxation effect from neighboring carbon spins for C1', pyrimidine C5 and C6 carbons. Until now such measurements were preferentially omitted for nucleic acids[120-122]; our selective labels[126, 133] make these readily accessible and will be useful for mapping the spectral density functions for RNA systems.[113]

3.2 Effect of selective pulse on steady-state ^{1}H-^{13}C hNOE measurement

Additionally, we applied selective proton saturation pulse for hNOE measurements to test whether such pulses can reduce the contribution from carbon-carbon interaction for uniformly labeled sample. For site-selectively labeled sample, NOE values with the selective proton saturation pulse and the 180° hard pulse were comparable indicating negligible effect of selective pulse when there is no adjacent carbon (Table 2.3). However, when the same selective pulse was applied to the uniformly labeled sample, the NOE values were reduced to values close to that of NOE^{s} (Table 2.3). This result suggests that the selective proton saturation pulse can be applied as an alternative method to minimize the effect of

	Selectively labeled sample			**Uniformly labeled sample**	
Carbon inversion pulse Proton saturation pulse	Hard Selective	Hard Hard	Selective Hard	Selective Selective	Selective Hard
Peak	NOE	NOE	NOE	NOE	NOE
1	1.20 ± 0.01	1.20 ± 0.02	1.21 ± 0.01	1.20 ± 0.01	1.26 ± 0.01
2	1.23 ± 0.02	1.25 ± 0.01	1.25 ± 0.01	1.22 ± 0.01	1.32 ± 0.01
3	1.17 ± 0.02	1.19 ± 0.01	1.18 ± 0.02	1.17 ± 0.01	1.25 ± 0.02
4	1.23 ± 0.02	1.23 ± 0.03	1.24 ± 0.01	1.22 ± 0.04	1.29 ± 0.01
5	1.18 ± 0.02	1.19 ± 0.01	1.20 ± 0.02	1.18 ± 0.02	1.27 ± 0.02
6	1.32 ± 0.03	1.33 ± 0.01	1.35 ± 0.02	1.27 ± 0.01	1.40 ± 0.02
7	1.28 ± 0.01	1.32 ± 0.04	1.29 ± 0.01	1.20 ± 0.04	1.35 ± 0.06
8	1.18 ± 0.02	1.19 ± 0.01	1.20 ± 0.01	1.16 ± 0.03	1.24 ± 0.02
9	1.21 ± 0.01	1.20 ± 0.01	1.22 ± 0.02	1.21 ± 0.01	1.28 ± 0.01
10	1.17 ± 0.05	1.20 ± 0.03	1.21 ± 0.02	1.17 ± 0.01	1.27 ± 0.03
11	1.14 ± 0.04	1.18 ± 0.01	1.19 ± 0.02	1.16 ± 0.01	1.24 ± 0.01
12	1.16 ± 0.05	1.18 ± 0.03	1.20 ± 0.02	1.13 ± 0.04	1.24 ± 0.02
13	1.28 ± 0.01	1.30 ± 0.03	1.28 ± 0.01	1.26 ± 0.06	1.38 ± 0.02
14	1.31 ± 0.02	1.33 ± 0.02	1.34 ± 0.01	1.25 ± 0.01	1.39 ± 0.02
15	1.14 ± 0.01	1.17 ± 0.03	1.17 ± 0.01	1.15 ± 0.01	1.22 ± 0.01
16	1.20 ± 0.01	1.21 ± 0.03	1.22 ± 0.01	1.19 ± 0.05	1.26 ± 0.01
17	1.18 ± 0.01	1.17 ± 0.03	1.19 ± 0.01	1.18 ± 0.04	1.26 ± 0.02

Table 2. 3 Effect of selective pulse on hNOE measurement. NOE values were measured with different pulse applied for carbon inversion and proton saturation for uniformly- and site-selectively labeled RNA.

carbon-carbon coupling on hNOE measurement when uniformly labeled sample is used.

4. ^{13}C $R_{1\rho}$ Relaxation Measurement

The transverse relaxation rate constant (R_2) is a measure of the rate at which transverse magnetization decays to zero. Unlike R_1, R_2 provides information on the chemical exchange process in biomolecules. Typically, Carr-Parcell-Meiboom-Gill (CPMG) pulse sequence is used to measure R_2 for proteins.[117] However, this is problematic for nucleic acids due to echo-modulation caused by the large J_{CC} (38-68 Hz) couplings during the ^{13}C relaxation delay.[116, 134] To avoid this problem, $R_{1\rho}$ is used as an alternative method of measuring R_2 of biomolecules. In an $R_{1\rho}$ experiment, a radio frequency field pulse spin-locks the magnetization in the rotating frame and relaxation rate constant along the effective field is measured. The observed $R_{1\rho}$ rates contain contribution from both longitudinal and transverse relaxation and the information on R_2 is easily extracted according to $R_{1\rho} = R_1 cos^2\theta + R_2 sin^2\theta$ in which $\theta = tan^{-1}\left(\frac{\omega_1}{\Omega}\right)$, ω_1 is the strength of the B_1 spin-lock field (Hz) and Ω is the resonance offset (Hz) from the spinlock carrier frequency. In contrast to R_1 simulation, which showed significant difference between R_1^u and R_1^s (Fig. 2.2), simulation of R_2 showed negligible difference between uniform (R_2^u) and site-selective (R_2^s) labeled sample (Fig. 2.6). The maximum difference predicted was less than 5 % for both purine and pyrimidine at all proton frequencies.

Therefore, we expected minimum differences in experimentally determined R_2^u and R_2^s as well.

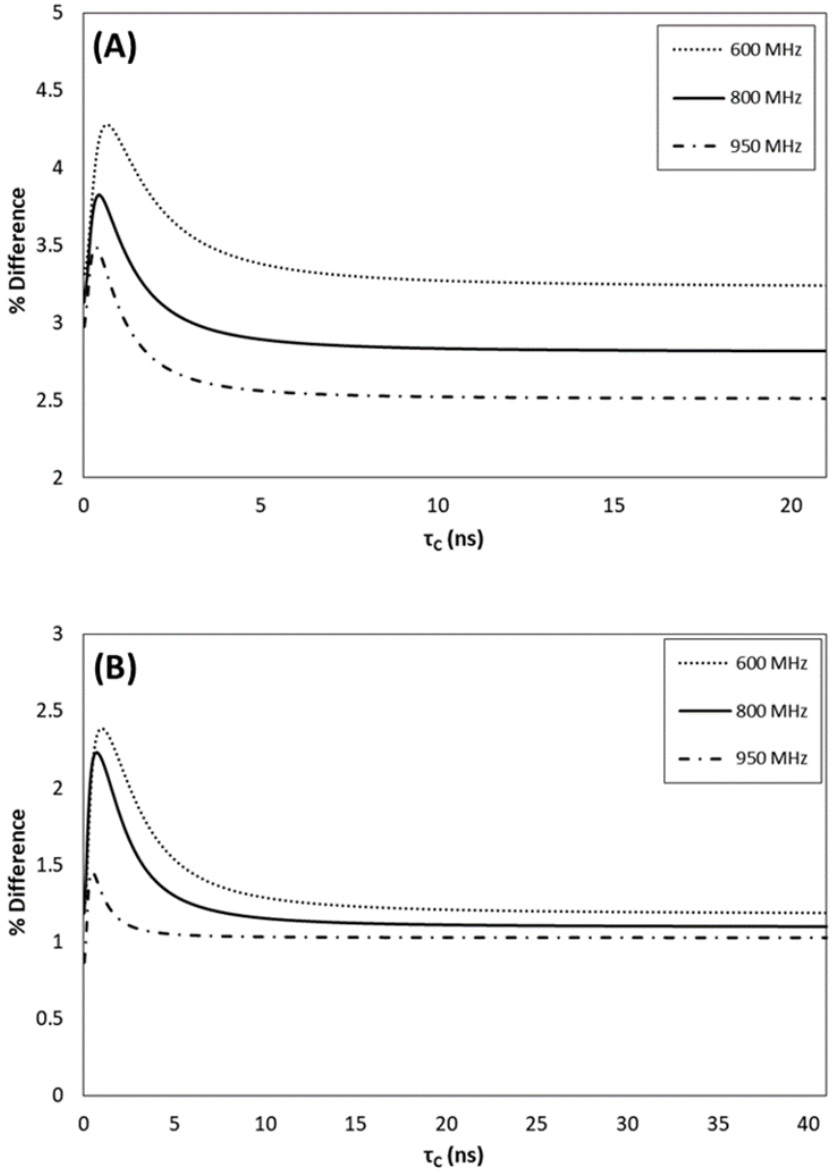


Figure 2. 6. Simulation of % differences between R_2^u and R_2^s for **(A)** UTP and **(B)** ATP. At all frequencies, the differences were less than ~ 5%. The internuclear distances used for simulation were same as the values used for the simulation in Figure 2.4.

R_2 relaxation rates and $R_{2,CSA}$ were given by:

$$R_{2,C_K,i} = \frac{1}{20} {D_{C_K,i}}^2 \left[3J\left(\omega_{C_K}\right) + 6J\left(\omega_{C_K} + \omega_i\right) + J\left(\omega_{C_K} - \omega_i\right) + 6J(\omega_i) + 4J(0)\right]$$

$$R_{2,CSA} = \frac{1}{20} {C_{C_K}}^2 \left[3J\left(\omega_{C_K}\right) + 4J(0)\right]$$

For $R_{1\rho}$ measurements, the potential errors caused by carbon – carbon interaction may come from the Hartmann – Hahn transfer during the spin-lock period for spins that have comparable chemical shifts. Work by Kay and co-workers indicated reasonable agreement between $R_{1\rho}$ rates for singly- and uniformly-labeled alanine indicating Hartmann – Hahn effects are negligible, except for Ser and Thr.[116] In addition, Hoogstraten and co-workers compared $R_{1\rho}$ rates of uniform and alternate-site labeled rAMP and observed agreement between the measurements under some combinations of resonance, spin-lock power and τ_C.[118]

In order to reduce the Hartmann – Hahn transfer between scalar coupled spins, the $R_{1\rho}$ pulse sequences[121] used for the experiment include selective spin-locks and therefore, minimum discrepancy was expected between $R_{1\rho}$ rate of uniformly labeled sample and site-selectively labeled sample. To investigate the effect of neighboring carbon, $R_{1\rho}$ experiments were performed at 800 MHz with 1.9 kHz B_1 field on-resonance for a 29 nt RNA and the R_2 rates were calculated as described earlier. The observed results showed a good agreement between R_2^u and R_2^s for C8 positions giving average difference of ~ 4% (Fig. 2.7). For C6 on pyrimidines and C1' on ribose, the average differences were ~7 % which is slightly higher than that of C8 (Fig. 2.7). Compared to the differences observed between R_1^u and R_1^s (Fig. 2.2, Table 2.1), R_2 measurements showed relatively good agreement between uniformly- and site-selectively labeled samples (Fig. 2.7). Comparable results were also observed for the R_2 rates of 61 nt RNA indicating that the increasing τ_C has no significant effect on the R_2 measurement (data not

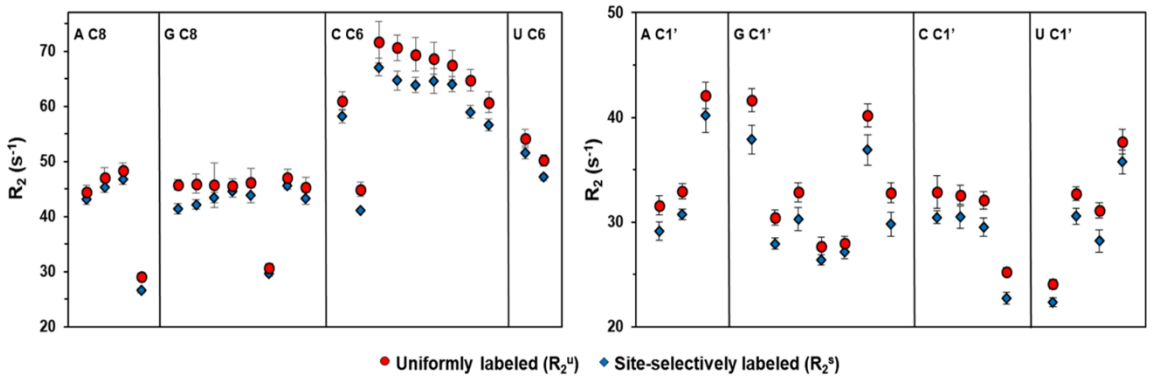


Figure 2. 7. R_2 rates of uniformly labeled sample (R_2^u) and site-selectively labeled sample (R_2^s) at 800 MHz. $R_{1\rho}$ experiments were performed at 1.9 kHz B_1 field on-resonance for a 29 nt RNA and the R_2 rates were extracted according to $R_{1\rho} = R_1 cos^2\theta + R_2 sin^2\theta$. The observed results showed a good agreement between R_2^u and R_2^s for C8 positions giving average difference of ~ 4%. For C6 on pyrimidines and C1' on ribose, the average differences were ~7 % which is slightly higher than that of C8. Compared to the discrepancies observed between R_1^u and R_1^s, R_2 measurements showed relatively good agreement between uniformly- and site-selectively labeled samples.

shown). These results are compatible with the R_2 simulation showing small difference between R_2^u and R_2^s (Fig. 2.6). The simulation also shows increasing τ_C has no notable effect on the % differences between R_2^u and R_2^s. In addition, the simulation predicts smaller % differences with higher frequencies, suggesting more accurate R_2 measurement at higher magnetic field (Fig. 2.6).

The observed results also suggest minimum Hartmann – Hahn transfer effect in $R_{1\rho}$ measurement for uniformly labeled samples. The maximum amount of net Hartmann – Hahn transfer between two spins 1 and 2 can be given by following equation[135]:

$$R_{1,2}^{max} = \frac{\left[{J_{1,2}}^2\{1+\cos(\theta_1-\theta_2)\}^2\right]}{\left[4\left(\nu_{1,eff}-\nu_{2,eff}\right)^2+{J_{1,2}}^2\{1+\cos(\theta_1-\theta_2)\}^2\right]} \qquad (2.9)$$

where $J_{1,2}$ is the coupling constants between two spins, θ_i is the spin-lock angle of spin i defined as described above and $\nu_{i,eff}$ is the effective field-strength on spin

i given by $\sqrt{\omega^2 + \Omega_i{}^2}$. Eq 2.9 predicts low transfer efficiency when the two spins have large difference in effective field strengths and relatively small coupling constants. Calculating maximum Hartmann – Hahn transfer for C1' and C2' using 43 Hz for a coupling constant, 1.9 kHz for a spin-lock field strength and 1.7 kHz for a difference in effective field results in maximum transfer efficiency < 1 ‰. Similar outcomes are observed when calculated for C6 and C5 supporting the experimental results showing minimum effect from adjacent carbon for uniformly labeled samples. Together with the earlier works[116, 118, 121], these results suggest that the effect from the neighboring carbon can be safely ignored for the $R_{1\rho}$ measurements.

5. Materials and Methods

5.1 Chemicals and oligonucleotides

Uniformly labeled rNTPs were purchased from Cambridge Isotope Laboratories, Inc (MA), and site-selectively labeled rNTPs were synthesized in-house using previously reported chemo-enzymatic methods[125, 126] An outline of the synthesis is described in Fig. 2.1. Synthetic DNA templates and the RNA polymerase promoter sequences (5′-CTA ATA CGA CTC ACT ATA G-3′) were purchased from Integrated DNA Technologies, Inc (CA). Two nucleotides at the 5' end of synthetic DNA templates were modified with 2'-O-methyl group to reduce transcription heterigenety.[136] Modified T7 RNA polymerases[94] used for *in vitro*

transcription were expressed in *Escherichia coli* BL21(DE3) and purified with a Ni-chelating Sepharose column (GE Healthcare).

5.2 RNA sample preparation

RNA samples were synthesized by *in vitro* transcription using T7 RNA polymerases. Sequences are provided in the supplementary materials. The transcription conditions used buffer containing 40 mM Tris-HCl (pH 8.0), 1 mM spermidine, 10 mM dithiothreitol (DTT), 0.01 % Triton X-100, 300 nM DNA template, 300 nM promoter sequence, 2 units/ml of thermostable inorganic pyrophosphatase (New England BioLabs, MA) and 0.1 mg/ml T7 polymerases for 10 ml reaction volume. Concentration of rNTPs and Mg^{2+}, ranging between 5 – 20 mM, were optimized for each RNAs to obtain maximum yields. After 3-4 hr of incubation at 37 °C, RNAs were purified by phenol:choloform extraction, ethanol precipitation, and denaturing polyacrylamide gel electrophoresis followed by electroelution. Eluted samples were solvent exchanged and concentrated into an appropriate buffer (buffer conditions provided in supplementary materials) using 3K MWCO tube (Fisher Scientific, NH). The samples were then folded by heating up to 95 °C for 3 min and snap-cooled on ice for 15 min. Finally, the RNA samples were lyophilized and re-suspended in 300 µl of 99.8 % D_2O resulting in a final concentration of 0.5 – 1 mM.

5.3 NMR Measurements

NMR experiments were performed at 298.15 K on a 800 MHz Avance III Bruker spectrometer equipped with a triple resonance cryogenic probe and collected data were analyzed using Bruker TopSpin 3.2. TROSY-detected pulse sequences for the relaxation measurements were adapted from earlier works.[116, 121] For R_1 and $R_{1\rho}$ experiments for uniform labeled samples, selective pulses were applied for decoupling as described in the previous work.[121] Shape pulses used for the on-resonance inversion, on-resonance refocusing, and off-resonance inversion were Q3, RSNOB, and Iburp2, respectively. The pulse length for each pulses were 937.5 µs, 1000 µs, and 450 µs, respectively. The offset for Iburp2 was -8000 ppm. For site-selective labeled samples, hard 180° pulse was applied instead during the second INEPT period. Applying hard pulse instead of the selective pulse did not affect the relaxation rates (data not shown). For both uniform and site-selective samples, carrier was set to 89.2 ppm for C1' with a sweep width of 6.4 ppm. For C6/C8, the carrier was set to 137.8 ppm with a sweep width of 10.4 ppm. For all experiments, a recycle delay of 2.5 s was used. Two-dimensional experiments were performed with 10 delay points including one repeated point for error estimation by Monte-Carlo simulation. RELAXFIT MATLAB program[137] was used to fit the data and calculate the rates and errors. For steady-state hNOE measurements, 8 s period of proton saturation was applied. Two-dimensional TROSY spectra were collected with and without the proton saturation. Each experiments were performed twice for error calculation. Peak positions and intensities were analyzed using TopSpin 3.2 (Bruker) and NOE was calculated

($\eta = (I_{sat} - I_{ref})/I_{ref}$) using the intensities determined from each spectrum. Uncertainties in NOE values were calculated by propagating the error in the peak intensities.

5.4 Electronic structure calculations

Calculations were carried out on 1-methyl-uracil and 9-methyl adenine using geometries optimized at the MP2/ccpVTZ level of theory. Chemical shielding tensors were computed (with GIAO orbitals) at the Hartree-Fock and MP2 levels, as well as with a "pure" density functional (OLYP)[138] and a "hybrid" model (PBE0)[139] that includes 25% of Hartree-Fock exchange. The pcSseg-2 basis, which is optimized for chemical shielding calculations[140], was used for calculations of the shielding tensors. To test basis set convergence, calculations other than MP2 were repeated at the pcSseg-3 basis set level. All calculations used the Gaussian-16 program.[141] The calculated CSA parameters varied slightly from what has been reported from Bax and co-workers[142]. Simulated relaxation rates showed marginal differences when calculated with the CSA parameter from Bax and co-workers[142]. However, it did not affect the % difference of relaxation rates between uniform and selectively labeled samples.

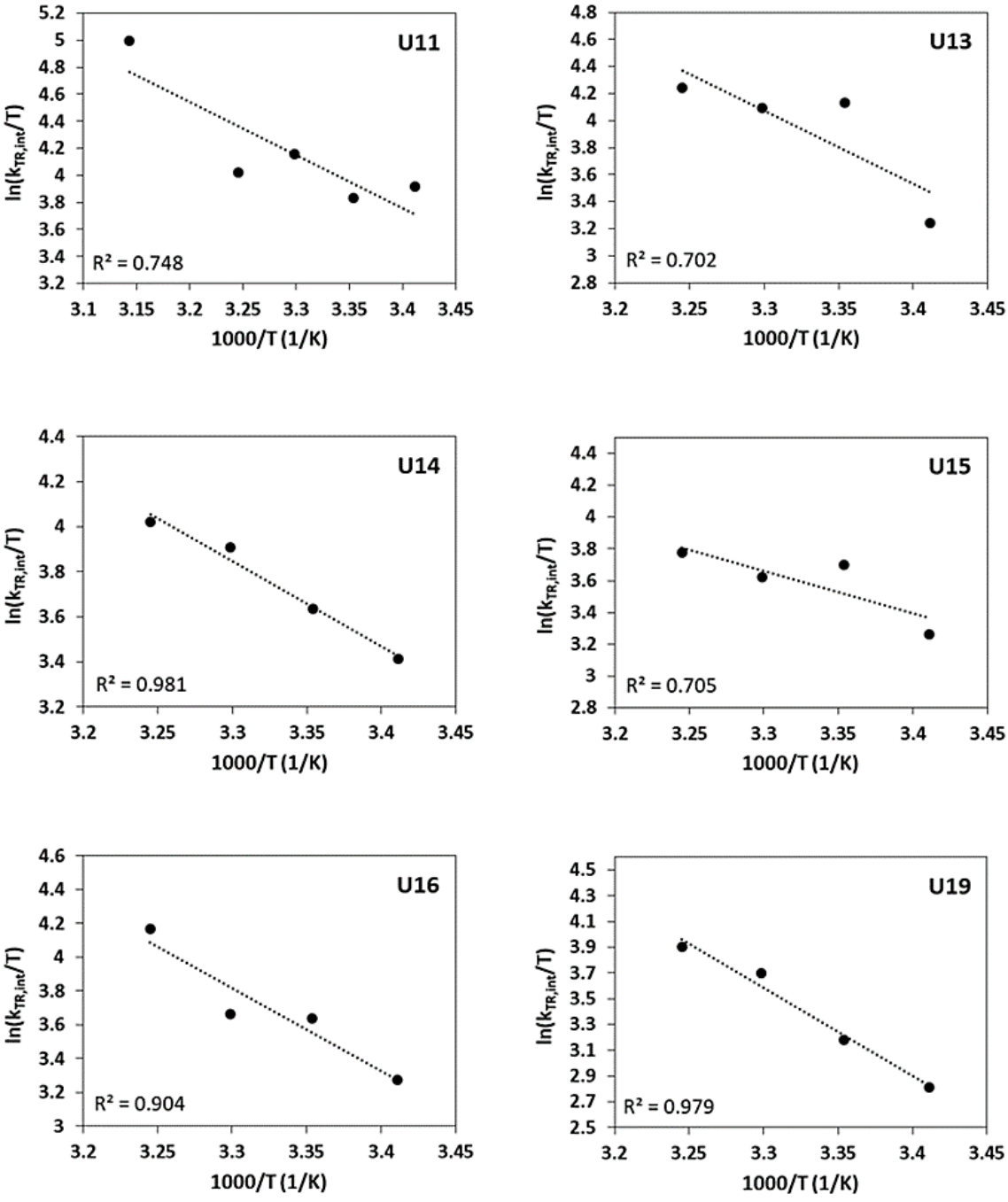
6. Conclusion

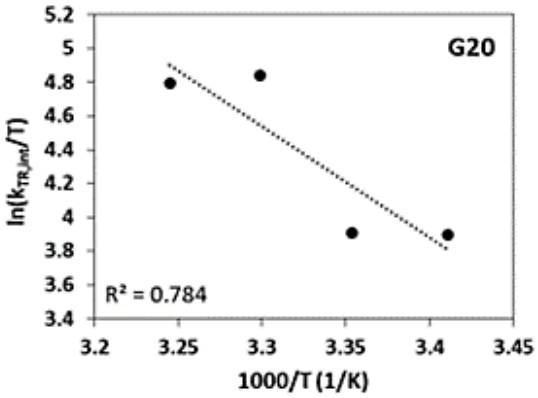
Although ^{13}C relaxation measurements provide a more complete coverage of RNA structural elements than ^{15}N measurements, a number of significant

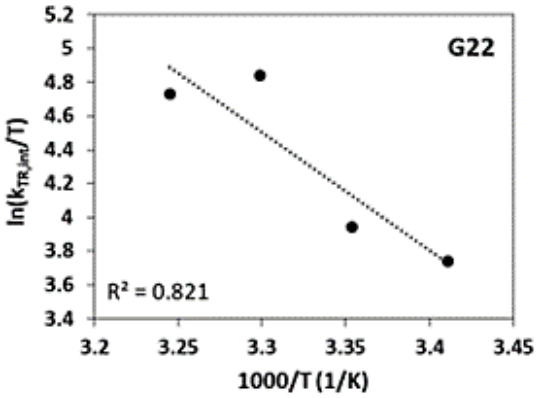
challenges render measurement and analysis of ^{13}C relaxation problematic. Foremost being the dipolar coupling network of interactions present in uniformly labeled RNA samples. We have therefore carried out dynamic studies, via and simulations and experimental NMR measurements on RNAs of varying sizes with different labeling patterns, to investigate the effect of adjacent carbon on the relaxation measurements. Not unexpectedly, simulations and measurements of R_1 of pyrimidine C6 and ribose C1' positions indicate large discrepancies exists between rates extracted using uniformly- and site-selectively labeled samples. Predictions, confirmed experimentally, suggest the carbon-carbon coupling effects from directly attached carbon hinder accurate dynamics measurements in uniformly labeled RNA. Furthermore, the contribution of the neighboring carbon on R_1 rates increase with increasing τ_C, necessitating more careful analysis of relaxation data with larger size of RNAs when the sample is uniformly labeled. Unexpectedly, simulations and measurements of R_1 on purine C8 position show larger discrepancy between R_1^u and R_1^s with increasing τ_C, indicating greater carbon-carbon coupling effects in large RNAs. This contrasts with prevailing assumptions that purine C8 forms an isolated spin system even within the network of carbon couplings found in uniformly labeled nucleobases. Moreover, the simulations suggest greater contribution on R_1 rates at higher magnetic field strengths. On the other hand, R_2^u and R_2^s of C6 and C1' extracted from $R_{1\rho}$ measurements showed relatively good agreement suggesting no dramatic effect from their attached carbon on $R_{1\rho}$ measurement.

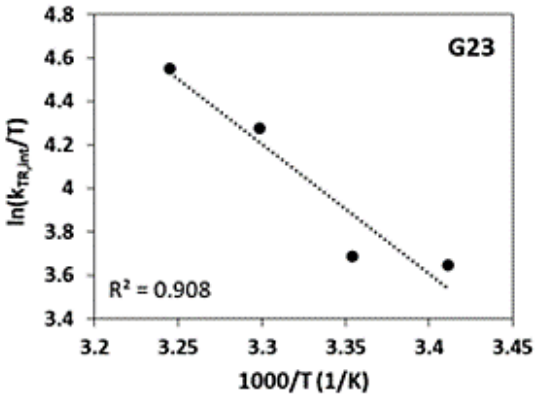
Appendices

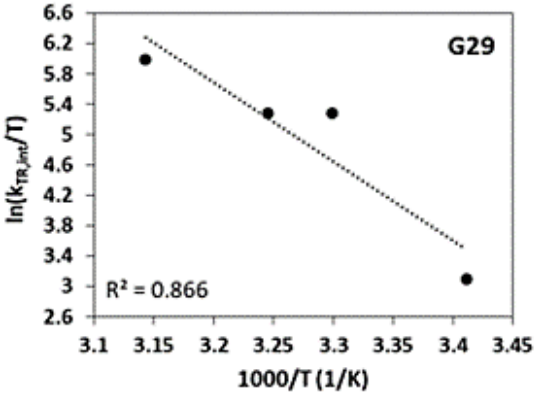
Figure A.1. Eyring plots of intrinsic catalysis rates $k_{tr,int}$ of residues in MALAT1 triple helix

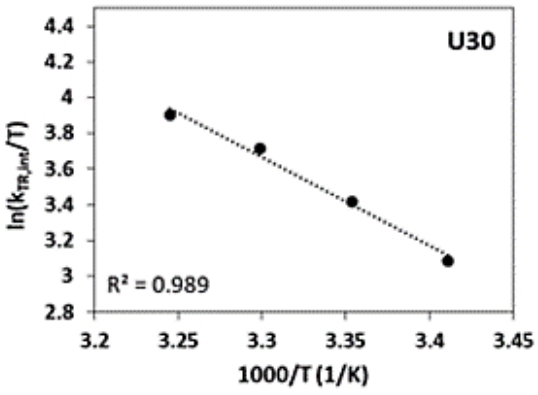


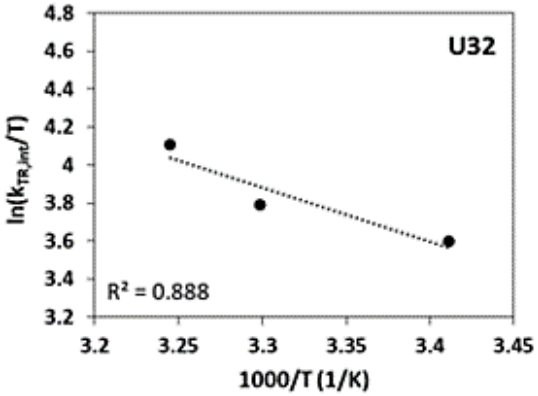

G20
R² = 0.784
ln(k_TR,int/T)
1000/T (1/K)

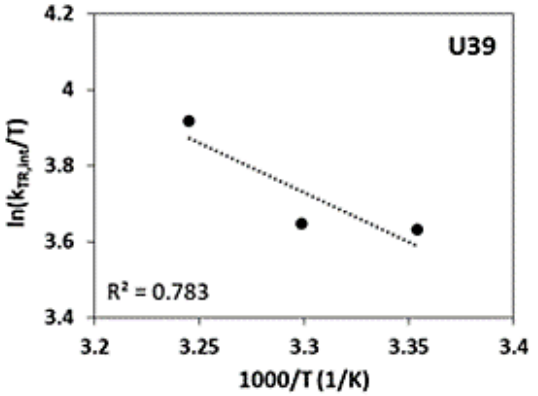

G22
R² = 0.821
ln(k_TR,int/T)
1000/T (1/K)

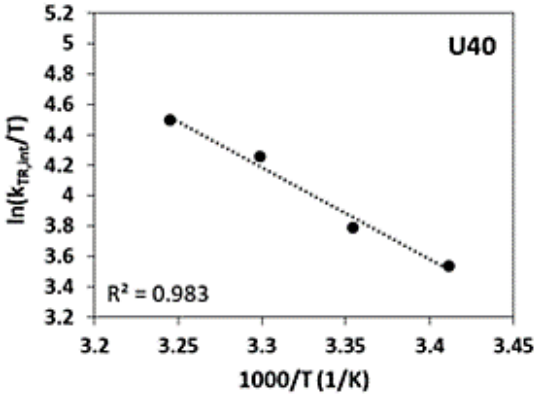

G23
R² = 0.908
ln(k_TR,int/T)
1000/T (1/K)

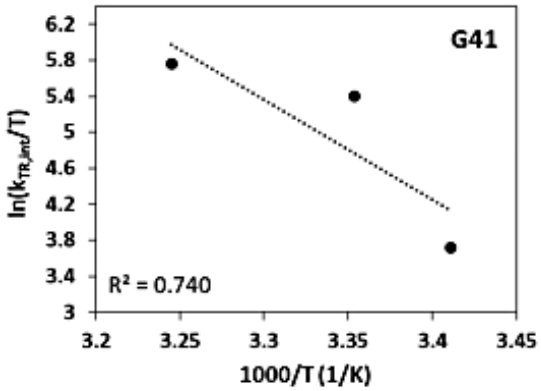

G29
R² = 0.866
ln(k_TR,int/T)
1000/T (1/K)

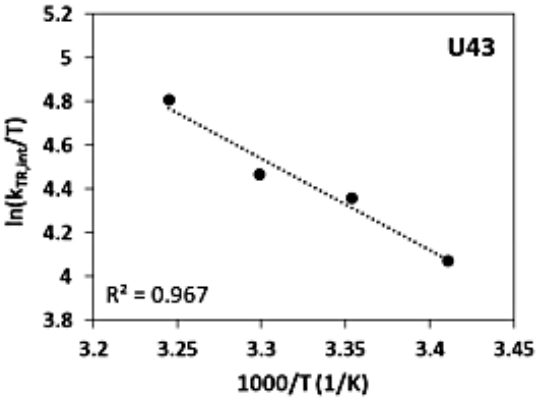

U30
R² = 0.989
ln(k_TR,int/T)
1000/T (1/K)

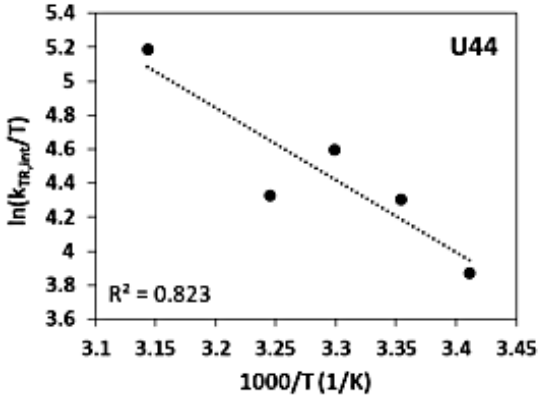

U32
R² = 0.888
ln(k_TR,int/T)
1000/T (1/K)


U39
R² = 0.783
ln(k_TR,int/T)
1000/T (1/K)


U40
R² = 0.983
ln(k_TR,int/T)
1000/T (1/K)


G41
R² = 0.740
ln(k_TR,int/T)
1000/T (1/K)


U43
R² = 0.967
ln(k_TR,int/T)
1000/T (1/K)


U44
R² = 0.823
ln(k_TR,int/T)
1000/T (1/K)

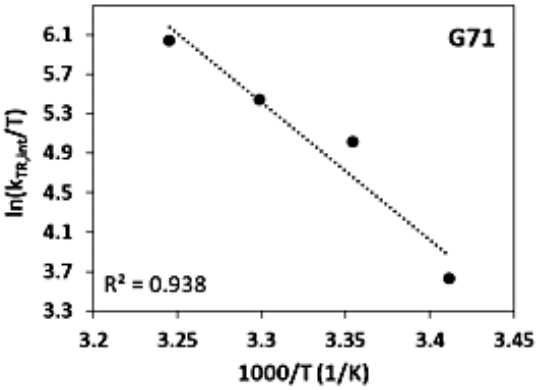
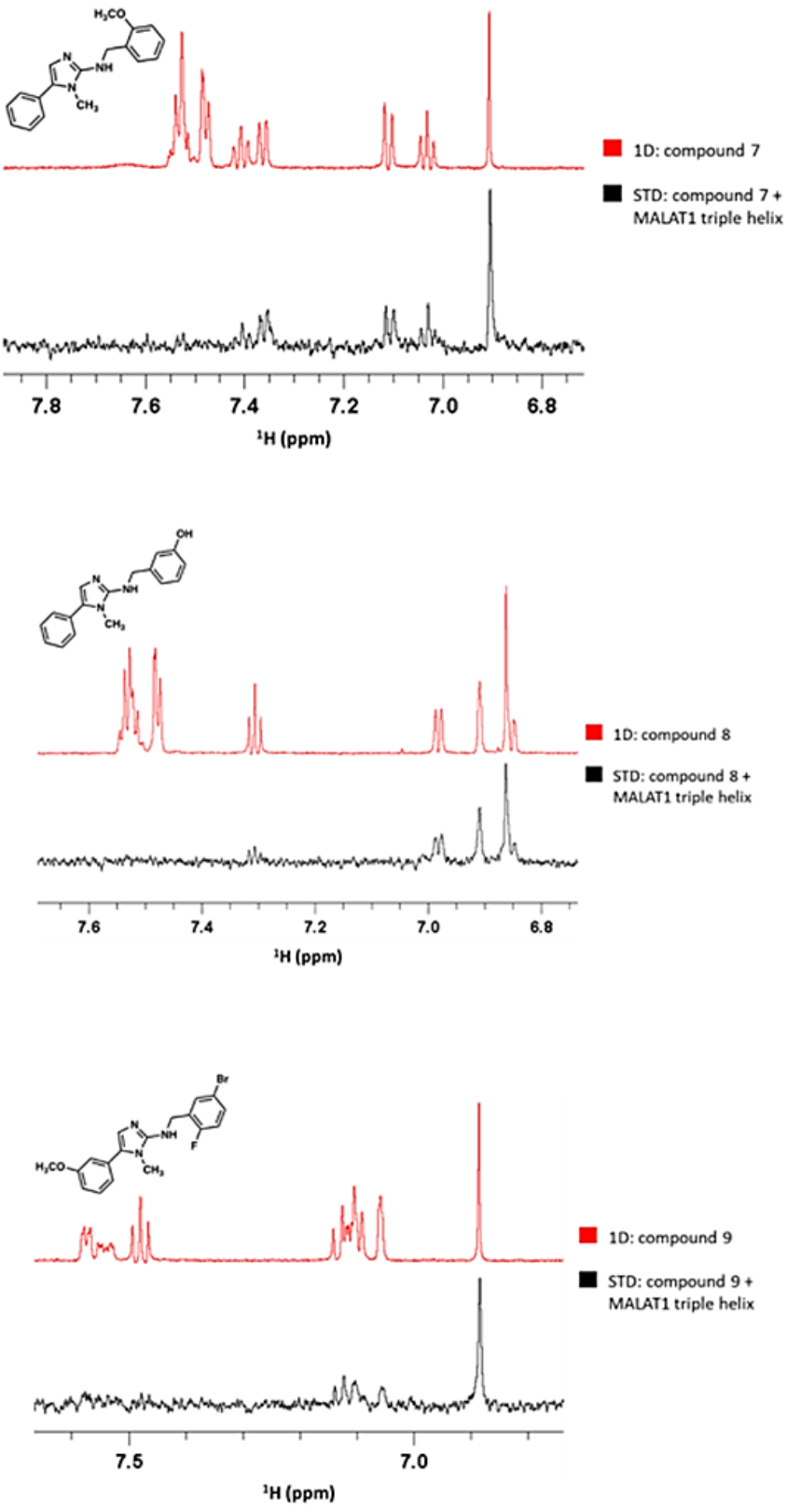
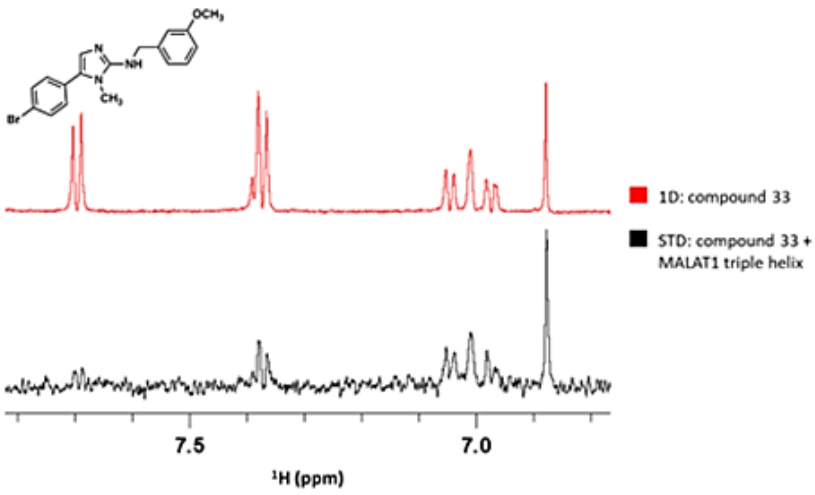
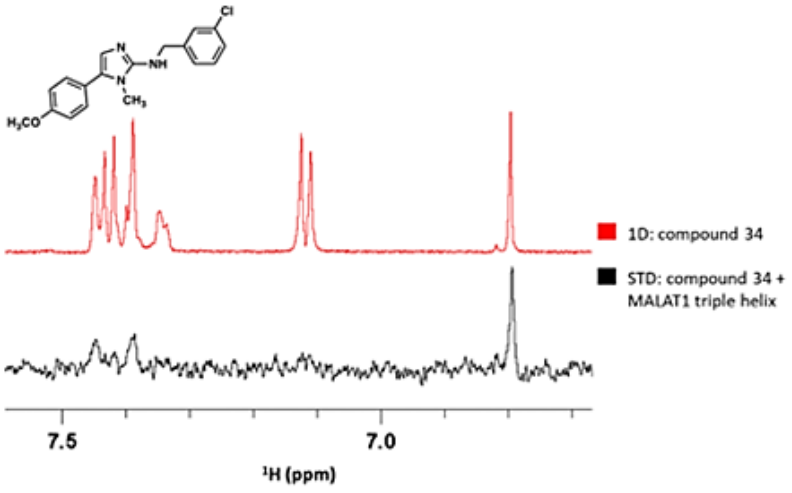
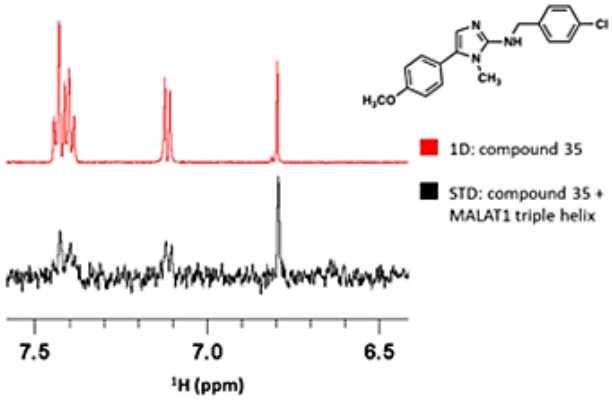

G71
R² = 0.938
ln(k_TR,int/T)
1000/T (1/K)

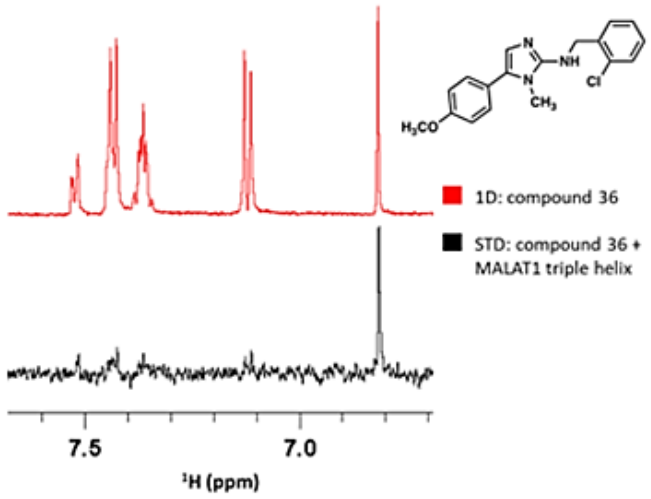
Figure A.2. STD results for compound 5 analogs

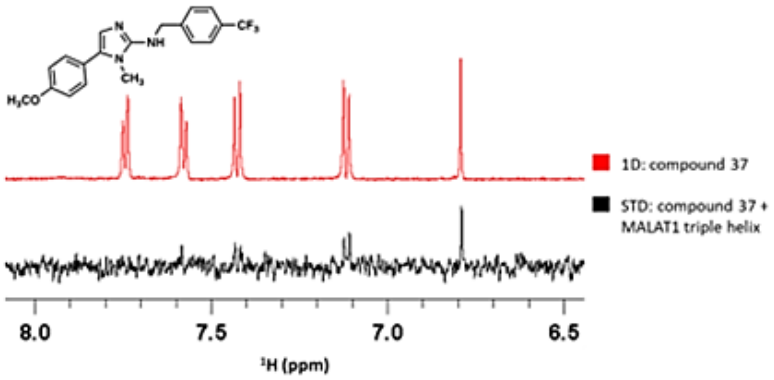


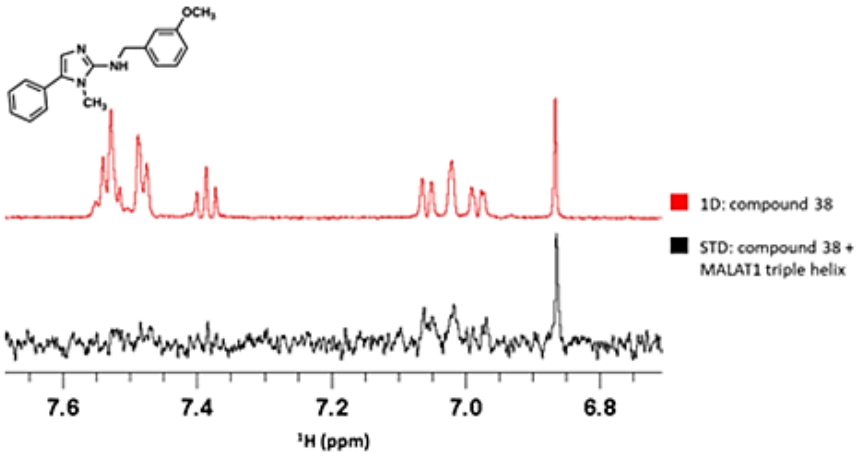

OCH3
NH
N
N
CH3
Br
1D: compound 33
STD: compound 33 + MALAT1 triple helix
7.5
7.0
1H (ppm)

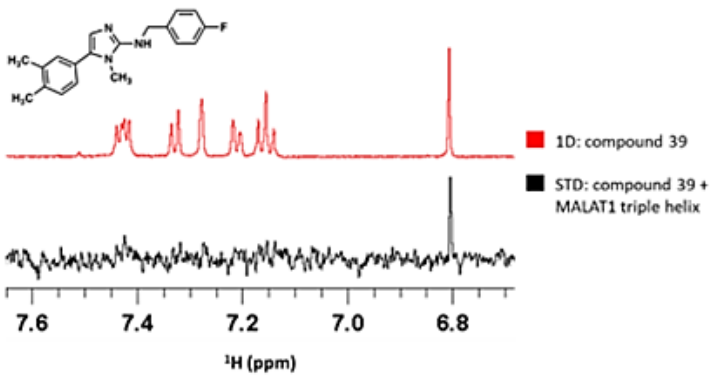

Cl
NH
N
N
CH3
H3CO
1D: compound 34
STD: compound 34 + MALAT1 triple helix
7.5
7.0
1H (ppm)

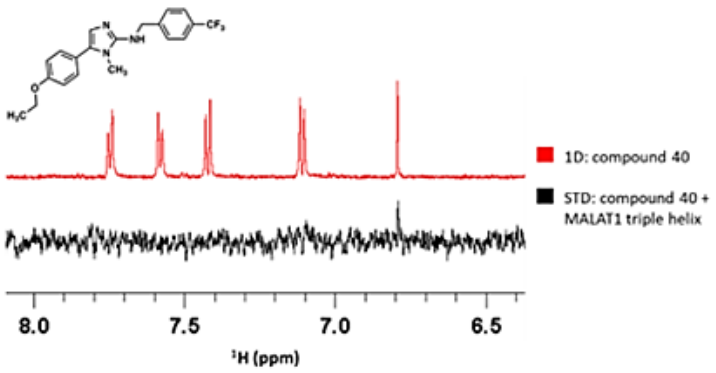

Cl
NH
N
N
CH3
H3CO
1D: compound 35
STD: compound 35 + MALAT1 triple helix
7.5
7.0
6.5
1H (ppm)

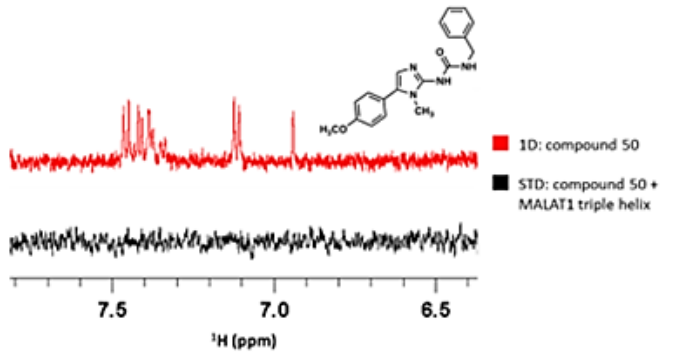

1D: compound 36
STD: compound 36 + MALAT1 triple helix
7.5
7.0
¹H (ppm)

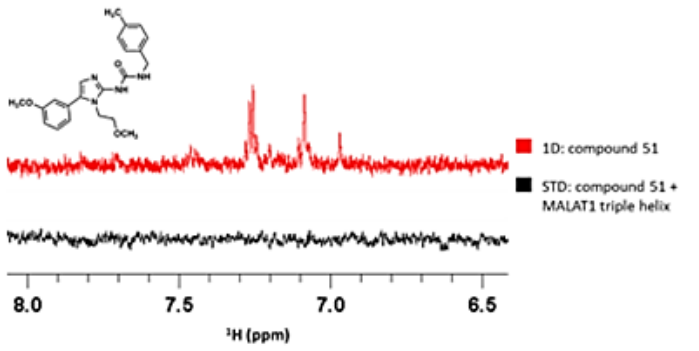

1D: compound 37
STD: compound 37 + MALAT1 triple helix
8.0
7.5
7.0
6.5
¹H (ppm)


1D: compound 38
STD: compound 38 + MALAT1 triple helix
7.6
7.4
7.2
7.0
6.8
¹H (ppm)


1D: compound 39
STD: compound 39 + MALAT1 triple helix
7.6
7.4
7.2
7.0
6.8
¹H (ppm)


1D: compound 40
STD: compound 40 + MALAT1 triple helix
8.0
7.5
7.0
6.5
¹H (ppm)


1D: compound 50
STD: compound 50 + MALAT1 triple helix
7.5
7.0
6.5
¹H (ppm)


1D: compound 51
STD: compound 51 + MALAT1 triple helix
8.0
7.5
7.0
6.5
¹H (ppm)

Bibliography

1. Wilusz, J. E.; Sunwoo, H.; Spector, D. L., Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* **2009,** *23* (13), 1494-504.
2. Li, G.; Zhang, H.; Wan, X.; Yang, X.; Zhu, C.; Wang, A.; He, L.; Miao, R.; Chen, S.; Zhao, H., Long noncoding RNA plays a key role in metastasis and prognosis of hepatocellular carcinoma. *Biomed Res Int* **2014,** *2014*, 780521.
3. Dey, B. K.; Mueller, A. C.; Dutta, A., Long non-coding RNAs as emerging regulators of differentiation, development, and disease. *Transcription* **2014,** *5* (4), e944014.
4. Qiu, M. T.; Hu, J. W.; Yin, R.; Xu, L., Long noncoding RNA: an emerging paradigm of cancer research. *Tumour Biol* **2013,** *34* (2), 613-20.
5. Li, Z. X.; Zhu, Q. N.; Zhang, H. B.; Hu, Y.; Wang, G.; Zhu, Y. S., MALAT1: a potential biomarker in cancer. *Cancer Manag Res* **2018,** *10*, 6757-68.
6. Amodio, N.; Raimondi, L.; Juli, G.; Stamato, M. A.; Caracciolo, D.; Tagliaferri, P.; Tassone, P., MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. *J Hematol Oncol* **2018,** *11* (1), 63.
7. Ji, P.; Diederichs, S.; Wang, W. B.; Boing, S.; Metzger, R.; Schneider, P. M.; Tidow, N.; Brandt, B.; Buerger, H.; Bulk, E.; Thomas, M.; Berdel, W. E.; Serve, H.; Muller-Tidow, C., MALAT-1, a novel noncoding RNA, and thymosin beta 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **2003,** *22* (39), 8031-8041.
8. Fidler, I. J., Critical factors in the biology of human cancer metastasis: twenty-eighth G.H.A. Clowes memorial award lecture. *Cancer Res* **1990,** *50* (19), 6130-8.
9. Schmidt, L. H.; Spieker, T.; Koschmieder, S.; Schaffers, S.; Humberg, J.; Jungen, D.; Bulk, E.; Hascher, A.; Wittmer, D.; Marra, A.; Hillejan, L.; Wiebe, K.; Berdel, W. E.; Wiewrodt, R.; Muller-Tidow, C., The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *J Thorac Oncol* **2011,** *6* (12), 1984-92.
10. Guo, F.; Guo, L.; Li, Y.; Zhou, Q.; Li, Z., MALAT1 is an oncogenic long non-coding RNA associated with tumor invasion in non-small cell lung cancer regulated by DNA methylation. *Int J Clin Exp Pathol* **2015,** *8* (12), 15903-10.
11. Jen, J.; Tang, Y. A.; Lu, Y. H.; Lin, C. C.; Lai, W. W.; Wang, Y. C., Oct4 transcriptionally regulates the expression of long non-coding RNAs NEAT1 and MALAT1 to promote lung cancer progression. *Mol Cancer* **2017,** *16* (1), 104.
12. Gutschner, T.; Hammerle, M.; Eissmann, M.; Hsu, J.; Kim, Y.; Hung, G.; Revenko, A.; Arun, G.; Stentrup, M.; Gross, M.; Zornig, M.; MacLeod, A. R.; Spector, D. L.; Diederichs, S., The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* **2013,** *73* (3), 1180-9.
13. Feng, T.; Shao, F.; Wu, Q.; Zhang, X.; Xu, D.; Qian, K.; Xie, Y.; Wang, S.; Xu, N.; Wang, Y.; Qi, C., miR-124 downregulation leads to breast cancer progression via LncRNA-MALAT1 regulation and CDK4/E2F1 signal activation. *Oncotarget* **2016,** *7* (13), 16205-16.
14. Zhao, Z.; Chen, C.; Liu, Y.; Wu, C., 17 β -Estradiol treatment inhibits breast cell proliferation, migration and invasion by decreasing MALAT-1 RNA level. *Biochem Biophys Res Commun* **2014,** *445* (2), 388-93.
15. Arun, G.; Diermeier, S.; Akerman, M.; Chang, K. C.; Wilkinson, J. E.; Hearn, S.; Kim, Y.; MacLeod, A. R.; Krainer, A. R.; Norton, L.; Brogi, E.; Egeblad, M.; Spector, D. L., Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev* **2016,** *30* (1), 34-51.

16. Huang, N. S.; Chi, Y. Y.; Xue, J. Y.; Liu, M. Y.; Huang, S.; Mo, M.; Zhou, S. L.; Wu, J., Long non-coding RNA metastasis associated in lung adenocarcinoma transcript 1 (MALAT1) interacts with estrogen receptor and predicted poor survival in breast cancer. *Oncotarget* **2016,** *7* (25), 37957-37965.
17. Lin, R.; Maeda, S.; Liu, C.; Karin, M.; Edgington, T. S., A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* **2007,** *26* (6), 851-8.
18. Lai, M. C.; Yang, Z.; Zhou, L.; Zhu, Q. Q.; Xie, H. Y.; Zhang, F.; Wu, L. M.; Chen, L. M.; Zheng, S. S., Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. *Med Oncol* **2012,** *29* (3), 1810-6.
19. Ma, X. Y.; Wang, J. H.; Wang, J. L.; Ma, C. X.; Wang, X. C.; Liu, F. S., Malat1 as an evolutionarily conserved lncRNA, plays a positive role in regulating proliferation and maintaining undifferentiated status of early-stage hematopoietic cells. *BMC Genomics* **2015,** *16*, 676.
20. Xie, Z. C.; Dang, Y. W.; Wei, D. M.; Chen, P.; Tang, R. X.; Huang, Q.; Liu, J. H.; Luo, D. Z., Clinical significance and prospective molecular mechanism of MALAT1 in pancreatic cancer exploration: a comprehensive study based on the GeneChip, GEO, Oncomine, and TCGA databases. *Onco Targets Ther* **2017,** *10*, 3991-4005.
21. Pang, E. J.; Yang, R.; Fu, X. B.; Liu, Y. F., Overexpression of long non-coding RNA MALAT1 is correlated with clinical progression and unfavorable prognosis in pancreatic cancer. *Tumour Biol* **2015,** *36* (4), 2403-7.
22. Zhou, Y.; Shan, T.; Ding, W.; Hua, Z.; Shen, Y.; Lu, Z.; Chen, B.; Dai, T., Study on mechanism about long noncoding RNA MALAT1 affecting pancreatic cancer by regulating Hippo-YAP signaling. *J Cell Physiol* **2018,** *233* (8), 5805-5814.
23. Jiao, F.; Hu, H.; Yuan, C.; Wang, L.; Jiang, W.; Jin, Z.; Guo, Z., Elevated expression level of long noncoding RNA MALAT-1 facilitates cell growth, migration and invasion in pancreatic cancer. *Oncol Rep* **2014,** *32* (6), 2485-92.
24. Zhang, H. M.; Yang, F. Q.; Chen, S. J.; Che, J.; Zheng, J. H., Upregulation of long non-coding RNA MALAT1 correlates with tumor progression and poor prognosis in clear cell renal cell carcinoma. *Tumour Biol* **2015,** *36* (4), 2947-55.
25. Ying, L.; Chen, Q.; Wang, Y.; Zhou, Z.; Huang, Y.; Qiu, F., Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition. *Mol Biosyst* **2012,** *8* (9), 2289-94.
26. Fan, Y.; Shen, B.; Tan, M.; Mu, X.; Qin, Y.; Zhang, F.; Liu, Y., TGF β -induced upregulation of malat1 promotes bladder cancer metastasis by associating with suz12. *Clin Cancer Res* **2014,** *20* (6), 1531-41.
27. Yang, L.; Bai, H. S.; Deng, Y.; Fan, L., High MALAT1 expression predicts a poor prognosis of cervical cancer and promotes cancer cell growth and invasion. *Eur Rev Med Pharmacol Sci* **2015,** *19* (17), 3187-93.
28. Guo, F.; Li, Y.; Liu, Y.; Wang, J.; Li, G., Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. *Acta Biochim Biophys Sin (Shanghai)* **2010,** *42* (3), 224-9.
29. Sun, R.; Qin, C.; Jiang, B.; Fang, S.; Pan, X.; Peng, L.; Liu, Z.; Li, W.; Li, Y.; Li, G., Down-regulation of MALAT1 inhibits cervical cancer cell invasion and metastasis by inhibition of epithelial-mesenchymal transition. *Mol Biosyst* **2016,** *12* (3), 952-62.
30. Yang, M. H.; Hu, Z. Y.; Xu, C.; Xie, L. Y.; Wang, X. Y.; Chen, S. Y.; Li, Z. G., MALAT1 promotes colorectal cancer cell proliferation/migration/invasion via PRKA kinase anchor protein 9. *Biochim Biophys Acta* **2015,** *1852* (1), 166-74.

31. Li, P.; Zhang, X.; Wang, H.; Wang, L.; Liu, T.; Du, L.; Yang, Y.; Wang, C., MALAT1 is associated with poor response to oxaliplatin-based chemotherapy in colorectal cancer patients and promotes chemoresistance through EZH2. *Mol Cancer Ther* **2017,** *16* (4), 739-751.
32. Tripathi, V.; Ellis, J. D.; Shen, Z.; Song, D. Y.; Pan, Q.; Watt, A. T.; Freier, S. M.; Bennett, C. F.; Sharma, A.; Bubulya, P. A.; Blencowe, B. J.; Prasanth, S. G.; Prasanth, K. V., The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **2010,** *39* (6), 925-38.
33. Wang, Y.; Zhou, Y.; Yang, Z.; Chen, B.; Huang, W.; Liu, Y.; Zhang, Y., MiR-204/ZEB2 axis functions as key mediator for MALAT1-induced epithelial-mesenchymal transition in breast cancer. *Tumour Biol* **2017,** *39* (7), 1010428317690998.
34. Hutchinson, J. N.; Ensminger, A. W.; Clemson, C. M.; Lynch, C. R.; Lawrence, J. B.; Chess, A., A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **2007,** *8*, 39.
35. Spector, D. L.; Lamond, A. I., Nuclear Speckles. *Cold Spring Harb Perspect Biol* **2011,** *3* (2).
36. Galganski, L.; Urbanek, M. O.; Krzyzosiak, W. J., Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res* **2017,** *45* (18), 10350-68.
37. Mintz, P. J.; Patterson, S. D.; Neuwald, A. F.; Spahr, C. S.; Spector, D. L., Purification and biochemical characterization of interchromatin granule clusters. *Embo j* **1999,** *18* (15), 4308-20.
38. Wilusz, J. E.; Freier, S. M.; Spector, D. L., 3 ' End Processing of a Long Nuclear-Retained Noncoding RNA Yields a tRNA-like Cytoplasmic RNA. *Cell* **2008,** *135* (5), 919-932.
39. Wilusz, J. E.; JnBaptiste, C. K.; Lu, L. Y.; Kuhn, C. D.; Joshua-Tor, L.; Sharp, P. A., A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev* **2012,** *26* (21), 2392-407.
40. Brown, J. A.; Valenstein, M. L.; Yario, T. A.; Tycowski, K. T.; Steitz, J. A., Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad Sci U S A* **2012,** *109* (47), 19202-7.
41. Brown, J. A.; Bulkley, D.; Wang, J.; Valenstein, M. L.; Yario, T. A.; Steitz, T. A.; Steitz, J. A., Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat Struct Mol Biol* **2014,** *21* (7), 633-640.
42. Mitton-Fry, R. M.; DeGregorio, S. J.; Wang, J.; Steitz, T. A.; Steitz, J. A., Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **2010,** *330* (6008), 1244-7.
43. 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* **2008,** *15* (2), 177-82.
44. Liberman, J. A.; Salim, M.; Krucinska, J.; Wedekind, J. E., Structure of a class II preQ$_1$ riboswitch reveals ligand recognition by a new fold. *Nat Chem Biol* **2013,** *9* (6), 353-5.
45. Theimer, C. A.; Blois, C. A.; Feigon, J., Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* **2005,** *17* (5), 671-82.
46. Ren, A.; Xue, Y.; Peselis, A.; Serganov, A.; Al-Hashimi, H. M.; Patel, D. J., Structural and Dynamic Basis for Low-Affinity, High-Selectivity Binding of L-Glutamine by the Glutamine Riboswitch. *Cell Rep* **2015,** *13* (9), 1800-13.
47. Sun, R.; Lin, S. F.; Gradoville, L.; Miller, G., Polyadenylylated nuclear RNA encoded by Kaposi sarcoma-associated herpesvirus. *Proc Natl Acad Sci U S A* **1996,** *93* (21), 11883-8.
48. Stadler, P. F., Evolution of the Long Non-Coding RNAs MALAT1 and MEN β/ε. *J Bioinf Comput Biol.* **2010,** *6268*, 1-12.
49. Conrad, N. K.; Shu, M. D.; Uyhazi, K. E.; Steitz, J. A., Mutational analysis of a viral RNA element that counteracts rapid RNA decay by interaction with the polyadenylate tail. *Proc Natl Acad Sci U S A* **2007,** *104* (25), 10412-7.

50. Sasaki, Y. T.; Ideue, T.; Sano, M.; Mituyama, T.; Hirose, T., MEN ε/ β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci U S A* **2009,** *106* (8), 2525-30.
51. Brown, J. A.; Kinzig, C. G.; Degregorio, S. J.; Steitz, J. A., Hoogsteen-position pyrimidines promote the stability and function of the MALAT1 RNA triple helix. *RNA* **2016,** *22* (5), 743-749.
52. Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H., NMR spectroscopy of RNA. *Chembiochem* **2003,** *4* (10), 936-62.
53. Cash, D. D.; Cohen-Zontag, O.; Kim, N. K.; Shefer, K.; Brown, Y.; Ulyanov, N. B.; Tzfati, Y.; Feigon, J., Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo. In *Proc Natl Acad Sci U S A*, 2013; Vol. 110, pp 10970-5.
54. Holland, J. A.; Hoffman, D. W., Structural features and stability of an RNA triple helix in solution. *Nucleic Acids Res* **1996,** *24* (14), 2841-8.
55. Misra, V. K.; Draper, D. E., On the Role of Magnesium Ions in RNA Stability. *Biopolymers* **1999,** *48*, 113-135.
56. Serra, M. J.; Baird, J. D.; Dale, T.; Fey, B. L.; Retatagos, K.; Westhof, E., Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA* **2002,** *8* (3), 307-23.
57. Lemay, J. F.; Penedo, J. C.; Tremblay, R.; Lilley, D. M.; Lafontaine, D. A., Folding of the adenine riboswitch. *Chem Biol* **2006,** *13* (8), 857-68.
58. 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* **2012,** *40* (7), 3117-30.
59. Bassi, G. S.; Murchie, A. I.; Lilley, D. M., The ion-induced folding of the hammerhead ribozyme: core sequence changes that perturb folding into the active conformation. *RNA* **1996,** *2* (8), 756-68.
60. Orr, J. W.; Hagerman, P. J.; Williamson, J. R., Protein and Mg^{2+}-induced conformational changes in the S15 binding site of 16 S ribosomal RNA. *J Mol Biol* **1998,** *275* (3), 453-64.
61. Coman, D.; Russu, I. M., Site-resolved stabilization of a DNA triple helix by magnesium ions. *Nucleic Acids Res* **2004,** *32* (3), 878-83.
62. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **1995,** *6* (2), 135-40.
63. Wohnert, J.; Dingley, A. J.; Stoldt, M.; Gorlach, M.; Grzesiek, S.; Brown, L. R., Direct identification of NH...N hydrogen bonds in non-canonical base pairs of RNA by NMR spectroscopy. *Nucleic Acids Res* **1999,** *27* (15), 3104-10.
64. Wijmenga, S. S.; van Buuren, B. N. M., The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1998,** *32*, 287-387.
65. Abulwerdi, F. A.; Xu, W.; Ageeli, A. A.; Yonkunas, M. J.; Arun, G.; Nam, H.; Schneekloth, J. S., Jr.; Dayie, T. K.; Spector, D.; Baird, N.; Le Grice, S. F. J., Selective small-molecule targeting of a triple helix encoded by the long noncoding RNA, MALAT1. *ACS Chem Biol* **2019,** *14* (2), 223-235.
66. Ageeli, A. A.; McGovern-Gooch, K. R.; Kaminska, M. M.; Baird, N. J., Finely tuned conformational dynamics regulate the protective function of the lncRNA MALAT1 triple helix. *Nucleic Acids Res* **2019,** *47* (3), 1468-1481.
67. Mott, M. L.; Berger, J. M., DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol* **2007,** *5* (5), 343-54.
68. Hahn, S., Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **2004,** *11* (5), 394-403.

69. Tapias, A.; Auriol, J.; Forget, D.; Enzlin, J. H.; Scharer, O. D.; Coin, F.; Coulombe, B.; Egly, J. M., Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *J Biol Chem* **2004,** *279* (18), 19074-83.
70. Gueron, M.; Leroy, J. L., Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol* **1995,** *261*, 383-413.
71. Rinnenthal, J.; Klinkert, B.; Narberhaus, F.; Schwalbe, H., Direct observation of the temperature-induced melting process of the Salmonella fourU RNA thermometer at base-pair resolution. *Nucleic Acids Res* **2010,** *38* (11), 3834-47.
72. Steinert, H. S.; Rinnenthal, J.; Schwalbe, H., Individual basepair stability of DNA and RNA studied by NMR-detected solvent exchange. *Biophys J* **2012,** *102* (11), 2564-74.
73. Wagner, D.; Rinnenthal, J.; Narberhaus, F.; Schwalbe, H., Mechanistic insights into temperature-dependent regulation of the simple cyanobacterial hsp17 RNA thermometer at base-pair resolution. *Nucleic Acids Res* **2015,** *43* (11), 5572-85.
74. Snoussi, K.; Leroy, J. L., Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry* **2001,** *40* (30), 8898-904.
75. Englander, S. W.; Kallenbach, N. R., Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* **1984,** *16*, 521-655.
76. Strazewski, P., Thermodynamic correlation analysis: hydration and perturbation sensitivity of RNA secondary structures. *J Am Chem Soc* **2002,** *124* (14), 3546-54.
77. Plum, G. E.; Park, Y. W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J., Thermodynamic characterization of the stability and the melting behavior of a DNA triplex: a spectroscopic and calorimetric study. *Proc Natl Acad Sci U S A* **1990,** *87* (23), 9436-40.
78. Hoyne, P. R.; Gacy, A. M.; McMurray, C. T.; Maher, L. J., 3rd, Stabilities of intrastrand pyrimidine motif DNA and RNA triple helices. *Nucleic Acids Res* **2000,** *28* (3), 770-5.
79. Coman, D.; Russu, I. M., Site-resolved energetics in DNA triple helices containing G*TA and T*CG triads. *Biochemistry* **2002,** *41* (13), 4407-14.
80. Warner, K. D.; Hajdin, C. E.; Weeks, K. M., Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* **2018,** *17* (8), 547-558.
81. Connelly, C. M.; Moon, M. H.; Schneekloth, J. S., Jr., The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **2016,** *23* (9), 1077-1090.
82. Abulwerdi, F. A.; Schneekloth, J. S., Jr., Microarray-based technologies for the discovery of selective, RNA-binding molecules. *Methods* **2016,** *103*, 188-95.
83. Venkitakrishnan, R. P.; Benard, O.; Max, M.; Markley, J. L., Use of NMR Saturation Transfer Difference Spectroscopy to Study Ligand Binding to Membrane Proteins. *Methods Mol Biol* **2012,** *914*, 47-63.
84. Xia, Y.; Zhu, Q.; Jun, K. Y.; Wang, J.; Gao, X., Clean STD-NMR spectrum for improved detection of ligand-protein interactions at low concentration of protein. *Magn Reson Chem* **2010,** *48* (12), 918-24.
85. Yang, H.; Liu, J.; Huang, Y.; Gao, R.; Tang, B.; Li, S.; He, J.; Li, H., Domain-specific interactions between MLN8237 and human serum albumin estimated by STD and WaterLOGSY NMR, ITC, spectroscopic, and docking techniques. *Sci Rep* **2017,** *7*, 45514.
86. Mayer, M.; Meyer, B., Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc* **2001,** *123* (25), 6108-17.
87. Mayer, M.; James, T. L., Detecting ligand binding to a small RNA target via saturation transfer difference NMR experiments in D_2O and H_2O. *J Am Chem Soc* **2002,** *124* (45), 13376-7.
88. Mayer, M.; James, T. L., NMR-based characterization of phenothiazines as a RNA binding scaffold. *J Am Chem Soc* **2004,** *126* (13), 4453-60.

89. Mayer, M.; Lang, P. T.; Gerber, S.; Madrid, P. B.; Pinto, I. G.; Guy, R. K.; James, T. L., Synthesis and testing of a focused phenothiazine library for binding to HIV-1 TAR RNA. *Chem Biol* **2006,** *13* (9), 993-1000.
90. Harner, M. J.; Mueller, L.; Robbins, K. J.; Reily, M. D., NMR in drug design. *Arch Biochem Biophys* **2017,** *628*, 132-147.
91. Peng, J. W.; Moore, J.; Abdul-Manan, N., NMR experiments for lead generation in drug discovery. *Prog Nucl Magn Reson Spetrosc* **2004,** *44*, 225-256.
92. Williamson, M. P., Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **2013,** *73*, 1-16.
93. Brown, J. A.; Kinzig, C. G.; DeGregorio, S. J.; Steitz, J. A., Methyltransferase-like protein 16 binds the 3 '-terminal triple helix of MALAT1 long noncoding RNA. *Proceedings of the National Academy of Sciences of the United States of America* **2016,** *113* (49), 14013-14018.
94. 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* **2005,** *102* (17), 5958-63.
95. Avis, J. M.; Conn, G. L.; Walker, S. C., Cis-acting ribozymes for the production of RNA in vitro transcripts with defined 5' and 3' ends. *Methods Mol Biol* **2012,** *941*, 83-98.
96. Green, H.; Freeman, R., Band-selective radiofrequency pulses. *J Magn Reson* **1991,** *93*, 93-141.
97. 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* **1995,** *112*, 275-279.
98. Donlic, A.; Morgan, B. S.; Xu, J. L.; Liu, A.; Roble, C., Jr.; Hargrove, A. E., Discovery of small molecule ligands for MALAT1 by tuning an RNA-binding scaffold. *Angew Chem Int Ed Engl* **2018,** *57* (40), 13242-13247.
99. Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M., The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol* **2019,** *20* (8), 474-489.
100. Marusic, M.; Schlagnitweit, J.; Petzold, K., RNA Dynamics by NMR Spectroscopy. *Chembiochem* **2019,** *20*, 1-27.
101. Montange, R. K.; Batey, R. T., Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys* **2008,** *37*, 117-33.
102. Furtig, B.; Nozinovic, S.; Reining, A.; Schwalbe, H., Multiple conformational states of riboswitches fine-tune gene regulation. *Curr Opin Struct Biol* **2015,** *30*, 112-124.
103. Alexander, R. W.; Eargle, J.; Luthey-Schulten, Z., Experimental and computational determination of tRNA dynamics. *FEBS Lett* **2010,** *584* (2), 376-86.
104. Doherty, E. A.; Doudna, J. A., Ribozyme structures and mechanisms. *Annu Rev Biophys Biomol Struct* **2001,** *30*, 457-75.
105. Dagenais, P.; Girard, N.; Bonneau, E.; Legault, P., Insights into RNA structure and dynamics from recent NMR and X-ray studies of the Neurospora Varkud satellite ribozyme. *Wiley Interdiscip Rev RNA* **2017,** *8* (5).
106. Glaeser, R. M., How Good Can Single-Particle Cryo-EM Become? What Remains Before It Approaches Its Physical Limits? *Annu Rev Biophys* **2019,** *48*, 45-61.
107. Murata, K.; Wolf, M., Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim Biophys Acta Gen Subj* **2018,** *1862* (2), 324-334.
108. Ray, S.; Widom, J. R.; Walter, N. G., Life under the Microscope: Single-Molecule Fluorescence Highlights the RNA World. *Chem Rev* **2018,** *118* (8), 4120-4155.
109. Stagno, J. R.; Liu, Y.; Bhandari, Y. R.; Conrad, C. E.; Panja, S.; Swain, M.; Fan, L.; Nelson, G.; Li, C.; Wendel, D. R.; White, T. A.; Coe, J. D.; Wiedorn, M. O.; Knoska, J.; Oberthuer, D.; Tuckey, R. A.; Yu, P.; Dyba, M.; Tarasov, S. G.; Weierstall, U.; Grant, T. D.; Schwieters, C. D.; Zhang, J.; Ferre-D'Amare, A. R.; Fromme, P.; Draper, D. E.; Liang, M.; Hunter, M. S.; Boutet, S.; Tan, K.; Zuo, X.; Ji,

X.; Barty, A.; Zatsepin, N. A.; Chapman, H. N.; Spence, J. C.; Woodson, S. A.; Wang, Y. X., Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **2017,** *541* (7636), 242-246.
110. Palmer, A. G., 3rd, NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **2004,** *104* (8), 3623-40.
111. Barnwal, R. P.; Yang, F.; Varani, G., Applications of NMR to structure determination of RNAs large and small. *Arch Biochem Biophys* **2017,** *628*, 42-56.
112. 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* **2007,** *365* (4), 930-44.
113. 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* **2013,** *57* (4), 333-52.
114. Wagner, G., NMR relaxation and protein mobility. *Curr Opin Struct Biol* **1993,** *3*, 748-754.
115. Xue, Y.; Gracia, B.; Herschlag, D.; Russell, R.; Al-Hashimi, H. M., Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch. *Nat Commun* **2016,** *7*, ncomms11768.
116. Yamazaki, T.; Muhandiram, R.; Kay, L. E., NMR experiments for the measurement of carbon relaxation properties in highly enriched, uniformly ^{13}C, ^{15}N-labeled proteins: application to $^{13}C^{\alpha}$ carbons. *J Am Chem Soc* **1994,** *116*, 8266-8278.
117. Shajani, Z.; Varani, G., NMR studies of dynamics in RNA and DNA by ^{13}C relaxation. *Biopolymers* **2007,** *86* (5-6), 348-59.
118. Johnson, J. E., Jr.; Julien, K. R.; Hoogstraten, C. G., Alternate-site isotopic labeling of ribonucleotides for NMR studies of ribose conformational dynamics in RNA. *J Biomol NMR* **2006,** *35* (4), 261-74.
119. Dayie, K. T.; Wagner, G., Carbonyl carbon probe of local mobility in ^{13}C, ^{15}N - enriched proteins using high-resolution nuclear magnetic resonance. *J Am Chem Soc* **1997,** *119* (33), 7797-7806.
120. Boisbouvier, J.; Wu, Z.; Ono, A.; Kainosho, M.; Bax, A., Rotational diffusion tensor of nucleic acids from ^{13}C NMR relaxation. *J Biomol NMR* **2003,** *27* (2), 133-42.
121. Hansen, A. L.; Al-Hashimi, H. M., Dynamics of large elongated RNA by NMR carbon relaxation. *J Am Chem Soc* **2007,** *129* (51), 16072-82.
122. Johnson, J. E., Jr.; Hoogstraten, C. G., Extensive backbone dynamics in the GCAA RNA tetraloop analyzed using ^{13}C NMR spin relaxation and specific isotope labeling. *J Am Chem Soc* **2008,** *130* (49), 16757-69.
123. Thakur, C. S.; Dayie, T. K., Asymmetry of ^{13}C labeled 3-pyruvate affords improved site specific labeling of RNA for NMR spectroscopy. *J Biomol NMR* **2012,** *52* (1), 65-77.
124. Thakur, C. S.; Luo, Y.; Chen, B.; Eldho, N. V.; Dayie, T. K., Biomass production of site selective $^{13}C/^{15}N$ nucleotides using wild type and a transketolase E. coli mutant for labeling RNA for high resolution NMR. *J Biomol NMR* **2012,** *52* (2), 103-14.
125. Alvarado, L. J.; Longhini, A. P.; LeBlanc, R. M.; Chen, B.; Kreutz, C.; Dayie, T. K., Chemo-enzymatic synthesis of selectively $^{13}C/^{15}N$-labeled RNA for NMR structural and dynamics studies. *Methods Enzymol* **2014,** *549*, 133-62.
126. Longhini, A. P.; LeBlanc, R. M.; Becette, O.; Salguero, C.; Wunderlich, C. H.; Johnson, B. A.; D'Souza, V. M.; Kreutz, C.; Dayie, T. K., Chemo-enzymatic synthesis of site-specific isotopically labeled nucleotides for use in NMR resonance assignment, dynamics and structural characterizations. *Nucleic Acids Res* **2016,** *44* (6), e52.
127. Keeler, J., *Understanding NMR Spectroscopy*. Second ed.; WILEY: Cambridge, UK, 2010.

128. Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., 3rd; Rance, M.; Skelton, N. J., *Protein NMR Spectroscopy*. 2nd ed.; Academic Press: 2006.
129. Fiala, R.; Czernek, J.; Sklenar, V., Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. *J Biomol NMR* **2000,** *16* (4), 291-302.
130. Brutscher, B.; Boisbouvier, J.; Pardi, A.; Marion, D.; Simorre, J., Improved sensitivity and resolution in ^{1}H-^{13}C NMR experiments of RNA. *J Am Chem Soc* **1998,** *120*, 11845-11851.
131. Goldman, M., Interference effects in the relaxation of a pair of unlike spin - 1/2 nuclei. *J Magn Reson* **1984,** *60*, 437-452.
132. Engelke, J.; Ruterjans, H., Determination of $^{13}C^{\alpha}$ relaxation times in uniformly $^{13}C/^{15}N$ -enriched proteins. *J Biomol NMR* **1995,** *5* (2), 173-82.
133. Alvarado, L. J.; LeBlanc, R. M.; Longhini, A. P.; Keane, S. C.; Jain, N.; Yildiz, Z. F.; Tolbert, B. S.; D'Souza, V. M.; Summers, M. F.; Kreutz, C.; Dayie, T. K., Regio-selective chemical-enzymatic synthesis of pyrimidine nucleotides facilitates RNA structure and dynamics studies. *Chembiochem* **2014,** *15* (11), 1573-7.
134. Peng, J. W.; Wagner, G., 2D Heteronuclear NMR measurements of spin-lattice relaxation times in the rotating frame of X nuclei in heteronuclear HX spin systems. *J Magn Reson* **1991,** *94*, 82-100.
135. Duchardt, E.; Schwalbe, H., Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by MNMR ^{13}C relaxation. *J Biomol NMR* **2005,** *32* (4), 295-308.
136. 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* **1999,** *5* (9), 1268-72.
137. Fushman, D.; Cahill, S.; Cowburn, D., The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of ^{15}N relaxation with monomer/dimer equilibration. *J Mol Biol* **1997,** *266* (1), 173-94.
138. Handy, N. C.; Cohen, A. J., Left-right correlation energy. *Mol. Phys.* **2001,** *99*, 403-412.
139. Adamo, C.; Barone, V., Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999,** *110*, 6158-69.
140. Jensen, F., Segmented contracted basis sets optimized for nuclear magnetic shielding. *J Chem Theory Comput* **2015,** *11* (1), 132-8.
141. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J., Gaussian 16, Revision A.03. *Gaussian, Inc., Wallingford CT* **2016**.
142. 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* **2006,** *128* (35), 11443-54.