Methods for Comprehensive RNA Structure and Dynamics Analysis using SHAPE Technologies

Kady-Ann Camele Steen-Burrell

A dissertation submitted to the faculty of the Univeristy of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill

2011

Approved by

Michael Jarstfer

Jeffrey Johnson

Linda Spremulli

Marcey Waters

Kevin Weeks

ABSTRACT

KADY-ANN CAMELE STEEN-BURRELL: Methods for Comprehensive RNA Structure and Dynamics Analysis using SHAPE Technologies
(Under the direction of Kevin M. Weeks)

The many important cellular functions of RNA molecules depend on formation of complex RNA secondary and tertiary structures. The formation of these structures is facilitated by the intrinsic motion of RNA nucleotides and can be influenced by various ligand or protein interactions. RNA SHAPE technology has made the determination of many RNA secondary structures facile. However, the applicability of traditional SHAPE technology to short RNAs (less than 100 nucleotides) in their native state is limited by primer extension detection. Additionally, it is often difficult to discern the structural context of constrained nucleotides in a traditional SHAPE experiment. In this work, I first develop an alternate SHAPE 2′-*O*-adduct detection method, termed RNase-detected SHAPE, which takes advantage of the RNA-specific activity of an exoribonuclease, RNase R. In the presence of a SHAPE 2′-*O*-adduct or adducts at the nucleotide base-pairing face, RNase R stops three or four nucleotides 3′ of the modification site, respectively. RNase-detected SHAPE allowed for the structural characterization of a small, biologically relevant riboswitch in its ligand-free state and identification of a bulge register shift that facilitates formation of the ligand-bound state. Second, I develop a method for the *de novo* identification of nucleotides involved in, or adjacent to, key RNA tertiary structure

interactions, termed differential SHAPE reactivity analysis. This method uses two SHAPE electrophiles, N-methylisatoic anhydride (NMIA) and 1-methyl-6-nitroisatoic anhydride (1M6) that detect slowly dynamic and one-sided stacking nucleotides, respectively. Together, both types of nucleotide behaviors provide a RNA tertiary structure "fingerprint" since both tend to be over-represented in tertiary structure interactions and motifs. Third, I develop a chemical method for the removal of SHAPE 2′-*O*-adducts by ester bond cleavage while limiting RNA phosphodiester backbone degradation which allows for the downstream manipulation of previously modified RNA. The extent of SHAPE 2′-*O*-adduct removal can be modulated by varying the concentration of hydroxylamine and reaction times. Finally, I evaluate the ability of SHAPE chemistry to detect ligand-induced conformational changes by comparing SHAPE reactivities and NMR measurements. Dynamics as measured by SHAPE reactivities and the NMR order parameter, S^2, correlate well for small molecule binding to RNA.

ACKNOWLEDGEMENT

First, I would like to thank my advisor, Kevin Weeks, for allowing me to work on this project and for giving me the opportunity to take the project in the direction that I wanted to go.

There are many people who have loved and supported me over the years. My family, many of whom I don't get to see very often, thanks for supporting me through your calls, visits, and prayers. My friends, both near and far, thank you all for your friendship, understanding and support. In particular, thanks to Vanessa and Jen for all the support over the years. Graduate school was a little easier because of you. Thanks to my labmates who were always there to bounce ideas and talk science with. I will truly miss you all.

Finally, to my husband, Kevin, for all you have done over the past years in support of my dreams and aspirations. I love you and I could not have done it without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

1M6	1-methyl-6-nitroisatoic anhydride
1M6Br	1-methyl-6-bromoisatoic anhydride
1M6M	1-methyl-6-methylisatoic anhydride
1M7	1-methyl-7-nitroisatoic anhydride
2′-OH	2′-hydroxyl
2-AP	2-aminopurine
A	adenine
AddC	adenosine-2′,3′-dideoxycytidine triphosphate
ATP	adenosine triphosphate
C	cytosine
cDNA	complementary deoxyribonucleic acid
Ci	curie
cm	centimeter
CSD	cold shock domain
°C	degree Celsius
dNTP	deoxyribonucleotide triphosphate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
G	guanosine
h	hour
H_2O	water
HEPES	N-2-hydroxyethylpiperazine-N′-2-ethanesulfonic acid

HIV	human immunodeficiency virus
KCl	potassium chloride
L	liter
M	molar
Mg^{2+}	magnesium ion
$MgCl_2$	magnesium chloride
min	minute
mg	milligram
mL	milliliter
mm	millimeter
mM	millimolar
μg	microgram
μL	microliter
μM	micromolar
N	nitrogen
NaCl	sodium chloride
NH_2OH	hydroxylamine
nM	nanomolar
NMIA	N-methylisatoic anhydride
NMR	nuclear magnetic resonance
NO_2	nitro
NOE	nuclear Overhauser effect
nt	nucleotide
NTP	ribonucleotide triphosphate
O	oxygen
P	phosphate
π-π	pi-pi

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pmol	picomole
PNK	polynucleotide kinase
QM	quantum mechanics
RNA	ribonucleic acid
RNase	ribonuclease
RMSD	root mean square deviation
RNB	ribonucleic acid binding
s	second
SAFA	semi-automated footprinting analysis
SCF	self-consistent field
SHAPE	Selective 2′-Hydroxyl Acylation analyzed by Primer Extension
σ_m	sigma meta
T_1	longitudinal relaxation rate
T_2	transverse relaxation rate
TAR	trans-activation region
Tat	trans-activator of transcription
TBE	90 mM Tris-borate, 2 mM EDTA
TE	10 mM Tris-HCl (pH 8), 1 mM EDTA
TPP	thiamine pyrophosphate
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
U	uridine
UTR	untranslated region
v	volume
V	volt

w	weight
W	watt

CHAPTER 1

Towards Comprehensive, Single Nucleotide Analysis of RNA Structure and Dynamics

1.1 Introduction

1.1.1 RNA structure and its importance to function

RNA serves dual roles as both an integral carrier of genetic information at the primary nucleotide sequence level, and as an important biomolecular machine through formation of higher-order structures that are central to almost every biological process in the cell [1-3]. Many of these critical cellular functions including protein synthesis, catalysis, gene expression and regulation, depend on the ability of RNA molecules to fold back on itself to form complex and specific three-dimensional structures [3-5]. RNA three-dimensional structures vary both in their size and complexity from small tRNAs to large ribosomal RNAs [6-8]. Formation of these three-dimensional structures requires base-pairing interactions to form secondary structures (Figure 1.1A) and long-range contacts including hydrogen bonding and nucleobase stacking interactions to form tertiary structures (Figure 1.1B) [9, 10]. Understanding the molecule architecture of RNA structure is a key first step towards explaining the functional role of RNA in the cell.

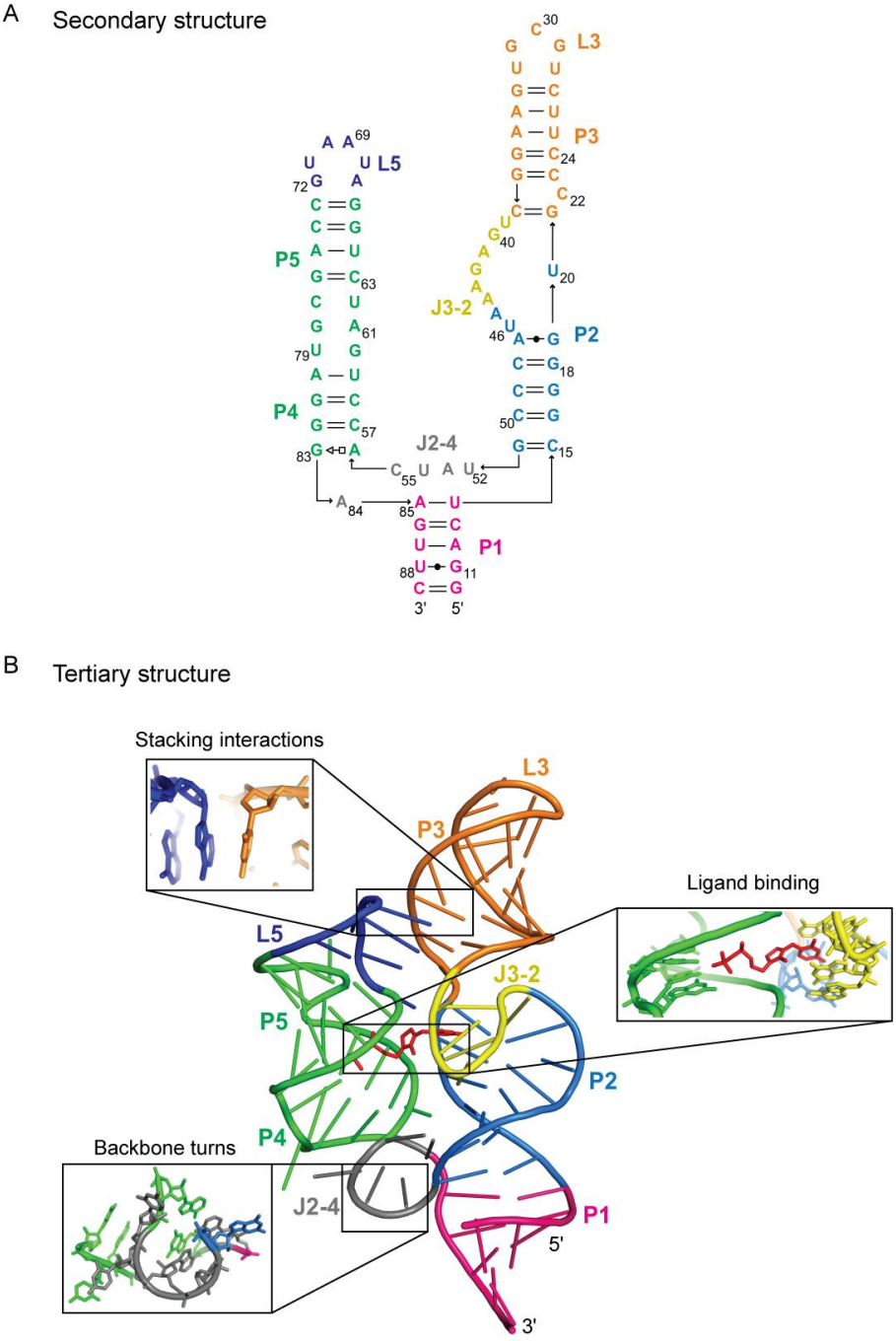


Figure 1.1 RNA structure levels. (A) Secondary structure highlighting base-pairing interactions and (B) tertiary structure highlighting the key tertiary motifs and interactions that stabilize the global fold. Illustrated is the aptamer domain of the *Escherichia coli* thiamine pyrophosphate (TPP) riboswitch [14, 15].

1.1.2 RNA ligand binding and dynamics

RNA structure, folding and function are also influenced by local nucleotide dynamics. Many cellular functions, such as gene regulation by riboswitches, require RNA molecule to undergo conformational changes in response to changes in ionic conditions or external cofactors such as small molecules and proteins (compare A and B, Figure 1.1) [3, 11]. Additionally, RNAs in many ribonucleoprotein complexes rely on protein binding to form the functional, active RNA conformation [12, 13]. These ligand-induced conformational changes are facilitated by the inherent motion of RNA nucleotides. RNA dynamics occur over a range of timescales from very fast bond vibrations to intermediate helical and domain motions to very slow structural rearrangements [16, 17]. The ability to biochemically evaluate ligand-induced conformational changes of individual RNA nucleotides regardless of RNA size and complexity is an unmet experimental challenge.

1.1.3 RNA structure determination methods

RNA secondary structure determination is a key first step towards comprehensive understanding of RNA structure-function relationships. Conventional biochemical methods of secondary structure determination rely on small molecules or ribonucleases that primarily modify or cleave the RNA in a base-specific manner including kethoxal, which modifies G residues, and RNase T1, which cleaves after G residues [18, 19]. Additionally, some ribonucleases cleave in a base non-specific manner such RNase V1, which cleaves base-paired nucleotides [18]. While these biochemical methods have provided valuable information about RNA structures, no single method is capable of completely mapping the structure of any given RNA of interest.

A relatively new chemical secondary structure probing method, Selective 2′-Hydroxyl

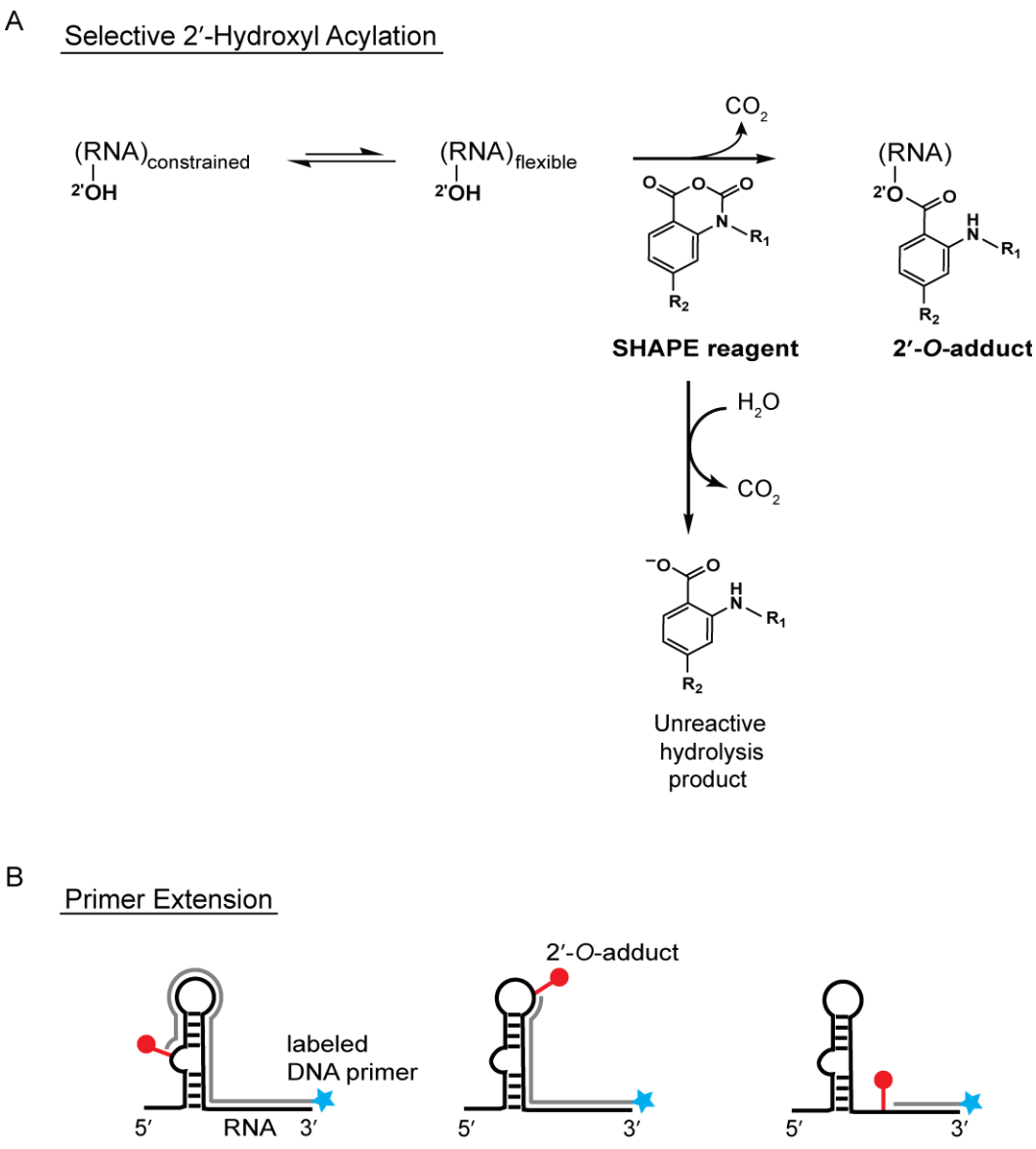


Figure 1.2 RNA SHAPE technology. (A) Mechanism of selective acylation reaction and concurrent hydrolysis reaction. (B) Primer extension detection of SHAPE 2′-*O*-adduct formation.

Acylation analyzed by Primer Extension (SHAPE) takes advantage of the intrinsic reactivity of the 2′-hydroxyl on the ribose sugar. This feature allows structural information to be obtained for all nucleotides in a single experiment. The ability of the 2′-hydroxyl to deprotonate and react with electrophiles is strongly influenced by local nucleotide flexibility [20]. Conformationally flexible nucleotides preferentially react with SHAPE electrophiles such as 1-methyl-7-nitroisatoic anhydride (1M7) resulting in the formation of ester adducts (2′-*O*-adducts) (Figure 1.2A). Base-paired or otherwise constrained nucleotides are less reactive towards SHAPE electrophiles [21, 22]. The resulting 2′-*O*-adducts are then identified as stops in a primer extension reaction using 5′-end labeled primers (radiolabeled or fluorescently labeled) that are annealed to the 3′-end of the RNA (Figure 1.2B). Reverse transcriptase produces cDNAs that terminate one nucleotide before the modification site resulting in a cDNA library whose fragment lengths correspond to the sites of modification and whose amounts correspond to degree of flexibility. The cDNA fragments are then separated by denaturing polyacrylamide gel electrophoresis (for radiolabeled fragments) or capillary electrophoresis (for fluorescently-labeled fragments) [23, 24]. Dideoxy-terminated sequencing reactions are performed concurrently in order to map the modification site to the RNA sequence.

SHAPE allows for quantitative, robust, single nucleotide resolution structural information to be obtained for most RNAs in a relatively short time using picomolar amounts of RNA. As local nucleotide flexibility correlates with SHAPE reactivity [25], SHAPE reactivities can be applied to the prediction of known and novel RNA structures by interpreting the reactivities as pseudo free energies that can be used to improve computer-based secondary structure prediction calculations [26, 27].

RNA tertiary structure determination has proven to be a more challenging undertaking. Biophysical methods such as NMR and x-ray crystallography have been primarily employed in the tertiary structure determination of RNA. X-ray crystallography provides high quality atomic resolution of RNA structure and allows for the observation of specific nucleotide conformations, tertiary structure motifs and contacts [8]. However, x-ray crystallography provides a snapshot of a static RNA state so information about RNA dynamics are difficult to detect. Additionally, many biologically relevant RNA states are too dynamic to be crystallized. The most widely used method for studying RNA local nucleotide dynamics is NMR spectroscopy [17, 28]. However, NMR spectroscopy is limited by the size and complexity of the RNA being studied.

Along with x-ray crystallography and NMR, biochemical methods have been employed to map the tertiary interactions in RNA. Current biochemical methods of RNA tertiary structure analysis are primarily based on cleavage of the RNA backbone by hydroxyl radicals. Hydroxyl radicals are usually generated *in situ* by reagents that are either free in solution or tethered to the RNA [19]. In-solution, free hydroxyl radical experiments provide information about solvent accessible nucleotides while tethered hydroxyl radical experiments provide information about nucleotide distance in three-dimensional space. However, some of these experiments tend to be difficult to implement due to the requirement of a modified RNA sequence for tethering the reagent [29] and the results from hydroxyl radical experiments can be technically challenging to interpret.

1.1.4 Research overview

Consequently, the overall goal of this project was to develop methods that address the problem of characterization of RNA structure and dynamics for any given RNA all within the framework of SHAPE technology.

SHAPE technology has revolutionized the RNA structure field by providing an accessible, relatively straightforward method for analyzing RNA structure. With improvements in reagent reaction time [22], electrophoresis separation [24] and data analysis [30], SHAPE technology has allowed for the probing of a number of RNAs [26, 31-33]. Traditional SHAPE technology relies on primer extension as the method of detecting modified nucleotides. Primer extension was ideal for probing long RNA structures, but the requirement of a primer binding site adjacent to the RNA sequence of interest prevented the application of traditional SHAPE to many important native short RNAs including riboswitches and nucleotides at the ends of all RNAs.

In Chapter 2, I describe an alternate method of 2′-*O*-adduct detection using the exoribonuclease *Mycoplasma genitalium* RNase R. This method, termed RNase-detected SHAPE, is simple to use and allows for the structural analysis of short, authentic RNAs and the ends of longer RNAs. RNase R is also capable of detecting covalent adducts on the base-pairing face of nucleotides. I then applied this new method to develop a secondary structure model for the ligand-free state of a biologically important RNA, the thiamine pyrophosphate-sensing (TPP) riboswitch.

In Chapter 3, I develop a new method, termed Differential SHAPE Reactivity Analysis, to identify nucleotides involved in key tertiary structure contacts in an RNA. This method relies on subtle biases in reactivity between two different SHAPE reagents, N-methylisatoic anhydride (NMIA) and 1-methyl-6-isatoic anhydride (1M6). NMIA reacts

preferentially with nucleotides undergoing slow dynamics while 1M6 reacts preferentially with nucleotides that possess an unoccupied nucleobase face for stacking interactions with 1M6. The combined preferential reactivities allows for the "fingerprinting" of RNA tertiary structure by identifying nucleotides involved in tertiary contacts in a simple, robust manner.

In Chapter 4, I develop a method to chemically cleave the ester linkage of the stable 2′-*O*-adduct while limiting phosphodiester backbone cleavage. This method results in complete removal of the 2′-*O*-adduct and allows for downstream biochemical manipulations of the previously modified RNA provided the structural information was previously encoded by either primer extension or RNase R detection.

Finally, in Chapter 5, I characterize the correlation between SHAPE and RNA dynamics for ligand-induced conformational changes. I compare SHAPE reactivities and NMR measurements for ligands bound to the trans-activating region (TAR) of HIV RNA. The absolute NMR measurements correlate strongly to SHAPE reactivities for binding of the small molecule argininamide, but the absolute measurements did not correlate with binding of a larger molecule, an 11 nucleotide linear peptide mimic of the Tat protein. However, both SHAPE chemistry and NMR reflected the same changes in local nucleotide flexibility upon linear peptide binding and suggests that SHAPE can detect changes in conformational dynamics upon ligand binding.

1.1.5 Perspective

The principles of molecular biology, biochemistry, organic chemistry and physical chemistry were integrated throughout this project to address the problem of accurate RNA structure determination regardless of RNA size or complexity. I demonstrate that SHAPE 2′-*O*-adducts and covalent adducts at the base-pairing face can be quantitatively detected using

an exoribonuclease; nucleotides involved in tertiary structure contacts can be detected using the differences in SHAPE reactivities between NMIA and 1M6; the ester linkage of SHAPE 2′-*O*-adducts can be specifically cleaved with limited RNA degradation using hydroxylamine; and changes in local nucleotide dynamics as a result of ligand binding can be detected using SHAPE chemistry.

It is my hope that each of the methods that I have developed will be applied to understanding structure-function relationships of RNAs in general, and specifically to predicting the secondary and tertiary structures of novel RNAs. I expect that these methods will contribute to the further development of SHAPE as a facile, high-throughput technology capable of interrogating RNA structures at the transcriptome scale.

1.2 References

1. Gesteland, R.F., Cech, T.R., and Atkins, J.F. *The RNA world*. 3rd ed. 2006, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

2. Strobel, S.A. and Cochrane, J.C. *RNA catalysis: ribozymes, ribosomes, and riboswitches.* Curr Opin Chem Biol, 2007. **11**(6): p. 636-43.

3. Breaker, R.R. *Riboswitches and the RNA world.* Cold Spring Harb Perspect Biol, 2010.

4. Korostelev, A. and Noller, H.F. *The ribosome in focus: new structures bring new insights.* Trends Biochem Sci, 2007. **32**(9): p. 434-41.

5. Sharp, P.A., *The centrality of RNA.* Cell, 2009. **136**(4): p. 577-80.

6. Rich, A. and RajBhandary, U.L. *Transfer RNA: molecular structure, sequence, and properties.* Annu Rev Biochem, 1976. **45**: p. 805-60.

7. Reiter, N.J., Chan, C.W., and Mondragon, A. *Emerging structural themes in large RNA molecules.* Curr Opin Struct Biol, 2011. **21**(3): p. 319-26.

8. Holbrook, S.R., *Structural principles from large RNAs.* Annu Rev Biophys, 2008. **37**: p. 445-64.

9. Tinoco, I., Jr. and Bustamante, C. *How RNA folds.* J Mol Biol, 1999. **293**(2): p. 271-81.

10. Butcher, S.E. and Pyle, A.M. *The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks.* Acc Chem Res, 2011.

11. Haller, A., Souliere, M.F., and Micura, R. *The dynamic nature of RNA as key to understanding riboswitch mechanisms.* Acc Chem Res, 2011.

12. Shajani, Z., Sykes, M.T., and Williamson, J.R. *Assembly of bacterial ribosomes.* Annu Rev Biochem, 2011. **80**: p. 501-26.

13. Duncan, C.D. and Weeks, K.M. *Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding.* Biochemistry, 2010. **49**(26): p. 5418-25.

14. Steen, K.A., Malhotra, A. and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease.* J Am Chem Soc, 2010. **132**(29): p. 9940-3.

15. Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J. *Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch.* Nature, 2006. **441**(7097): p. 1167-71.

16. Al-Hashimi, H.M. and Walter, N.G. *RNA dynamics: it is about time.* Curr Opin Struct Biol, 2008. **18**(3): p. 321-9.

17. Rinnenthal, J., Buck, J., Ferner, J., Wacker, A., Furtig, B., and Schwalbe, H. *Mapping the landscape of RNA dynamics with NMR spectroscopy.* Acc Chem Res, 2011.

18. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P., and Ehresmann, B. *Probing the structure of RNAs in solution.* Nucleic Acids Res, 1987. **15**(22): p. 9109-28.

19. Weeks, K.M., *Advances in RNA structure analysis by chemical probing.* Curr Opin Struct Biol, 2010. **20**(3): p. 295-304.

20. Merino, E.J., Wilkinson, K.A. Coughlan, J.L. and Weeks, K.M. *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).* J Am Chem Soc, 2005. **127**(12): p. 4223-31.

21. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. *RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts.* J Am Chem Soc, 2005. **127**(13): p. 4659-67.

22. Weeks, K.M. and Mauger, D.M. *Exploring RNA structural codes with SHAPE chemistry.* Acc Chem Res, 2011.

23. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.* Nat Protoc, 2006. **1**(3): p. 1610-6.

24. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. *High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states.* PLoS Biol, 2008. **6**(4): p. e96.

25. Gherghe, C.M., Shajani, Z., Wilkinson, K.A., Varani, G., and Weeks, K.M. *Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA.* J Am Chem Soc, 2008. **130**(37): p. 12244-5.

26. Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. *Accurate SHAPE-directed RNA structure determination.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 97-102.

27. Low, J.T. and Weeks, K.M. *SHAPE-directed RNA secondary structure prediction.* Methods, 2010. **52**(2): p. 150-8.

28. Latham, M.P., Brown, D.J., McCallum, S.A., and Pardi, A. *NMR methods for studying the structure and dynamics of RNA.* Chembiochem, 2005. **6**(9): p. 1492-505.

29. Gherghe, C.M., Leonard, C.W., Ding, F., Dokholyan, N.V., and Weeks, K.M. *Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics.* J Am Chem Soc, 2009. **131**(7): p. 2541-6.

30. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. *ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis.* RNA, 2008. **14**(10): p. 1979-90.

31. Mortimer, S.A. and Weeks, K.M. *Time-resolved RNA SHAPE chemistry.* J Am Chem Soc, 2008. **130**(48): p. 16178-80.

32. Duncan, C.D. and Weeks, K.M. *SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA.* Biochemistry, 2008. **47**(33): p. 8504-13.

33. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L., and Weeks, K.M. *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-6.

CHAPTER 2

Selective 2′-Hydroxyl Acylation analyzed by Protection from Exoribonuclease (RNase-detected SHAPE)

2.1 Introduction

The three-dimensional structures of RNAs play direct roles in gene expression and regulation. The sizes of important regulatory elements vary enormously from large catalytic and structural RNAs [1] to the small, compact structures of microRNA precursors, tRNAs, and riboswitches [2-4]. The function of each of these motifs is dependent on the specific local structures and dynamics that characterize each nucleotide.

The local structural environment at most nucleotides in large RNAs can be probed using selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) [5, 6]. SHAPE yields quantitative, nucleotide-resolution, structural information for RNAs ranging in size from small tRNAs to entire RNA genomes [6-8]. SHAPE chemistry takes advantage of the discovery that the reactivity of the 2′-hydroxyl position is highly sensitive to the precise conformation of a given nucleotide. Flexible nucleotides adopt many different conformations, a subset of which increases the nucleophilicity of the 2′-hydroxyl group while

constrained nucleotides sample fewer conformations [9, 10]. Dynamic or conformationally flexible nucleotides react preferentially with electrophilic SHAPE reagents to form 2′-*O*-adducts (Figure 2.1).

In principle, local nucleotide dynamics and flexibility can be assessed rapidly in a simple chemical interrogation step since all four RNA nucleotides react similarly with the acylating reagent [11] and the extent of reaction is quantitatively proportional to local nucleotide dynamics [10, 12]. The sites of 2′-*O*-adduct formation are then detected by primer extension [13].

The use of primer extension to detect 2′-*O*-adducts represents a weakness of SHAPE technology, especially as applied to important short RNAs, and is also a shortcoming for many other probing approaches for analyzing RNA modification chemistries and structure. Using primer extension, RNA adducts are detected by annealing a labeled DNA primer to the 3′ end of a modified RNA and then a reverse transcriptase enzyme is used to extend these primers to the sites of modification. Chemical information is thus read out indirectly, as the lengths and frequency of a given cDNA product, instead of by direct analysis of the RNA fragment. No structural data is obtained for the 40-60 nucleotides at the 3′ end of the RNA due to overlap with the primer binding site and to incomplete enzyme processivity in the early stages of extension. Structural data for 10-20 nucleotides at the 5′ end of the RNA are also obscured due to overlap with full-length extension products. These two features make it impossible to analyze the structures of biologically important short RNAs in their native forms.

The limitations of primer extension-based structure probing are also evident in longer RNA sequences where structural data is not obtained at the ends [6, 14]. A useful solution to

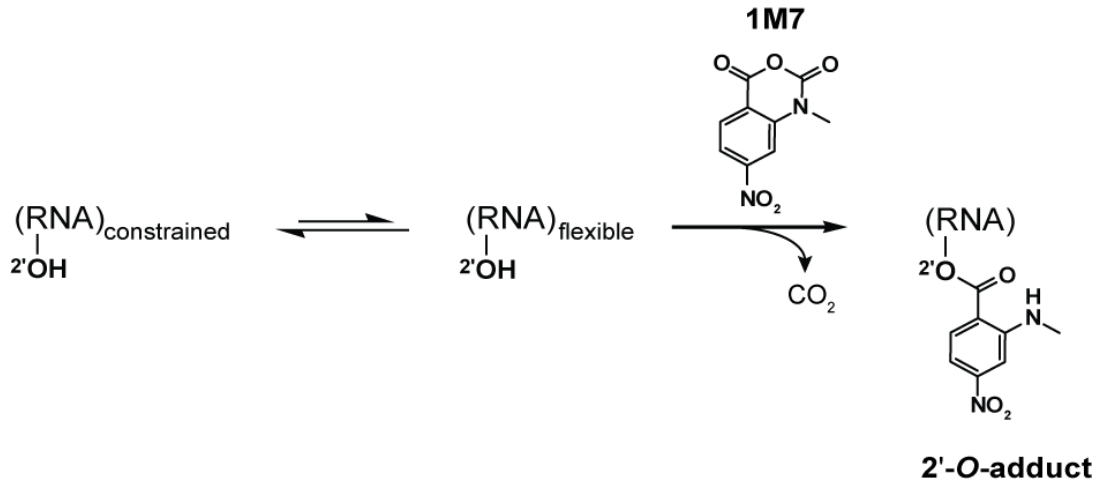


Figure 2.1 Schematic for SHAPE chemistry.

this problem involves appending non-native flanking sequences, a "structure cassette", on both ends of the RNA to move the region of interest to the readable center of the RNA [13]. Finally, although powerful and highly quantitative, the primer extension process requires multiple biochemical manipulations, RNA-specific primer design, and optimization of annealing and extension conditions.

I developed a *direct* method for detecting covalent adducts in RNA, including the 2′-*O*-adducts created by SHAPE, based on adduct-selective protection from exoribonuclease degradation. The RNase R family of exoribonucleases processively and nonspecifically hydrolyzes RNA in the 3′→5′ direction to release 5′-nucleotide monophosphates [15-17]. I screened RNase R enzymes from three organisms and determined that the RNase R enzyme from *Mycoplasma genitalium* degrades structured RNAs, but cannot proceed past 2′-*O*-methyl modifications [18], and is readily inactivated by heat treatment. I then evaluated whether RNase R degradation could be used to detect sites of 2′-*O*-adduct formation in the aptamer domain of the *Escherichia coli thiM* thiamine pyrophosphate (TPP) riboswitch RNA, a small RNA (80 nts) that contains many features common to structured RNAs, such as canonical and non-canonical base-pairing, local stacking, and long-range docking interactions [19].

2.2 Results

2.2.1 RNase R detects sites of 2′-*O*-adduct formation in the TPP riboswitch RNA

I modified a 5′-end labeled TPP riboswitch using 1-methyl-7-nitroisatoic anhydride (1M7) [20] under conditions that stabilize the native fold and in the absence of ligand (100 mM HEPES-NaOH (pH 8.0), 100 mM NaCl, 10 mM $MgCl_2$,; at 37 °C). The modified RNA was then subjected to degradation by RNase R (20 mM Tris-HCl (pH 8.0), 100 mM KCl,

0.25 mM $MgCl_2$; 30 min, 50 °C) and the resulting end-labeled RNA fragments were resolved by gel electrophoresis (Figure 2.2A). Following the degradation step, the RNase R enzyme can be completely heat inactivated by incubation at elevated temperature (95 °C, 3 min). The half-life of *M. genitalium* RNase R is 18.4 s at 95 °C (Figure 2.3). To facilitate comparison with conventional primer extension analysis of the sites of modification, these initial experiments were performed using a TPP riboswitch RNA containing non-native 5′ and 3′ "structure cassette" sequences of 14 and 43 nucleotides, respectively [13].

The RNase R enzyme efficiently degraded the unmodified riboswitch RNA. In contrast, when the RNA was treated with 1M7 and then incubated with RNase R, a strong pattern of banding was observed [compare (–) and (+) 1M7 lanes, Figure 2.2A]. The lengths of these 1M7-modified fragments were determined by comparison with sequencing ladders generated by iodine-mediated cleavage of phosphorothioate-substituted RNA [21] and by RNase R degradation of kethoxal-modified RNA (sequencing lanes, Figure 2.2A). Unexpectedly, bands corresponding to guanosine residues in the two sequencing reactions were offset by 5 nucleotide positions on the gel. Phosphorothioate cleavage results in a 2′,3′-cyclic phosphate fragment [21] that is one nucleotide shorter than the guanosine-terminated fragment, when visualized using 5′-labeled RNA. Taking into account this offset for phosphorothioate cleavage, the net offset for the phosphorothioate- and kethoxal-mediated sequencing reactions is 4 nucleotides.

To determine if RNase R degradation, in fact, recapitulated RNA 1M7 reactivity and to understand the large offset in the sequencing reactions, I performed identical SHAPE reactions on the TPP riboswitch but analyzed the results by conventional primer extension. Absolute reactivities were calculated by subtracting the no-reagent intensities from the 1M7-

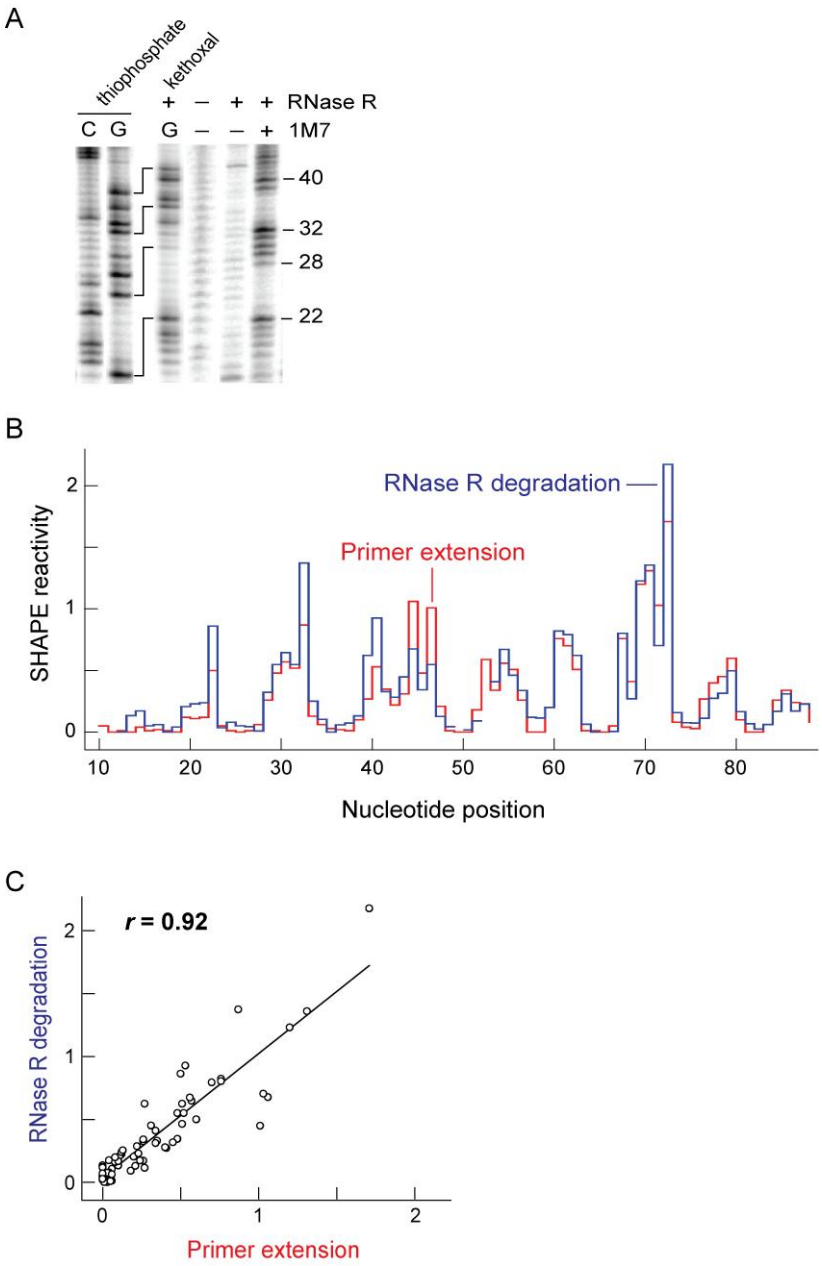


Figure 2.2 Quantitative detection of 2′-*O*-adducts by RNase R degradation. (A) Comparison of RNase R degradation of 5′-end labeled RNAs containing 1M7 adducts with C and G sequencing ladders, visualized by gel electrophoresis. Sequencing ladders were generated by either I_2-mediated cleavage of phosphorothioate-containing RNAs or as kethoxal-mediated stops to RNase R degradation. The kethoxal marker lane fragments (G) are one nucleotide shorter than those in the 1M7 lanes. (B) Histograms comparing absolute SHAPE reactivities as determined by RNase R degradation and primer extension. Nucleotides that could not be analyzed due to high background are indicated by breaks in the plot. (C) Correlation between SHAPE reactivities detected by RNase R degradation and primer extension. Pearson's linear *r*-value is shown.

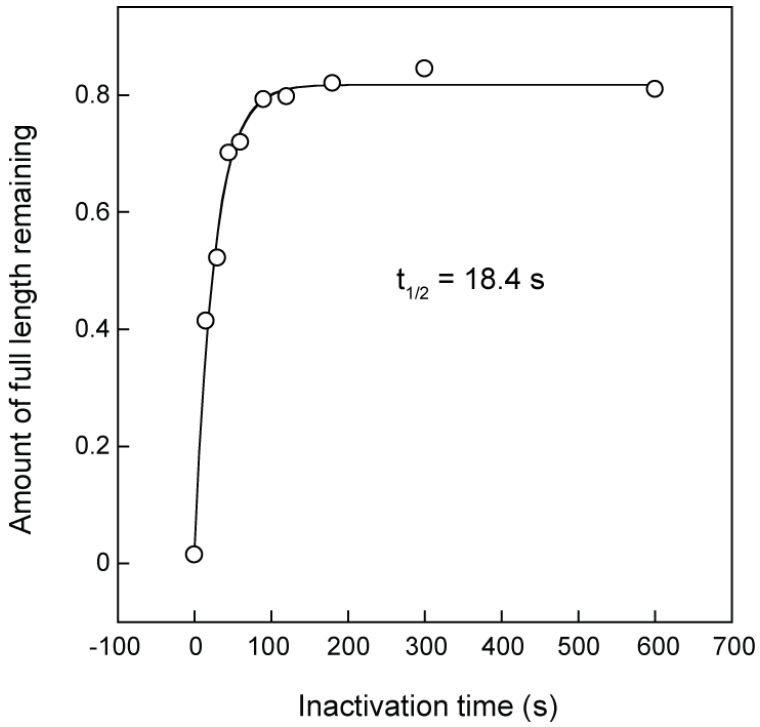


Figure 2.3 Heat inactivation of RNase R at 95 °C and resulting half-life of inactivation.

modified intensities for reactions analyzed by both RNase R degradation and primer extension, (blue and red, Figure 2.2B). Absolute reactivities detected by RNase R degradation were almost identical to those obtained by primer extension (Pearson's linear r = 0.92; Figure 2.2C). However, in order to superimpose the reactivity profiles, RNase R detected bands in the 1M7 reaction and in kethoxal-mediated sequencing required 3- and 4-nucleotide offsets, respectively, to yield agreement with the primer extension-detected reactions.

2.2.2 RNase R detects covalent 2′-*O*- and kethoxal adducts in a two-site model in the enzyme catalytic site

To understand the chemical basis of these offsets, I created a homology model [22, 23] for the *M. genitalium* RNase R enzyme based on the known structures of two close homologs, *E. coli* RNase II (2ix0, 2ix1) [24] and *S. cerevisiae* Rrp44 (2vnu) [25]. The *M. genitalium* RNase R enzyme consists of four major domains: two N-terminal cold shock domains, CSD1 and CSD2; a central, highly conserved, RNA binding (RNB) domain; and a C-terminal S1 domain (Figure 2.4A). The RNB domain contains the RNA substrate-binding channel and the active site for hydrolytic degradation of RNA (Figure 2.4B). The RNA strand is threaded into the RNB domain through an opening between the CSD1 and RNB domains (black strand, Figure 2.4A); makes numerous contacts with protein residues in the RNB domain (Figure 2.4B) and ultimately occupies the active site where phosphate cleavage occurs (see green sphere, Figure 2.4B).

Critically, the enzyme also makes several contacts with RNA nucleotides 5′ of the active site. First, the RNase R enzyme forms a hydrogen bond between the 2′-OH group at nucleotide N-3 (red sphere, Figure 2.4B) and glutamic acid 463, consistent with studies [26] showing this residue is essential for RNA cleavage specificity. Second, serine 433 forms a

hydrogen bond with the base-pairing face of nucleotide N-4 (blue nucleotide, Figure 2.4B). Thus, the RNase structure is consistent with a model in which a 1M7-mediated 2′-*O*-adduct or a kethoxal-mediated cyclic adduct at guanosine [27] cause RNase R degradation to stop either 3 or 4 nucleotides, respectively, 3′ of the site of modification (Figure 2.4B). In sum, RNase R-mediated degradation of end-labeled RNA yields quantitative detection (Figures 2.2B and C) of covalent adducts at both the ribose 2′-OH position in the backbone and at the base-pairing face of guanosine. Sites of adduct formation can be readily assigned by noting that kethoxal-mediated sequencing bands are exactly 1 nucleotide shorter than the corresponding 2′-*O*-adduct (Figure 2.4B). RNase R degradation thus provides a novel, direct and efficient one-step approach for detecting covalent adducts in RNA.

2.2.3 RNase-detected SHAPE yields quantitative structural information for the folding of the TPP riboswitch

Crystallographic and biochemical analyses have yielded a wealth of information about the secondary and tertiary structures of the ligand-bound state of the TPP riboswitch [19, 28-32]. However, a single nucleotide resolution structure of the ligand-free state and the changes in RNA dynamics that occur upon ligand binding are unknown. I, therefore, used RNase-detected SHAPE to analyze the structure of the *native* TPP riboswitch aptamer domain in the absence and presence of the TPP ligand. Significant differences in SHAPE reactivities for the free and ligand-bound states were observed (Figure 2.5). Kethoxal-mediated sequencing at guanosine nucleotides were used to assign bands observed in the (+) and (–) SHAPE reagent lanes (Figure 2.5A). Bands at the top of the gel correspond to undigested full-length RNA and larger RNA fragments that are not fully resolved in this particular electrophoresis run (Figure 2.5A). Bright bands at the bottom of the gel are short

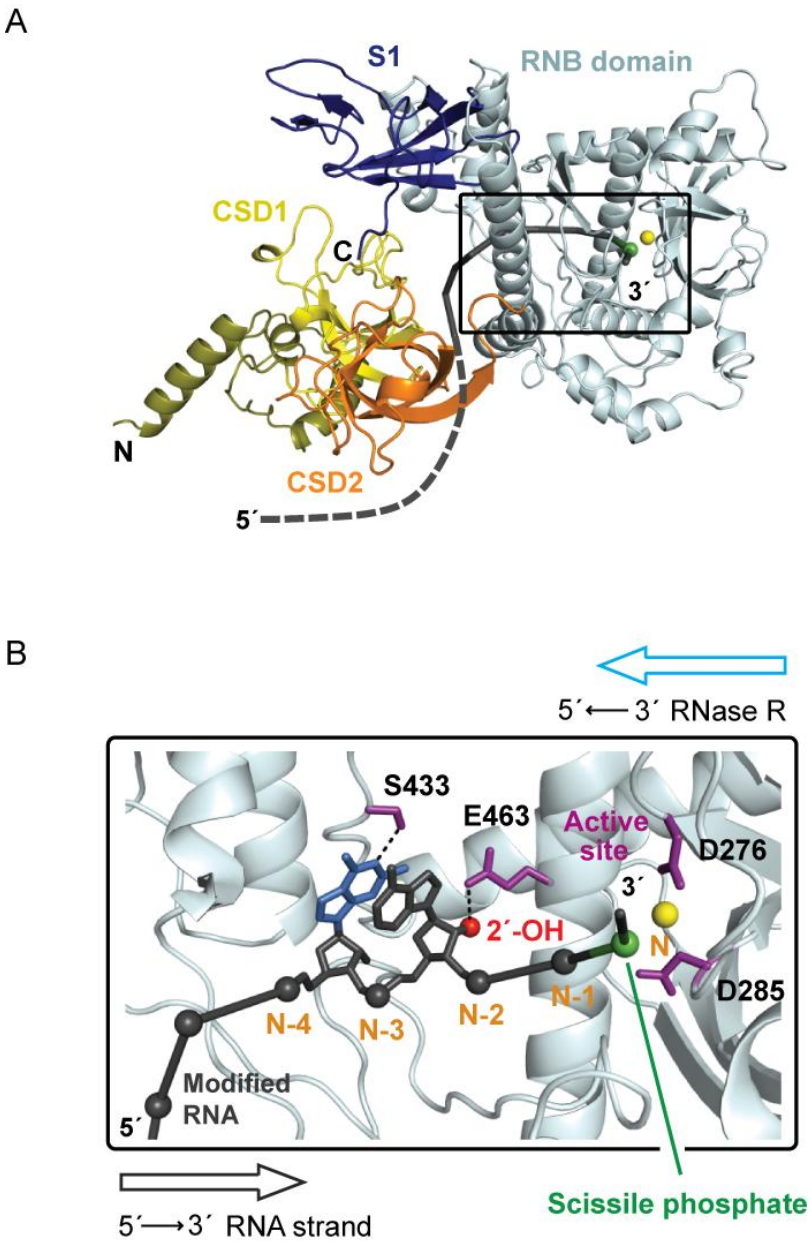


Figure 2.4 Model of *M. genitalium* RNase R and the interactions that mediate covalent adduct detection in RNA. (A) The path of the RNA strand (in dark gray) is shown relative to the major enzyme domains. (B) The substrate-binding channel of the ribonuclease (RNB) domain. Modification of a 2′-hydroxyl group prevents exoribonuclease digestion; the modified residue is shown as a red sphere at N-3. The nucleobase whose base-pairing face is recognized by hydrogen bonding with serine 433 is shown in blue at N-4. The site of RNA strand hydrolysis is shown as a green sphere and a catalytic Mg^{2+} ion is yellow. The homology model was generated using I-TASSER [23].

oligonucleotide fragments that reflect the short RNA "handle" by which RNase R binds RNA, and correspond to the end products of 3′→5′ exoribonuclease digestion. Using RNase R-detected SHAPE, SHAPE reactivities for both the ligand-free and the ligand-bound riboswitch RNA states were resolved and quantified (Figure 2.5B).

Absolute SHAPE reactivities for the ligand-free state were used to create an experimentally supported secondary structure model using RNAStructure [14] (Figure 2.6A). The ligand-free secondary structure is characterized by three significant changes relative to the ligand-bound state. First, all loops in the ligand-free RNA were highly flexible. Second, with the exception of the G19-A47 and A56-G83 base pairs, non-canonical base pairs did not form stably in the ligand-free structure. Third, the P3 helix showed a significant register shift in the ligand-free relative to the ligand-bound state. In the TPP-bound state, nucleotides in P3 have SHAPE reactivities that are exactly consistent with their crystallographically visualized structure: C24 is reactive while all base-paired positions are unreactive (see C24 in Figure 2.6B). In contrast, in the ligand-free state, C24 is unreactive while C22 is reactive, indicating the latter nucleotide is the bulged nucleotide (Figure 2.6A). In sum, the SHAPE data emphasize that, in the ligand-free state, the TPP riboswitch consists of five well-formed and stable helices that are linked by highly dynamic single-stranded regions, suggestive of a structure with little or no stable tertiary interactions.

Upon addition of TPP, large changes in local nucleotide dynamics occur throughout the riboswitch RNA, especially in P3, J3-2, J2-4, L5, and nucleotides 60-62 (compare panels, Figure 2.5B). Several regions that became more constrained in the ligand-bound state reflect direct interactions between the ligand (specifically with the TPP aminopyrimidine and pyrophosphate moieties) and the RNA. These elements include J3-2 and nucleotides 60-61

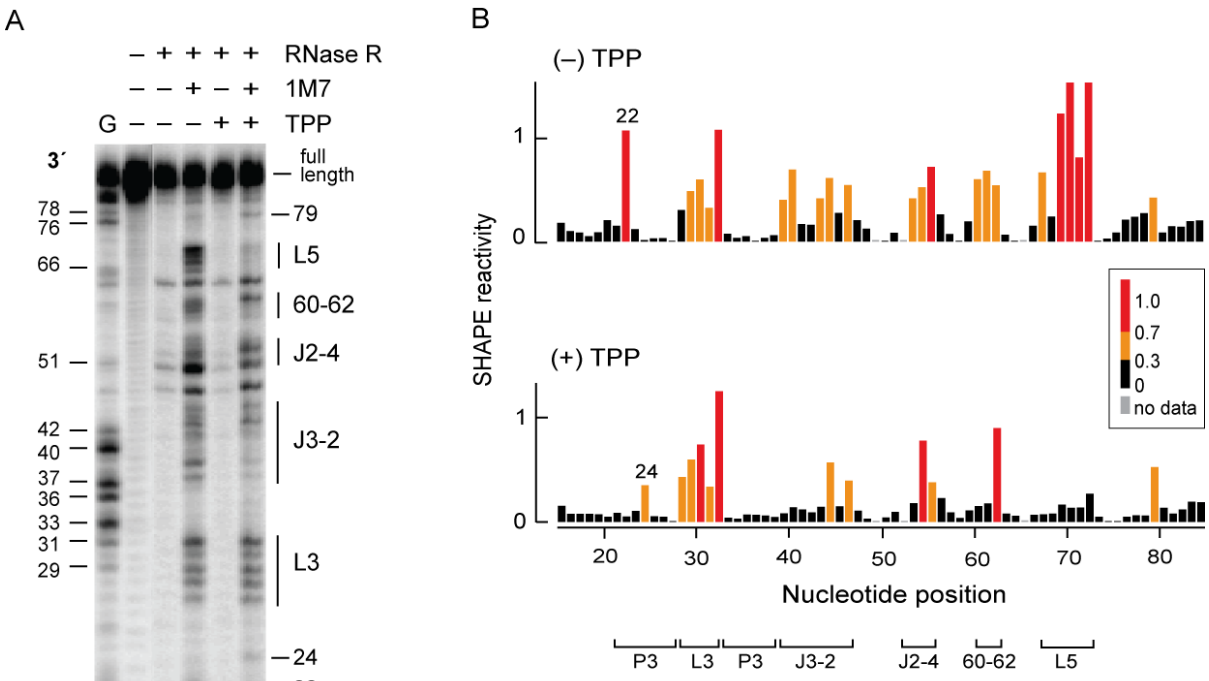


Figure 2.5 Representative RNase-detected SHAPE experiment. (A) SHAPE reactions, performed using the 1M7 reagent, for the (80 nucleotide) TPP riboswitch domain and kethoxal-mediated sequencing (indicated with a G) resolved by denaturing polyacrylamide gel electrophoresis. The guanosine sequencing marker is one nucleotide shorter than bands corresponding to the (–) and (+) 1M7 reactions. Guanosine nucleotides are indicated at left; structural landmarks in the RNA are highlighted on the right. (Lanes are from the same gel) (B) Absolute SHAPE reactivities in the absence (top) and presence (bottom) of TPP ligand. Columns are colored by individual nucleotide SHAPE reactivities (see scale). SHAPE data are normalized to a scale in which zero indicates no reactivity and 1.0 is defined as the average intensity of highly reactive positions and the offsets are corrected.

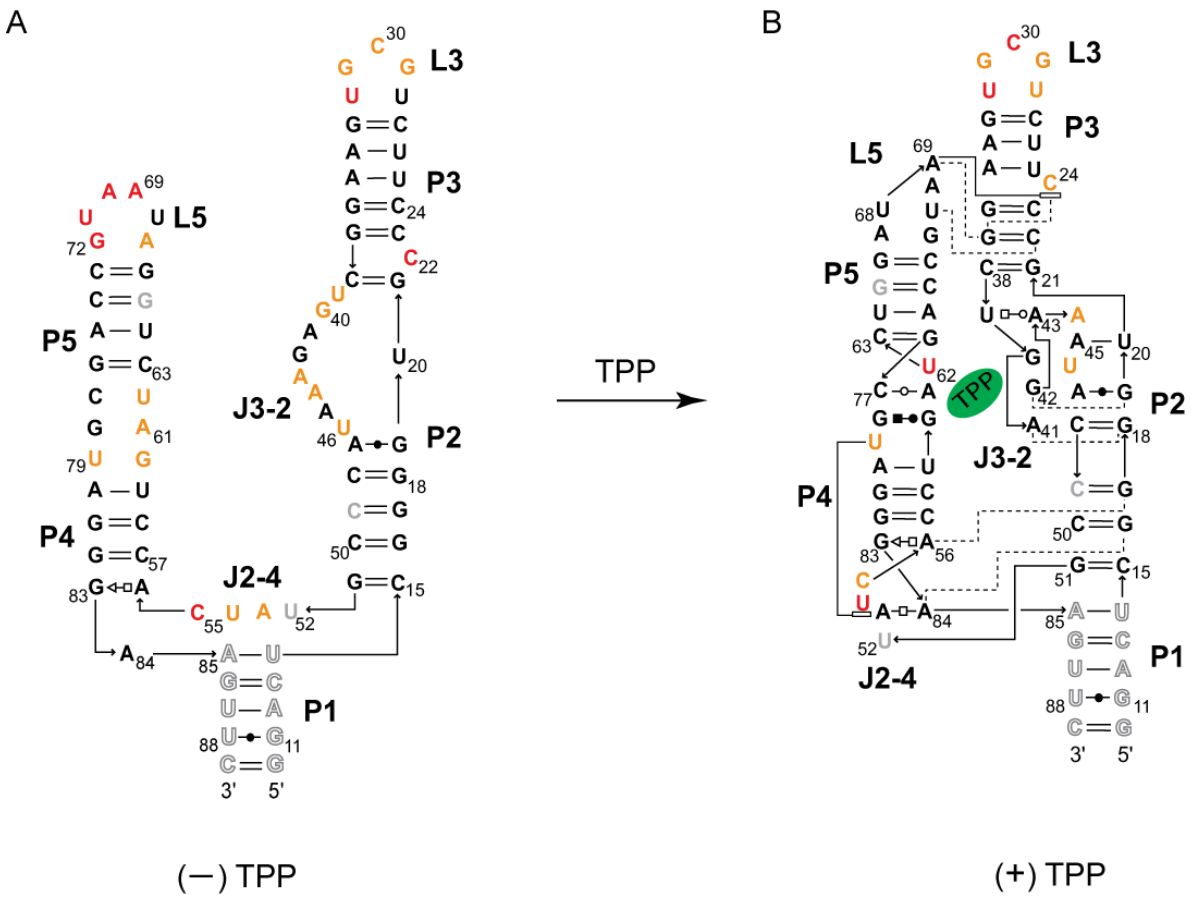


Figure 2.6 Absolute SHAPE reactivities determined by RNase R degradation superimposed on the secondary structure models for the 80 nucleotide TPP riboswitch in the (A) absence and (B) presence of TPP ligand. Nucleotides are colored by SHAPE reactivity using the scale shown in Figure 2.5. All nucleotides in the RNA are shown; the small number of positions for which no data were obtained are gray.

(Figures 2.6B and 2.7A). A few nucleotides near the TPP binding site became more flexible in the ligand-bound state. For example, the SHAPE reactivity of nucleotide 62 increased, consistent with becoming extrahelical to form the pyrophosphate binding pocket for TPP [19, 28] (Figures 2.5B and 2.7B).

Critically, regions that are spatially distant from the ligand binding site also showed large changes in SHAPE reactivity. L5 was highly dynamic in the ligand-free state but became completely constrained in the ligand-bound state, consistent with L5 docking into the minor groove of P3 (Figure 2.5B). The register shift in P3 results in the bulged nucleotide migrating from C22 to C24 and allows C24 to stack with A69 (Figure 2.6B). The L5-P3 interaction forms early during TPP binding [29] and the C24-A69 stacking interaction makes a significant contribution to the stability of the RNA-ligand complex [31]. The large change in SHAPE reactivity observed in L5 nucleotides thus implies that the C22 to C24 bulge migration functions to direct riboswitch folding and to stabilize the structure of the ligand-bound state.

The J2-4 region also undergoes large conformational changes to form interactions that stabilize the three-way junction (Figures 2.5 and 2.6). A53 becomes less reactive in the ligand-bound state, consistent with formation of the A53-G84 non-canonical base pair that stabilizes the three-way junction [19]. The crystallographic data indicate that U54 and U79 form a stacking interaction in the ligand-bound state [19]; SHAPE data indicated that these nucleotides are highly flexible in both free and ligand-bound states. The flexibility in this stacking interaction suggests that, while the interaction may occur transiently, it does not stabilize the ligand-bound state in the isolated aptamer domain.

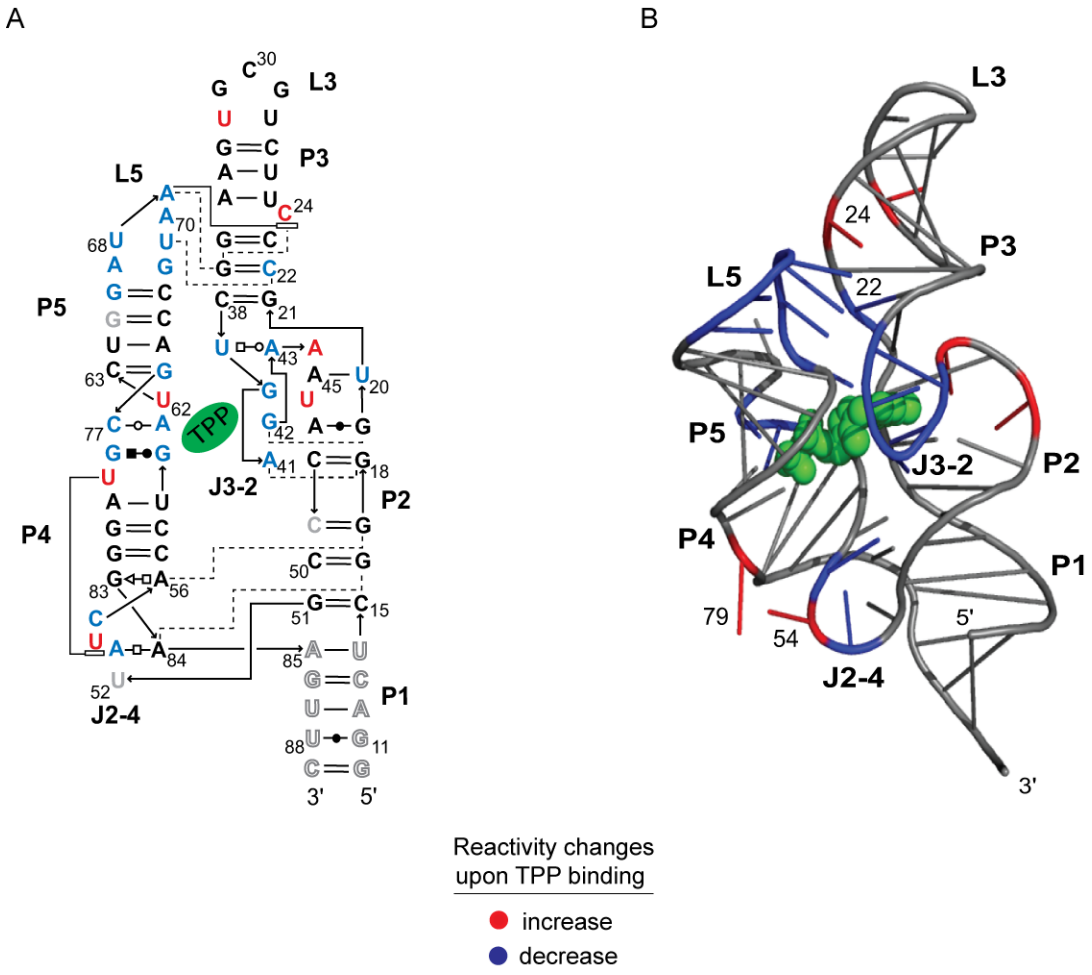


Figure 2.7 Base-pairing (A) and tertiary structure (B) of the ligand-bound TPP riboswitch showing structural features whose constituent nucleotides increase (red) or decrease (blue) in SHAPE reactivity when TPP binds. (PDB 2GDI)

2.3 Discussion

Selective 2′-hydroxyl acylation analyzed by protection from exoribonuclease (RNase-detected SHAPE) combines quantitative and robust structure-selective RNA acylation with a simple one-tube exoribonuclease degradation step for detecting sites of covalent modification in RNA. Degradation by *M. genitalium* RNase R was inhibited by 2′-*O*-adducts and adducts at the base-pairing face of guanosine due to specific, but distinct, interactions in the substrate binding channel of the enzyme (Figure 2.4). RNase-detected SHAPE is thus likely to be broadly useful for identifying and quantifying many additional classes of chemical adducts in RNA. Using this technology, I generated single-nucleotide resolution secondary structure models for both the ligand-free and ligand-bound states of the 80 nt native sequence TPP riboswitch aptamer domain (Figures 2.6A and B); determined the nucleotides undergoing the largest conformational changes upon ligand binding, and characterized a single nucleotide bulge shift from C22 in the ligand-free state to C24 in the ligand-bound state that is partly responsible for stabilizing the ligand-bound RNA structure.

The usefulness of RNase-detected SHAPE is general and it is anticipated that this approach will make possible structural analysis of miRNAs and their precursors, riboswitches, and small non-coding RNAs in their native forms. RNase-detected SHAPE will also facilitate complete analysis of functionally important structures at the 5′ and 3′ ends of large RNAs, including the genomes of RNA viruses.

2.4 Experimental

2.4.1 Screening of candidate RNases

Recombinant RNase R enzymes from *Escherichia coli* [33], *Mycoplasma genitalium* [18] and *Aquifex aeolicus* were evaluated to identify an enzyme that efficiently and non-

specifically degrades RNA (including highly structured RNA), is quantitatively inhibited by 2′-*O*-adducts, and can be fully and permanently inactivated. All three enzymes efficiently degraded structured RNA. The *E. coli* and *M. genitalium* enzymes are additionally readily inactivated by a simple heating step (95 °C, 3 min) whereas the *A. aeolicus* enzyme is very difficult to heat inactivate. 2′-*O*-adducts caused significantly stronger stops with the *M. genitalium* enzyme than with the *E. coli* version. For these reasons, I focused on using the *M. genitalium* enzyme for detection of covalent adducts in RNA. RNase R from *M. genitalium* was purified as described [18] and used for all subsequent experiments. Enzyme purification should be performed using equipment separate from that used for RNA-based research.

2.4.2 Heat inactivation of RNase R

RNase R enzymes require Mg^{2+} for activity and can be immediately inactivated by addition of excess EDTA. The enzyme is also permanently inactivated by heat denaturation. To assess the thermal inactivation of *M. genitalium* RNase R, individual 1 μL aliquots of the enzyme (1.5 μg/μL stock) were incubated at 95 °C [in 20 mM Tris-HCl (pH 8.0), 100 mM KCl]. Each aliquot was removed at specific time points ranging from 15 s to 10 min and placed on ice. To the heat-treated RNase, 8 μL of 5′-[^{32}P]-labeled RNA (~ 1 pmol) in the same buffer containing 0.25 mM $MgCl_2$ were added. Reactions were subjected to the standard exoribonuclease digestion step [20 mM Tris-HCl (pH 8.0), 100 mM KCl, 0.25 mM $MgCl_2$ and 0.45 μg/μL *M. genitalium* RNase R; incubation at 50 °C for 30 min] and resolved on a 10% denaturing polyacrylamide gel. RNase R activity was quantified as the fraction of remaining full length RNA. Although the RNase-detected SHAPE experiment does not require that the enzyme be inactivated, it is recommended that all RNase R-containing solutions are heated for 3 min at 95 °C (~10 half-lives) to prevent introduction of RNase to

other RNA experiments in the laboratory.

2.4.3 Synthesis of *E. coli* thiamine pyrophosphate (TPP) riboswitch RNAs

DNA templates for the aptamer domain of the *E. coli* thiamine pyrophosphate (TPP) riboswitch [19] both with and without 5′ and 3′ structure cassette flanking sequences [13] were generated by PCR [1 mL; 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 250 nM each forward and reverse primer (IDT), 40 nM template (IDT), and 0.025 units/μL Taq polymerase; denaturation at 95 °C, 45 s; annealing at 55 °C, 30 s; elongation at 72 °C, 1 min; 35 cycles]. The PCR product was recovered by ethanol precipitation and resuspended in 200 μL TE [10 mM Tris-HCl (pH 8.0), 1 mM EDTA]. RNA constructs were synthesized by *in vitro* transcription [1 mL; 40 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 10 mM dithiothreitol, 2 mM spermidine, 0.01% (v/v) Triton X-100, 4% (w/v) poly(ethylene) glycol 8000, 2 mM each NTP, 40 μL PCR-generated template, 0.1 mg/mL T7 RNA polymerase; 37 °C; 4 h]. Phosphorothioate-containing RNAs for sequencing were synthesized with the same protocol in 100 μL volumes using 10 μL PCR-generated template and contained 0.2 mM guanosine or cytidine α-thiotriphosphate (Glen Research). RNAs were purified by denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 7 M urea, 29:1 acrylamide:bisacrylamide, 0.4 mm × 28.5 cm × 23 cm; 32 W, 1.5 h), excised from the gel, recovered by overnight passive elution at 4 °C, and precipitation with ethanol. Purified RNAs were resuspended in TE and stored at -20 °C.

Purified RNAs were 5′-[^{32}P]-radiolabeled by: (1) dephosphorylation [300 μL; 50 mM Tris-HCl (pH 8.5), 0.1 mM EDTA, 10 μM RNA (TPP RNA) or 1 μM (phosphorothioate-containing TPP RNA), 300 units SUPERase-In (Ambion), 200 units alkaline phosphatase (Roche); 50 °C; 1 h]; (2) phenol:chloroform:isoamyl alcohol extraction, ethanol

precipitation, and resuspension in TE (storage at -20 °C); and (3) treatment with T4 polynucleotide kinase [20 μL; 80 pmol dephosphorylated RNA, 70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT, 2 μL T4 polynucleotide kinase (NEB, 10,000 units/mL), 80 μCi [γ-^{32}P]-ATP; 37 °C; 30 min]. Phosphorothioate RNA radiolabeling reactions contained 20-30 pmol RNA, 1 μL T4 polynucleotide kinase (10,000 units/mL), and 20 μCi [γ-^{32}P]-ATP. Radiolabeled RNAs were purified by denaturing (8%) gel electrophoresis, excised from the gel, and recovered by overnight passive elution at 4 °C. The purified 5′-[^{32}P]-labeled RNAs were precipitated with ethanol, resuspended in 10 mM HEPES-NaOH (pH 8.0), and stored at -20 °C.

2.4.4 Structure-selective RNA modification

Unlabeled RNA with flanking sequences (5 pmol) or 5′-[^{32}P]-labeled RNA with and without flanking sequences (~1 pmol) in 5 μL 1/2× TE (for modified RNAs analyzed by primer extension) or sterile water (for modified RNAs analyzed by RNase R degradation) was heated at 95 °C for 2 min, cooled on ice, treated with 3 μL 3.3× folding buffer [333 mM HEPES-NaOH (pH 8.0), 333 mM NaCl, 33.3 mM $MgCl_2$], and incubated at 37 °C for 10 min. The ligand (1 μL; 50 μM TPP) or sterile water was added and incubated at 37 °C for 20 min. After incubation, 9 μL of the folded RNA (+/– TPP) was added to 1 μL 80 mM 1M7 (in DMSO) [20] and incubated at 37 °C for 2 min. No-reagent control reactions were performed with 1 μL neat DMSO. The RNA was recovered by ethanol precipitation. For the 5′-[^{32}P]-labeled RNA, a five-fold molar excess of EDTA was added to chelate Mg^{2+} before ethanol precipitation. Unlabeled RNA was resuspended in 10 μL 1/2× TE and 5′-[^{32}P]-labeled RNA was resuspended in 8 μL sterile water after being washed twice with 70% ethanol.

2.4.5 RNase R digestion

5′-[^{32}P]-labeled RNA from the 1M7 or kethoxal modification reaction (8 μL) was supplemented with 1 μL 10× reaction buffer [200 mM Tris-HCl (pH 8.0), 1 M KCl, 2.5 mM $MgCl_2$] and 1 μL *M. genitalium* RNase R [18] (4.5 μg/μL) and incubated at 50 °C for 30 min. [Note: *M. genitalium* RNase R activity is very sensitive to Mg^{2+} concentration.] The RNase R enzyme was inactivated by the addition of 1 μL 100 mM EDTA followed by incubation at 95 °C for 3 min. The RNA fragments were recovered by precipitation with 2.5 vol ethanol plus 1 vol isopropanol, washed once with 70% ethanol, and resuspended in 2 μL sterile water and 7 μL stop dye [96% formamide, 1 mM EDTA (pH 8.0), bromophenol blue and xylene cyanol]. RNA fragments were separated by denaturing polyacrylamide gel electrophoresis (10% polyacrylamide, 29:1 acrylamide:bisacrylamide, 7 M urea, 1× TBE, 0.75 mm × 31 cm × 38.5 cm; 70 W) and visualized by phosphorimaging. Fragments were run for both 1.5 and 3.5-4 h to resolve most nucleotide positions.

2.4.6 Kethoxal modification

5′-[^{32}P]-labeled RNA (1 μL) in 15 μL sterile water was heated at 95 °C for 2 min, cooled on ice, mixed with 2 μL 1 M HEPES-NaOH (pH 8.0), and incubated at 70 °C for 3 min. The RNA was then treated with 2 μL 20 mM kethoxal (in sterile water, from USB) and incubated at 70 °C for 5 min [34]. The reaction was quenched with 20 μL 10 mM boric acid followed by ethanol precipitation, washed twice with 70% ethanol, and resuspended in 8 μL sterile water.

2.4.7 Cleavage of phosphorothioate-containing RNA

5′-[^{32}P]-labeled phosphorothioate-containing RNA was treated with 3 μL 85 mM iodine in ethanol for 3 min at room temperature [21]. The reaction was quenched with stop

dye, and used directly for sequencing.

2.4.8 Primer extension

DNA primers were 5′-end labeled with VIC or NED fluorophores (from Applied Biosystems). Unlabeled RNA from the 1M7 modification reaction (10 μL) was added to a fluorescently labeled DNA primer (5′-VIC-labeled GAA CCG GAC CGA AGC CCG; 3 μL, 0.3 μM) and allowed to anneal at 65 °C for 6 min and then cooled on ice. Reverse transcription buffer [6 μL; 167 mM Tris-HCl (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] and Superscript III (1 μL, 200 units) were added and incubated at 45 °C for 2 min, 52 °C for 20 min then 65 °C for 5 min. The reactions were quenched with 4 μL 50 mM EDTA. The cDNAs were recovered by ethanol precipitation, washed twice with 70% ethanol, dried in a SpeedVac for 5 min, and resuspended in 10 μL deionized formamide. Dideoxy sequencing ladders were produced using unlabeled, unmodified RNA, annealing a 5′-NED-labeled fluorescently labeled DNA primer (same sequence as above) (3 μL, 0.3 μM), and by adding 1 μL 2′,3′-dideoxycytosine triphosphate (10 mM) before addition of Superscript III. cDNA fragments were separated by capillary electrophoresis using an Applied Biosystems 3130 DNA sequencing instrument.

2.4.9 Data analysis

Band intensities of fragments separated by gel electrophoresis were quantified using SAFA (Semi-Automated Footprinting Analysis) [35] and raw capillary electrophoresis traces were analyzed using SHAPEFinder [36]. SHAPE reactivity profiles for both the RNase R degradation- and primer extension-detected experiments were obtained by subtracting the no-reagent background from the (+) reaction intensities. All data sets were normalized by excluding the top 2% of the reactive nucleotides, averaging the next 10% of reactive

nucleotides, and then dividing all intensities by this averaged value.

2.4.10 RNase R structure modeling

A homology model for the *M. genitalium* RNase R was generated using I-TASSER [22, 23, 37] with 2ix0 [24] and 2vnu [25] as the template structures. The center-most model (by RMSD) is shown in Figure 2.4. The RNA strand was generated from the RNAs in the 2ix1 [24] and 2vnu structures; the adenosine residue visualized in the 2ix1 structure at position N-4 was changed to a guanosine residue. RNase R likely uses a two Mg^{2+} mechanism for hydrolytic cleavage. Only one Mg^{2+} ion has been visualized crystallographically and is shown, along with its coordinating aspartic acid residues, in Figure 2.4. Images were composed using Pymol (Delano Scientific). The secondary structure for the ligand-bound TPP riboswitch aptamer domain was adapted from Lescoute and Westhof [38].

2.5 References

1. Holbrook, S.R., *Structural principles from large RNAs.* Annu Rev Biophys, 2008. **37**: p. 445-64.

2. He, L. and Hannon, G.J. *MicroRNAs: small RNAs with a big role in gene regulation.* Nat Rev Genet, 2004. **5**(7): p. 522-31.

3. Rich, A. and RajBhandary, U.L. *Transfer RNA: molecular structure, sequence, and properties.* Annu Rev Biochem, 1976. **45**: p. 805-60.

4. Montange, R.K. and Batey, R.T. *Riboswitches: emerging themes in RNA structure and function.* Annu Rev Biophys, 2008. **37**: p. 117-33.

5. Duncan, C.D. and Weeks, K.M. *SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA.* Biochemistry, 2008. **47**(33): p. 8504-13.

6. Watts, J.M., Dang, K.K., Gorelick, R.J. Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L., and Weeks, K.M. *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-6.

7. Wang, B., Wilkinson, K.A., and Weeks, K.M. *Complex ligand-induced conformational changes in tRNA(Asp) revealed by single-nucleotide resolution SHAPE chemistry.* Biochemistry, 2008. **47**(11): p. 3454-61.

8. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. *High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states.* PLoS Biol, 2008. **6**(4): p. e96.

9. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. *RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts.* J Am Chem Soc, 2005. **127**(13): p. 4659-67.

10. Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).* J Am Chem Soc, 2005. **127**(12): p. 4223-31.

11. Wilkinson, K.A., Vasa, S.M. Deigan, K.E., Mortimer, S.A., Giddings, M.C., and Weeks, K.M. *Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA.* RNA, 2009. **15**(7): p. 1314-21.

12. Gherghe, C.M., Z. Shajani, K.A. Wilkinson, G. Varani, and K.M. Weeks, *Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA.* J Am Chem Soc, 2008. **130**(37): p. 12244-5.

13. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.* Nat Protoc, 2006. **1**(3): p. 1610-6.

14. Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. *Accurate SHAPE-directed RNA structure determination.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 97-102.

15. Vincent, H.A. and Deutscher, M.P. *Substrate recognition and catalysis by the exoribonuclease RNase R.* J Biol Chem, 2006. **281**(40): p. 29769-75.

16. Vincent, H.A. and Deutscher, M.P. *Insights into how RNase R degrades structured RNA: analysis of the nuclease domain.* J Mol Biol, 2009. **387**(3): p. 570-83.

17. Matos, R.G., Barbas, A., and Arraiano, C.M. *RNase R mutants elucidate the catalysis of structured RNA: RNA-binding domains select the RNAs targeted for degradation.* Biochem J, 2009. **423**(2): p. 291-301.

18. Lalonde, M.S., Zuo, Y., Zhang, J., Gong, X., Wu, S., Malhotra, A., and Li, Z. *Exoribonuclease R in Mycoplasma genitalium can carry out both RNA processing and degradative functions and is sensitive to RNA ribose methylation.* RNA, 2007. **13**(11): p. 1957-68.

19. Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J. *Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch.* Nature, 2006. **441**(7097): p. 1167-71.

20. Mortimer, S.A. and Weeks, K.M. *A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry.* J Am Chem Soc, 2007. **129**(14): p. 4144-5.

21. Heidenreich, O., Pieken, W., and Eckstein, F. *Chemically modified RNA: approaches and applications.* FASEB J, 1993. **7**(1): p. 90-6.

22. Wu, S., Skolnick, J., and Zhang, Y. *Ab initio modeling of small proteins by iterative TASSER simulations.* BMC Biol, 2007. **5**: p. 17.

23. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.

24. Frazao, C., McVey, C.E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C.M., and Carrondo, M.A. *Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex.* Nature, 2006. **443**(7107): p. 110-4.

25. Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. *Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family.* Mol Cell, 2008. **29**(6): p. 717-28.

26. Barbas, A., Matos, R.G., Amblar, M., Lopez-Vinas, E., Gomez-Puertas, P., and Arraiano, C.M. *Determination of key residues for catalysis and RNA cleavage specificity: one*

mutation turns RNase II into a "SUPER-ENZYME". J Biol Chem, 2009. **284**(31): p. 20486-98.

27. Shapiro, R. and Hachmann, J. *The reaction of guanine derivatives with 1,2-dicarbonyl compounds.* Biochemistry, 1966. **5**(9): p. 2799-807.

28. Edwards, T.E. and Ferre-D'Amare, A.R. *Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition.* Structure, 2006. **14**(9): p. 1459-68.

29. Lang, K., Rieder, R., and Micura, R. *Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach.* Nucleic Acids Res, 2007. **35**(16): p. 5370-8.

30. Rentmeister, A., Mayer, G., Kuhn, N., and Famulok, M. *Conformational changes in the expression domain of the Escherichia coli thiM riboswitch.* Nucleic Acids Res, 2007. **35**(11): p. 3713-22.

31. Kulshina, N., Edwards, T.E., and Ferre-D'Amare, A.R. *Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch.* RNA, 2010. **16**(1): p. 186-96.

32. Ali, M., Lipfert, J. Seifert, S., Herschlag, D., and Doniach, S. *The ligand-free state of the TPP riboswitch: a partially folded RNA structure.* J Mol Biol, 2010. **396**(1): p. 153-65.

33. Cheng, Z.F. and Deutscher, M.P. *Purification and characterization of the Escherichia coli exoribonuclease RNase R. Comparison with RNase II.* J Biol Chem, 2002. **277**(24): p. 21624-9.

34. Mortimer, S.A., Johnson, J.S., and Weeks, K.M. *Quantitative analysis of RNA solvent accessibility by N-silylation of guanosine.* Biochemistry, 2009. **48**(10): p. 2109-14.

35. Das, R., Laederach, A., Pearlman, S.M., Herschlag, D., and Altman, R.B. *SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments.* RNA, 2005. **11**(3): p. 344-54.

36. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. *ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis.* RNA, 2008. **14**(10): p. 1979-90.

37. Zhang, Y., *Template-based modeling and free modeling by I-TASSER in CASP7.* Proteins, 2007. **69 Suppl 8**: p. 108-17.

38. Lescoute, A. and Westhof, E. *The interaction networks of structured RNAs.* Nucleic Acids Res, 2006. **34**(22): p. 6587-604.

CHAPTER 3

Fingerprinting RNA Tertiary Structure by Differential SHAPE Reactivity Analysis

3.1 Introduction

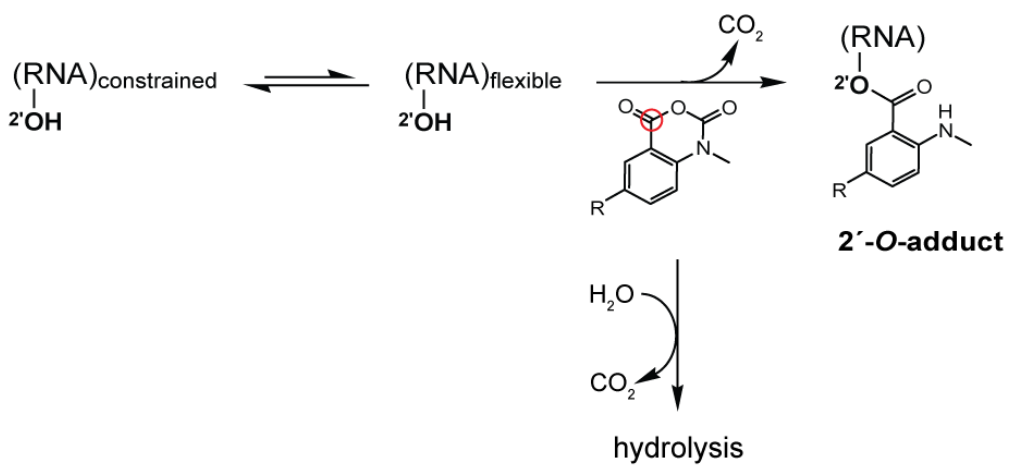
RNA molecules are involved in essentially every aspect of cellular information transfer and regulation [1, 2]. The information encoded in RNA, through which it carries out these functions, is encoded in both the primary sequence and the higher-order secondary and tertiary structure of the RNA [3]. RNA three-dimensional structure is comprised of defined secondary structure elements typically held together by a few key tertiary interactions [4, 5]. These tertiary interactions include long-range stacking interactions, loop-loop and loop-helix contacts, pseudoknots, and well-defined turns in the RNA backbone [5-8]. The regions of an RNA that contain significant tertiary structures play numerous important functional roles, as elegantly revealed in the dramatic increase in available high resolution structures [7, 8]. However, *de novo* detection of nucleotides involved in specific tertiary structure interactions, especially for long RNAs in their native biological environments, remains an unmet challenge.

Nucleotides that participate in either base-pairing or higher-order tertiary structure interactions can often be detected using chemical probing strategies. In general, nucleotides that base-pair or form key tertiary structure motifs tend to be unreactive to solution phase reagents while single-stranded and relatively unstructured elements are reactive [9]. Selective 2′-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) has become an especially informative approach for probing the structure of RNA [9-11]. SHAPE chemistry exploits the discovery that the reactivity of the ribose 2′-hydroxyl is highly sensitive to local nucleotide flexibility (Figure 3.1A). Flexible nucleotides sample many conformations, a few of which preferentially react with hydroxyl-selective, electrophilic reagents to form a 2′-*O*-adduct (Figure 3.1A). The strong relationship between SHAPE reactivity and molecular motion [12] makes it possible to use this chemistry to achieve accurate secondary structure predictions, to monitor RNA dynamics and folding, and to explore RNA-protein interactions [10, 13-15].

SHAPE thus provides a robust measure of molecular motion and makes it straightforward to detect local conformational changes due to base-pairing, formation of tertiary interactions, or induced by protein or small molecule binding. However, in the absence of additional information, it is often not obvious whether a given constraining interaction reflects base-pairing or higher-order tertiary interactions. Consequently, there is a pressing need for a biochemical method that precisely and reliably identifies specific nucleotides involved in tertiary interactions.

To this end, I have developed an approach for rapidly “fingerprinting” RNA tertiary structure derived from additional, orthogonal, chemical reactivity information based on the SHAPE framework. The key insight is that nucleotides with unusual features tend to be

A



B

Reagent	hydrolysis $t_{1/2}$ (s)	R
1M6M	304	CH_3
NMIA	260	H
1M6Br	76	Br
1M6	31	NO_2

increasing electron-withdrawing

Figure 3.1 (A) The mechanism of RNA SHAPE chemistry with the concurrent hydrolysis reaction. The red circle denotes the reactive center of the reagent. (B) The SHAPE reagents used in this chapter are highlighted with their corresponding hydrolysis half-lives.

highly overrepresented in RNA motifs that form tertiary structures. Such features include unusual backbone and stacking geometries [5, 7], prevalence of *syn* nucleotides [16], and nucleotides that undergo local conformational changes on slow timescales [14]. Thus, identification of chemical probes that react selectively with nucleotides in these diagnostic categories would make possible *de novo* discovery of tertiary structures in arbitrary RNAs. Here I explore two such distinctive features using RNAs of known structures.

First, nucleotides that undergo local conformational changes on slow timescales can be selectively detected using SHAPE reagents that also react slowly. These nucleotides are usually in the relatively rare C2′-endo conformation [13] and, in some cases, govern the folding of entire RNA domains [14]. Second, nucleotide base moieties typically form strong, stacking interactions that are almost fully saturated in well folded RNAs [5]. Nucleotides whose bases stack on only one side (one-sided stackers) can be robustly detected using a SHAPE reagent that binds at the unoccupied stacking face.

Using two reagents, N-methylisatoic anhydride (NMIA) and 1-methyl-6-nitroisatoic anhydride (1M6), that differ from each other only in the identity of a single functional group (compare NMIA and 1M6, Figure 3.1B), nucleotides displaying slow, local dynamics and one-sided stacking interactions can be selectively detected. The reactivities of these two reagents can be combined to make possible very concise analysis and discovery of tertiary structure elements by simply calculating the difference in their reactivities. The resulting differential reactivities provide an information-rich "fingerprint" for RNA tertiary structure, as exemplified by the ability to recapitulate tertiary structure information for four RNAs of known structure. Differential SHAPE reactivity analysis was then applied to detect sites of preformed tertiary structure in the *E. coli* thiamine pyrophosphate (TPP) riboswitch as a

function of ionic environment and ligand binding and to understand the formation of tertiary structure contacts in loosely structured RNAs.

3.2 Results

3.2.1 Differential SHAPE analysis of the TPP riboswitch

I initially screened potential SHAPE reagents with the ability to "fingerprint" RNA tertiary structure motifs using the aptamer domain of the TPP riboswitch in the ligand-bound state. The TPP riboswitch has been extensively characterized by x-ray crystallography [17-19], in-solvent dynamics [20-22] and SHAPE chemistry [23]. This RNA contains tertiary structure features that are common to highly structured RNAs including stacking interactions and long-range docking interactions, especially in the ligand binding pocket [17, 19].

Two reagents proved especially promising. NMIA, one of the first reagents used in the SHAPE approach [24], reacts slowly with RNA and correspondingly undergoes the self-inactivating hydrolysis reaction with a half-life of 260 s (Figure 3.1B). The second reagent, 1M6, differs from NMIA by the addition of a nitro (NO_2) group at one end of the double ring system (Figure 3.1B). This modification changes the chemical behavior of the reagent in two ways relative to that of NMIA. First, addition of the electron-withdrawing group increases the electrophilicity of 1M6. Consequently, 1M6 reacts fairly rapidly with both RNA and water with a hydrolysis half-life of 31 s. Second, the NO_2 group significantly changes the electronic distribution of the double ring system. I will show below that this change in electronic profile increases the ability of 1M6 to stack with RNA nucleobases.

When the folded, ligand-bound state of the TPP riboswitch was allowed to react with the slow SHAPE reagent, NMIA, the observed reactivities recapitulate the known structure for the ligand-bound TPP riboswitch (NMIA panels, Figures 3.2 and 3.3A). Similarly, when

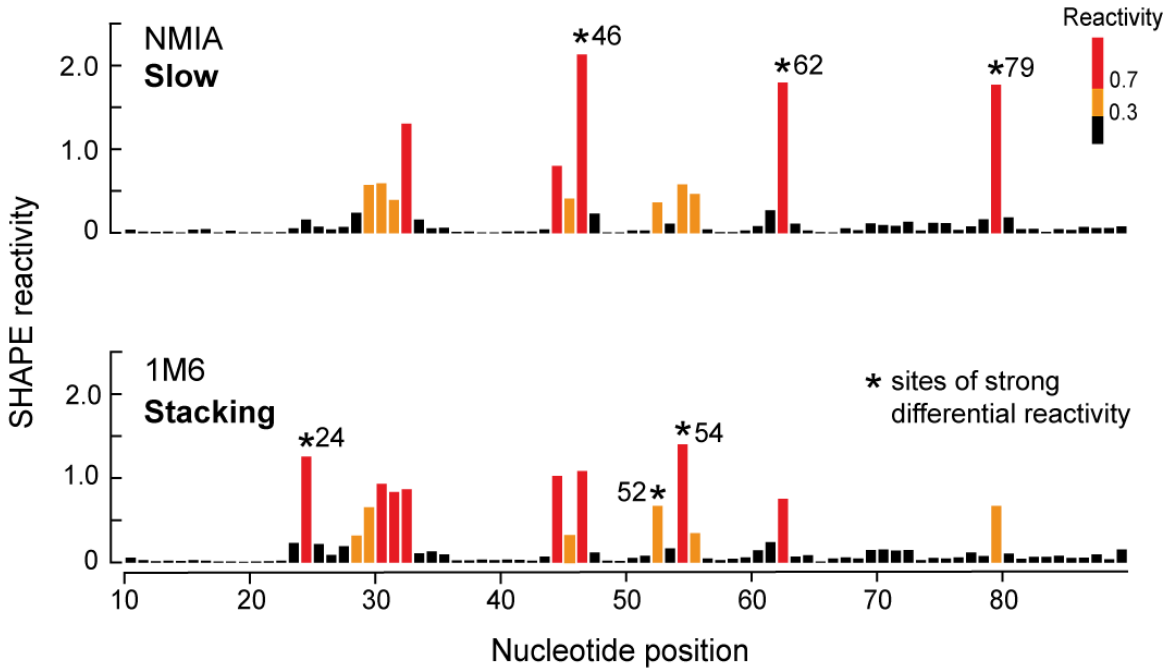


Figure 3.2 Absolute SHAPE reactivities for the ligand-bound state of the TPP riboswitch analyzed by primer extension. Absolute SHAPE reactivities resulting from reaction with NMIA (top) and 1M6 (bottom). Columns are colored by individual nucleotide reactivities (see scale). Nucleotides with asterisks indicate sites of strong differential reactivity between the two reagents.

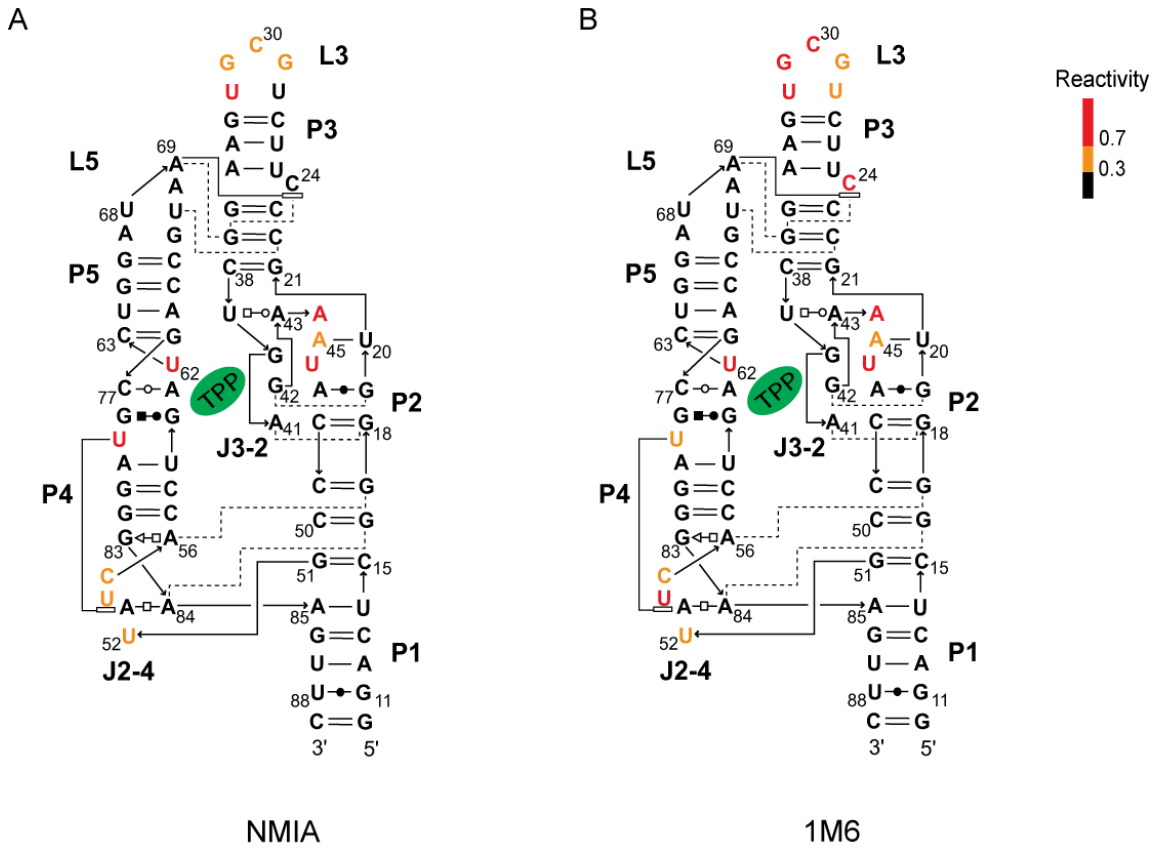


Figure 3.3 Absolute SHAPE reactivities for the ligand-bound TPP riboswitch resulting from reaction with (A) NMIA and (B) 1M6 are superimposed on the planar representation of the three-dimensional structure.

the same RNA was treated with the fast, stacking SHAPE reagent, 1M6, the overall SHAPE reactivity profile was very similar to that for NMIA (1M6 panels, Figures 3.2 and 3.3B). In particular, all base-paired nucleotides are unreactive and many single-stranded nucleotides exhibit similar reactivity towards both reagents (compare structures, Figures 3.3A and B). However, a few nucleotides exhibit strongly enhanced reactivity towards one of the two reagents (asterisks, Figure 3.2). SHAPE chemistry is highly quantitative and reagent-specific reactivities can be readily characterized by simply subtracting one profile from another. After excluding nucleotides that participate in crystal contacts or have poorly-defined electron densities in the crystal structure (gray columns, Figure 3.4), I identified six nucleotides that exhibited significant differential reactivities with either NMIA or 1M6 (green and blue columns respectively, Figure 3.4).

Nucleotides that react preferentially with NMIA over the time course of a SHAPE experiment are best explained as undergoing slow conformational dynamics [13, 14]. SHAPE reagents undergo a concurrent hydrolysis reaction with water (Figure 3.1A); the hydrolysis rate of the reagent determines how long the reagent remains active. Nucleotides that sample SHAPE-reactive conformations slowly react preferentially with NMIA ($t_{1/2}$ = 260 s) because the slow reagent remains in solution longer than does 1M6 ($t_{1/2}$ = 31 s). In addition, all three nucleotides that react preferentially with NMIA occur in the relatively rare C2′-endo conformation (Figure 3.5A).

The mechanism by which nucleotides might react preferentially with 1M6 has not been previously explored. Intriguingly, the three nucleotides exhibiting strong enhancements when treated with 1M6 all occur in diverse local structural contexts but share the feature that one face of the base moiety in the nucleotide is open for potential interactions with the

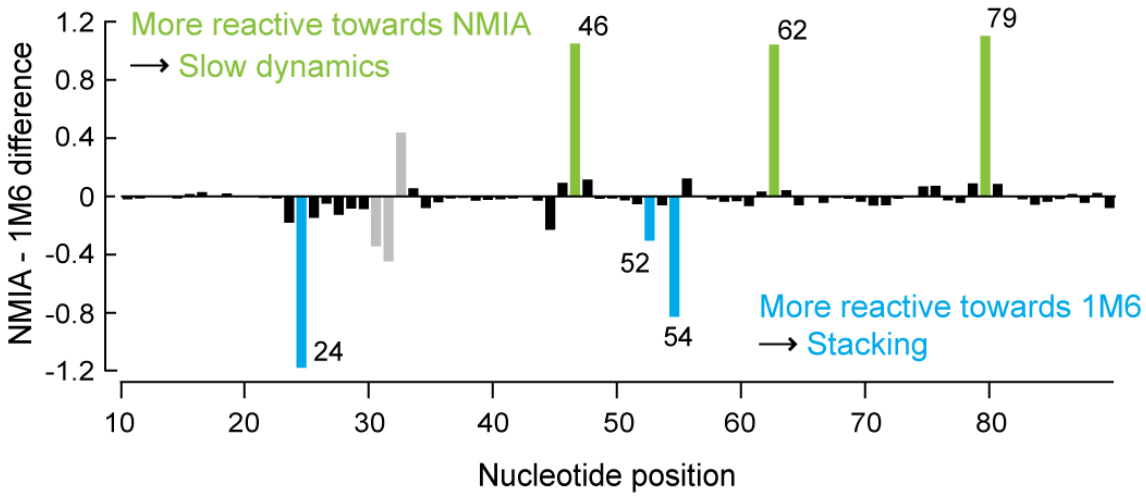


Figure 3.4 Differential reactivities for the ligand-bound state of the TPP riboswitch. Column plot showing the differential SHAPE reactivities calculated by subtracting the 1M6 profile from NMIA. Columns corresponding to nucleotides that exhibit preferential reactivity towards NMIA and 1M6 are colored in green and blue respectively. Gray columns represent nucleotides that exhibit strong differential reactivity but the nucleotide conformations could not be confirmed due to crystal contacts or lack of electron density.

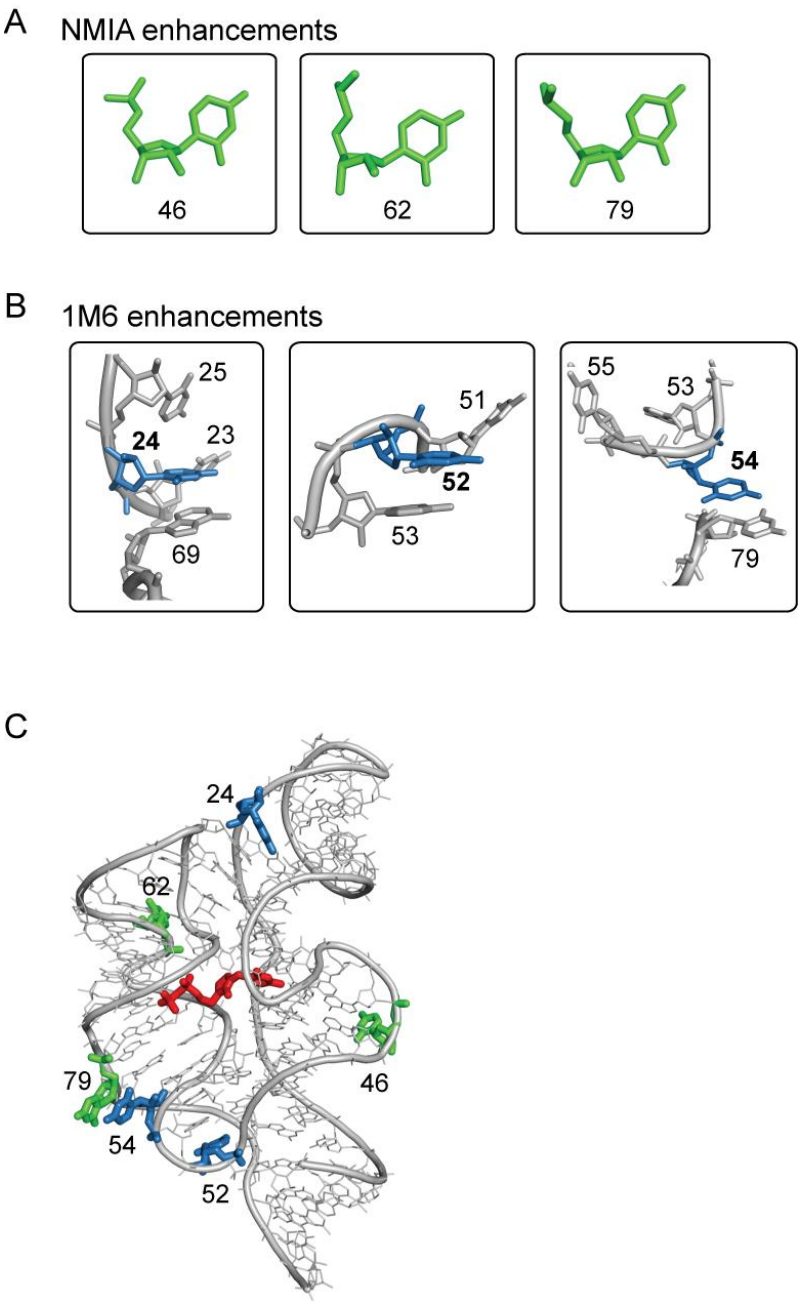


Figure 3.5 Nucleotide conformations and structural context for all the differential reactivities in the ligand-bound state of the TPP riboswitch. (A) NMIA enhancements correspond to C2′-endo nucleotide conformation. (B) 1M6 enhancements correspond to one-sided stacking conformations. (C) NMIA (green) and 1M6 (blue) enhancements are superimposed on the three-dimensional riboswitch structure with TPP shown in red (PDB 2GDI [17]).

reagent (Figure 3.5B). This conformation is unusual because in both A-form helices and most highly folded RNAs, base-base stacking is nearly fully saturated [5, 7]. However, a few nucleotides especially bulged nucleotides, nucleotides involved in backbone turns, and nucleotides at the terminus of some helices form "one- sided" stacking interactions such that one side of the base is available for stable π-π stacking interactions with a small molecule such as 1M6.

These exploratory experiments suggest that the differential reactivity of NMIA vs. 1M6 provides a direct experimental measure of unusual nucleotide conformations that are typically located in tertiary structure motifs in folded RNAs (emphasized with green and blue nucleotides, Figure 3.5C). Each of the three NMIA enhancements corresponds to anucleotide in the C2′-endo conformation. Furthermore, each maps to sites where the RNA backbone has an unusual geometry that contributes either to stabilizing the RNA fold (nucleotide 79, Figure 3.5C) or is adjacent to the ligand-binding pocket (nucleotides 46 and 62, Figure 3.5C). All three 1M6 enhancements report nucleotides that form one-sided stacks. Furthermore, the nucleotides correspond to local structures that form long-range stacking interactions (nucleotides 24 and 54, Figure 3.5C) or turns in the backbone (nucleotide 52, Figure 3.5C).

These differential reactivities create an RNA tertiary structure "fingerprint" by highlighting nucleotides that form relatively rare, but structurally critical, interactions (Figures 3.4 and 3.5C).

3.2.2 Analysis of the mechanism of differential 1M6 reactivity

The observation that 1M6 reacts preferentially at nucleotides that form one-sided stacking interactions is consistent with a model in which the reagent might be binding preferentially at those nucleotides. In this view, the primary contribution of the $–NO_2$

substituent would be to polarize the two-ring system to increase the favorable free-energy contribution of the reagent-nucleobase stacking interaction. If this model is correct, then varying the electron-withdrawing ability of the ring functional group should alter the reagent-nucleobase stacking interaction and thus the SHAPE reactivity.

I probed the ligand-bound state of the TPP riboswitch using structurally similar reagents with substituents that varied from a methyl group (slightly electron-donating), to bromine (moderately electron-withdrawing), to a nitro group (strongly electron-withdrawing) (Figure 3.6A). Increasing the electron-withdrawing ability of the functional group, results in a more electrophilic reactive center due to inductive effect and changes to the overall electrostatic profile of the reagent (compare reagent electrostatic potential maps, Figure 3.6A). The SHAPE reactivity of the one-sided stacking nucleotide, C24, increases monotonically with increasing electron-withdrawing ability of the reagent substituents as reflected by the Hammett coefficient (σ_m) [25] for each functional group (Pearson's linear r = 0.97; Figure 3.6B). The increasing reactivity reflects increasingly favorable reagent-nucleobase stacking interactions and is consistent with a model that as the Hammett coefficient increases, the electron density associated with the aromatic ring decreases and results in decreased π-electron repulsive forces [26, 27].

Since the $-NO_2$ substituent influences the ability of 1M6 to stack preferentially with one-sided stacking nucleotides, it should be possible to calculate a thermodynamic value associated with this interaction. To determine whether the 1M6-nucleobase stacking interaction is more thermodynamically favorable than the NMIA-nucleobase interaction, Greggory, a graduate student, calculated the energy associated with the formation of a stacking complex between NMIA or 1M6 and each of the four RNA nucleotide types (Figure

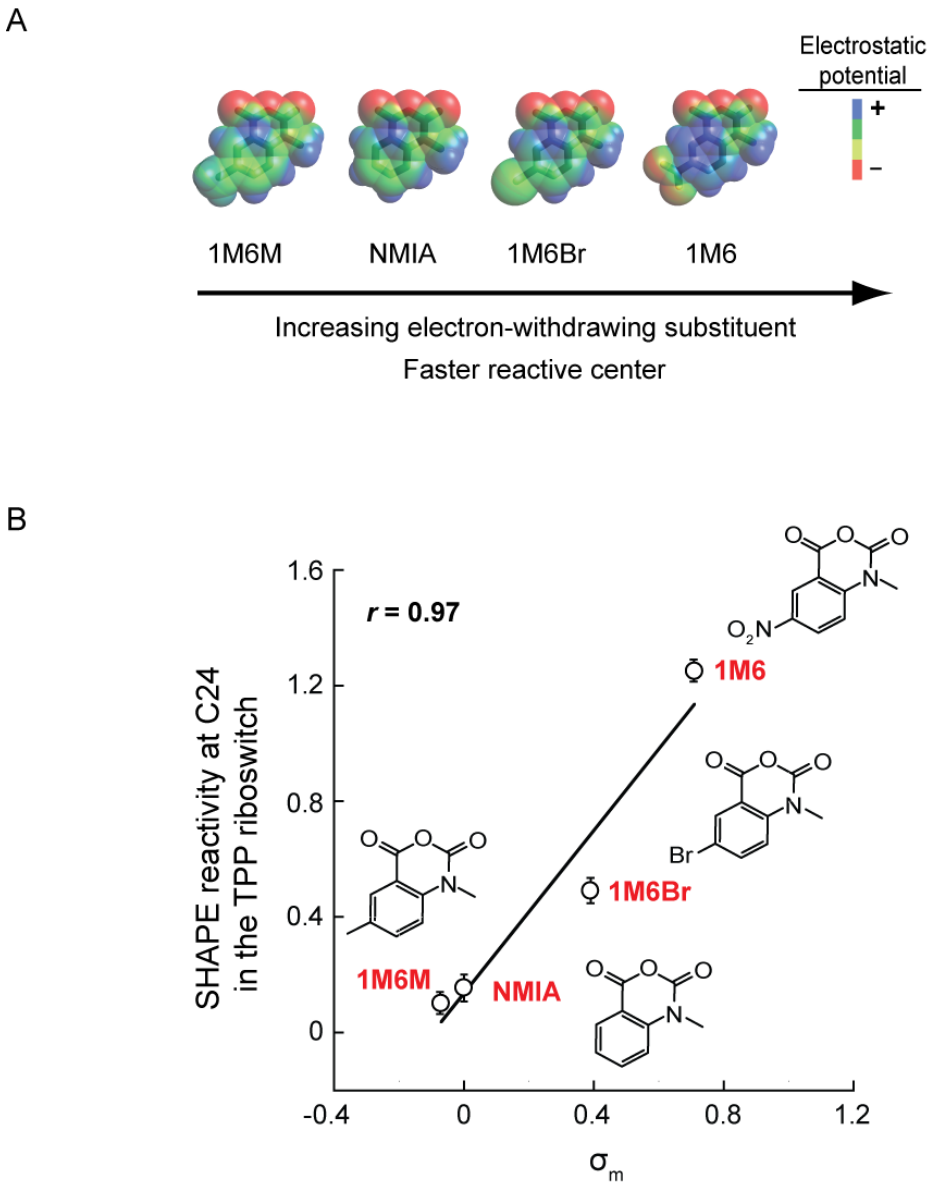


Figure 3.6 Effect of varying electron-withdrawing substituents on the SHAPE reactivity at "one-sided" stacking nucleotides. (A) Electrostatic potential maps for each reagent were calculated using Avogadro [28]. (B) Correlation between SHAPE reactivities at C24 in the ligand-bound TPP riboswitch (three replicates) and the σ_m value of the R group of the reagent. Pearson's linear *r*-value is shown.

A

Cytidine-1M6 complex

Cytidine-NMIA complex

B

	Sugar pucker	1M6 complex enegy (kcal/mol)	NMIA complex energy (kcal/mol)	Stabilization energy (1M6-NMIA)
Cytidine	C3′-endo	-23.0	-18.3	-4.7
Guanosine	C3′-endo	-28.6	-24.0	-4.6
Adenosine	C2′-endo	-21.5	-17.8	-3.7
Uridine	C2′-endo	-20.2	-18.1	-2.1

Figure 3.7 Quantum mechanics calculations for the 1M6-nucleotide and NMIA-nucleotide complexes. (A) The most stable calculated stacking conformations for the cytidine-1M6 and cytidine-NMIA complexes are shown. (B) The more common ribose sugar pucker for each nucleotide type, stacking complex energy for each reagent-nucleotide type, and the stabilization energy [(1M6-complex) – (NMIA-complex)] of each nucleotide type are shown.

	Slow	Stacking
TPP riboswitch	3	3
Lysine riboswitch	2	1
Adenine riboswitch	2	0
RNase P (specificity domain)	3	4
Total number of enhancements from all 4 RNAs	**10**	**8**

Table 3.1 Summary of all the RNAs studied for differential SHAPE reactivity analysis. The total numbers of each type of nucleotide behavior (slow and one-sided stacking) are shown for each RNA.

3.7). I determined the more common ribose sugar pucker (C2′ or C3′-endo) for each nucleotide type (Figure 3.7B) by comparing all one-sided stacking nucleotides observed in the four RNAs investigated in this study (Table 3.1). 1M6-nucleotide complex formation possessed more favorable energies than NMIA-nucleotide complexes across all four nucleotide types (Figure 3.7B). If the stabilization energy is considered, which is the difference between the 1M6- and NMIA-nucleotide complex energies for each nucleotide type, a range of differences between approximately -2 and -5 kcal/mol were observed. On average, the 1M6-nucleotide complexes are -3.8 kcal/mol more stable than the NMIA-nucleotide complexes. This net stabilization energy is especially significant when compared to the approximate net stabilization energy of a two base pair stack of -1.9 kcal/mol [29]. Additionally, while there is a range in the stabilization energies (Figure 3.7B), the 1M6-nucleotide interaction is relatively insensitive to nucleobase identity since all four nucleotide types form more favorable stacking complexes with 1M6 in comparison to NMIA.

In sum, preferential 1M6 reactivity is consistent with a model where one-sided stacking nucleotides form favorable stacking interactions with 1M6. This interaction is relatively insensitive to base identity, quantitative as evidenced by quantum mechanics calculations and can be modulated by decreasing the electron-withdrawing ability of the ring substituent. The most favorable stacking geometries are not likely to position 1M6 for optimal nucleophilic attack by the 2'-OH group; rather stacking appears to enhance reactivity by increasing the effective concentration of the reagent at nucleotides at which one face is available for the one-sided stacking interaction.

3.2.3 Analysis of diverse RNAs by differential SHAPE reactivities

Differential NMIA vs. 1M6 SHAPE reactivities provided a tertiary structure "fingerprint" for the TPP riboswitch by accurately highlighting nucleotides that were involved in key tertiary interactions. To determine whether differential SHAPE reactivity analysis was robustly and broadly applicable to diverse RNA structures, I probed three additional RNAs with well-characterized structures: the RNase P specificity domain [30], and the adenine [31], and lysine [32] riboswitches. When differential SHAPE reactivity analysis was performed on all three RNAs, a unique pattern of preferential reactivities or "fingerprint" was observed that corresponds to both slow nucleotide dynamics and one-sided stacking conformations (green and blue nucleotides and columns respectively, Figure 3.8). For each RNA, the highlighted nucleotides all mapped to specific important tertiary structure interactions. Differential reactivities in the adenine riboswitch are located at the ligand-binding pocket and the crucial loop-loop interaction (nucleotides 47 and 63 respectively, Figure 3.8A). In the lysine riboswitch, differential reactivities were localized to the P2a-L2 turn motif, a structure similar to a canonical kink-turn motif [32], that is important for the formation of the folded RNA structure [33] (nucleotides 38, 40 and 52, Figure 3.8B). In the RNase P specificity domain, while there were many differential reactivities observed, the well-resolved nucleotides were located at key long-range stacking interactions (nucleotides 130 and 194, Figure 3.8C); or at, or adjacent to, tertiary structure motifs (nucleotide 142; sarcin-ricin motif and nucleotide 186, T-loop [34], Figure 3.8C).

Analysis of all four RNA structures revealed a total of 18 nucleotides that displayed reagent-specific preferential reactivties (Table 3.1) and all map to nucleotides that are involved in, or adjacent to, key tertiary structure interactions. Ten nucleotides displayed

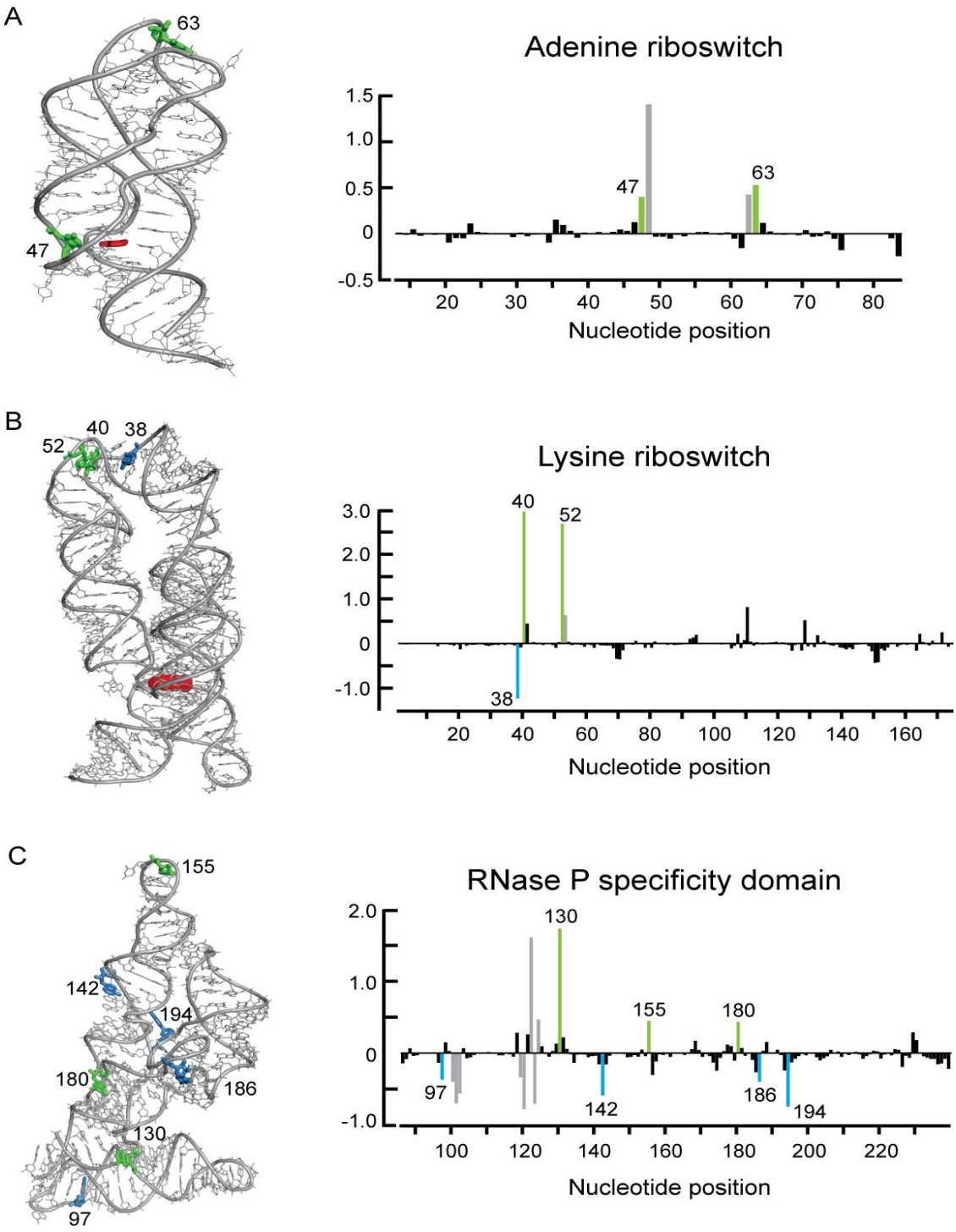


Figure 3.8 Reagent differential reactivities superimposed on the three-dimensional structures (left) and differential reactivity plots (right) of the (A) adenine ribsowitch (PDB 1Y26 [31]), (B) lysine riboswitch (PDB 3DIL [32]), and (C) specificity domain of RNase P (PDB 1NBS [30]). Nucleotides displaying strong differential reactivities are numbered on the tertiary structures and column plots using the same color scheme as Figure 3.4.

preferential reactivity with NMIA and represent slow local nucleotide dynamics. Additionally, nine of the ten nucleotides that exhibit preferential NMIA reactivities were found to be in the C2′-endo ribose conformation. Eight nucleotides displayed preferential reactivities with 1M6 and represent nucleotides that form favorable stacking interactions with 1M6. One of the eight 1M6 preferential stacking nucleotides was found in a partial "one-sided" stacking conformation in which only a fraction of the nucleobase is available for stacking interactions with 1M6. The 18 nucleotides that displayed reagent-specific preferential reactivities were detected out of approximately 470 RNA nucleotides probed. This detection specificity highlights the specialized ability of differential SHAPE reactivity analysis for detecting unique and important nucleotide conformations and dynamics.

3.2.4 Differential reactivities as a function of the extent of RNA structure

Since SHAPE differential reactivity analysis highlights sites of tertiary structure contacts, it was hypothesized that both the magnitude and number of reagent-specific preferential reactivites should decrease as a function of decreased tertiary structure. To test our hypothesis, I probed the less structured, ligand-free state of the TPP riboswitch with both NMIA and 1M6. In the previous chapter, I proposed that the structure of the TPP ligand-free state exists in an open, Y-shaped conformation with little evidence of long-range tertiary interactions [23]. However, structure and folding studies of the TPP riboswitch suggests that the J3-2 region of the TPP binding pocket may exist in a preformed state [1, 18, 20], and other studies suggest that the ligand-free states of many riboswitches are more structured than previously assumed [1, 35].

Two features became evident when the "fingerprints" of the ligand-free and the ligand-bound states of the TPP riboswitch were compared (compare Figures 3.4 and 3.9A).

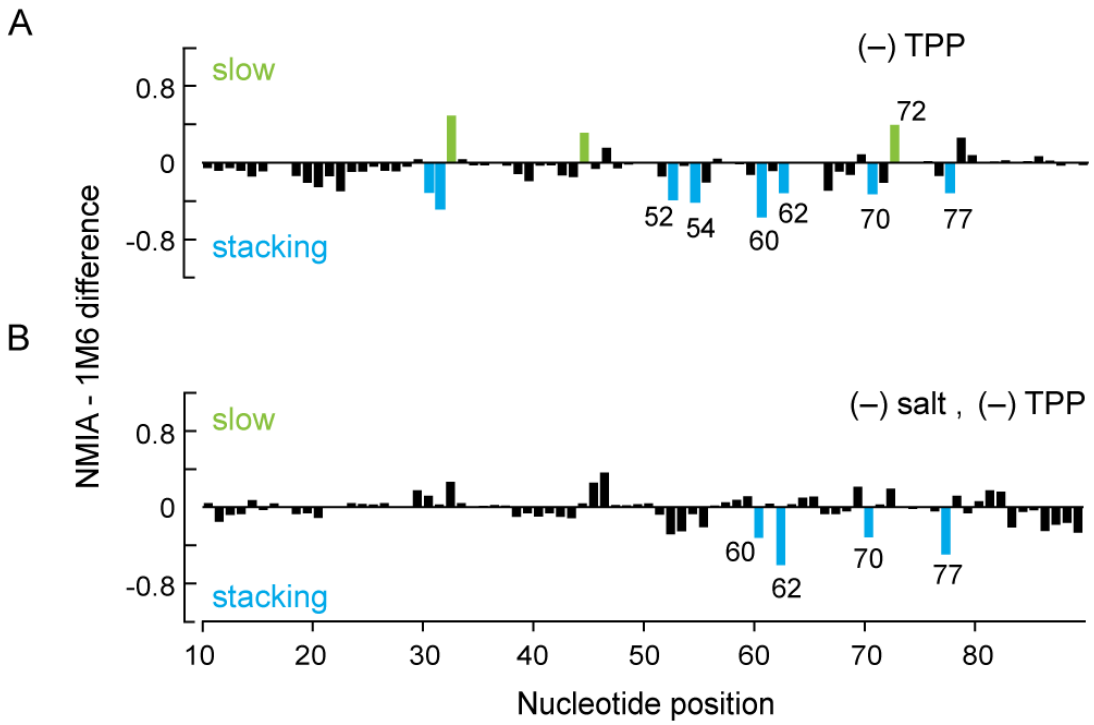


Figure 3.9 Differential SHAPE reactivities as a function of the extent of RNA structure. Differential SHAPE reactivity analysis of the (A) ligand-free and (B) ion-free, ligand-free states of the TPP riboswitch. Nucleotides that show significant enhancements towards NMIA or 1M6 are highlighted (green and blue respectively).

First, the magnitudes of the reagent-specific preferential reactivities in the ligand-free state are much smaller than those in the ligand-bound state (compare green and blue column heights; Figures 3.4 and 3.9A) and are consistent with the initial hypothesis. The intensity or magnitude of the preferential reactivities may provide information about the stability or importance of the tertiary structure motif where smaller enhancements may suggest transiently formed conformations [36] while larger enhancements may suggest a more stable interaction. Surprisingly, however, the number of reagent-specific preferential reactivties in the ligand-free state, especially the one-sided stacking nucleotides, outnumbered those observed in the ligand-bound state (compare numbered columns; Figures 3.4 and 3.9A).

To understand the significance of the observed differential reactivities in the less structured, ligand-free TPP state, I mapped the nucleotide enhancements onto the proposed secondary structure model of the TPP ligand-free state (Figure 3.10B). The nucleotide enhancements of both the ligand-bound state and the ligand-free state in the context of the secondary structures were then compared (compare Figures 3.10A and B). In the less structured, ligand-free state, the differential reactivities correspond to nucleotides that are in two different structural contexts. First, some of the differential reactivities were located in, or adjacent to, the ligand-binding pocket (nucleotides 44, 60, 62 and 77, Figure 3.10B). Second, differential reactivties were also located at nucleotides that form key tertiary structure motifs and interactions in the fully folded ligand-bound structure (nucleotides 52, 54, 70 and 72, Figure 3.10B).

NMIA and 1M6 preferential reactivities in the ligand-free state suggest that both sides of the ligand-binding pockets are preformed (compare J3-2 and P4-5, Figures 3.10A and B). A preformed ligand binding pocket would likely result in rapid recognition and binding of

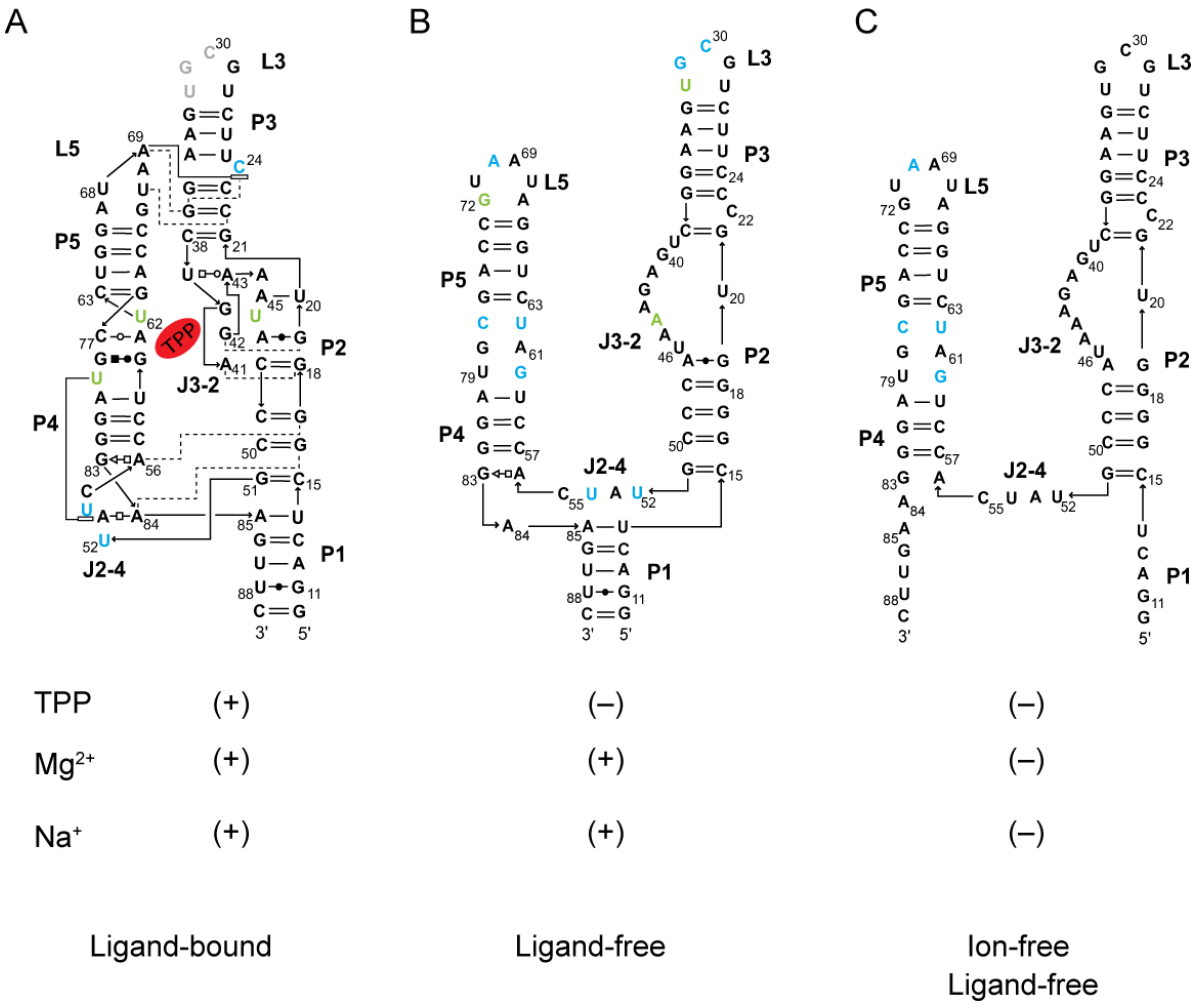


Figure 3.10 NMIA and 1M6 reagent enhancements are superimposed on (A) the planar representation of the three-dimensional structure of the ligand-bound; (B) secondary structure model of the ligand-free and (C) secondary structure model of the ion-free, ligand-free states of the TPP riboswitch. (Color scheme from Figure 3.4)

the TPP ligand, providing a functional explanation for the observed "fingerprint" in the binding pocket. This interpretation is consistent with 2-aminopurine (2AP) fluorescence experiments which probed the folding of the TPP riboswitch. and confirmed that in the presence of TPP ligand, the TPP ligand-free state undergoes fast ligand binding and docking of the L5 nucleotides (k = ~12 × 10^4 M^{-1} s^{-1}) [20]. Additionally, the large number of preferential reactivities in the TPP ligand-free state is likely a feature of riboswitch structure and folding in general. RNA folding and ligand binding in a riboswitch require that the ligand-free RNA adopts a specific structure. The formation of this "ligand-accessible structure" [37] is likely timed by the observed slow nucleotides (nucleotides 44 and 72, Figure 3.10B). Therefore, the numbers and magnitudes of NMIA and 1M6 preferential reactivities can provide information about unknown RNA structures in regards to the stability and possible role of the detected interactions.

To further investigate the effect of decreasing structure on differential SHAPE reactivities, I probed the ligand-free TPP riboswitch in the absence of monovalent and divalent ions (using EDTA chelation) with both NMIA and 1M6. Consistent with the previous hypothesis, the ion-free, loosely structured TPP RNA exhibits fewer reagent-specific preferential reactivities that are smaller in magnitude in comparison to both the ligand-bound and ligand-free TPP structures (compare numbered columns, Figures 3.4, 3.9A and B). However, there are four significant 1M6 one-sided stacking nucleotides observed in the ion-free, loosely structured state that are also present in the ligand-free state of the TPP riboswitch (compare blue columns, Figures 3.9A and B).

To determine whether the four 1M6 stacking nucleotides observed in the ion-free TPP RNA are in the same structural context as those in the ligand-free RNA, a secondary

structure model of the loosely structured TPP RNA was required. I used the absolute NMIA and 1M6 SHAPE reactivities to generate an experimentally supported secondary structure model of the loosely structured RNA using RNAStructure [10] (Figure 3.10C). A comparison of the secondary structures of both ligand-free structures in the presence and absence of ions (monovalent and divalent) reveals that the major structural difference is localized to the formation of the P1 helix (compare P1 structures, Figures 3.10B and C). Since the structure of the ligand-free TPP RNA in the absence and presence of ions is very similar, the four one-sided stacking nucleotides (60, 62, 70 and 77) in the ion-free, loosely structured riboswitch likely serve the same functional role in the riboswitch. These observations suggest that the ligand binding pocket is preformed and formation of this preformed pocket is independent of monovalent or divalent ions.

These experiments with loosely structured TPP RNAs indicate that, as the degree of tertiary structure decreased, the magnitude and number of reagent-specific preferential reactivities also decrease. NMIA and 1M6 preferential reactivities suggest that nucleotides in loosely structured RNAs can adopt specific conformations that allow for favorable but possibly transient, tertiary structure interactions.

3.3 Discussion

This study demonstrates how differences in the reactivities between two SHAPE reagents can be used as robust indicators of local, higher-order, tertiary structure interactions. In this study, across four highly structured RNAs: RNase P specificity domain, TPP, adenine and lysine riboswitches, eighteen nucleotides were detected that were directly involved in, or adjacent to, key tertiary structure contacts or motifs. Differential SHAPE reactivity analysis was then applied to predict sites of preformed tertiary structure in RNAs that were previously

thought to lack higher-order structure. Differential SHAPE reactivity analysis is significant because it allows for any RNA, regardless of size or dynamics, to be probed in a typical laboratory setting. Additionally, the experiment provides relatively straightforward results which can be interpreted by the user.

The novelty of this technique lies in the choice of reagents. By changing the identity of a single functional group in the double ring system, two desired effects were achieved. First, the hydrogen atom on the double ring system in NMIA results in decreased electrophilicity of the reactive center and a slower reacting reagent. Second, the addition of a strongly electronegative functional group in 1M6 results in an increased ability of the reagent to stack with RNA nucleobases. By using both reagents to probe a single RNA state, two different types of nucleotide behaviors can be detected: slowly dynamic and "one-sided" stacking nucleotides. Combined, both nucleotide behaviors provide specific information about nucleotides involved in RNA tertiary structure interactions.

The quantum mechanical calculations indicate that the 1M6-nucleotide interaction is a favorable, quantitative interaction. A reagent that is capable of forming detectable, stable, stacking interactions with RNA nucleotides at specific locations has important implications for drug discovery. The 1M6 scaffold could be used as a starting point for the design of compounds that target RNA nucleotides at key tertiary structure locations that could result in disruption of the folded RNA structure.

Since differential SHAPE reactivities are based on the SHAPE chemistry framework, various tertiary structure motifs, such as pseudoknots, may be silent because they are comprised of highly constrained nucleotides. Differential SHAPE reactivity analysis does not detect all tertiary structure contacts or motifs. However, the nucleotides highlighted by this

method are significant and always map to, or are adjacent to, important tertiary structure interactions.

Differential SHAPE reactivity analysis makes it possible to rapidly probe nucleotides that are involved in key tertiary interactions in diverse RNA structures, especially large RNAs. Differential SHAPE reactivities can provide a starting point for proposing tertiary structure interactions and motifs especially in RNAs not amenable to high resolution structure techniques. I anticipate that differential reactivities and the resulting RNA tertiary structure "fingerprint" can be used to aid the *de novo* prediction of both secondary and tertiary structures. Additionally, differential reactivities may be applied to the prediction of potential ligand-binding and RNA-protein contacts.

3.4 Experimental

3.4.1 Synthesis of RNA constructs

DNA templates for the aptamer domains of the *E. coli* thiamine pyrophosphate (TPP) riboswitch [17], *Vibrio vulnificus* adenine riboswitch [31], *Thermotoga maritima* lysine riboswitch [32], and the specificity domain of the *Bacillus subtilis* RNase P ribozyme [30] with 5′ and 3′ structure cassette flanking sequences [38] were generated by PCR [1 mL; 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 500 nM each forward and reverse primer (IDT), 40 nM template (IDT), and 0.025 units/μL Taq polymerase; denaturation at 95 °C, 45 s; annealing at 55 °C, 30 s; elongation at 72 °C, 1 min; 35 cycles; 72 °C, 10 min]. The PCR products were precipitated with ethanol and resuspended in 200 μL TE [10 mM Tris-HCl (pH 8.0), 1 mM EDTA]. *In vitro* transcription of all RNA constructs [1 mL; 40 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 10 mM dithiothreitol, 2 mM spermidine, 0.01% (v/v) Triton X-100, 4% (w/v) poly(ethylene) glycol 8000, 2 mM each

NTP, 50 μL PCR-generated template, 0.1 mg/mL T7 RNA polymerase; 37 °C; 4 h]. RNAs were purified by denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 7 M urea, 29:1 acrylamide:bisacrylamide, 0.4 mm × 28.5 cm × 23 cm; 32 W, 1.5-2.5 h (based on the length of the RNA)), excised from the gel, recovered by overnight passive elution at 4 °C into 1× TE, and precipitated with ethanol. The purified RNAs were resuspended in 50 μL 1×TE and stored at -20 °C.

3.4.2 RNA folding and modification

The adenine, lysine and TPP riboswitch RNA constructs (5 pmol in 5 μL 1/2× TE) were denatured at 95 °C for 2 min, cooled on ice, treated with 3 μL 3.3× folding buffer [333 mM HEPES-NaOH (pH 8.0), 333 mM NaCl (333 mM KCl for lysine riboswitch), 33.3 mM $MgCl_2$], and incubated at 37 °C for 10 min. The corresponding ligand (1 μL; 50 μM adenine, lysine, or TPP) was added and incubated at 37 °C for 20 min. For RNase P, 5 pmol in 6 μL 1/2× TE were denatured at 95 °C for 2 min, cooled on ice, treated with 3 μL 3.3× folding buffer [333 mM HEPES-NaOH (pH 8.0), 333 mM NaCl, 200 mM $MgCl_2$], and incubated at 37 °C for 20 min. After folding, RNAs were treated with reagent [1 μL; 80 mM for RNase P, adenine and TPP riboswitches (50 mM for lysine riboswitch); in anhydrous DMSO] and allowed to react at 37 °C for 3 min [1-methyl-6-nitroisatoic anhydride (1M6)], 7 min [1-methyl-6-bromoisatoic anhydride (1M6Br)], 22 min [N-methylisatoic anhydride (NMIA)], or 26 min [1-methyl-6-methylisatoic anhydride (1M6M)]. No-reagent control reactions were performed with 1 μL neat DMSO. The RNAs were recovered by ethanol precipitation and resuspended in 10 μL 1/2× TE.

3.4.3 Primer extension

Resuspended RNA from the previous modification reaction (TPP and adenine

riboswitches and RNase P) was added to a fluorescently labeled DNA primer (5′-VIC-labeled GAA CCG GAC CGA AGC CCG; 3 μL, 0.3 μM) and annealed at 65 °C for 6 min and then cooled on ice. Reverse transcription buffer [6 μL; 167 mM Tris-HCl (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] and Superscript III (1 μL, 200 units) were added and incubated at 45 °C for 2 min, 52 °C for 20 min then 65 °C for 5 min. For lysine riboswitch, after addition of the fluorescently labeled primer (5′-VIC-labeled GAA CCG GAC CGA AGC CCG; 3 μL, 0.3 μM), the primer was allowed to anneal at 65 °C for 6 min, 35 °C for 5 min and cooled on ice. Reverse transcription buffer [6 μL; 167 mM Tris-HCl (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] and Superscript III (1 μL, 200 units) were added and incubated at 40 °C for 2 min, 52 °C for 1 h then 65 °C for 5 min. The reactions were quenched with 4 μL 50 mM EDTA. The cDNAs were recovered by ethanol precipitation, washed twice with 70% ethanol, dried and resuspended in 10 μL deionized formamide. Dideoxy sequencing ladders were produced using unlabeled, unmodified RNA, annealing a 5′-NED-labeled fluorescently labeled DNA primer (same primer sequence as above) (3 μL, 0.3 μM) and by adding 1 μL 2′,3′-dideoxycytidine triphosphate (10 mM for adenine and TPP riboswitches and RNase P; 30 mM for lysine riboswitch) before addition of Superscript III. cDNA fragments were separated by capillary electrophoresis using Applied Biosystems 3130 and 3730 DNA sequencing instruments.

3.4.4 Data analysis

Raw capillary electrophoresis traces were analyzed using SHAPEFinder [39]. SHAPE reactivities for each experiment were obtained by subtracting the no-reagent background integrated areas from the (+) reaction integrated areas. All datasets were normalized by excluding the top 2% of the reactive nucleotides, and dividing all nucleotides by the average

of the next 10% of reactive nucleotides. Each experiment was performed in triplicate and the average SHAPE reactivity for each nucleotide for each reagent was determined.

3.4.5 Differential SHAPE analysis

After obtaining averaged SHAPE reactivities for both the 1M6 and NMIA reactions, the 1M6 reactivities were subtracted from the NMIA reactivities to obtain the mean difference for each nucleotide. The student's T-test was performed for each nucleotide position in order to obtain absolute differential reactivities. Nucleotides that exhibit a *p*-value < 0.05 and a mean difference of ≥ 0.3 or ≤ -0.3 SHAPE units were determined to be statistically significant based on the reactivity cutoff for unreactive nucleotides on the SHAPE reactivity scale.

3.4.6 Quantum mechanics calculations

Models of SHAPE reagents and nucleic acids were created separately using the modeling program Avogadro [28]. The simplified nucleic acid model system consisting of a nucleoside with a phosphate at the 3′ position and an alcohol at the 5′ position of the sugar was created for each of the four RNA bases. The charge of the base was neutralized at the phosphate to simplify the gas phase optimization. Nucleotide and SHAPE molecule complexes were based on the modeling of the 1M6 stacking interaction at C24 in the TPP riboswitch [17]. The orientation of the SHAPE molecule was chosen to maximize the overlap of the ring systems and substituent interactions between the molecules, which has been shown to enhance the stacking interactions [40]. Structures were optimized using the default method implemented in the Gaussian 09 package [41] in gas phase using M06-2X/6-311G*. The M06-2X functional has been shown to be robust enough to model stacking interactions of aromatic systems [42] and consistently recover the CCSD(T) CBS π-π interaction energy

[43]. This 6-311G* basis set was chosen for its computational efficiency for a system of this size. Once the models were optimized, a high level single point energy calculation on the optimized structure using M06-2X/6-311+G(2d,p) was done to get accurate high level structure energies. Interaction energies for different SHAPE complexes were calculated by subtracting the SCF energies of the nucleotide and SHAPE structures from the complex structure to get the energy difference at infinite distance allowing for direct comparison between different complexes:

$$E_{interaction} = E_{complex} - E_{SHAPE} - E_{Nucleoside}$$

3.5 References

1. Montange, R.K. and Batey, R.T. *Riboswitches: emerging themes in RNA structure and function.* Annu Rev Biophys, 2008. **37**: p. 117-33.

2. Sharp, P.A., *The centrality of RNA.* Cell, 2009. **136**(4): p. 577-80.

3. Gesteland, R.F., Cech, T.R., and Atkins, J.F. *The RNA world*. 3rd ed. 2006, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

4. Leontis, N.B., Lescoute, A., and Westhof, E. *The building blocks and motifs of RNA architecture.* Curr Opin Struct Biol, 2006. **16**(3): p. 279-87.

5. Butcher, S.E. and Pyle, A.M. *The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks.* Acc Chem Res, 2011.

6. Lescoute, A. and Westhof, E. *The interaction networks of structured RNAs.* Nucleic Acids Res, 2006. **34**(22): p. 6587-604.

7. Holbrook, S.R., *Structural principles from large RNAs.* Annu Rev Biophys, 2008. **37**: p. 445-64.

8. Reiter, N.J., Chan, C.W. and Mondragon, A. *Emerging structural themes in large RNA molecules.* Curr Opin Struct Biol, 2011. **21**(3): p. 319-26.

9. Weeks, K.M., *Advances in RNA structure analysis by chemical probing.* Curr Opin Struct Biol, 2010. **20**(3): p. 295-304.

10. Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. *Accurate SHAPE-directed RNA structure determination.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 97-102.

11. Weeks, K.M. and Mauger, D.M. *Exploring RNA structural codes with SHAPE chemistry.* Acc Chem Res, 2011.

12. Gherghe, C.M., Shajani, Z., Wilkinson, K.A., Varani, G., and Weeks, K.M. *Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA.* J Am Chem Soc, 2008. **130**(37): p. 12244-5.

13. Gherghe, C.M., Mortimer, S.A., Krahn, J.M., Thompson, N.L., and Weeks, K.M. *Slow conformational dynamics at C2'-endo nucleotides in RNA.* J Am Chem Soc, 2008. **130**(28): p. 8884-5.

14. Mortimer, S.A. and Weeks, K.M. *C2'-endo nucleotides as molecular timers suggested by the folding of an RNA domain.* Proc Natl Acad Sci U S A, 2009. **106**(37): p. 15622-7.

15. Duncan, C.D. and Weeks, K.M. *Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding.* Biochemistry, 2010. **49**(26): p. 5418-25.

16. Sokoloski, J.E., Godfrey, S.A., Dombrowski, S.E., and Bevilacqua, P.C. *Prevalence of syn nucleobases in the active sites of functional RNAs.* RNA, 2011. **17**(10): p. 1775-87.

17. Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J. *Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch.* Nature, 2006. **441**(7097): p. 1167-71.

18. Edwards, T.E. and Ferre-D'Amare, A.R. *Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition.* Structure, 2006. **14**(9): p. 1459-68.

19. Kulshina, N., Edwards, T.E., and Ferre-D'Amare, A.R. *Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch.* RNA, 2010. **16**(1): p. 186-96.

20. Lang, K., Rieder, R., and Micura, R. *Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach.* Nucleic Acids Res, 2007. **35**(16): p. 5370-8.

21. Baird, N.J. and Ferre-D'Amare, A.R. *Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis.* RNA, 2010. **16**(3): p. 598-609.

22. Ali, M., Lipfert, J., Seifert, S., Herschlag, D., and Doniach, S. *The ligand-free state of the TPP riboswitch: a partially folded RNA structure.* J Mol Biol, 2010. **396**(1): p. 153-65.

23. Steen, K.A., Malhotra, A., and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease.* J Am Chem Soc, 2010. **132**(29): p. 9940-3.

24. Merino, E.J., Wilkinson, K.A. Coughlan, J.L., and Weeks, K.M. *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).* J Am Chem Soc, 2005. **127**(12): p. 4223-31.

25. Hansch, C. and Leo, A. *Substituent constants for correlation analysis in chemistry and biology*. 1979, NY: Wiley-Interscience.

26. Hunter, C.A., Lawson, K.R., Perkins, J., and Urch, C.J. *Aromatic interactions.* J. Chem. Soc. Perkin Trans., 2001. **2**(5): p. 651-669.

27. Mignon, P., Loverix, S., Steyaert, J., and Geerlings, P. *Influence of the pi-pi interaction on the hydrogen bonding capacity of stacked DNA/RNA bases.* Nucleic Acids Res, 2005. **33**(6): p. 1779-89.

28. *Avogadro: an open-source molecular builder and visualization tool. Version 1.0.1.* . Available from: <http://avogadro.openmolecules.net/>.

29. Petersheim, M. and Turner, D.H. *Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp.* Biochemistry, 1983. **22**(2): p. 256-63.

30. Krasilnikov, A.S., Yang, X., Pan, T., and Mondragon, A. *Crystal structure of the specificity domain of ribonuclease P.* Nature, 2003. **421**(6924): p. 760-4.

31. Serganov, A., Yuan, Y.R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T., Hobartner, C., Micura, R., Breaker, R.R., and Patel, D.J. *Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs.* Chem Biol, 2004. **11**(12): p. 1729-41.

32. Serganov, A., Huang, L., and Patel, D.J. *Structural insights into amino acid binding and gene control by a lysine riboswitch.* Nature, 2008. **455**(7217): p. 1263-7.

33. Blouin, S. and Lafontaine, D.A. *A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control.* RNA, 2007. **13**(8): p. 1256-67.

34. Krasilnikov, A.S. and Mondragon, A. *On the occurrence of the T-loop RNA folding motif in large RNA molecules.* RNA, 2003. **9**(6): p. 640-3.

35. Haller, A., Souliere, M.F., and Micura, R. *The dynamic nature of RNA as key to understanding riboswitch mechanisms.* Acc Chem Res, 2011.

36. Al-Hashimi, H.M. and Walter, N.G. *RNA dynamics: it is about time.* Curr Opin Struct Biol, 2008. **18**(3): p. 321-9.

37. Garst, A.D. and Batey, R.T. *A switch in time: detailing the life of a riboswitch.* Biochim Biophys Acta, 2009. **1789**(9-10): p. 584-91.

38. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.* Nat Protoc, 2006. **1**(3): p. 1610-6.

39. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. *ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis.* RNA, 2008. **14**(10): p. 1979-90.

40. Florián, J., Šponer, J., and Warshel, A. *Thermodynamic parameters for stacking and hydrogen bonding of nucleic acid bases in aqueous solution: ab initio/Langevin dipoles study.* J Phys. Chem. B, 1999. **103**(5): p. 884-892.

41. Frisch, M.J.T.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V.

G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. , *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford CT.

42. Churchill, C.D. and Wetmore, S.D. *Developing a computational model that accurately reproduces the structural features of a dinucleoside monophosphate unit within B-DNA.* Phys Chem Chem Phys, 2011. **13**(36): p. 16373-83.

43. Rutledge, L.R. and Wetmore, S.D. *The assessment of density functionals for DNA-protein stacked and T-shaped complexes.* Canadian Journal of Chemistry, 2010. **88**: p. 815-830.

CHAPTER 4

Hydroxylamine Cleavage of the SHAPE 2′-*O*-adduct Ester Linkage

4.1 Introduction

The cellular functions of most RNAs are influenced by their structure [1-3]. A key first step towards understanding RNA structure-function relationships is accurate secondary structure determination. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) has emerged as a powerful biochemical tool in the determination of RNA secondary structures [4]. SHAPE chemistry is unique because it takes advantage of the inherent reactivity of the ribose 2′-hydroxyl and this feature allows SHAPE to be applicable to almost every nucleotide in an RNA. The ability of the 2′-hydroxyl to deprotonate and react with electrophiles is strongly influenced by both local nucleotide flexibility and chemical environment [4, 5]. Conformationally flexible nucleotides sample many conformations, a few of which increase the nucleophilicity of the 2′-hydroxyl group towards electrophilic SHAPE reagents such as 1-methyl-7-nitroisatoic anhydride (1M7) [6-8]. Constrained nucleotides sample fewer conformations and are less reactive towards SHAPE reagents. The acylation reaction results in the formation of an ester bond between the 2′-oxygen of the ribose and the

reactive carbon center of the reagent (bold line, Figure 4.1). The sites of 2′-*O*-adduct formation are then detected by primer extension. The resulting SHAPE reactivities can be used to generate accurate secondary structure models of RNAs [9, 10].

SHAPE chemistry provides quantitative single nucleotide structural information about local nucleotide flexibility in an RNA and has been very useful for probing the structure of a number of RNAs [7, 9, 11, 12]. The covalent 2′-*O*-adduct is stable and this feature is useful for performing SHAPE experiments in a wide variety of biological environments including *in vitro, in virio*, and *in vivo* [13]. Additionally, the stability of the 2′-*O*-adduct is especially important in biological applications where purification methods such as phenol/chloroform extraction are required before primer extension detection of adducts [9, 13].

While the stability of the 2′-*O*-adduct has been beneficial for our structure probing applications, the modified RNA of interest cannot be used in further biochemical experiments after the SHAPE experiment due to the presence of these stable adducts. For some downstream biochemical manipulations such as reverse transcription coupled with PCR, it would be advantageous to be able to remove this 2′-*O*-adduct while retaining the integrity of the RNA phosphodiester backbone. The stability of the 2′-*O*-adduct ester bond coupled with the inherent ability of RNA to spontaneously undergo phosphodiester backbone breakage [14] makes the development of a chemical cleavage method very challenging.

There are several chemical methods to cleave an ester bond, the most common being hydrolysis. However, initial attempts using acid and base hydrolysis were unsuccessful or in the case of base hydrolysis, resulted in phosphodiester bond cleavage in addition to adduct removal. Additionally, biochemical approaches such as esterases and lipases were also

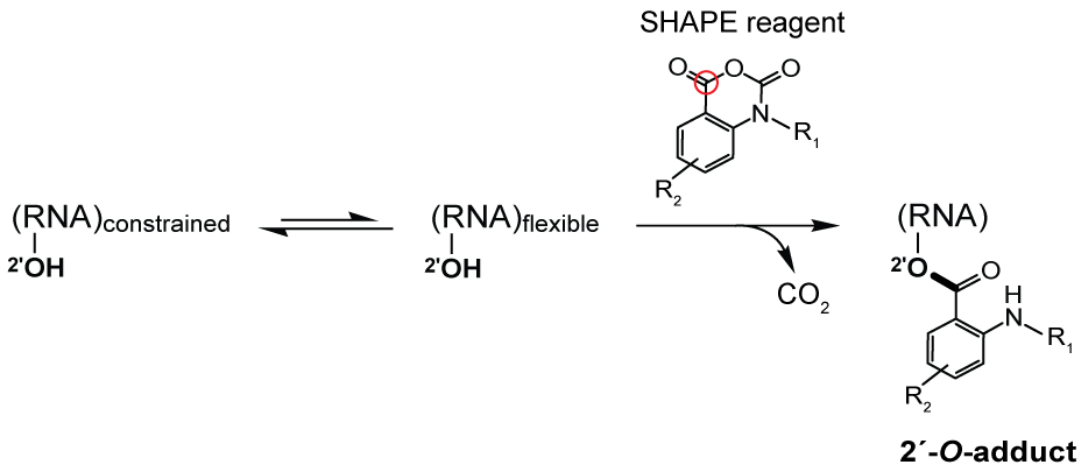


Figure 4.1 Mechanism of RNA SHAPE reaction showing formation of the ester linkage (bold line) between the 2′-hydroxyl of flexible nucleotides and the SHAPE reagent. The reactive center of the SHAPE reagent is circled in red.

unsuccessful (results not shown). Reversal of the 2′-*O*-adduct necessitates a method that selectively cleaves the ester bond but leaves the phosphodiester bond intact.

The ideal 2′-*O*-adduct removal method would (1) be experimentally easy to implement, (2) use a single reagent that is effective over a wide pH range and in various buffer conditions, (3) result in limited RNA backbone cleavage, and (4) be quenched to allow modulation of the extent of adduct removal. In this chapter, I present the development of a method that employs the highly nucleophilic reagent, hydroxylamine (NH_2OH) for cleavage of the ester bond. This simple reagent is a highly reactive, ambient α-nucleophile with two nucleophilic centers, N and O [15-17]. NH_2OH is thought to exist in a specific, dominant tautomeric state depending on the pH, with $^{+}NH_3OH$ at low pH, NH_2OH at neutral pH and NH_2O^{-} at high pH [18]. Additionally, at or near neutral pH, the neutral NH_2OH species exists in equilibrium with its zwitterionic form $H_3^{+}N–O^{-}$ [19-21]. The two nucleophilic centers along with the different tautomeric states account for the high nucleophilicity and wide pH range for nucleophilic attack. In this chapter, I show that the ester bond of SHAPE 2′-*O*-adducts can be chemically cleaved using hydroxylamine (NH_2OH) in biological buffer conditions while retaining RNA backbone integrity. I also show that NH_2OH cleavage of 2′-*O*-adducts can be modulated by the reagent concentration and reaction time.

4.2 Results

4.2.1 Hydroxylamine solution cleaves the ester bond of SHAPE 2′-*O*-adducts

To investigate the ability of hydroxylamine to selectively cleave the ester linkage of the 2′-*O*-adduct while retaining the phosphodiester bond integrity, I first treated the dinucleotide, AddC, with the standard SHAPE electrophile, 1M7. AddC has only one free hydroxyl group which, when modified, can be easily separated from the unmodified

nucleotide on a gel. The addition of 1M7 results in the formation of a covalent 2′-*O*-adduct (see upper band in 1M7 lane, Figure 4.2). In the absence of 1M7, no adducts are observed in the AddC only and (-) 1M7 treated lanes (Figure 4.2). The modified AddC was treated with hydroxylamine solution and resulted in the disappearance of the labeled 2′-*O*-adduct (see NH_2OH treated lane, Figure 4.2). This experiment suggests that the adduct reversal reaction is almost quantitative and proceeds to near completion. Also, the lack of additional bands below the unmodified AddC band suggests that hydroxylamine cleavage does not result in phosphodiester bond cleavage or that reaction with an ester is more favorable than reaction with the phosphodiester bond.

4.2.2 Hydroxylamine cleavage of 2′-*O*-adducts in TPP riboswitch

I then evaluated whether hydroxylamine cleavage of SHAPE 2′-*O*-adducts was applicable to a complex RNA system using the aptamer domain of the TPP riboswitch [22]. The TPP RNA was previously structurally characterized using SHAPE chemistry (chapters 2 and 3) [23] so it was the ideal RNA model system for this work.

The TPP riboswitch was folded under conditions that stabilized the native ligand-free state and then probed with 1M7. Sites of 2′-*O*-adducts were identified as stops to primer extension and the resulting cDNAs were resolved using capillary electrophoresis. The capillary electropherogram for the 1M7-modified RNA has two important features (Figure 4.3A). First, the (-) 1M7 peaks, commonly referred to as background peaks, have low fluorescent intensities and are a result of the processivity of the reverse transcriptase enzyme (green trace, Figure 4.3A). Second, the fluorescent intensities of the (+) 1M7 peaks are distinctly higher than that of the background peaks (compare blue and green traces, Figure 4.3A) and correspond to potential sites of 2′-*O*-adduct formation.

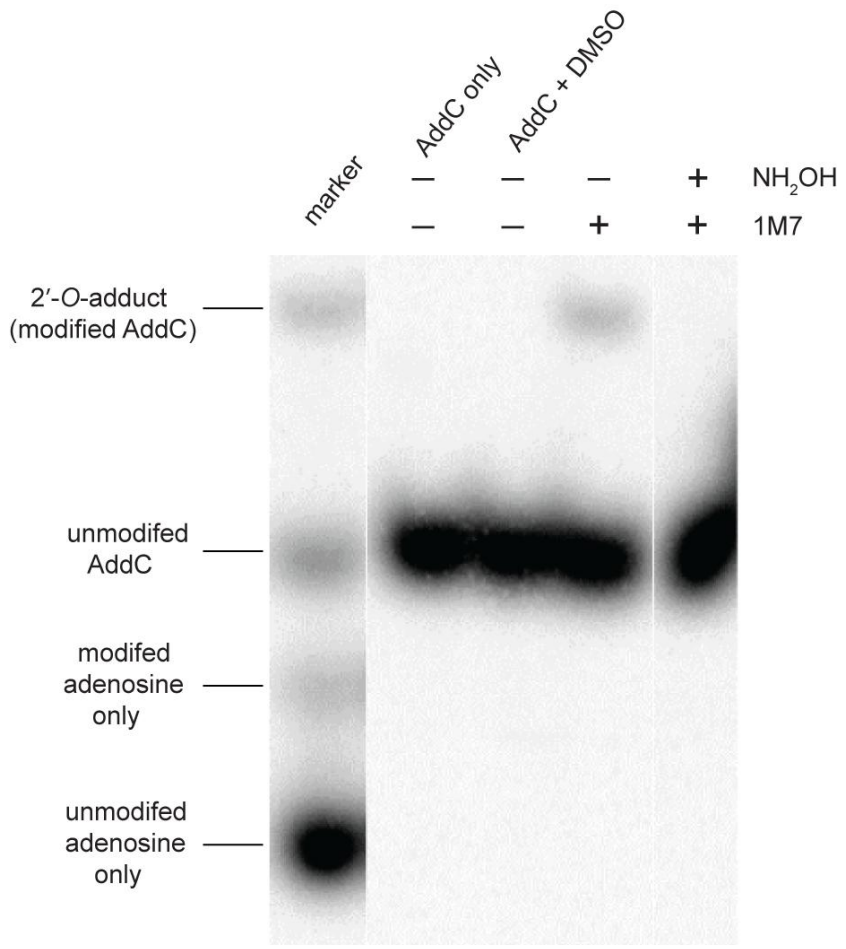


Figure 4.2 Gel showing modification of a model nucleotide, AddC, with 1M7 and removal of 2′-*O*-adduct by hydroxylamine. The left marker lane indicates the position of important bands on the gel corresponding to unmodified and modified mono- and dinucleotide species. (Lanes shown are from the same experiment but cut to show different experimental conditions adjacent to each other.)

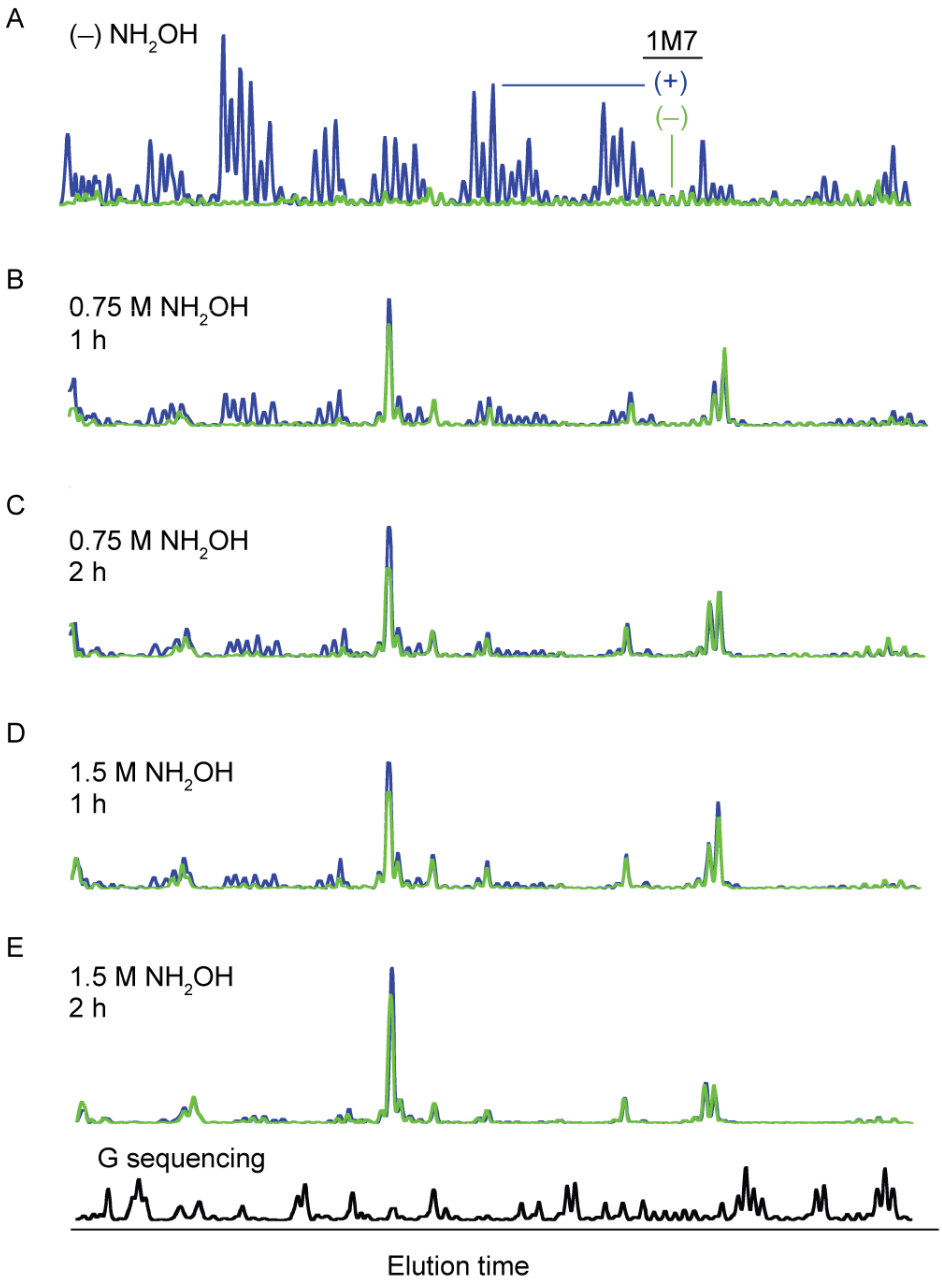


Figure 4.3 Capillary electropherograms showing fluorescent signal corresponding to detection of 2′-*O*-adduct (blue) and background stops (green) as a function of hydroxylamine (NH_2OH) concentration and reaction time. Concentration of NH_2OH added and reaction time (A) 0 M NH_2OH; (B) 0.75 M, 1 h; (C) 0.75 M, 2 h; (D) 1.5 M, 1 h; and (E) 1.5 M, 2 h.

When NH_2OH was added to 1M7-modified TPP riboswitch RNA, the intensities of the (+) 1M7 peaks decreased relative to that of the no NH_2OH experiment (compare blue traces, Figures 4.3A and B-E). As both the concentration of hydroxylamine and reaction time increased, the observed intensities of the (+) 1M7 peaks above background also decreased. In particular, for the 1.5 M NH_2OH reaction with 1M7-modified RNA for 2 h, only a few (+) 1M7 peaks above background were observed (compare blue and green traces, Figure 4.3E). Combined, these observations suggest that (1) NH_2OH can almost quantitatively cleave the ester bond of SHAPE 2′-*O*-adducts in a complex RNA system and (2) the extent of 2′-*O*-adduct removal can be modulated by increasing the NH_2OH concentration and reaction time.

A small increase in the background trace was observed at a few positions in the RNA (Figures 4.3B-E). The increased backbone peaks may indicate a small amount of phosphodiester bond cleavage since NH_2OH has been shown to react with phosphate esters [17, 19]. However, the number of nucleotides with increased background peaks is relatively small, independent of base identity and exhibits little or no additional 1M7 reactivity above background. These results suggest that NH_2OH cleavage results in SHAPE 2′-*O*-adduct ester bond cleavage with limited RNA backbone cleavage.

4.3 Discussion

Although hydroxylamine and the acylation reaction have been studied for many years [15, 16, 24-26], the mechanism of hydroxylamine action is still inconclusive. I sought to propose a plausible mechanism for the reaction of SHAPE 2′-*O*-adducts with NH_2OH that results in the RNA 2′-hydroxyl as the leaving group and is consistent with previous studies. Due to the presence of two nucleophilic centers, hydroxylamine can perform nucleophilic attack via the N or O. At neutral pH, the predominant species is the neutral NH_2OH but

studies indicate that the zwitterionic form, $H_3N^+–O^-$ is as much as ~20% in solution [20]. The presence of this zwitterion may facilitate the O-acylation pathway which is more kinetically favorable [16]. While the O-acylated product is the kinetically favorable product, the N-acylated product is proposed to be the more thermodynamically stable product because in the presence of excess NH_2OH, the O-acylated product converts to the N-acylated product [16, 24]. The second factor in this reaction mechanism is the rate-limiting step. This step is proposed to be influenced by the pKa of the leaving group. Since the pKa of the ribose leaving group is likely greater than 7, the rate-limiting step is thought to be the decomposition of the tetrahedral intermediate [16].

Under our experimental conditions, I propose that ester bond cleavage using NH_2OH occurs mainly via O-acylation in a stepwise mechanism with the rate-limiting step being the decomposition of the tetrahedral intermediate (Figure 4.4). However, since the neutral NH_2OH is the predominant species under our buffer conditions, the neutral NH_2OH may predominately function to replenish the reacted zwitterionic form $H_3N^+–O^-$ to maintain the equilibrium. Additionally, the N-acylation reaction pathway may also occur.

Using the simple nucleophilic reagent, NH_2OH, I was able to remove the SHAPE 2′-*O*-adduct by ester bond cleavage. The extent of ester removal can be modulated by varying both the NH_2OH concentration and the reaction time. Hydroxylamine cleavage of 2′-*O*-adducts is simple to implement and results in very limited increase in RNA background degradation.

In particular, this method is ideal for producing intact RNA molecules for downstream biochemical applications. One immediate application of this method is in the coupling of SHAPE technology with next generation high-throughput sequencing (seq)

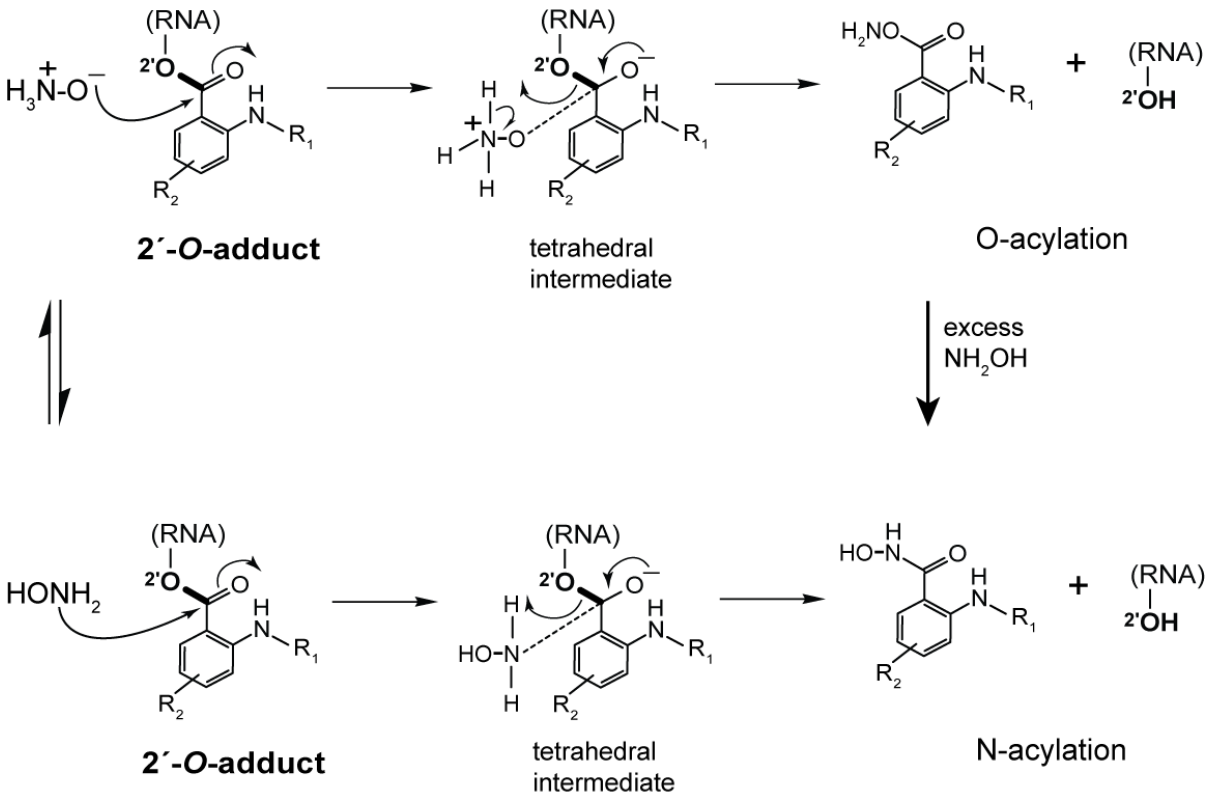


Figure 4.4 Proposed mechanism consistent with hydroxylamine reactivity and SHAPE 2′-*O*-adduct ester bond cleavage.

readouts. In the proposed SHAPE-seq experiment, previously modified RNA is treated with NH_2OH cleavage and results in 2′-*O*-adduct removal. The resulting RNA fragments can then be ligated with known RNA adapter sequences which is then followed by reverse transcription and PCR [27]. This procedure would result in a library that is compatible with RNA next generation high-throughput sequencing methods.

4.4 Experimental

4.4.1 Synthesis of [^{32}P]-labeled AddC

Adenosine-2′,3′-dideoxycytidine (18.75 μM final) (Dharmacon) was 5′-[^{32}P]-labeled [10 μL; 70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 1.5 μL T4 polynucleotide kinase (PNK, 10,000 units/mL), 60 μCi [γ-^{32}P]-ATP; 37 °C; 1 h]. The radiolabeled AddC dinucleotide was purified by gel electrophoresis (30% polyacrylamide, 29:1 acrylamide:bisacrylamide, 0.4 mm × 28.5 cm × 23 cm; 30 W, 1 h), excised from the gel, recovered by overnight passive elution at 4 °C into 300 μL TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and separated from solid acrylamide by microfiltration (EZ spin columns, Millipore).

4.4.2 Detection of 2′-*O*-adduct cleavage in AddC

2′-*O*-adducts were formed by reacting 5′-[^{32}P]-labeled AddC (10,000 cpm/μL) in 1.1× buffer [111 mM HEPES-NaOH (pH 8.0), 111 mM NaCl, 11 mM $MgCl_2$] with 1M7 (1 μL, 100 mM in anhydrous DMSO) at 37 °C for 2 min. Hydroxylamine solution (1 μL, 15.1 M) was added to the 1M7-treated AddC for 30 min. Products were resolved by gel electrophoresis (30% polyacrylamide, 29:1 acrylamide:bisacrylamide, 0.4 mm × 28.5 cm × 23 cm; 30 W, 1 h); and visualized by phosphorimaging.

4.4.3 Synthesis of thiamine pyrophosphate riboswitch RNA

The DNA template for the aptamer domain of the *E. coli* TPP riboswitch with 5′ and 3′ structure cassette flanking sequences [6] were generated by PCR [1 mL; 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 250 nM each forward and reverse primer (IDT), 40 nM DNA template (IDT), and 0.025 units/μL Taq polymerase; denaturation at 95 °C, 45 s; annealing at 55 °C, 30 s; elongation at 72 °C, 1 min; 35 cycles]. The PCR product was recovered by ethanol precipitation and resuspended in 200 μL 1× TE [10 mM Tris-HCl (pH 8.0), 1 mM EDTA]. RNA constructs were synthesized by *in vitro* transcription [1 mL; 40 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 10 mM dithiothreitol, 2 mM spermidine, 0.01% (v/v) Triton X-100, 4% (w/v) poly(ethylene) glycol 8000, 2 mM each NTP, 40 μL PCR-generated template, 0.1 mg/mL T7 RNA polymerase; 37 °C; 4 h]. RNAs were purified by denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 7 M urea, 29:1 acrylamide:bisacrylamide, 0.4 mm × 28.5 cm × 23 cm; 32 W, 1.5 h), excised from the gel, recovered by overnight passive elution at 4 °C, and precipitation with ethanol. Purified RNAs were resuspended in 50 μL 1× TE and stored at -20 °C.

4.4.4 RNA 2′-*O*-adduct cleavage in TPP riboswitch

TPP RNA (5 pmol) in 6 μL 1/2× TE was heated at 95 °C for 2 min, cooled on ice, treated with 3 μL 3.3× folding buffer [333 mM HEPES-NaOH (pH 8.0), 333 mM NaCl, 33.3 mM $MgCl_2$], and incubated at 37 °C for 30 min. After incubation, 9 μL of the folded RNA was added to 1 μL 80 mM 1M7 (in anhydrous DMSO) and incubated at 37 °C for 2 min. No-reagent control reactions were performed with 1 μL neat DMSO. The RNA was recovered by ethanol precipitation and resuspended in 10 μL 100 mM HEPES-NaOH (pH 8.0) after being washed twice with 70% ethanol. The RNA from the modification step (10 μL) was treated

with either 0.75 M or 1.5 M (final) of HCl-buffered hydroxylamine solution (pH 6.9) and incubated at room temperature for either 1 or 2 hours. Reactions were quenched using ethyl acetate (3 μL, 10.2 M). The RNA was recovered by ethanol precipitation, washed twice with 70% ethanol and resuspended in 10 μL 1/2× TE.

4.4.5 Primer extension

DNA primers were 5′-end labeled with VIC or NED fluorophores (from Applied Biosystems). RNA from the 2′-*O*-adduct cleavage reaction (10 μL) was added to a fluorescently labeled DNA primer (5′-VIC-labeled GAA CCG GAC CGA AGC CCG; 3 μL, 0.3 μM) and allowed to anneal at 65 °C for 6 min and then cooled on ice. Reverse transcription buffer [6 μL; 167 mM Tris-HCl (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] and Superscript III (1 μL, 200 units) were added and incubated at 45 °C for 2 min, 52 °C for 20 min then 65 °C for 5 min. The reactions were quenched with 4 μL 50 mM EDTA. The cDNAs were recovered by ethanol precipitation, washed twice with 70% ethanol, dried at 65 °C for 2 min, and resuspended in 10 μL deionized formamide. Dideoxy sequencing ladders were produced using unlabeled, unmodified RNA, annealing a 5′-NED-labeled fluorescently labeled DNA primer (same sequence as above) (3 μL, 0.3 μM), and by adding 1 μL 2′,3′-dideoxycytidine (10 mM) triphosphate before addition of Superscript III. cDNA fragments were separated by capillary electrophoresis using an Applied Biosystems 3130 DNA sequencing instrument.

4.5 References

1. Holbrook, S.R., *Structural principles from large RNAs.* Annu Rev Biophys, 2008. **37**: p. 445-64.

2. Gesteland, R.F., Cech, T.R., and Atkins, J.F. *The RNA world*. 3rd ed. 2006, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

3. Korostelev, A. and Noller, H.F. *The ribosome in focus: new structures bring new insights.* Trends Biochem Sci, 2007. **32**(9): p. 434-41.

4. Weeks, K.M. and Mauger, D.M. *Exploring RNA structural codes with SHAPE chemistry.* Acc Chem Res, 2011.

5. Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).* J Am Chem Soc, 2005. **127**(12): p. 4223-31.

6. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.* Nat Protoc, 2006. **1**(3): p. 1610-6.

7. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. *RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts.* J Am Chem Soc, 2005. **127**(13): p. 4659-67.

8. Mortimer, S.A. and Weeks, K.M. *A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry.* J Am Chem Soc, 2007. **129**(14): p. 4144-5.

9. Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. *Accurate SHAPE-directed RNA structure determination.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 97-102.

10. Low, J.T. and Weeks, K.M. *SHAPE-directed RNA secondary structure prediction.* Methods, 2010. **52**(2): p. 150-8.

11. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L., and Weeks, K.M. *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-6.

12. Duncan, C.D. and Weeks, K.M. *SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA.* Biochemistry, 2008. **47**(33): p. 8504-13.

13. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. *High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states.* PLoS Biol, 2008. **6**(4): p. e96.

14. Soukup, G.A. and Breaker, R.R. *Relationship between internucleotide linkage geometry and the stability of RNA.* RNA, 1999. **5**(10): p. 1308-25.

15. Simanenko, S., Popov, A.F., Prokop'eva, T.M., Karpichev, E.A., Savelona, V.A., Suprun, I.P., and Bunton, C.A. *Inorganic anionic oxygen-containing α-nucleophiles-effective acyl group acceptors: hydroxylamine ranks first among the α-nucleophile series.* Russ J Org Chem, 2002. **38**(9): p. 1286-1298.

16. Mazera, D.J., Gesser, J.C., and Pliego, J.R. *On the mechanism of the reaction between aryl acetates and hydroxylamine.* ARKIVOC, 2007. **15**: p. 199-214.

17. Kirby, A.J., Lima, M.F., da Silva, D., Roussev, C.D., and Nome, F. *Efficient intramolecular general acid catalysis of nucleophilic attack on a phosphodiester.* J Am Chem Soc, 2006. **128**(51): p. 16944-52.

18. Kashima, C., Yoshiwara, N., and Omote, Y. *Alkylation of aminohydroxy anion, dissociated species of hydroxylamine.* Tetrahedron Lett, 1982. **23**(29): p. 2955-2956.

19. Kirby, A.J., Davies, J.E., Brandao, T.A., da Silva, P.F., Rocha, W.R., and Nome, F. *Hydroxylamine as an oxygen nucleophile. Structure and reactivity of ammonia oxide.* J Am Chem Soc, 2006. **128**(38): p. 12374-5.

20. Kirby, A.J., Davies, J.E., Fox, D.J., Hodgson, D.R., Goeta, A.E., Lima, M.F., Priebe, J.P., Santaballa, J.A., and Nome, F. *Ammonia oxide makes up some 20% of an aqueous solution of hydroxylamine.* Chem Commun (Camb), 2010. **46**(8): p. 1302-4.

21. Medeiros, M., Souza, B.S., Orth, E.S., Brandao, T.A., Rocha, W., Kirby, A.J., and Nome, F. *The reaction of hydroxylamine with aspirin.* ARKIVOC, 2011. **7**: p. 461-476.

22. Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J. *Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch.* Nature, 2006. **441**(7097): p. 1167-71.

23. Steen, K.A., Malhotra, A., and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease.* J Am Chem Soc, 2010. **132**(29): p. 9940-3.

24. Jencks, W.P. *The reaction of hydroxylamine with activated acyl groups. I. Formation of O-acylhydroxylamine.* J Am Chem Soc, 1958. **80**(17): p. 4581-4584.

25. Jencks, W.P. *The reaction of hydroxylamine with activated acyl groups. II. Mechanism of the reaction.* J Am Chem Soc, 1958. **80**(17): p. 4585-4588.

26. Jencks, W.P. *The mechanism of the reaction of hydroxylamine with activated acyl groups.* Biochim Biophys Acta, 1958. **27**(2): p. 417-8.

27. Weeks, K.M. *RNA structure probing dash seq.* Proc Natl Acad Sci U S A, 2011. **108**(27): p. 10933-4.

CHAPTER 5

Effects of Ligand Binding on RNA Nucleotide Dynamics as Detected by SHAPE Chemistry and NMR

5.1 Introduction

Many important biological functions, such as ribonuceloprotein assembly and riboswitch function, depend on the ability of RNA molecules to bind small molecules [1, 2]. Diverse ligands such as metabolites and proteins have been shown to modulate RNA structure through binding events [1-4]. The binding of external cofactors results in specific conformational changes that are made possible by the intrinsic motion of RNA molecules [4]. Additionally, some RNA molecules are fairly dynamic even when folded and allow the biopolymer to sample a range of structural and nucleotide conformations. The formation of a ligand-accessible structure, which is the optimal conformation for ligand binding, is also facilitated by RNA conformational sampling [5-7].

NMR spectroscopy has been the technique of choice for studying RNA motion and ligand-induced nucleotide dynamics in solution [4, 7, 8] because it allows for a wide range of RNA motions to be characterized on a per residue basis [4, 7]. The experimental

measurements (T_1, T_2 and NOE) obtained from an NMR experiment can then be interpreted using model-free analysis in terms of the generalized order parameter, S^2, for each nucleotide [9, 10]. S^2 describes the extent of internal motion of a nucleotide and ranges in value from 0 to 1, where 0 represents a completely disordered nucleotide and 1 represents a completely ordered nucleotide [7]. NMR has provided key insights into ligand-induced RNA conformational changes in RNA-small molecule [11, 12] and RNA-protein complexes [13]. However, while larger structures have been studied using NMR (up to 100 nucleotides), NMR measurements are typically limited to RNAs of approximately 70 nucleotides or less [14, 15]. Consequently, there is a need for a biochemical method that detects ligand-induced RNA conformational dynamics in solution at single nucleotide resolution regardless of RNA size or complexity.

The degree of RNA nucleotide flexibility can be biochemically evaluated using SHAPE. SHAPE chemistry takes advantage of the intrinsic motion of flexible nucleotides. Flexible nucleotides can adopt many conformations, a subset of which is preferentially reactive with electrophilic SHAPE reagents. As a result, flexible nucleotides display high reactivities with SHAPE reagents while non-dynamic nucleotides are less reactive (Figure 5.1) [16, 17].

SHAPE chemistry is well-established and has been used to characterize a diverse set of RNA structures ranging from small tRNAs to large viral genomes [18, 19]. Additionally, SHAPE chemistry has been employed to detect changes in both local nucleotide flexibility and RNA global structure upon small molecule and protein binding [20-22]. Previously, the relationship between SHAPE chemistry and S^2 flexibility measurements had been established by comparing the SHAPE reactivities and S^2 values for three RNA structures [23]. It was

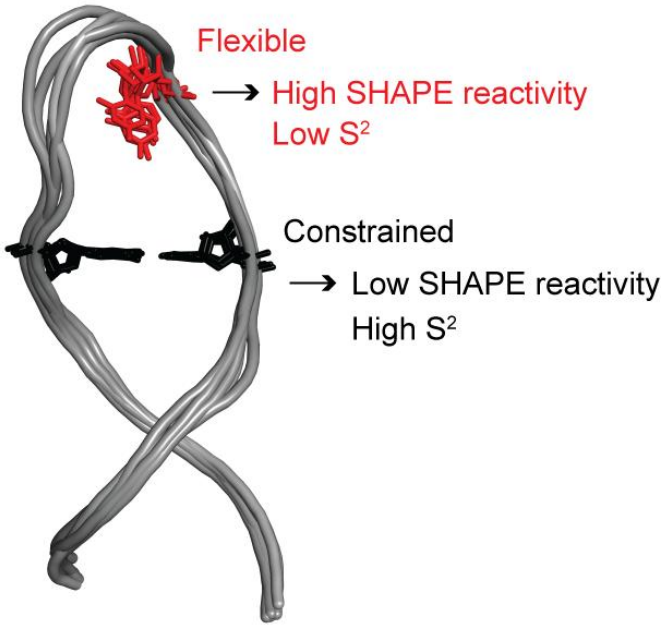


Figure 5.1 Schematic for the relationship between SHAPE and NMR flexibility measurements.

observed, as expected, that SHAPE reactivity is inversely proportional to S^2 value since flexible nucleotides possess high SHAPE reactivities and low S^2 values (Figure 5.1) [23]. Based on this previously established relationship and the ability of both methods to detect RNA ligand binding events, it should be possible to determine whether ligand-induced RNA dynamics as measured by both SHAPE chemistry and NMR are correlated.

In this chapter, I explored the ability of SHAPE chemistry to capture changes in RNA nucleotide dynamics due ligand binding in a manner similar to S^2 values derived from NMR measurements. Ligand binding to the TAR RNA element of the HIV-1 RNA, a conserved stem-loop structure found in the 5′ untranslated region (UTR) of the viral genome, wasevaluated in the context of both SHAPE chemistry and NMR. The TAR RNA element is responsible for promoting efficient transcriptional elongation via the recruitment of proteins including the Tat protein [24]. The TAR RNA element (Figure 5.2A) was chosen as the model system because a number of biochemical and biophysical studies have been devoted to characterizing the structure and dynamics of both the ligand-free and ligand-bound states of this RNA [11, 25-27].

SHAPE and NMR analyses were performed in parallel on the TAR RNA element in three states: (1) ligand-free, for direct comparison to the ligand-bound states; (2) bound to argininamide, a small molecule known to bind the TAR RNA at the UCU bulge; and (3) bound to an 11 nucleotide linear Tat peptide mimic, also proposed to bind to the UCU bulge of TAR. For each of the RNA states studied, the SHAPE reactivities were compared to the S^2 values separately in order to determine a correlation. Additionally, for each of the ligand-bound states, the SHAPE reactivities of each ligand-bound state were compared to the ligand-free state in order to confirm ligand binding. It was determined that the inverse

correlation between SHAPE and S^2 values is robust for the ligand-free state (as previously determined) and, importantly, also upon small molecule binding.

5.2 Results

5.2.1 Correlation between SHAPE and S^2 values for the ligand-free HIV-1 TAR RNA

To confirm the correlation between SHAPE and S^2 values in the context of the NMR RNA construct, I performed SHAPE analysis on the 29 nucleotide TAR RNA construct used in the NMR experiments in context of the 5′ and 3′ flanking sequences (Figure 5.2A). Previous studies on the correlation between SHAPE and the NMR order parameter, S^2, found a Pearson's linear *r*-value of 0.78 for the TAR RNA element [23]. However, SHAPE reactivities for this previous correlation were obtained in the context of a 976 nucleotide HIV-1 RNA transcript [23, 28] while the NMR measurements were obtained in the context of the 29 nucleotide TAR RNA construct [23]. I performed SHAPE modification under conditions similar to those used in the NMR experiments [50 mM HEPES-NaOH (pH 8.0), 50 mM KCl] using 1-methyl-7-nitroisatoic anhydride (1M7) [29]. Sites of modification were detected by primer extension and SHAPE reactivities were calculated [30]. SHAPE reactivities were adjusted to account for the difference in 1M7 reactivity due to nucleotide identity because absolute SHAPE reactivities were being compared to S^2 values [31]. S^2 values at the C1′ ribose position were calculated for most nucleotides in the ligand-free TAR RNA construct by our collaborators at the University of Washington and were previously published in ref 11.

Since SHAPE and S^2 are inversely correlated, I calculated "1- S^2" values at each nucleotide position in order to facilitate direct comparison between SHAPE reactivities and

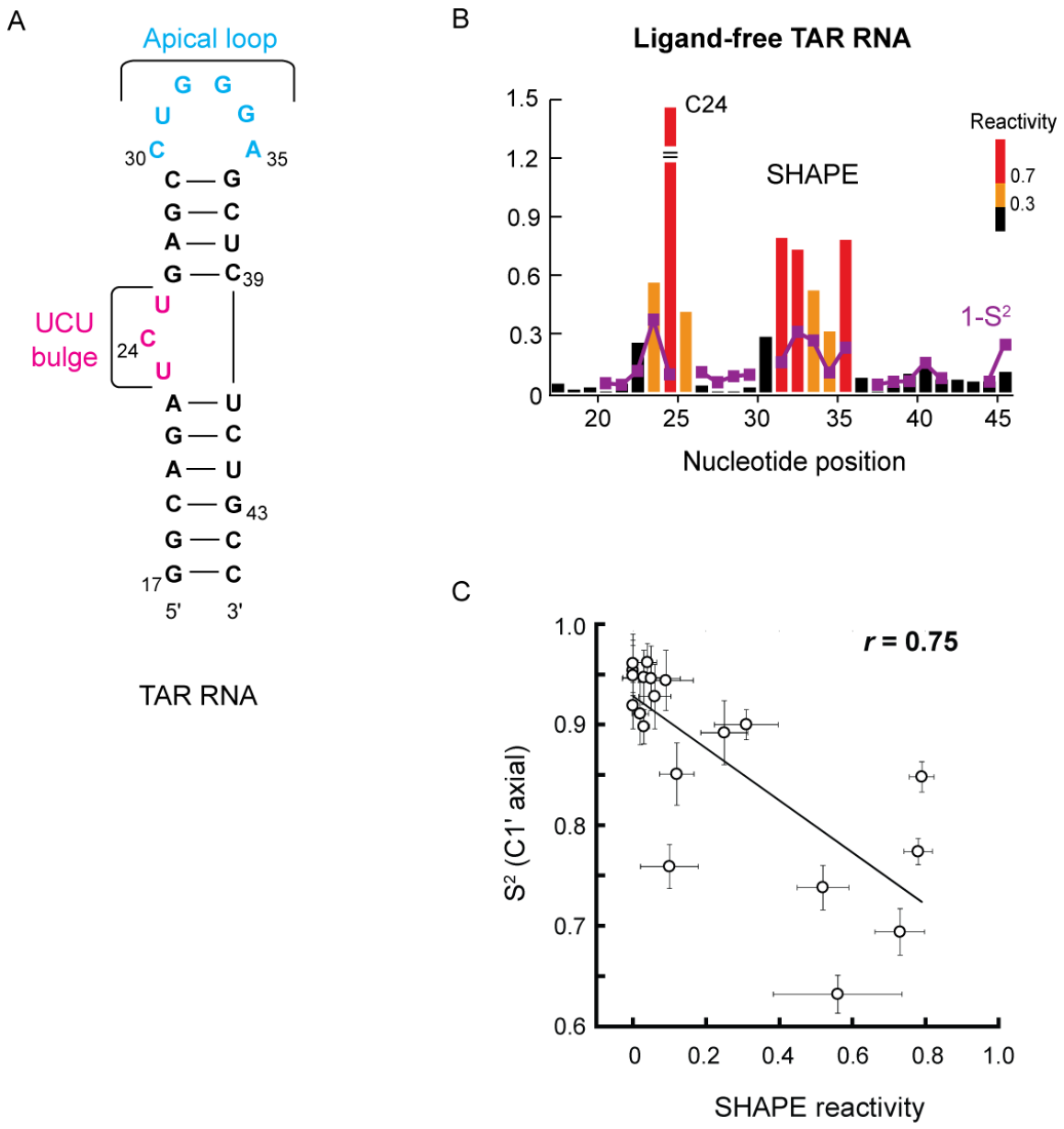


Figure 5.2 Ligand-free TAR RNA from the HIV-1 RNA. (A) Secondary structure with structural features highlighted. (B) Histograms of SHAPE reactivities with 1-S^2 values superimposed (purple squares). (C) Correlation between SHAPE reactivities and S^2 values (Pearson's linear *r*-value shown).

S^2, and overlaid these data in Figure 5.2B. SHAPE data are shown for all nucleotides while 1-S^2 values are displayed for all well-resolved nucleotides in the ligand-free TAR RNA. Examination of this plot reveals that regions of RNA structure displaying low 1- S^2 values tend to correlate with low SHAPE reactivities, as anticipated. Conversely, high 1- S^2 values correlate to high SHAPE reactivities and correspond to nucleotides in the flexible UCU bulge and apical loop regions of the TAR RNA (Figures 5.2A and B). For this RNA construct, SHAPE reactivities and S^2 values are inversely correlated with a Pearson's linear *r*-value of 0.75 (Figure 5.2C), which is consistent with the previous study [23]. A more stringent calculation was also performed where nucleotides involved in canonical base pairs are excluded from the correlation. This was done because canonical base paired nucleotides are usually ordered and display good correlation between SHAPE and S^2, therefore eliminating these nucleotides from the correlation reports directly on the more dynamic regions of the RNA. A correlation value of 0.70 was observed under these constraints; a value that is somewhat lower than that in the previous study [23]. However this new correlation is based on a more complete S^2 dataset for comparison to SHAPE reactivities.

Notably, the SHAPE reactivity and 1-S^2 values do not correlate well at C24 (Figure 5.2B). NMR measurements suggest that this nucleotide is highly ordered while the SHAPE reactivity is extremely high for this nucleotide (>2) and suggests a highly dynamic nucleotide. This discrepancy can be explained by examining the NMR structure of the ligand-free TAR RNA. In this structure, the O2 moiety of C24 is in close enough proximity to the 2′-hydroxyl of the ribose to facilitate hydrogen bonding. This interaction results in the oxygen of the 2′-hydroxyl becoming more nucleophilic, a trend that favors the general nucleophilic attack mechanism behind SHAPE modification [32]. This specific nucleotide is

constrained in a conformation that is particularly reactive to SHAPE chemistry [32] and, as a result, C24 is excluded from all correlation calculations.

In sum, local nucleotide flexibility as described by both SHAPE reactivities and S^2 values, correlate strongly for the ligand-free 29 nucleotide TAR RNA construct and this correlation is comparable to the previously determined value.

5.2.2 SHAPE detects argininamide binding to HIV-1 TAR RNA and correlates to S^2

Next I explored the ability of SHAPE to quantitatively detect ligand binding to RNA in a manner similar to NMR using the TAR RNA model system. I first probed the TAR RNA construct in the presence of 10 mM argininamide and observed a distinct *decrease* in SHAPE reactivity for the argininamide-bound TAR RNA at the UCU bulge in comparison to the ligand-free RNA (compare red and blue plots, Figure 5.3). This decrease in SHAPE reactivity corresponds to the known binding site for argininamide at U23 and reflects the loss of flexibility concomitant with binding [33-35]. This data indicates that SHAPE analysis can be used to detect ligand binding events to an RNA.

Using the previously published S^2 data for the argininamide-bound TAR RNA [11], I calculated $1\text{-}S^2$ values and compared them to SHAPE reactivities (Figure 5.4A) to evaluate how well both SHAPE and NMR report on this binding event. In general, regions that exhibit high SHAPE reactivities (especially the apical loop) correspond to higher $1\text{-}S^2$ values and are consistent with a high degree of disorder for these nucleotides. Overall, a correlation of 0.72 (Pearson's linear *r* value) was obtained between S^2 and SHAPE for the argininamide-bound TAR RNA (Figure 5.4B). The more stringent evaluation of the S^2 and SHAPE correlation was determined by omitting canonical base pairs and resulted in a correlation of 0.84 for

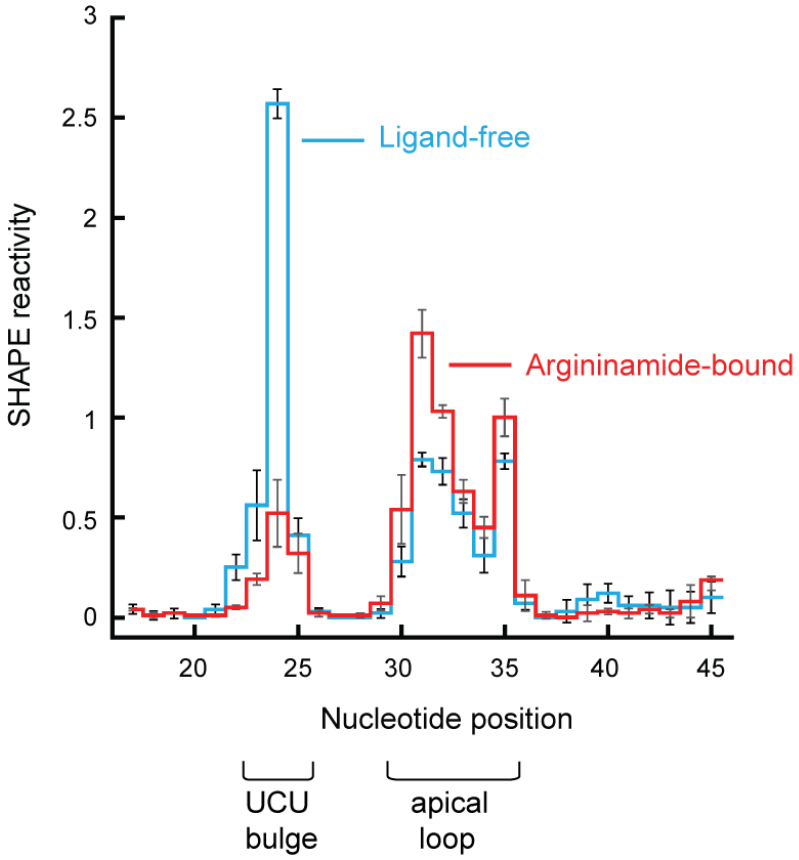


Figure 5.3 SHAPE reactivities of the free (blue) and argininamide-bound (red) TAR RNA.

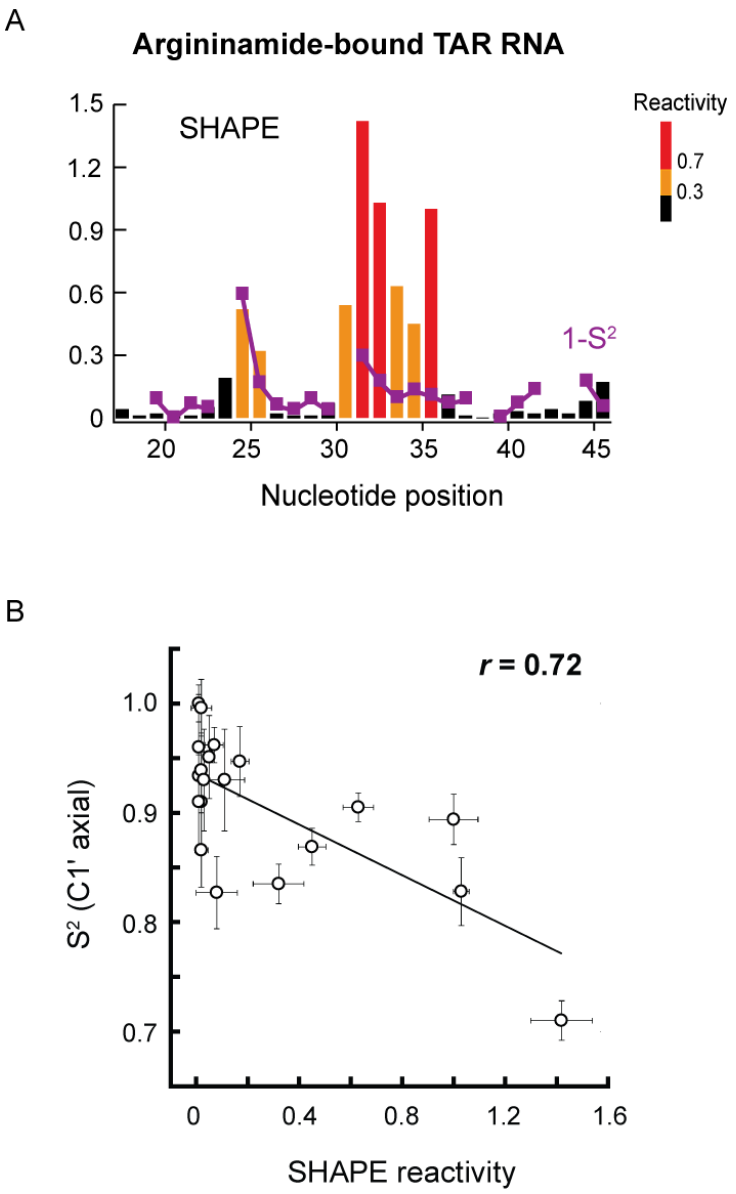


Figure 5.4 SHAPE and S^2 values for the argininamide-bound TAR RNA (A) Histograms of SHAPE reactivities with 1-S^2 values superimposed (purple squares). (B) Quantitative correlation between SHAPE reactivity and S^2 at the C1′ position.

argininamide binding. Since argininamide binding occurs in the flexible UCU bulge, this correlation is particularly significant because a good correlation at flexible nucleotides suggests that SHAPE chemistry is able to detect changes in dynamics associated with small molecule binding to specific regions of an RNA.

Notably, a S^2 value for U23 was not calculated because there is direct interaction between this nucleotide and the ligand that makes accurate calculations of S^2 experimentally difficult. Despite this difficulty, NMR relaxation times obtained for U23 indicate that U23 becomes more rigid upon argininamide binding [11]. This increased rigidity is consistent with our observed decrease in SHAPE reactivity at U23, suggesting that U23 becomes more constrained upon argininamide binding (Figure 5.5A). These results suggest that SHAPE analysis can be used to detect the local disorder of almost all nucleotides in an RNA including ones that interact directly with ligands or proteins.

A decrease in SHAPE reactivity at C24 was also observed (Figure 5.4A) and suggests that C24 becomes more constrained in the presence of argininamide. However, the decrease in flexibility at C24 is inconsistent with an observed extrahelical flip of this nucleotide in the NMR structure of the argininamide-bound state (see C24, Figure 5.5B). Typically, an extrahelical flip results in an increase in SHAPE reactivity due to increased flexibility at that position [36]. However, since C24 exists in an ordered but highly SHAPE-reactive conformation in the ligand-free state, changes in the SHAPE reactivity at this nucleotide are difficult to interpret as true changes in local nucleotide flexibility.

Finally, significant increases in the SHAPE reactivities of some nucleotides in the apical loop were observed (see Figure 5.3). This increase in reactivity suggests that binding of argininamide at the UCU bulge affects the flexibility of distal nucleotides in the TAR

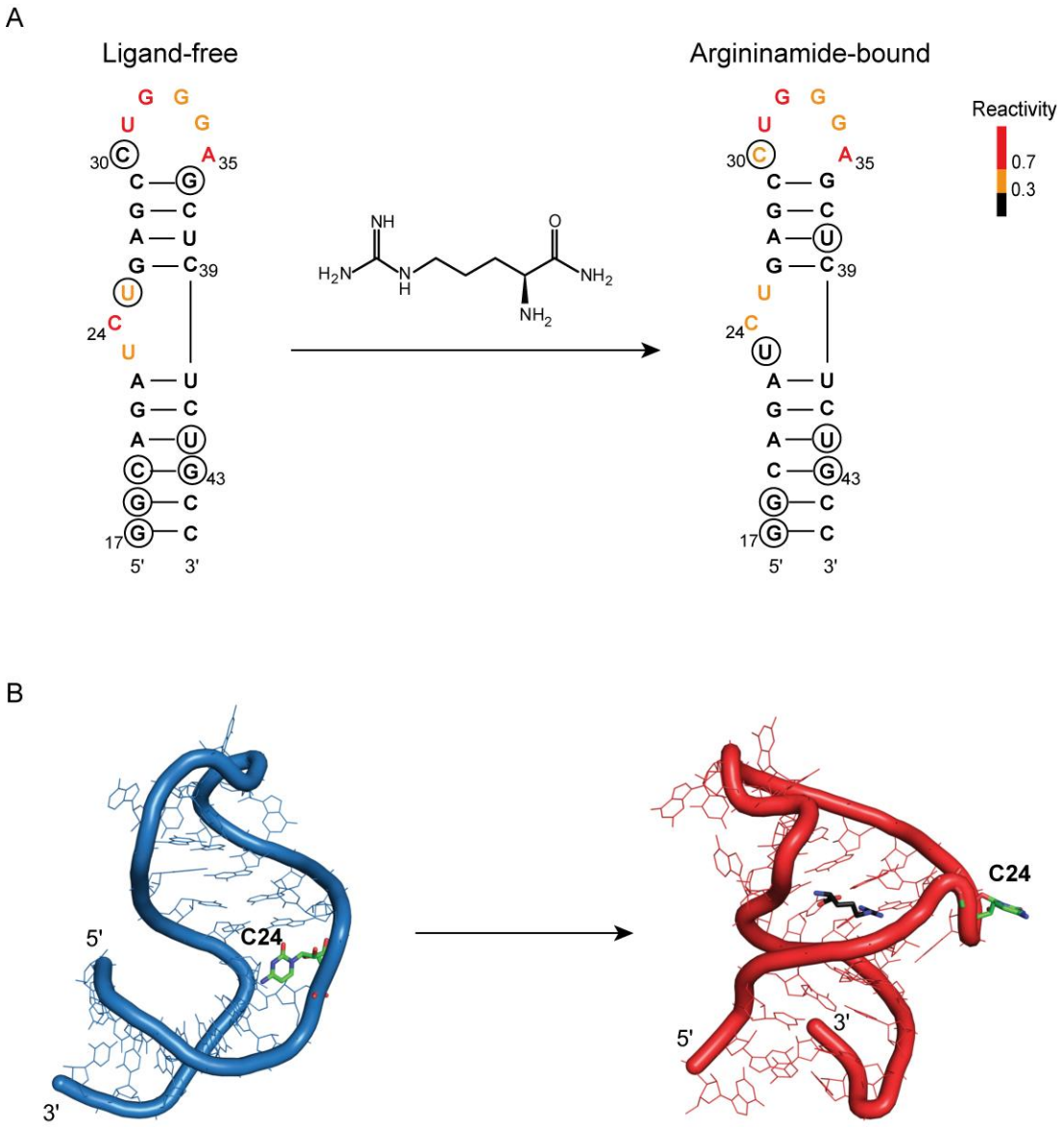


Figure 5.5 Structures of the free and argininamide-bound TAR RNA. (A) Absolute SHAPE reactivities superimposed on the secondary structure. Positions for which S^2 values were not calculated are shown as circled nucleotides. (B) Three-dimensional structures showing the position of C24 in both structures.

RNA structure. In contrast, the 1-S^2 values for the apical loop do *not* indicate a significant increase in the degree of flexibility in the presence of argininamide (compare 1-S^2 plots on Figures 5.2B and 5.4A). The SHAPE-detected increase in flexibility is likely a better representation of the local structure and dynamics of the apical loop than the S^2 values due to the unique dynamics that many of these loop nucleotides exhibit known as conformational exchange [11]. Conformational exchange occurs when a nucleotide exhibits slower motions than the typical fast NMR timescales (μs to ms rather than ps to ns) and results in inaccuracies in the T_1/T_2 ratios that S^2 calculations are not sufficiently able to account for and typically results in a more ordered value for S^2 [37, 38]. To support this theory, I examined the NMR structures of both the ligand-free and argininamide-bound states (Figure 5.5B) and observed that the location of the apical loop relative to the UCU bulge is significantly different between both RNA states. The position of the apical loop suggests that binding of argininamide results in a significant movement of the apical loop nucleotides and reflects a higher degree of flexibility than the S^2 values suggest. Therefore, SHAPE analysis detects small molecule ligand binding and conformational changes in the TAR RNA and correlates to the established quantitative measure of local order, S^2.

5.2.3 SHAPE chemistry detects linear peptide binding to the TAR RNA construct

After demonstrating that SHAPE chemistry was able to detect small molecule binding, the binding of a larger ligand to TAR RNA was investigated. To determine whether SHAPE chemistry could detect the binding and dynamic changes associated with large ligand binding, I probed the TAR RNA construct in the presence of an 11 nucleotide linear peptide mimic of the Tat protein [YRGKKRRQRRR; 60 nM]. A large decrease in SHAPE reactivity at the UCU bulge was observed, which is the proposed Tat protein-binding site (Figure

5.6A). Binding of the linear Tat peptide is expected to have similar effects on the TAR RNA element as argininamide binding [11] and this is consistent with the observed decrease in SHAPE reactivities at the UCU binding site (compare Figures 5.3 and 5.6A). Additionally, increases in the SHAPE reactivities of the apical loop nucleotides were observed, which suggest an increase in flexibility that is consistent with the distal effects noted upon argininamide binding (Figures 5.6A).

To evaluate SHAPE and NMR detection of large ligand binding, I then compared the SHAPE reactivities and 1-S^2 values of the linear Tat peptide-bound TAR RNA. In general, a strong correlation between the 1-S^2 values and SHAPE reactivities was not observed for the UCU bulge or the apical loop. The NMR measurements suggest that the UCU bulge is more flexible than our observed SHAPE reactivities indicate. Additionally, the apical loop nucleotides exhibit lower 1-S^2 values than our corresponding higher SHAPE reactivities (compare apical loop reactivities, Figure 5.7A). The S^2 data suggests that the apical loop is more ordered while our SHAPE reactivities suggest a higher degree of flexibility. When the SHAPE reactivities were compared to the calculated S^2 values, a poor correlation was obtained for linear Tat peptide binding (Pearson's linear $r = 0.20$, plot not shown).

Since S^2 and absolute SHAPE reactivities did not correlate well for linear Tat peptide binding, I examined the direction of changes (increases or decreases) for both the SHAPE reactivities and S^2 values separately. The rationale for this analysis stems from noting that the S^2 values for the linear peptide complex were calculated without the aid of a three dimensional structure and resulted in a lower degree of accuracy than the argininamide-bound calculations due to the increased assumptions in the calculations [11]. By comparing the free and peptide-bound TAR RNA states for SHAPE and S^2 separately, the effects of

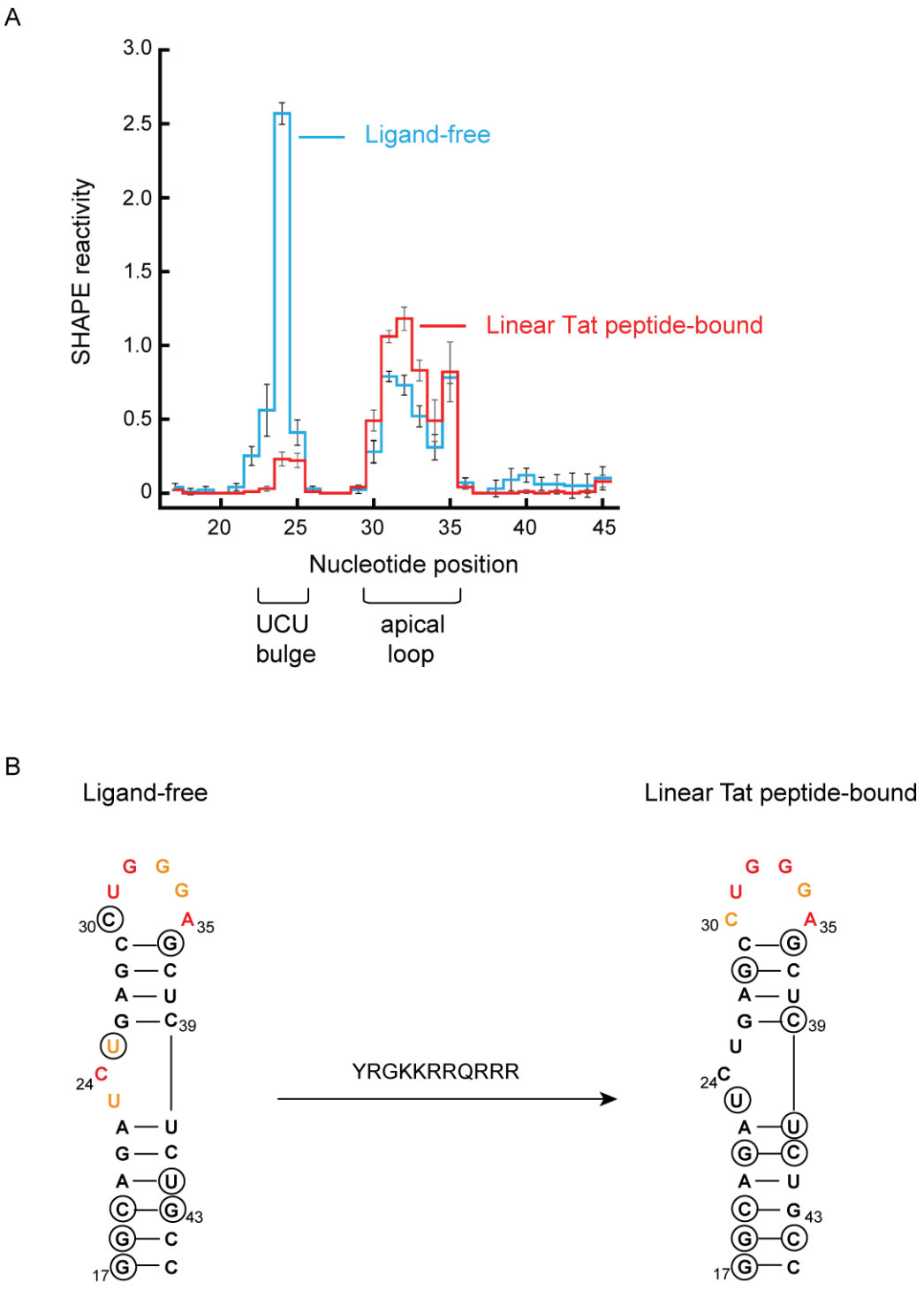


Figure 5.6 SHAPE reactivities for the free and linear peptide-bound TAR RNA (A) Plots for the free (blue) and linear peptide-bound (red) TAR RNA. (B) Absolute SHAPE reactivities are superimposed on the secondary structure. Positions for which S^2 values were not calculated are shown as circled nucleotides.

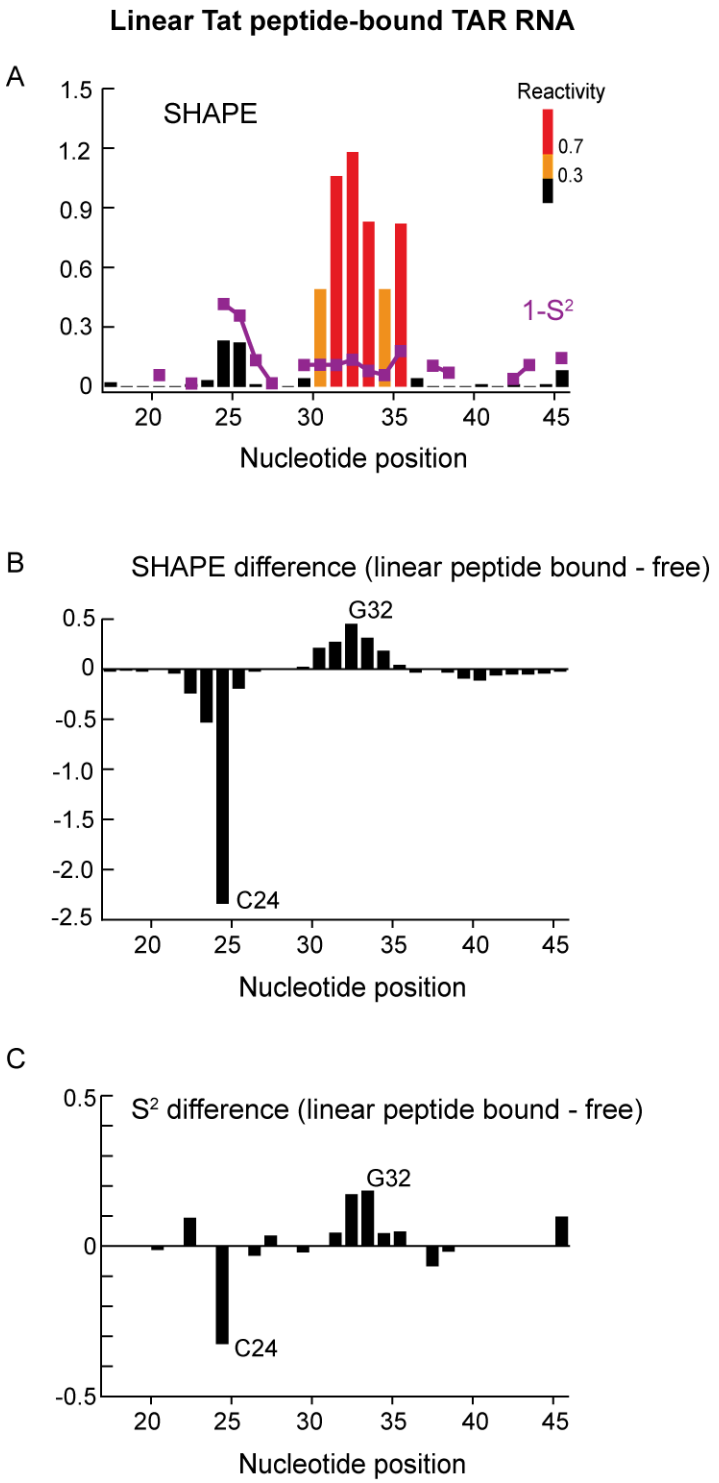


Figure 5.7 SHAPE and S^2 values for the linear peptide-bound TAR RNA (A) Histograms of SHAPE reactivities with 1-S^2 values superimposed (purple squares). Difference plots for (B) absolute SHAPE reactivities and (C) S^2 at the C1′ position.

ligand binding on both detection methods can be determined in a qualitative manner. I created difference plots for both SHAPE and S^2 to compare the changes in both measurements between the ligand-free and linear peptide-bound states (Figures 5.7B and C). In general, decreases were observed in SHAPE reactivities *and* S^2 upon linear Tat peptide binding at the UCU bulge while increases were also observed in both measurements at the apical loop nucleotides upon ligand binding. These changes in SHAPE reactivity and S^2 are similar to those observed upon argininamide-binding and are consistent with previous studies of both ligands and their effects on TAR RNA [11].

While both S^2 and SHAPE measurements reveal similar changes in RNA nucleotide flexibility upon large ligand binding, a good correlation between the absolute values for S^2 and SHAPE reactivities was not obtained. SHAPE chemistry and S^2 correlates quantitatively for ligand-free and small molecule binding to TAR RNA but only a qualitative correlation is obtained between SHAPE and S^2 for large ligand binding to TAR RNA.

5.3 Discussion

In this chapter, I investigated whether SHAPE chemistry correlated with the widely accepted biophysical order parameter, S^2, for the detection of changes in RNA nucleotide dynamics upon ligand binding to the HIV-1 TAR RNA. Both SHAPE chemistry and S^2 measurements detect the binding of a small molecule, argininamide, to the TAR RNA element and correlate strongly for argininamide binding. The binding of a larger linear peptide ligand to the TAR element was detected using both SHAPE and S^2, but these values correlated poorly for linear peptide binding.

While S^2 is a NMR biophysical indicator of order, the calculation of accurate S^2 values is highly dependent on the availability of quality NMR measurements. Due to spectral

overlap, the number of well-resolved RNA nucleotides decreases as a function of increased RNA-ligand complex size and complexity (compare circled nucleotides, Figures 5.5A and 5.6B) [8]. In fact, S^2 values were obtained for only 62% of the nucleotides in the linear peptide-bound TAR RNA in contrast to 76% obtained for the argininamide-bound TAR RNA. The lack of S^2 measurements is likely one factor that contributes to the poor correlation of SHAPE and S^2 for linear peptide binding. Additionally, the slower motions (μs to ms timescale) exhibited by some nucleotides in both the ligand-free and ligand-bound states, known as conformational exchange, is inadequately accounted for in the S^2 calculations since S^2 is based on faster timescales [3]. Finally, some RNA nucleotides become disordered on even slower timescales (seconds) [39, 40]; these motions are not detectable by these NMR experiments and would also result in a weaker correlation between SHAPE and S^2 values. Consequently, SHAPE chemistry likely provides a more accurate indication of nucleotide flexibility at many nucleotide positions where SHAPE and S^2 do not correlate due to the availability of SHAPE measurements for every nucleotide in the TAR RNA regardless of RNA-ligand complexity or size.

One important limitation of SHAPE chemistry for studying nucleotide dynamics is exemplified by the behavior of C24 in the ligand-free state. Some nucleotides are capable of adopting rare, catalytic conformations that are constrained yet highly reactive to SHAPE reagents. These nucleotides are almost always located in flexible regions of an RNA so they do not impact the predictive power of SHAPE but result in poor correlation between S^2 and SHAPE due to the high degree of order that the nucleotide exhibits.

The ability to detect local conformational changes associated with RNA ligand-binding has important implications for RNA drug discovery. Currently, RNA drug binding is

primarily studied using time-intensive x-ray crystallography, supplemented with biochemical experiments. SHAPE chemistry could be used to rapidly report ligand-induced conformational changes in the context of potential drug screening before crystallographic studies are undertaken.

The ability to rapidly probe RNA structure and dynamics at single nucleotide resolution in any RNA allows for diverse RNA sizes to be studied in the context of small molecule or protein-bound states. Due to the broad applicability of SHAPE chemistry, conformational changes associated with ligand binding can be studied in many important RNA structures in a variety of environments such as *in vitro*, *in virio* and *in vivo*.

5.4 Experimental

5.4.1 Synthesis of TAR RNA construct

The TAR RNA construct (5′- GGCAG AUCUG AGCCU GGGAG CUCUC UGCC-3′) with 5′ and 3′ structure cassette flanking sequences [41] were synthesized via *in vitro* transcription by: (1) PCR [1 mL; 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 250 nM each forward and reverse primer (IDT), 40 nM template (IDT), and 0.025 units/μL Taq polymerase; denaturation at 95 °C, 45 s; annealing at 55 °C, 30 s; elongation at 72 °C, 1 min; 35 cycles]; recovered by ethanol precipitation and resuspended in 200 μL 1× TE [10 mM Tris-HCl (pH 8.0), 1 mM EDTA]; and (2) treated with T7 RNA polymerase [1 mL; 40 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 10 mM dithiothreitol, 2 mM spermidine, 0.01% (v/v) Triton X-100, 4% (w/v) poly(ethylene) glycol 8000, 2 mM each NTP, 40 μL PCR-generated template, 0.1 mg/mL T7 RNA polymerase; 37 °C; 4 h]. RNAs were purified by denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 7 M urea, 29:1 acrylamide:bisacrylamide, 0.4 mm × 28.5 cm × 23 cm; 32 W, 1.5 h), excised from

the gel, recovered by overnight passive elution in 1× TE at 4 °C, and precipitation with ethanol. Purified RNAs were resuspended in 50 μL 1× TE and stored at -20 °C.

5.4.2 Structure-selective RNA SHAPE modification

The TAR RNA construct (5 pmol) in 7 μL H_2O was heated at 95 °C for 2 min, cooled on ice, treated with 1 μL 10× folding buffer [500 mM HEPES-NaOH (pH 8.0), 500 mM KCl], and incubated at 37 °C for 10 min followed by incubation at room temperature for 2 min. The ligand (1 μL; 100 mM argininamide or 600 nM Tat 11 linear peptide [YRGKKRRQRRR] [11]) or sterile water was added to the RNA and incubated at room temperature for an additional 20 min. After incubation, 9 μL of the folded RNA (+/– ligand) was added to 1 μL 100 mM 1M7 (in anhydrous DMSO) [29] and incubated at room temperature for 5 min. No-reagent control reactions were performed with 1 μL neat DMSO. The RNAs were recovered by phenol:chloroform:isoamyl alcohol extraction followed by ethanol precipitation and resuspension in 10 μL 1/2× TE.

5.4.3 Primer extension

Sites of 2′-*O*-adduct formation were detected by primer extension using DNA primers that were 5′-end labeled with VIC or NED fluorophores (from Applied Biosystems). RNA from the 1M7 modification reaction (10 μL) was added to a fluorescently labeled DNA primer (5′-VIC-labeled GAA CCG GAC CGA AGC CCG; 3 μL, 0.3 μM) and allowed to anneal at 65 °C for 6 min and then cooled on ice. Reverse transcription buffer [6 μL; 167 mM Tris-HCl (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] and Superscript III (1 μL, 200 units) were added and incubated at 45 °C for 2 min, 52 °C for 20 min then 65 °C for 5 min. The reactions were quenched with 4 μL 50 mM EDTA. The cDNAs were recovered by ethanol precipitation, washed twice with 70% ethanol, dried at 65 °C for 5 min,

and resuspended in 10 μL deionized formamide. Dideoxy sequencing ladders were produced using unlabeled, unmodified RNA, annealing a 5′-NED-labeled fluorescently labeled DNA primer (same sequence as above) (3 μL, 0.3 μM), and by adding 1 μL 2′,3′-dideoxycytidine (10 mM) triphosphate before addition of Superscript III. cDNA fragments were separated by capillary electrophoresis using Applied Biosystems 3130 and 3730 DNA sequencing instruments.

5.4.4 SHAPE data analysis

Raw capillary electrophoresis traces were analyzed using SHAPEFinder [30]. SHAPE reactivities were obtained by subtracting the no-reagent background from the (+) reagent intensities and normalized by excluding the top 2% of the reactive nucleotides, averaging the next 10% of reactive nucleotides, and then dividing all intensities by this averaged value. The normalized SHAPE reactivities were adjusted based on their nucleotide identity and extent of reactivity with 1M7 [31]. All SHAPE experiments were performed in triplicate and the SHAPE reactivities for each nucleotide were expressed as averages while the error was expressed as a standard deviation.

5.4.5 NMR relaxation experiments and calculation of the generalized order parameter, S^2

The NMR protocols and measurements for the free, argininamide-bound and linear peptide-bound TAR RNAs were previously described in ref 11.

5.5 References

1. Woodson, S.A. *RNA folding and ribosome assembly.* Curr Opin Chem Biol, 2008. **12**(6): p. 667-73.

2. Haller, A., Souliere, M.F., and Micura, R. *The dynamic nature of RNA as key to understanding riboswitch mechanisms.* Acc Chem Res, 2011.

3. Al-Hashimi, H.M. *Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy.* Chembiochem, 2005. **6**(9): p. 1506-19.

4. Al-Hashimi, H.M. and Walter, N.G. *RNA dynamics: it is about time.* Curr Opin Struct Biol, 2008. **18**(3): p. 321-9.

5. Shajani, Z. and Varani, G. *13C NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein.* J Mol Biol, 2005. **349**(4): p. 699-715.

6. Shajani, Z. and Varani, G. *NMR studies of dynamics in RNA and DNA by 13C relaxation.* Biopolymers, 2007. **86**(5-6): p. 348-59.

7. Rinnenthal, J., Buck, J., Ferner, J., Wacker, A., Furtig, B., and Schwalbe, H. *Mapping the landscape of RNA dynamics with NMR spectroscopy.* Acc Chem Res, 2011.

8. Latham, M.P., Brown, D.J., McCallum, S.A., and Pardi, A. *NMR methods for studying the structure and dynamics of RNA.* Chembiochem, 2005. **6**(9): p. 1492-505.

9. Lipari, G. and Szabo, A. *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity.* J Am Chem Soc, 1982. **104**(17): p. 4546-4559.

10. Lipari, G. and Szabo, A. *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results.* J Am Chem Soc, 1982. **104**(17): p. 4559-4570.

11. Bardaro, M.F., Shajani, Z., Patora-Komisarska, K., Robinson, J.A., and Varani, G. *How binding of small molecule and peptide ligands to HIV-1 TAR alters the RNA motional landscape.* Nucleic Acids Res, 2009. **37**(5): p. 1529-40.

12. Buck, J., Furtig, B., Noeske, J., Wohnert, J., and Schwalbe, H. *Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution.* Proc Natl Acad Sci U S A, 2007. **104**(40): p. 15699-704.

13. Shajani, Z., Drobny, G., and Varani, G. *Binding of U1A protein changes RNA dynamics as observed by 13C NMR relaxation studies.* Biochemistry, 2007. **46**(20): p. 5875-83.

14. D'Souza, V., Dey, A., Habib, D., and Summers, M.F. *NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus.* J Mol Biol, 2004. **337**(2): p. 427-42.

15. Furtig, B., Richter, C., Wohnert, J., and Schwalbe, H. *NMR spectroscopy of RNA.* Chembiochem, 2003. **4**(10): p. 936-62.

16. Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).* J Am Chem Soc, 2005. **127**(12): p. 4223-31.

17. Weeks, K.M. and Mauger, D.M. *Exploring RNA structural codes with SHAPE chemistry.* Acc Chem Res, 2011.

18. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. *RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts.* J Am Chem Soc, 2005. **127**(13): p. 4659-67.

19. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L., and Weeks, K.M. *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-6.

20. Wang, B., Wilkinson, K.A., and Weeks, K.M. *Complex ligand-induced conformational changes in tRNA(Asp) revealed by single-nucleotide resolution SHAPE chemistry.* Biochemistry, 2008. **47**(11): p. 3454-61.

21. Duncan, C.D. and Weeks, K.M. *Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding.* Biochemistry, 2010. **49**(26): p. 5418-25.

22. Steen, K.A., Malhotra, A., and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease.* J Am Chem Soc, 2010. **132**(29): p. 9940-3.

23. Gherghe, C.M., Shajani, Z., Wilkinson, K.A., Varani, G., and Weeks, K.M. *Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA.* J Am Chem Soc, 2008. **130**(37): p. 12244-5.

24. Das, A.T., Harwig, A., and Berkhout, B. *The HIV-1 Tat protein has a versatile role in activating viral transcription.* J Virol, 2011. **85**(18): p. 9506-16.

25. Aboul-ela, F., Karn, J., and Varani, G. *The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein.* J Mol Biol, 1995. **253**(2): p. 313-32.

26. Aboul-ela, F., Karn, J., and Varani, G. *Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge.* Nucleic Acids Res, 1996. **24**(20): p. 3974-81.

27. Raghunathan, D., Sanchez-Pedregal, V.M., Junker, J., Schwiegk, C., Kalesse, M., Kirschning, A., and Carlomagno, T. *TAR-RNA recognition by a novel cyclic aminoglycoside analogue.* Nucleic Acids Res, 2006. **34**(12): p. 3599-608.

28. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. *High-throughput SHAPE analysis reveals structures*

in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol, 2008. **6**(4): p. e96.

29. Mortimer, S.A. and Weeks, K.M. *A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry.* J Am Chem Soc, 2007. **129**(14): p. 4144-5.

30. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. *ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis.* RNA, 2008. **14**(10): p. 1979-90.

31. Wilkinson, K.A., Vasa, S.M., Deigan, K.E., Mortimer, S.A., Giddings, M.C., and Weeks, K.M. *Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA.* RNA, 2009. **15**(7): p. 1314-21.

32. McGinnis, J.L., Dunkle, J.A., Cate, J.H., and Weeks, K.M. *The mechanism of RNA SHAPE chemistry*. submitted

33. Brodsky, A.S. and Williamson, J.R. *Solution structure of the HIV-2 TAR-argininamide complex.* J Mol Biol, 1997. **267**(3): p. 624-39.

34. Aboul-ela, F. and Varani, G. *Recognition of HIV-1 TAR RNA by Tat protein and Tat-derived peptides.* J Mol Struct, 1998. **423**(1-2): p. 29-39.

35. Pitt, S.W., Majumdar, A., Serganov, A., Patel, D.J., and Al-Hashimi, H.M. *Argininamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg2+-induced conformational stabilization.* J Mol Biol, 2004. **338**(1): p. 7-16.

36. Souliere, M.F., Haller, A., Rieder, R., and Micura, R. *A powerful approach for the selection of 2-aminopurine substitution sites to investigate RNA folding.* J Am Chem Soc, 2011. **133**(40): p. 16161-7.

37. Kroenke, C.D., Loria, J.P., Lee, L.K., Rance, M., and Palmer, A.G. *Longitudinal and transverse 1H-15N dipolar/15N chemical shift anisotropy relaxation interference: unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules.* J Am Chem Soc, 1998. **120**(31): p. 7905-7915.

38. De Alba, E., Baber, J.L., and Tjandra, N. *The use of residual dipolar coupling in concert with backbone relaxation rates to identify conformational exchange by NMR.* J Am Chem Soc, 1999. **121**(17): p. 4282-4283.

39. Mortimer, S.A. and Weeks, K.M. *Time-resolved RNA SHAPE chemistry.* J Am Chem Soc, 2008. **130**(48): p. 16178-80.

40. Mortimer, S.A. and Weeks, K.M. *C2'-endo nucleotides as molecular timers suggested by the folding of an RNA domain.* Proc Natl Acad Sci U S A, 2009. **106**(37): p. 15622-7.

41. Wilkinson, K.A., Merino, E.J., and Weeks, K.M. *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.* Nat Protoc, 2006. **1**(3): p. 1610-6.