Methods for High-Throughput Analysis of RNA Structure and Dynamics

Stefanie Ann Ward Mortimer

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

Chapel Hill
2008

Approved By
Michael Jarstfer
Jeffrey Johnson
Gary Pielak
Marcy Waters
Kevin Weeks

ABSTRACT

Stefanie Ann Ward Mortimer

Methods for High-Throughput Analysis of RNA Structure and Dynamics

(Under the direction of Kevin M Weeks)

RNA sequences fold back on themselves to form secondary and tertiary structures that are difficult to predict. Knowledge of these structures and the time-resolved mechanism by which RNA molecules fold into these structures are necessary for a full understanding of structure-function relationships in RNA biology. A newly developed technology, called SHAPE, provides accurate and quantitative RNA structure information. SHAPE chemistry was developed using N-methylisatoic anhydride (NMIA), which is only moderately electrophilic and requires tens of minutes to form ribose 2'-*O*-adducts. In this work I design and evaluate several classes of significantly more useful reagents for SHAPE chemistry and create a new way to assess RNA tertiary structure in an experimentally straightforward way. First, I design and synthesize a faster reacting reagent for SHAPE chemistry, 1-methyl-7-nitroisatoic anhydride (1M7), based on the NMIA scaffold. With 1M7, single nucleotide resolution interrogation of RNA structure is complete in 70 seconds and appears to be the ideal reagent for equilibrium analysis of RNA structure. Second, I apply SHAPE reagents of varying electrophilicity to identify a class of nucleotides with slow conformational dynamics. The observation is that select C2´-endo nucleotides undergo extremely slow

conformational changes on the order of ~10^{-4} sec^{-1}. Third, I extend SHAPE chemistry to a benzoyl cyanide scaffold to make possible facile time-resolved kinetic studies of RNA in ~1 s snapshots. I then use SHAPE chemistry to follow the time-dependent folding of an RNase P specificity domain RNA and identify a slow folding region in the RNA. I am able to attribute this slow folding step to the conformational dynamics of a single nucleotide. Finally, I show that show that *N*,*N*-(dimethylamino)dimethylchlorosilane (DMAS-Cl) reacts selectively at the guanosine N2 position. Critically, DMAS-Cl reactivity yields a near-perfect measure ($r \geq 0.82$) of solvent accessibility for this position in a folded RNA. This silane-based chemistry represents a significant improvement over classical approaches that employ carbon electrophiles for probing solvent accessibility at the base pairing face of guanosine in RNA.

AKNOWLEDGEMENT

I first and foremost want to thank Kevin Weeks. Not only did he give me the opportunity to work on this great project, but also he is an outstanding mentor. I am the scientist I am today because of his guidance and insight.

There are so many people who have helped me along the way and mean so much to me. Jeremy, this past year has been one of the greatest in my life because you have been in it; thank you for being so supportive of me and my aspirations. My family, thanks for being a constant source of support even though I have been far from home for a long time. I know I can do anything with you behind me. I want to thank my friends, those whom I have known and those whom I have met along the way in graduate school, for all the good times we shared and will hopefully continue to share in the future.

Lastly, I want to thank all the members of the lab for making it such a great place to work. I couldn't ask to be among a better group of scientists and friends, I will truly miss you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1M7	1-methyl-7-nitroisatoic anhydride
4NIA	4-nitro isatoic anhydride
A	adenine
BzCN	benzoyl cyanide
C	cytosine
Ci	curie
°C	degree Celsius
DMSO	dimethylsulfoxide
DMAS-Cl	*N,N*-(dimethylamino)dimethylchlorosilane
DNA	deoxyribonucleic acid
DTT	dithiotreitol
EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol
G	guanosine
GMPS	guanosine monophosphorothioate
h	hour
H_2O	water
HEPES	N-2-hydroxyethylpiperazine-N´-2-ethanesulfonic acid
IA	isatoic anhydride
KCl	potassium chloride

k_{open}	opening rate
k_{close}	closing rate
$k_{hydrolysis}$	hydrolysis rate
k_{adduct}	adduct rate
L	liter
Mg^{2+}	magnesium ion
$MgCl_2$	magnesium chloride
min	minute
mg	milligram
mRNA	messenger RNA
ms	millisecond
µg	microgram
µL	microliter
µM	micromolar
NMIA	N-methyl isatoic anhydride
NaCl	sodium chloride
nM	nanomolar
NMR	nuclear magnetic resonance
ns	nanosecond
nt	nucleotide
NTP	nucleotide triphosphate
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction

Phe	phenylalanine
pmol	picomol
ps	picosecond
RNA	ribonucleic acid
RNase	Ribonuclease
s	second
SHAPE	selective 2´-hydroxyl acylation analyzed by primer extension
TBE	90 mM Tris-borate, 2 mM EDTA
TE	10 mM Tris (pH 7.5), 1 mM EDTA
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
U	uridine
v	volume
V	volt
w	weight
W	watt

CHAPTER 1

Towards Chemical Methods for Efficient and Accurate Analysis of RNA Structure and Dynamics

1. Introduction

1.1.1. RNA structure and function. RNA molecules carry out many important roles in the cell besides being a passive carrier of genetic information (mRNAs) [1]: RNA is essential for protein translation (tRNAs and ribosomes), the splicing of mRNA (group I and II introns and the splicosome), and in genetic regulation (riboswitches, siRNAs, and miRNAs). RNA molecules often form complex three-dimensional structures in order carry out important cellular mechanisms. Higher-order RNA structures are built up from base-paired secondary and long-range tertiary interactions (Figure 1.1) [2, 3]. Accurate and comprehensive knowledge of these interactions is essential to understanding the function of almost all RNAs.

1.1.2. RNA SHAPE chemistry and secondary structure determination. An important first step towards understanding the function of an RNA is determination of the secondary structure. Chemically assisted RNA structure analysis utilizes small molecules to react with specific functional group in the nucleobases [4, 5]. As conventionally used, chemically assisted RNA structure analysis has significant disadvantages. Most reagents react with only a small subset of nucleotides requiring the use of multiple reagents to probe different bases in RNA. Also, these reagents exhibit poor discrimination between paired and unpaired nucleotides [6] and often, the observed reactivity pattern does not correlate well with known structures which results in poor structure prediction [7]. Therefore, using conventional approaches, it is necessary to use a complementary and overlapping series of structural probes to gain a comprehensive understanding of RNA structure. A significant recent advance, the

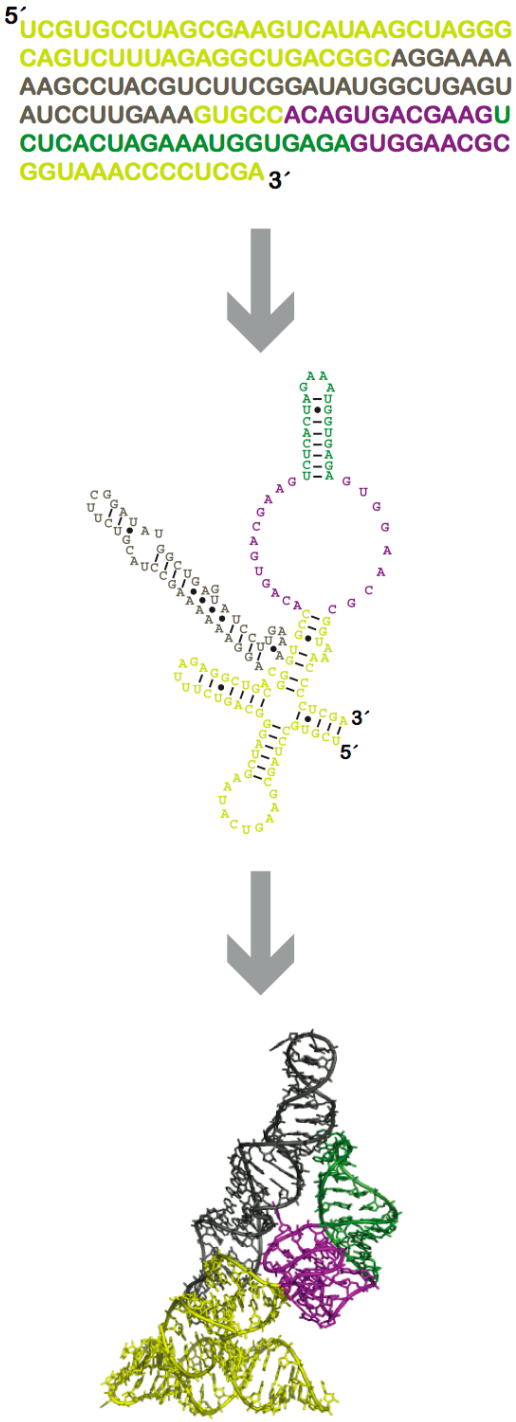


Figure 1.1. The hierarchical nature of RNA structure. RNA molecules are built up from a primary sequence to a secondary, base paired, structure and finally to a three-dimensional tertiary structure. Illustrated here is the specificity domain of the Ribonuclease P ribozyme from *Bacillus subtilis* [8-10].

development of SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) chemistry, has radically simplified and improved analysis of RNA secondary structure [11].

SHAPE is an alternate high-throughput approach to chemically map RNA structure that exploits the discovery that the nucleophilic reactivity of the ribose 2′-hydroxyl group is strongly gated by the underlying nucleotide flexibility (Figure 1.2A) [11-13]. Flexible nucleotides preferentially adopt conformations that react with a hydroxyl-selective electrophile to form a 2′-*O*-adduct, while base paired or otherwise constrained nucleotides are unreactive (Figure 1.2B). Sites of 2′-*O*-adduct formation cause reverse transcriptase to stop exactly one nucleotide prior to the modified base. The primers extended in the reverse transcriptase reaction are labeled with a fluorophore to allow the cDNA transcripts to be visualized when they are separated on a capillary DNA sequencer (Figure 1.2C). The length and amount of a cDNA transcript correlates with the position and degree of modification at each position in the RNA. The length of each cDNA can be assigned by comparison with a primer extension reaction in which a dideoxy nucleotide is incorporated, which causes the termination of extension at specific nucleotides (Figure 1.2D).

SHAPE chemistry provides quantitative, reproducible, single-nucleotide resolution data for the equilibrium structures of RNAs spanning a few nucleotides to thousands of nucleotides [12, 14-16]. Because almost all ribonucleotides possess a free 2′-hydroxyl, each position in the RNA is interrogated, obviating the need to use several different reagents. SHAPE data can be converted to a pseudo-energy function in an RNA

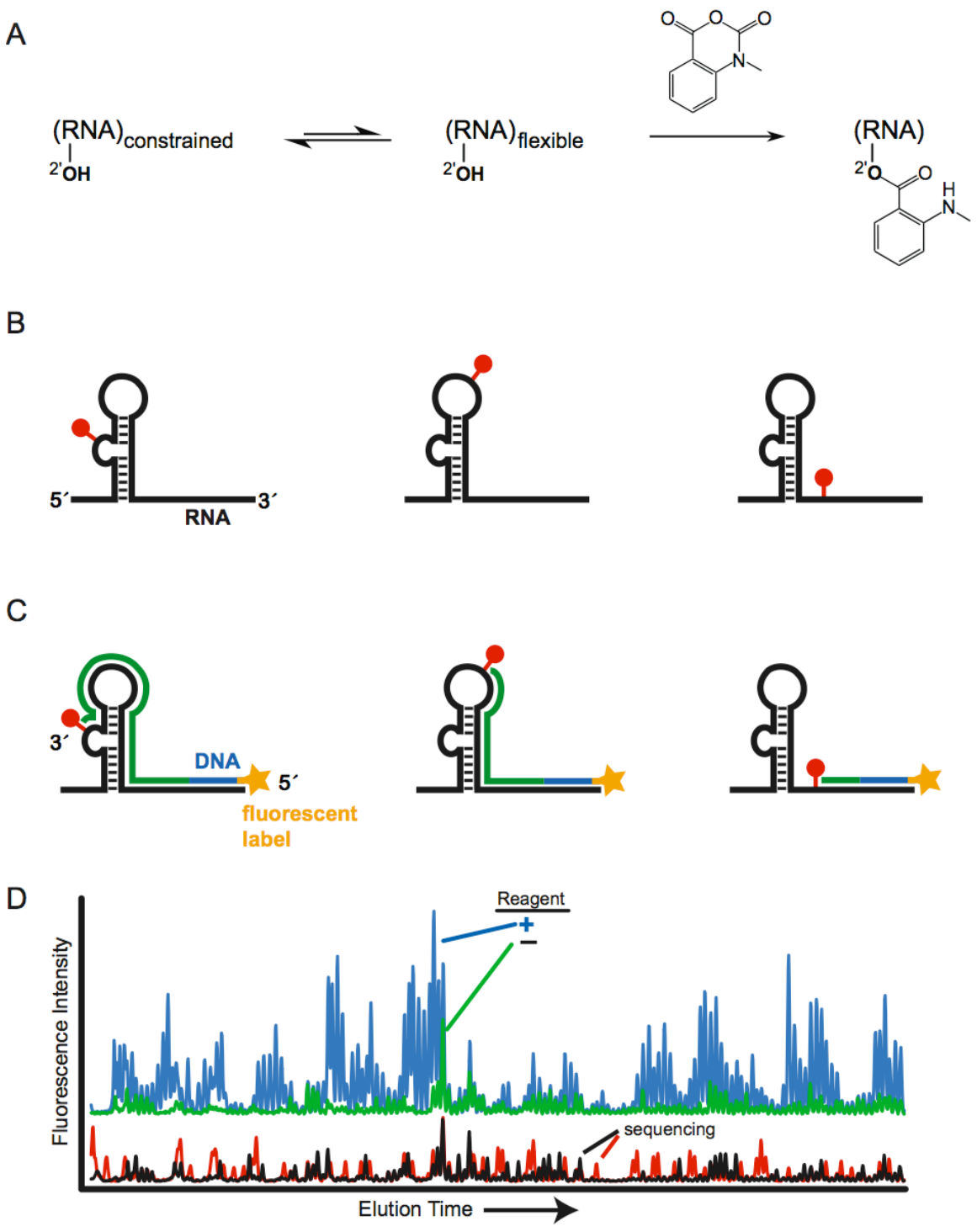


Figure 1.2. High-throughput RNA SHAPE chemistry. (A) Mechanism of RNA SHAPE chemistry. (B) Selective 2´-hydroxyl acylation and (C) analysis by primer extension. (D) Resulting capillary electropherogram.

secondary structure prediction program, called RNAstructure, to constrain or limit the structures predicted. Current algorithms correctly predict 50-70% of known base pairs on average [17]. Incorporating SHAPE reactivity information dramatically improves the prediction of RNA secondary structures, especially for large RNAs, with accuracies usually reaching 90% or greater [18].

The initial electrophile developed for SHAPE, NMIA, is not very reactive and therefore RNA structures must be probed over periods of tens of minutes (Figure 1.2A). This represented a major disadvantage for the initial version of SHAPE chemistry. Much of my research work focused on creating superior reagents for SHAPE chemistry including, those that are faster reacting and capable of interrogating RNA structure on the second timescale.

1.1.3. RNA tertiary structure and solvent accessibility. The analysis of RNA tertiary structure in solution is a necessary step to understand the function of an RNA. An important experimental objective is the ability to map solvent accessibility in RNA at single nucleotide resolution. At present there are very few ways to do this in a direct and quantitative way.

Solvent accessibility at the RNA backbone can be estimated using hydroxyl radical footprinting [19], which makes use of an Fe(II)-EDTA^{2-} catalyst and an oxygen source to induce cleavage in the RNA backbone. The hydroxyl radical is an aggressive reagent which reacts at multiple backbone positions [20] to induce strand cleavage; in addition, the Fe(II)-EDTA moiety carries a net –2 charge. Thus, both steric and electrostatic factors contribute to reactivity of this reagent. Absolute correlations

between hydroxyl radical reactivity and solvent accessibility are good, but imperfect [21, 22].

I am interested in developing new chemical probes of RNA structure that would react with RNA to form stable covalent adducts at solvent accessible positions. These reagents would provide a means to define RNA solvent accessibility in solution in a way that is not convoluted by electrostatic and other factors; an experimental objective that to date has not been accomplished.

1.1.4. RNA folding and dynamics. A full understanding of structure-function relationships in RNA biology requires a nucleotide resolution view of local nucleotide dynamics and the time-resolved mechanisms by which RNA molecules fold, interconvert between distinct states and function in ribonucleoprotein complexes. Currently, the only method able to capture RNA dynamics at nucleotide resolution is NMR spectroscopy, which is limited to RNAs ~100 nts or smaller [23, 24]. Local nucleotide dynamics likely play important roles in governing RNA-RNA, RNA-protein, and RNA-ligand interactions. Previously existing methods for probing the kinetics of RNA folding at nucleotide resolution include DMS and hydroxyl radical footprinting [25-28]. Both DMS and hydroxyl radical footprinting experiments require significant optimization including the addition of a rapid quench step [26, 28]. Furthermore, DMS only probes a subset of nucleotides [5] and can exhibit poor discrimination between paired and unpaired nucleotides [6]. To date, it has experimentally challenging to evaluate folding states and dynamics at nucleotide resolution for large RNAs, thus making it difficult to study the dynamics of folded RNAs and associated folding pathways.

1.1.5. Research overview. The overarching goal of my research has been to create efficient and information–rich experiments that allow one to understand as completely as possible the structure of any RNA. My focus is to develop new chemical methods that make it possible to go directly and accurately from the primary sequence to the secondary structure and eventually from the secondary structure to the tertiary structure.

It is becoming increasingly evident that SHAPE represents a major advance relative to other ways of looking at RNA structure. SHAPE allows accurate and comprehensive RNA structural information to be achieved in a single experiment. The initial electrophile used for SHAPE, NMIA, is not very reactive and therefore RNA structures must be probed over periods of tens of minutes. In Chapter 2, I design and synthesize a significantly faster and more useful SHAPE reagent, 1-methyl-7-nitroisatoic anhydride (1M7). My "second generation" reagent, 1M7, is capable of taking snapshots of RNA structure over 70 s periods, is insensitive to solution conditions like Mg^{2+}, and provides highly accurate secondary structure models. In Chapter 3, Costin Gherghe and I apply SHAPE reagents of varying electrophilicity to monitor conformational dynamics over distinct time domains. We observed that select C2′-endo nucleotides in both a model RNA and a large RNA with a complex tertiary structure undergo extraordinarily slow conformational dynamics. This discovery is the first observation of extremely slow motions in RNA.

In Chapter 4, I identify a "third generation" SHAPE reagent, benzoyl cyanide (BzCN), which reacts with RNA 2′-hydroxyl groups in a structure–selective way and provides 1 second snapshots of RNA structure. This reagent makes possible

straightforward and very accessible technologies for following RNA structure in a time-dependent manner. I use this chemistry to study the time-dependent folding of the ribonuclease P specificity domain. Specifically, I identify a slow folding region in the RNA and, in Chapter 5, I attribute this slow folding step to the slow conformational dynamics at a single C2′-endo nucleotide in the RNA. This discovery was the first evidence that extremely slow conformational dynamics at certain C2′-endo nucleotides can play a central role in RNA folding.

Finally, in Chapter 6, I create a new way to assess RNA tertiary structure in an experimentally straightforward way by developing a reagent, *N*,*N*-(dimethylamino)dimethylsilyl chloride (DMAS-Cl), that reacts with the N2 position of guanosine in a manner that is rigorously proportional to solvent accessibility. DMAS-Cl yields near-perfect measures of solvent accessibility in RNA and we believe that this may now be the best reagent choice for analyzing the solvent accessibility of the base-pairing face of guanosine in RNA.

1.1.6. Perspective. In this work I utilize principles of physical organic chemistry to improve upon and create new technologies for the straightforward and accurate interrogation of RNA structure and dynamics. I show that, with these newly developed technologies, I am able to predict secondary structures with greater than 90% accuracy, discover slow conformational dynamics at certain nucleotides, study the time-dependent folding of an RNA, uncover a new motif in RNA regulation, and probe the solvent accessibility in an RNA with near-perfect accuracy. It is my hope that the technologies that I have created will continue to find wide application in the field of RNA biology and for understanding RNA structure-function relationships.

1.2 References

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

2. Brion, P. and Westhof, E., *Hierarchy and Dynamics of RNA folding*. Annu. Rev. Biophys. Biomol. Struct., 1997. **26**: p. 113-137.

3. Tinoco, I. and Bustamante, C., *How RNA folds*. J. Mol. Biol., 1999. **293**: p. 271-281.

4. Peattie, D.A. and Gilbert, W., *Chemical Probes for higher-order structure in RNA*. Proc. Natl. Acad. Sci. U.S.A., 1980. **77**: p. 4679-4682.

5. Ehresmann, C., et al., *Probing the structure of RNAs in solution*. Nucleic Acids Res., 1987. **15**: p. 9109-9128.

6. Lavery, R. and Pullman, A., *A New Theoretical Index of Biochemical Reactivity Combining Steric and Electrostatic Factors - an Application to Yeast Transfer Rnaphe*. Biophys. Chem., 1984. **19**(2): p. 171-181.

7. Mathews, D.H., et al., *Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure*. Proc. Natl. Acad. Sci. U. S. A., 2004. **101**(19): p. 7287-92.

8. Krasilnikov, A.S., et al., *Crystal structure of the specificity domain of ribonuclease P*. Nature, 2003. **421**(760-764).

9. Brown, J.W., et al., *Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements*. Proc. Natl. Acad. Sci. U. S. A., 1996. **93**(7): p. 3001-6.

10. Massire, C., Jaeger, L., and Westhof, E., *Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis*. J. Mol. Biol., 1998. **279**(4): p. 773-93.

11. Merino, E.J., et al., *RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE)*. J. Am. Chem. Soc., 2005. **127**: p. 4223-4231.

12. Wilkinson, K.A., et al., *High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states*. PLoS Biol., 2008. **6**(4): p. e96.

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**: p. 1610-1616.

14. Wilkinson, K.A., Merino, E.J., and Weeks, K.M., *RNA SHAPE chemistry reveals non-hierarchical interactions dominate equilibrium structural transitions in $tRNA^{Asp}$ transcripts*. J. Am. Chem. Soc., 2005. **127**: p. 4659-4667.

15. 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**: p. 3454-3461.

16. Duncan, C.D.S. 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**: p. 8504-8513.

17. Mathews, D.H. and Turner, D.H., *Prediction of RNA secondary structure by free energy minimization*. Curr. Opin. Struct. Biol., 2006. **16**: p. 270-278.

18. Deigan, K.E., et al., *Accurate SHAPE-Directed RNA Structure Determination* Proc. Natl. Acad. Sci. U. S. A., 2008. **submitted**.

19. Tullius, T.D. and Dombroski, B.A., *Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein*. Proc. Natl. Acad. Sci. U. S. A., 1986. **83**(15): p. 5469-73.

20. Balasubramanian, B., Pogozelski, W.K., and Tullius, T.D., *DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone*. Proc. Natl. Acad. Sci. U. S. A., 1998. **95**(17): p. 9738-43.

21. Cate, J.H., et al., *Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing*. Science, 1996. **273**: p. 1678-1685.

22. Adams, P.L., et al., *Crystal Structure of a Group I Intron Splicing Intermediate*. RNA, 2004. **10**: p. 1867-1887.

23. Chen, C., et al., *Structural energetics and base-pair opening dynamics in sarcin-ricin domain RNA*. Biochemistry, 2006. **45**(45): p. 13606-13.

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

25. Hennelly, S.P., et al., *A time-resolved investigation of ribosomal subunit association*. J. Mol. Biol., 2005. **346**(5): p. 1243-58.

26. Tijerina, P., Mohr, S., and Russell, R., *DMS footprinting of structured RNAs and RNA-protein complexes*. Nat. Protoc., 2007. **2**(10): p. 2608-2623.

27. Sclavi, B., et al., *RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting*. Science, 1998. **279**(5358): p. 1940-1943.

28. Shcherbakova, I. and Brenowitz, M., *Monitoring structural changes in nucleic acids with single residue spatial and millisecond time resolution by quantitative hydroxyl radical footprinting*. Nat. Protoc., 2008. **3**(2): p. 288-302.

CHAPTER 2

A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry

2.1 Introduction

RNA sequences fold back on themselves to form structures that are difficult to predict, especially if only a single sequence is known [1-3]. Current algorithms correctly predict 50-70% of known base pairs on average [4, 5]. Predicted secondary structure models achieving 50-70% accuracy tend to have regions in which the overall topology differs significantly from the correct one, making it difficult or impossible to develop robust biological hypotheses. Knowledge of which nucleotides are likely to be paired or single-stranded can significantly improve prediction accuracies [6-10].

The ideal technology for chemically-assisted RNA structure analysis would (1) be experimentally straightforward, (2) use a single reagent that reacts generically with all four nucleotides, (3) employ a self-quenching reagent, (4) involve short reaction times, and (5) be proven to yield accurate results with complex RNAs of known structure.

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) chemistry [11-13] takes advantage of the discovery that the nucleophilic reactivity of a ribose 2′-hydroxyl group is gated by local nucleotide flexibility. At nucleotides constrained by base pairing or tertiary interactions, the 3′-phosphodiester anion and other interactions reduce reactivity of the 2′-hydroxyl [11]. In contrast, flexible positions preferentially adopt conformations (Figure 2.1A) that react with N-methlyisatoic anhydride (**1**, NMIA) to form a 2′-*O*-adduct (Figure 2.1B). NMIA reacts generically with all four nucleotides and the reagent undergoes a parallel, self-inactivating, hydrolysis reaction (Figure 2.1B). Thus SHAPE chemistry meets the first three of the five criteria outlined above.

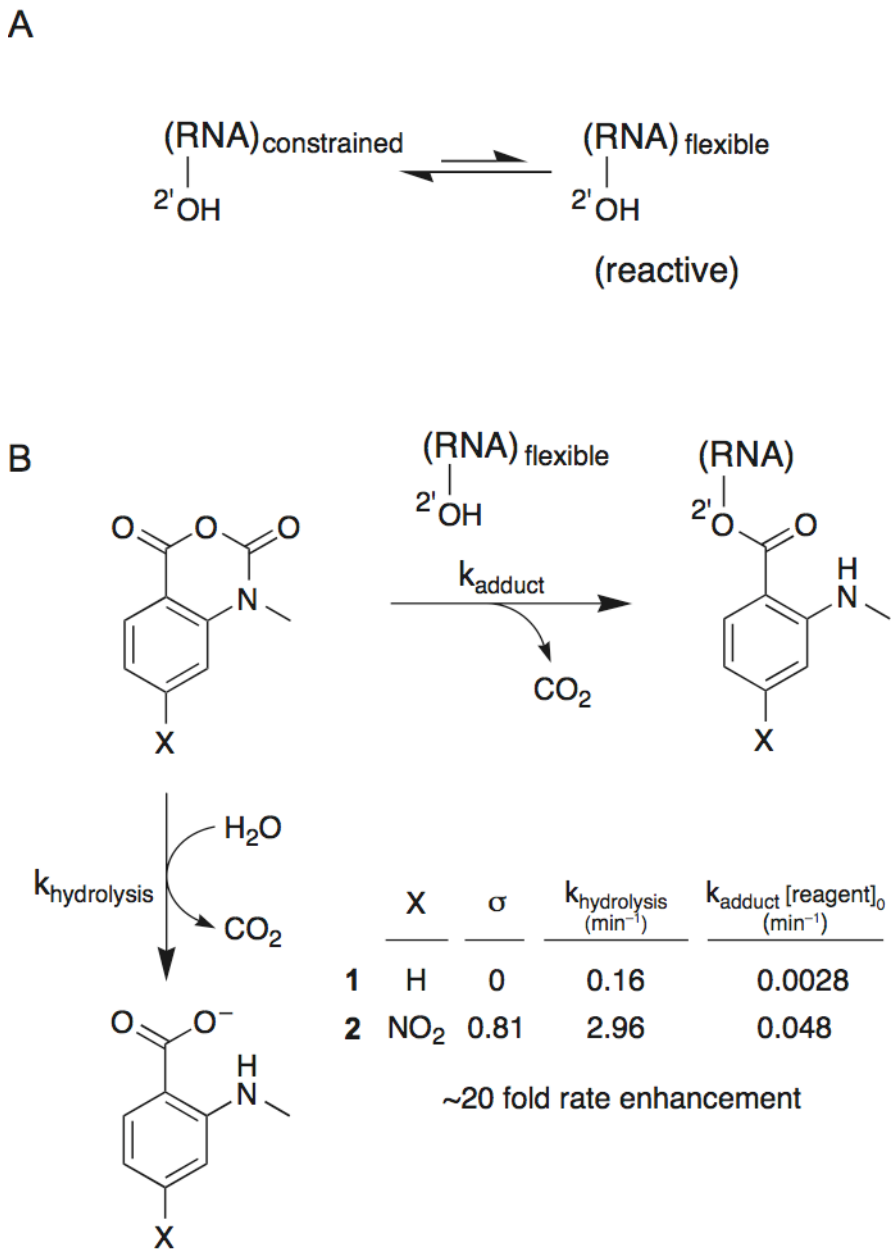


Figure 2.1. Mechanism of RNA SHAPE Chemistry. (A) The nucleophilc reactivity of the 2′-hydroxyl group is selectively enhanced at flexible positions [11]. (B) Parallel reaction of *N*-methylisatoic anhydride derivatives with RNA 2′-hydroxyl groups and with water. $[reagent]_0$ was 5 mM.

However, NMIA is relatively unreactive and requires tens of minutes to react to completion. To address the final two criteria for a near-ideal RNA structure interrogation technology, we design a significantly more useful, fast acting, reagent for SHAPE chemistry. We then show that structural constraints obtained using this reagent allow the secondary and tertiary structure of a large RNA to be assessed with high accuracy.

2.2 Results

2.2.1 Comparative Reactivity of 1M7 and NMIA. The reactive carbonyl of NMIA is in the benzylic position relative to the aromatic ring system and should be sensitive to substituents in the *meta* or the *para* positions, due to a direct resonance effect. We therefore evaluated the 2′-*O*-adduct-forming and hydrolysis activities of 1-methyl-7-nitroisatoic anhydride (**2**, 1M7) (Figure 2.1B). The *para* nitro substituent is strongly electron-withdrawing (σ_p = 0.81) [14] and should increase adduct formation and hydrolysis rates in two ways: via a ground state effect by increasing the electrophilicity of the reactive carbonyl and via a transition state effect by stabilizing the negative charge in the developing tetrahedral reaction intermediate. We first monitored reagent hydrolysis as the increase in UV absorbance of the aminobenzoate products. 1M7 is significantly more labile towards hydrolysis than NMIA. 1M7 undergoes hydrolysis with a half-life of 14 sec and therefore the reaction is complete in ~70 sec; whereas, NMIA requires over 20 minutes to react to completion (left panel, Figure 2.2).

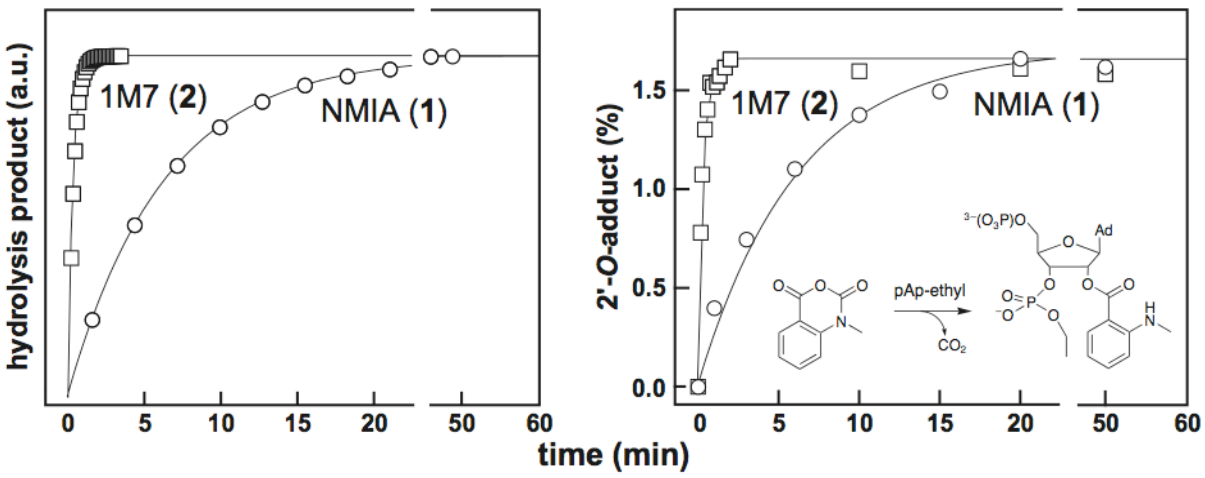


Figure 2.2. Comparative reactivity of 1M7 and NMIA via hydrolysis (left) and 2′-*O*-adduct formation with pAp-ethyl (right)

We next evaluated the ability of each compound to react with 3′-phosphoethyl-5′-adenosine monophosphate (pAp-ethyl). pAp-ethyl contains a 2′-hydroxyl and 3′-phosphodiester monoanion and is a good analogue for an unstructured RNA nucleotide [11]. 1M7 reacts significantly more rapidly with pAp-ethyl than does NMIA; however, the final extent of 2'-*O*-adduct formation for the two compounds is identical, within error (right panel,Figure 2.2).

Identical extents of reaction for NMIA and 1M7, despite the much faster reactivity of 1M7, indicate that the rates of hydrolysis and of 2'-hydroxyl acylation have increased by precisely the same 20-fold increment. These experiments indicate that 1M7 has the ideal chemical characteristics for a fast acting and self-quenching reagent for RNA SHAPE chemistry.

2.2.2 Accurate and quantitative RNA structure analysis using 1M7. We next evaluated the extent to which 1M7 provides accurate and quantitative information regarding RNA structure using the specificity domain of the *Bacillus subtilis* RNase P enzyme. This domain was chosen because it is a large (154 nt) RNA with a known structure [15] that does not contain pseudoknots, which are not well predicted by current algorithms. This RNA spans numerous typical base-pairing and stacking interactions, a tetraloop-receptor tertiary interaction (involving L12 and P10.1) common to many large RNAs, and two large internal loops (J11/12 and J12/11) stabilized by an extensive series of non-canonical interactions [15].

A SHAPE experiment was performed on the RNase P domain under conditions that stabilize the native tertiary fold (6 mM $MgCl_2$, 100 mM NaCl, pH 8.0) by treating the RNA with 6.5 mM 1M7. Sites of 2′-*O*-adduct formation were identified as stops to

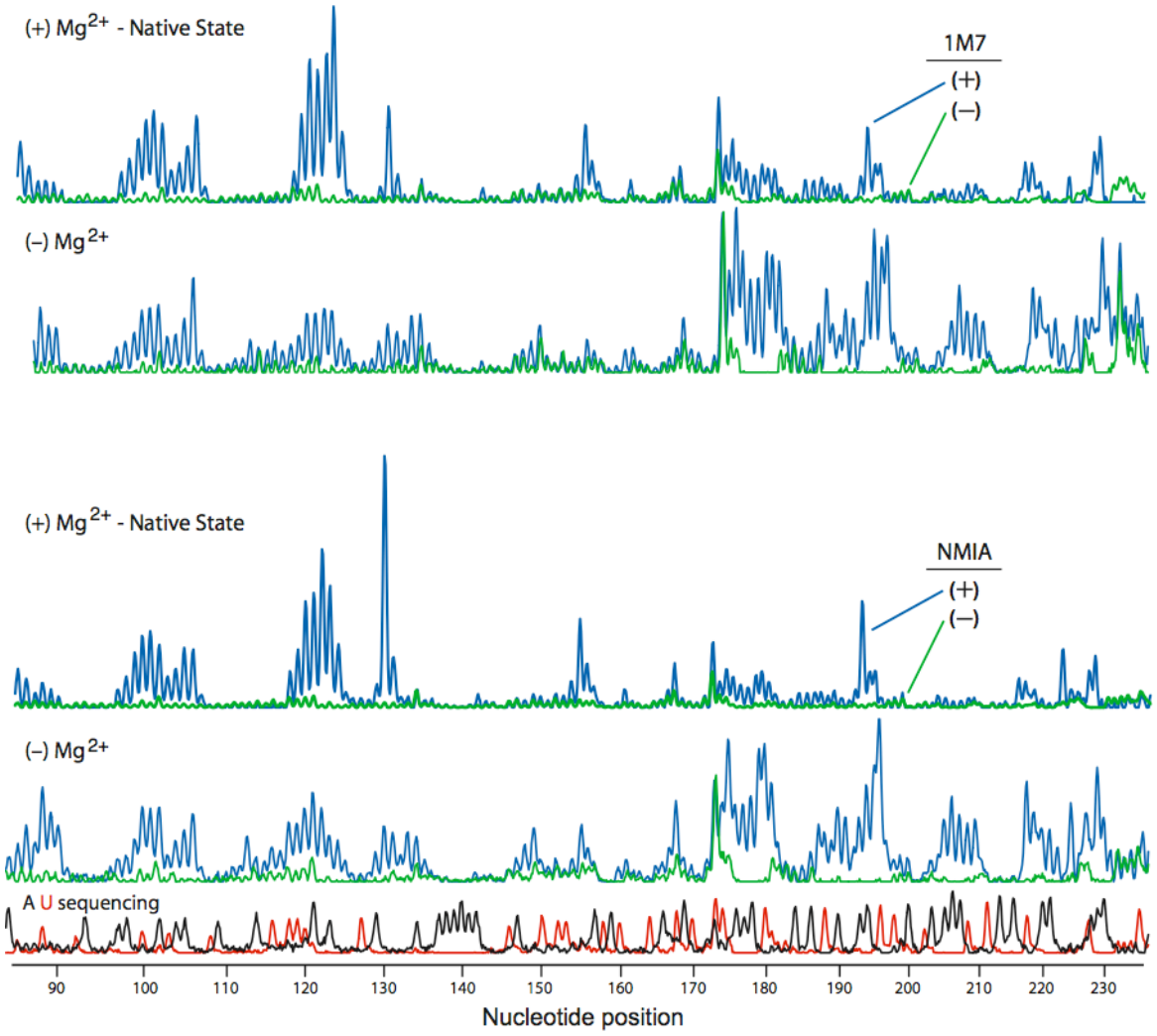


Figure 2.3. Comparison of processed capillary electrophoresis traces for SHAPE experiments performed using 1M7 (upper panels) and NMIA (lower panels). Peak areas correspond to position and relative extent of 2′-*O*-adduct formation at each nucleotide. Sequencing ladders (A & U) were used to assign peak positions. Both (+) and (–) Mg^{2+} experiments are shown for each reagent.

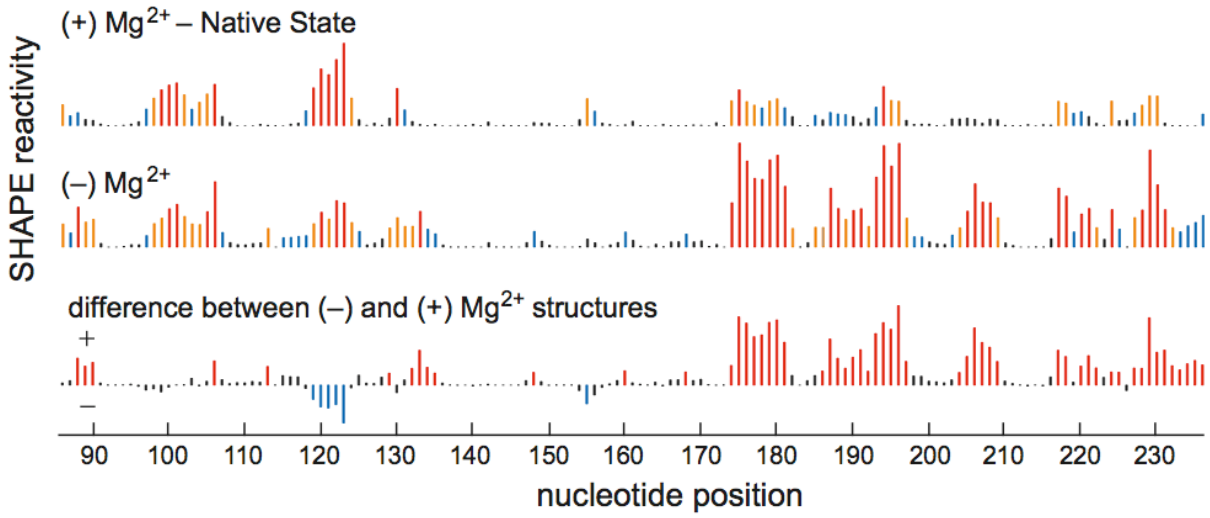


Figure 2.4. Histograms and difference plot of absolute reactivities for SHAPE experiments performed with 1M7 in the presence and absence of 6 mM Mg^{2+}.

primer extension, using fluorescently labeled DNA primers, resolved by capillary electrophoresis [16] (Figure 2.3). Absolute SHAPE reactivities were calculated by subtracting the background observed in no-reagent control experiments that omitted 1M7. Reactivity at each nucleotide was classified as high, medium, low, or near-zero (red, orange, blue, and black columns, respectively, in Figure 2.4).

Superposition of the quantitative reactivity information on a secondary structure diagram [15] for the RNase P specificity domain shows that a 70 sec reaction with 1M7 accurately reports the known secondary and tertiary structure for this RNA (Figure 2.5A). Essentially all nucleotides involved in Watson-Crick base-pairs are unreactive; moreover, many non-canonical, but stable, U•G, A•A, and A•G pairs are unreactive. Nucleotides in P10.1 and in L12 that form the tetraloop-receptor tertiary structure motif are also unreactive. In contrast, nucleotides in loops or adjacent to bulges or other irregularities are reactive. Nucleotides in the structurally idiosyncratic module involving J11/12 and J12/11 show a wide range of reactivities. Strikingly, the most highly conserved nucleotides in this module (A187, A191, G219-G220, A222), which participate in stabilizing tertiary interactions [15], also show the lowest SHAPE reactivities using 1M7.

We performed a similar SHAPE experiment in the absence of magnesium ion (middle histogram, Figure 2.4). Control experiments show that both reaction with the model nucleotide, pAp-ethyl, and 1M7 hydrolysis are independent of Mg^{2+} concentration (Figure 2.6). This Mg^{2+}-independence represents an additional significant improvement over the parent compound, NMIA, whose reactivity is strongly dependent on ionic strength (Figure 2.6). Thus, observed changes in SHAPE reactivity with 1M7

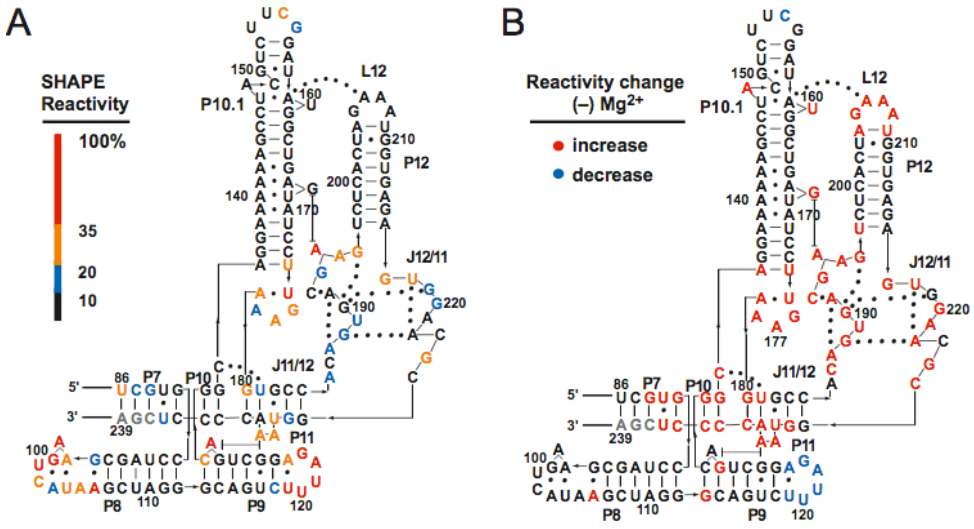


Figure 2.5. Base pairing and tertiary interactions for the specificity domain of *Bacillus subtilis* RNase P .

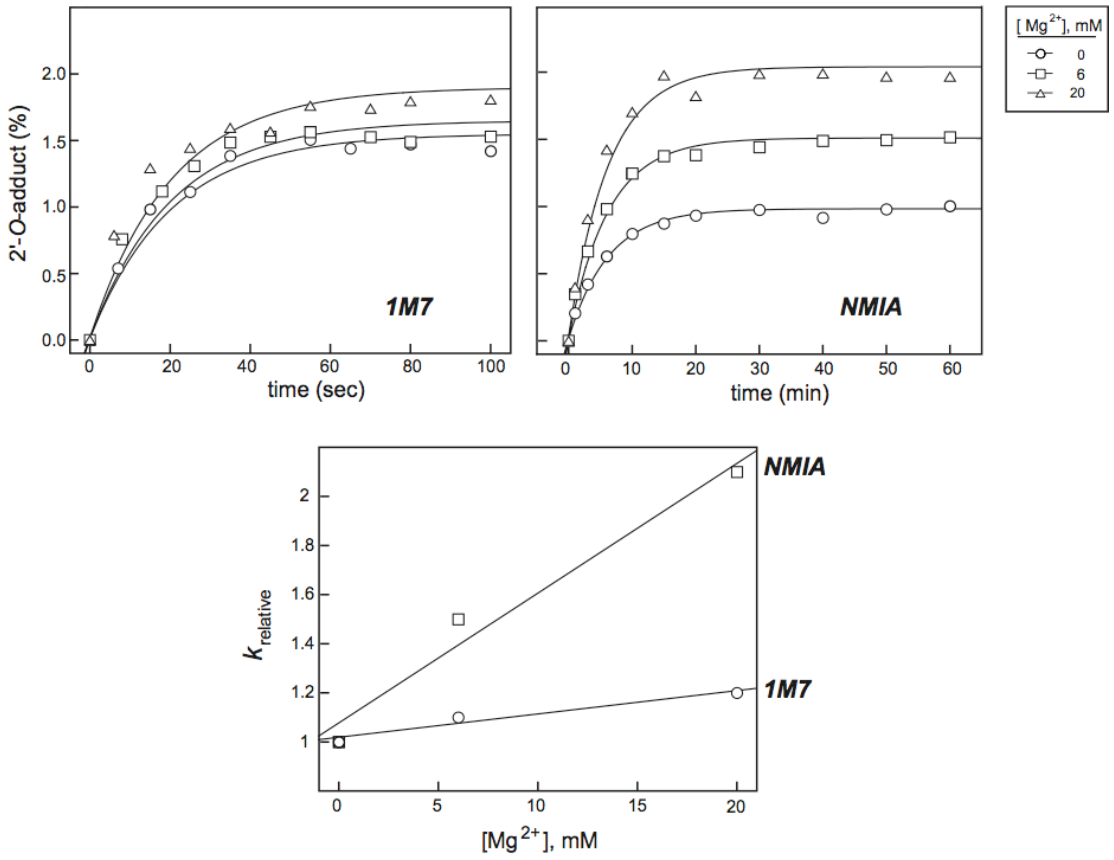


Figure 2.6. Reaction between the model nucleotide, pAp-ethyl, and 1M7 is independent of Mg^{2+} concentration over the range 0-20 mM; whereas, reaction of pAp-ethyl with the parent compound, NMIA, is not. The dependence of reaction rate on Mg^{2+} concentration is shown by both the absolute rate and the extent of 2´-*O*-adduct formation at long time points for 0, 6, and 20 mM Mg^{2+} (top). The change in rate from 0 to 20 mM Mg^{2+} for 1M7 is negligible, while for NMIA the change is greater than 2-fold (bottom).

reflect changes in RNA secondary and tertiary structure and not Mg^{2+}-induced differences in reagent properties.

We quantified the effect of Mg^{2+} on the structure of the RNA using a difference plot in which nucleotide reactivities in the (+) Mg^{2+} experiment were subtracted from the (–) Mg^{2+} experiment. Positive and negative peaks thus indicate an increase or decrease in local nucleotide flexibility in the absence of Mg^{2+}, respectively (lower panel, Figure 2.4). Many sites in the (–) Mg^{2+} experiment show increased SHAPE reactivity. Strikingly, increased reactivity occurs precisely at nucleotides that participate in *tertiary* interactions in the RNase SHAPE reactivity also shows that the irregularly stacked P7-P10-P11 helical domain unfolds when Mg^{2+} is removed (Figure 2.5B).

2.2.3. Prediction of highly accurate secondary structure models using 1M7. We then evaluated how well SHAPE information can be used to constrain the output of an RNA secondary structure prediction algorithm. We calculated prediction accuracies both using the native secondary structure (Figure 2.5A) as the target and using a modified structure that excluded the Mg^{2+}-dependent base pairs in the P7-P10-P11 domain. When the specificity domain of *Bacillus subtilis* RNase P is folded in RNAstructure [9], the lowest free energy structure contains 52% of the correct pairs and features an overall topology that is radically different from the correct structure (Figure 2.7A). When SHAPE reactivity information is added to constrain single-stranded and non-internal base pairs, the lowest free energy structure is 76% correct using the native secondary structure as the target and 91% correct when base pairs in the P7-10-P11 domain (which do not form in the absence of native tertiary interactions) are excluded (Figure 2.7B). Using either target structure, the SHAPE-constrained

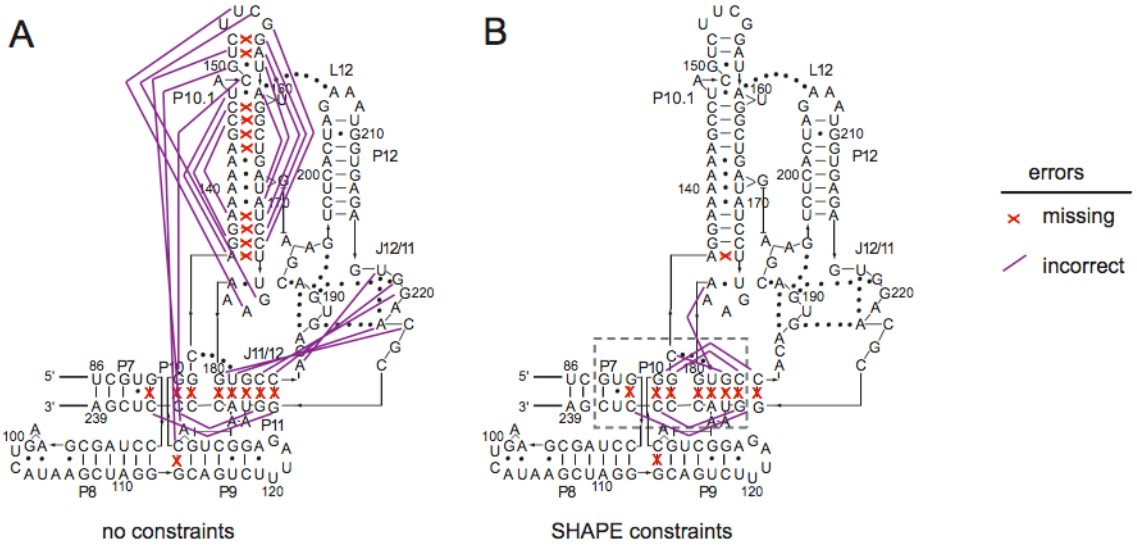


Figure 2.7. Comparison of the known secondary structure for the *B. subtilis* RNase P specificity domain with (A) simple algorithmic prediction based on a thermodynamic model alone and (B) as constrained by RNA SHAPE reactivity using 1M7. Both calculations were performed using the RNAstructure program. Missing base pairs and incorrectly predicted paired nucleotides are illustrated with red x's and purple lines, respectively. Dashed box indicated region that becomes unpaired in the absence of Mg^{2+}.

prediction features an overall topology that closely resembles the correct structure.

2.3 Discussion

SHAPE chemistry performed with 1M7 accurately reports the known structure of the RNase P specificity domain under native conditions. 1M7 reactivity detects nucleotides constrained both by base pairing and by idiosyncratic, non-canonical tertiary interactions (Figure 2.5). SHAPE chemistry enables very precise analysis of the differences between two structures, such as Mg^{2+}-dependent tertiary interactions. 1M7 is easily handled in the laboratory and enables analysis of large RNA structures at single nucleotide resolution in less than 70 seconds.

2.4 Experimental

2.4.1. Synthesis of [^{32}P]-labeled pAp-ethyl. Adenosine-3′-(*O*-ethyl)-phosphate prescursor (10 μM final) was 5′-[^{32}P]-labeled using T4 polynucleotide kinase [10 μL; containing 70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 1 μL T4 PNK (10,000 units/mL), 60 μCi [γ-^{32}P]ATP; 37 °C for 1 hr]; purified by gel electrophoresis (30% polyacrylamide, 29:1 acrylamide:bisacrylamide, 0.4 mm × 28.5 cm × 23 cm; 30 W; 1 hr); excised from the gel; passively eluted into 300 μL HE (10 mM Hepes pH 8.0, 1 mM EDTA; overnight at 4 °C); and separated from solid acrylamide by microfiltration (EZ spin columns, Millipore).

2.4.2. Synthesis of 1-methyl-7-nitroisatoic anhydride (1M7). To a suspension of 0.1656 g (4.14 mmoles) of sodium hydride (60% in mineral oil) in 20 mL DMF was added a solution of 0.6584 g (3.16 mmoles) of 4-nitroisatoic anhydride in 20 mL DMF. After stirring a few minutes at room temperature, a clear orange solution formed. To the reaction 0.2615 g (3.2 mmoles) of methyl iodide were added, and the

mixture was stirred at room temperature for 4 hours. The reaction was poured into 50 mL of cold 1 N HCl, and the resulting bright orange precipitate was filtered and washed sequentially with water and then ether to give 608.3 mg (86%) of product. ^{1}H NMR ($CO(CD_3)_2$, 400 MHz,) δ 3.69 (s, 3 H, -NC*H*$_3$-), 8.12 (dd, *J* = 8.8 Hz, 2 Hz, 1 H, Ar*H*), 8.2 (d, *J* = 2 Hz, 1H, Ar*H*), 8.34 (d, *J* = 8.4 Hz, 1 H, Ar*H*).

2.4.3. NMIA and 1M7 hydrolysis and 2′-*O*-adduct Formation. Hydrolysis was followed by adding (1.5 mM NMIA or 2.0 mM 1M7 in 300 μL DMSO) reagent to 1.1× buffer [2.7 mL, 6.7 mM $MgCl_2$, 111 mM NaCl, 111 mM HEPES (pH 8.0)] equilibrated at 37 °C in a cuvette. Pseudo-first-order rates were obtained by monitoring the absorbance of the hydrolysis product (at 360 nm for 2-methylaminobenzoate and 430 nm for 2-methylamino-4-nitrobenzoate). Rates of adduct formation for [^{32}P]-labeled pAp-ethyl (10,000 cpm/μL) were obtained by adding 10% (v/v) reagent (5 mM final NMIA or 1M7 in DMSO) to 1.1× reaction buffer; quenching the reaction with 1 vol 250 mM dithiothreitol; resolution by gel electrophoresis (30% polyacrylamide; 29:1 acrylamide:bisacrylamide; 0.4 mm × 28.5 cm × 23 cm; 30 W; 45 min); and quantifying by phosphorimaging. Reaction rates were obtained using an equation that accounts for parallel reaction of NMIA or 1M7 by 2′-*O*-adduct formation (k_{adduct}) and by hydrolysis ($k_{\text{hydrolysis}}$): fraction product = $1 - \exp[(k_{\text{adduct}}[\text{reagent}]_0/(k_{\text{hydrolysis}}))(e^{-(k\text{hydrolysis})t} - 1)]$ [17].

2.4.4. Synthesis of *Bacillus subtilis* RNase P RNA. A DNA template for transcription of the specificity domain of the *B. subtilis* RNase P, inserted in the context of a 5′ and 3′ flanking structure cassette [12, 13], was generated by PCR [1 mL; containing 20 mM Tris (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 200 μM each dNTP, 500 nM

each forward and reverse primer, 5 pM template, and 0.025 units/μL Taq polymerase; denaturation at 94 °C, 45 s; annealing 55 °C, 30 s; and elongation 72 °C, 1 min; 38 cycles]. The PCR product was recovered by ethanol precipitation and resuspended in 150 μL of TE [10 mM Tris (pH 8.0), 1 mM EDTA]. Transcription reactions (1.5 mL, 37 °C, 4 h) contained 40 mM Tris (pH 8.0), 10 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 0.01% (v/v) Triton X-100, 4% (w/v) poly(ethylene) glycol 8000, 2 mM each NTP, 50 μL of PCR-generated template, and 0.1 mg/mL of T7 RNA polymerase. The RNA product was purified by denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 7 M urea, 29:1 acrylamide:bisacrylamide, 32 W, 2 h), excised from the gel, and recovered by electroelution and ethanol precipitation. The purified RNA (~4 nmol) was resuspended in 100 μL TE.

2.4.5. Structure-Selective RNA Modification. RNA (2 pmol) in 5 μL 1/2× TE was heated at 95 °C for 2 min, cooled on ice, treated with 3 μL of 3× folding buffer [333 mM NaCl, 333 mM Hepes (pH 8.0), 33.3 mM $MgCl_2$ (or no $MgCl_2$)], and incubated at 37 °C for 20 min. The RNA solution was treated with 1M7 or NMIA (1 μL, 65 mM in anhydrous DMSO), allowed to react for 70 sec (equal to five 1M7 hydrolysis half-lives, accompanied by a colorimetric change from pale yellow-orange to deep orange-brown upon completion) or 25 min (five NMIA hydrolysis half-lives). No-reagent control reactions contained 1 μL of DMSO. Modified RNA was recovered by ethanol precipitation [90 μL sterile H_2O, 5 μL NaCl (5 M), 1 μL Glycogen (20 mg/mL), 400 μL ethanol; 30 min at –80 °C] and resuspended in 10 μL of TE.

2.4.6. Primer Extension. The general procedure is that outlined by Wilkinson et al [16]. A fluorescently labeled DNA primer (5′-Cy5 or Cy5.5-labelled GAA CCG GAC CGA

AGC CCG; 3 μL, 0.4 μM) was added to the RNA (10 μL, from the previous step) by heating to 65 °C (6 min) and 35 °C (20 min). Reverse transcription buffer [6 μL; 167 mM Tris (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] was added; the RNA was heated to 52 °C; Superscript III (1 μL, 200 units) was added and reactions were incubated at 52 °C for 30 min. Primer extension reactions were quenched by addition of 4 μL of an equal mixture of EDTA (100 mM) and sodium acetate (3 M, pH 5.2). The resulting cDNAs were recovered by ethanol precipitation, washed twice with 70% ethanol, dried in a SpeedVac for 10 min, and resuspended in 40 μL de-ionized formamide. Dideoxy sequencing markers were generated using unmodified RNA and primers labeled with unique fluorophores (D2 or IR800, 1 μM), and by adding 1 μL of 3′-deoxythymidine (10 mM) or 3′-dideoxyadenosine (2 mM) triphosphate after addition of reverse transcription buffer. The cDNA extension products were separated by capillary electrophoresis using a Beckman Coulter CEQ 2000XL DNA Analysis System.

2.4.7. Data Analysis. Raw traces from the CEQ 2000XL were processed using a custom software package [18]. Reactivities for comparison of the (+) Mg^{2+} and (–) Mg^{2+} experiments were normalized to intensities at positions 101 and 102; all negative intensities were set to zero. The percent reactivity for each nucleotide was obtained by averaging the highest reactivities, corresponding to positions 123 and 196 for the (+) Mg^{2+} and (–) Mg^{2+} traces, respectively, and dividing all intensities by this average reactive value. On this scale, SHAPE reactivities are reproducible to ±5%. The intensities for the (+) Mg^{2+} experiment were normalized for the purpose of defining constraints for the RNAstructure program [9] by excluding the top 2% of reactive

nucleotides (3 nts), averaging the next 8% of reactive nucleotides (12 nts), and then dividing all intensities by this average high value. This gives intensities from 0 to slightly greater than 2. In RNAstructure [9], nucleotides with reactivities greater than 0.75 were required to be single stranded and positions with reactivities greater than 0.35 were prohibited from forming internal Watson-Crick pairs.

2.5 References

1. Tinoco, I. and Bustamante, C., *How RNA folds*. J. Mol. Biol., 1999. **293**: p. 271-281.

2. Eddy, S.R., *How do RNA folding algorithms work?* Nature Biotechnol., 2004. **22**(11): p. 1457-1458.

3. Doshi, K.J., et al., *Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction*. BMC Bioinf., 2004. **5**: p. 105.

4. Dowell, R.D. and Eddy, S.R., *Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction*. BMC Bioinf., 2004. **5**: p. 71.

5. Mathews, D.H. and Turner, D.H., *Prediction of RNA secondary structure by free energy minimization*. Curr. Opin. Struct. Biol., 2006. **16**: p. 270-278.

6. Peattie, D.A. and Gilbert, W., *Chemical Probes for higher-order structure in RNA*. Proc. Natl. Acad. Sci. U.S.A., 1980. **77**: p. 4679-4682.

7. Ehresmann, C., et al., *Probing the structure of RNAs in solution*. Nucleic Acids Res., 1987. **15**: p. 9109-9128.

8. Knapp, G., *Enzymic approaches to probing of RNA secondary and tertiary structure*. Methods Enzymol., 1989. **180**: p. 192-212.

9. Mathews, D., et al., *Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure*. Proc. Natl. Acad. Sci. U. S. A., 2004. **101**(19): p. 7287-7292.

10. Badorrek, C.S. and Weeks, K.M., *RNA flexibility in the dimerization domain of a gamma retrovirus*. Nat. Chem. Biol., 2005. **1**(2): p. 104-11.

11. Merino, E.J., et al., *RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE)*. J. Am. Chem. Soc., 2005. **127**: p. 4223-4231.

12. Wilkinson, K.A., Merino, E.J., and Weeks, K.M., *RNA SHAPE chemistry reveals non-hierarchical interactions dominate equilibrium structural transitions in tRNA^{Asp} transcripts*. J. Am. Chem. Soc., 2005. **127**: p. 4659-4667.

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. Exner, O., *Correlation Analysis in Chemistry*, ed. Chapmen, N.B. and J.Shorter. 1978, New York: Plenum Press.

15. Krasilnikov, A.S., et al., *Crystal structure of the specificity domain of ribonuclease P*. Nature, 2003. **421**: p. 760-764.

16. Wilkinson, K.A., et al., *High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states*. PLoS Biol., 2008. **6**(4): p. e96.

17. Chamberlin, S.I. and Weeks, K.M., *Mapping local nucleotide flexibility by selective acylation of 2 '-amine substituted RNA*. J. Am. Chem. Soc., 2000. **122**(2): p. 216-224.

18. Vasa, S.M., et al., *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.

CHAPTER 3

Slow Conformational Dynamics at C2´-endo Nucleotides in RNA

3.1 Introduction

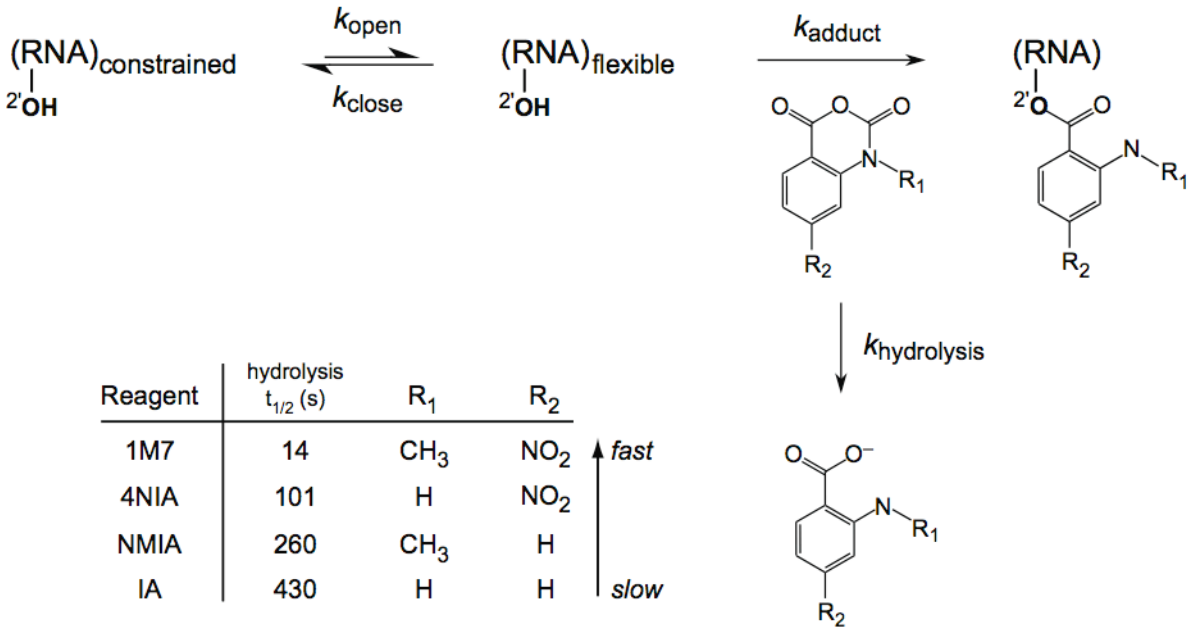
Local and global dynamics in folded RNAs occur over broad timescales spanning picoseconds to minutes [1, 2]. Slow motions likely play predominant roles in governing RNA folding and ribonucleoprotein assembly reactions. However, slow local motions are extremely difficult to detect in a routine way, especially for large RNAs with complex structures.

The local environment and degree of flexibility can be evaluated at nucleotide resolution for RNAs of any size using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) chemistry [3-5]. RNA nucleotides exist in equilibrium between constrained (closed) and flexible (open) states. The 2'-OH group in flexible nucleotides preferentially adopts an open, reactive, conformation that facilitates reaction with electrophilic reagents to form a 2'-*O*-adduct (Figure 3.1). SHAPE experiments work well using electrophiles based on the isatoic anhydride (IA) scaffold [3, 6]. Positions that form 2'-*O*-adducts are detected by primer extension [3-5].

IA derivatives both react with the RNA 2'-OH group and also undergo concurrent degradation by hydrolysis (Figure 3.1). 2'-hydroxyl reactivity is thus conveniently monitored by allowing a reaction to proceed until the reagent has been consumed, either by hydrolysis or by reaction with RNA. At this endpoint, the fraction adduct [7] at any nucleotide (f) is:

$$f \approx 1 - e^{-(k_{\mathrm{obs}}/k_{\mathrm{hydrolysis}})} \qquad (1)$$

where



Reagent	hydrolysis $t_{1/2}$ (s)	R_1	R_2	
1M7	14	CH_3	NO_2	*fast*
4NIA	101	H	NO_2	
NMIA	260	CH_3	H	
IA	430	H	H	*slow*

Figure 3.1. Mechanistic framework for RNA SHAPE chemistry

$$k_{obs} = \frac{k_{open} k_{adduct}[reagent]}{k_{open} + k_{close} + k_{adduct}[reagent]} \quad (2)$$

And the rate of hydrolysis has been shown to be proportional to the rate of adduct formation [4, 6],

$$k_{adduct}/k_{hydrolysis} = \beta \quad (3)$$

These relationships lead to two limits. In Limit 1, $k_{open} + k_{close} >> k_{adduct}[reagent]$,

$$f = 1 - e^{-\frac{k_{open}}{k_{open}+k_{close}}\beta[reagent]} \quad (4)$$

For Limit 2, $k_{open} + k_{close} << k_{adduct}[reagent]$,

$$f = 1 - e^{-k_{open}/k_{hydrolysis}} + A \quad (5),$$

where A represents a plateau term that corresponds to a small fraction of RNAs that are unfolded, therefore reactive towards all reagents.

In this chapter, Costin Gherghe and I show that is possible to monitor local nucleotide dynamics in RNA under conditions where Limit 2 applies by varying the reactivity (or $k_{hydrolysis}$) of the hydroxyl-selective electrophile. IA has a hydrolysis half-life ($t_{1/2}$) of 430 sec at 37 °C (table, Figure 3.1). Electron withdrawing substituents at the cyclic amine (R_1) or in the benzene ring (R_2) enhance reagent reactivity. Compared to IA, N-methyl isatoic anhydride (NMIA), 4-nitro-isatioc anhydride (4NIA) and 1-methyl 7-nitroisatoic anhydride (1M7) [6] have progressively shorter hydrolysis half-lives (Figure 3.1).

3.2 Results

3.2.1 SHAPE analysis of C2´-endo nucleotides. To investigate if distinct local nucleotide dynamics can be captured by varying the SHAPE electrophile, we analyzed

local nucleotide dynamics at an important variation in RNA structure: the C2'-endo conformation. Although C2'-endo nucleotides are relatively rare, they are highly overrepresented in important RNA tertiary interactions and in catalytic active sites [8]. Local structure at tandem G•A mismatches depends on the local sequence context [9-11]. Guanosine nucleotides in G•A pairs adopt the C2'-endo conformation in the sequences $(UGAA)_2$ [10] and $(GGAU)_2$ [11], the C3'-endo conformation typical of standard A-form helix geometry in $(CGAG)_2$ [9], and a mixture of C2'-endo/C3'-endo conformations in $(UGAG)_2$ [11]. Costin Gherghe constructed a simple hairpin RNA (termed the C2'-endo RNA) containing each of these sequences. Because these G•A mismatch-containing sequences are palindromic, there are two equivalent examples of each G•A pair in the RNA, including four total C2'-endo nucleotides (in red, Figure 3.2A).

When the C2'-endo RNA was subjected to SHAPE analysis using the fastest reagent, 1M7, flexible nucleotides in the apical loop (nts 45-48) are reactive, while positions constrained by base pairing are unreactive, regardless of the sugar pucker (top panel, Figure 3.2B). When an otherwise identical experiment was performed with IA, which reacts 30-fold more slowly, nucleotides in the apical loop were again reactive while most of the base paired nucleotides were unreactive, similar to their reactivities with 1M7. In strong contrast, the four G nucleotides that adopt the C2'-endo conformation were highly reactive, even more so than some nucleotides in the flexible loop (Figure 3.2B, bottom panel; red bars at positions 19, 40, 52, 73). For the two reagents with intermediate reactivities, 4NIA and NMIA, the C2'-endo positions are moderately reactive (Figure 3.2B). Nucleotides constrained in these C2'-endo conformations are thus unreactive towards fast-reacting electrophiles but are highly

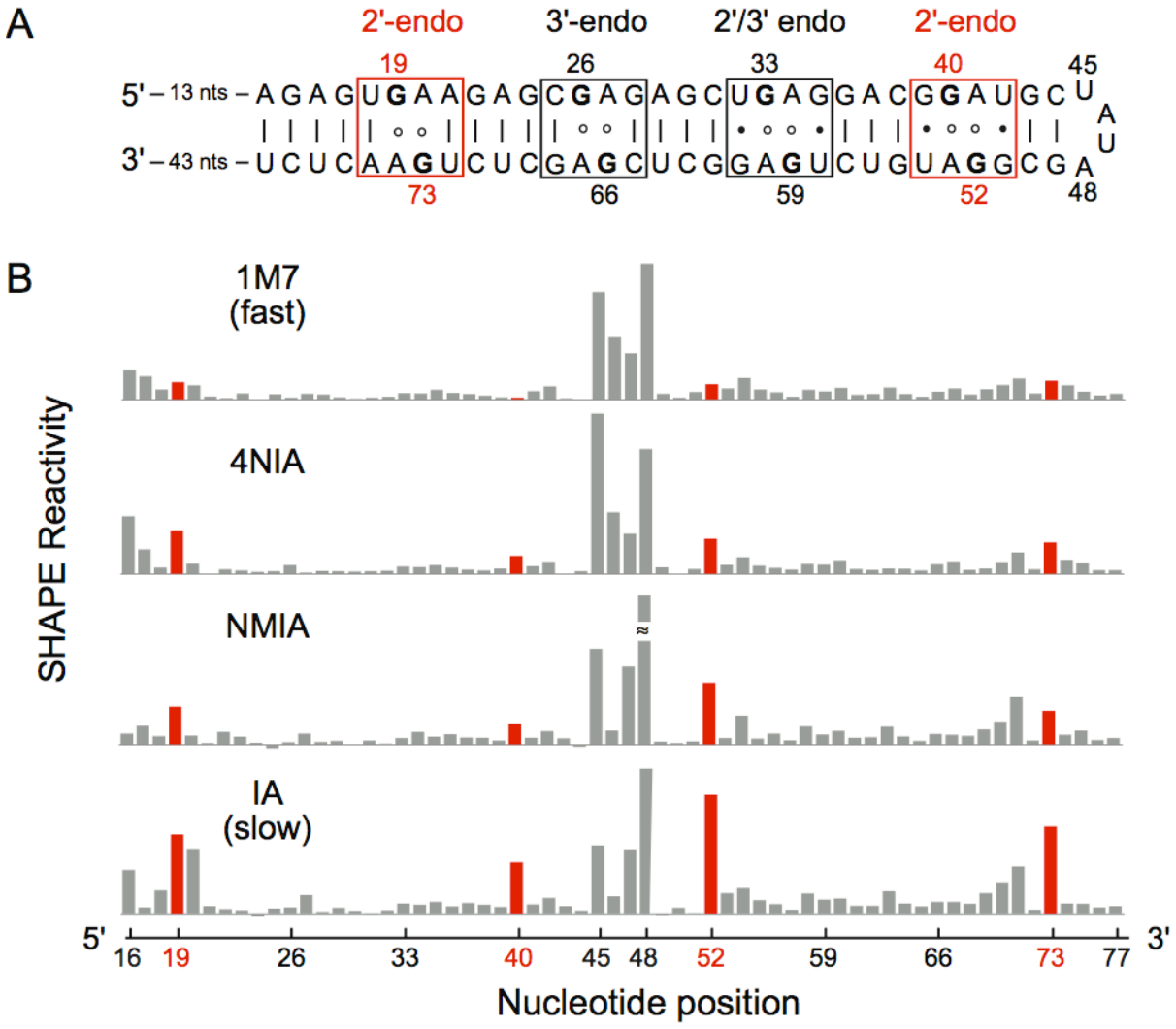


Figure 3.2. (A) The C2′-endo RNA construct contains tandem G•A base pairs that adopt either the C2′- or C3′-endo ribose conformations. (B) Absolute SHAPE reactivities as a function of reagent electrophilicity.

reactive towards the slower reagents.

Limit 1 (Eqn. 4) predicts that adduct formation is proportional to the equilibrium constant for formation of the open state [$k_{open}/(k_{open} + k_{close})$] and is independent of the reagent hydrolysis rate. Most nucleotides thus far analyzed by SHAPE, including in tRNA [3, 4] and in an RNase P RNA [6], are characterized by limit 1.

In contrast, the observation of a strong dependence of adduct formation on $k_{hydrolysis}$ suggests that Limit 2 applies to the C2'-endo nucleotides in Figure 2A. Limit 2 also implies (1) that the extent of reaction at C2'-endo nucleotides will be independent of reagent concentration and (2) that k_{obs} reports k_{open} (compare Eqns. 4 & 5). Costin Gherghe analyzed the concentration dependence for reaction at positions 52 and 73 using isatoic anhydride and found, as predicted by Limit 2, that adduct formation is independent of reagent concentration under conditions where reaction of the unconstrained model nucleotide, pAp-ethyl, showed a clear concentration dependence (Figure 3.3).

Costin Gherghe estimated the magnitude of k_{open} at the C2'-endo nucleotides at positions 19/73 and 40/52 by fitting the extent of 2'-*O*-adduct formation as a function of $k_{hydrolysis}$ to Eqn. 5 (Figure 3.4). In both cases, k_{open} is 4×10^{-5} s^{-1}. In contrast, reactivities for both pAp-ethyl and flexible loop nucleotides are independent of $k_{hydrolysis}$ (Figure 3.5).

Critically, some C2'-endo nucleotides thus experience extraordinarily slow local dynamics to form conformations reactive towards isatoic anhydride-based electrophiles.

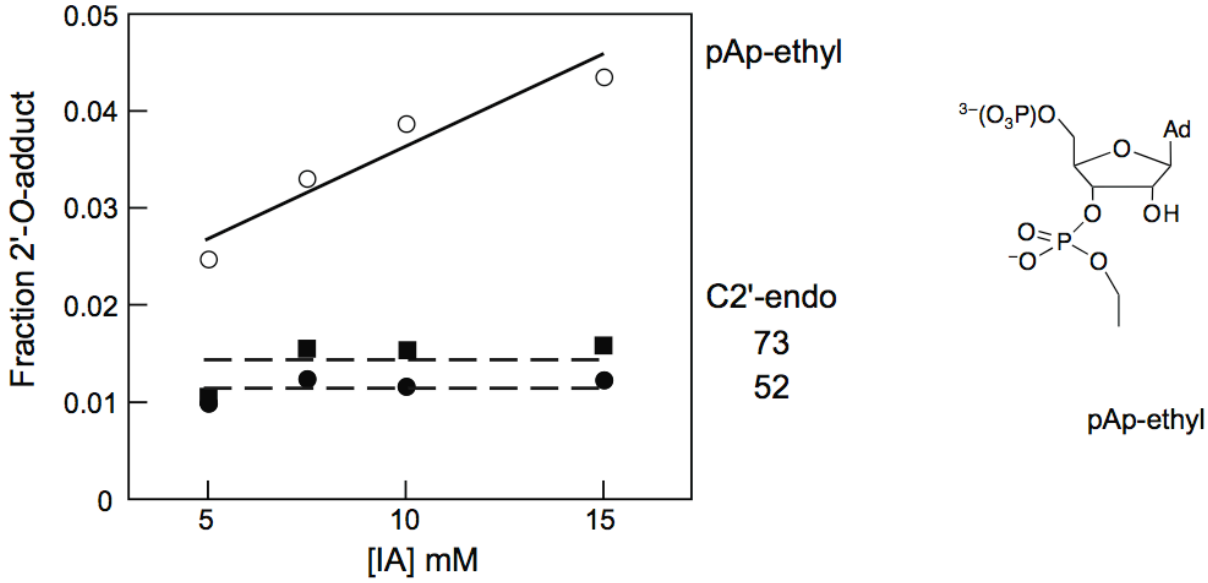


Figure 3.3. Concentration dependence for reaction at positions 52 and 73 in the C2´-endo RNA construct and for the (unconstrained) model nucleotide pAp-ethyl.

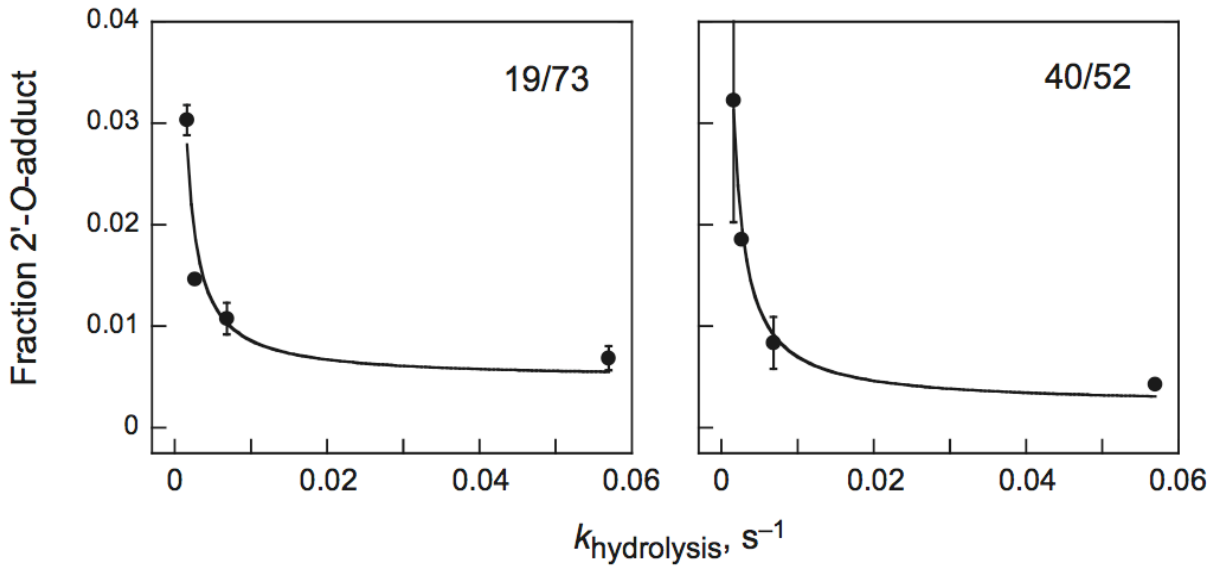


Figure 3.4. Determination of k_{open} for C2′-endo position 19/73 and 40/52. Lines represent a fit to Eqn. 5; rate constants are ~ 4×10^{-5} s^{-1}. Error bars indicate standard deviations from two independent measurements.

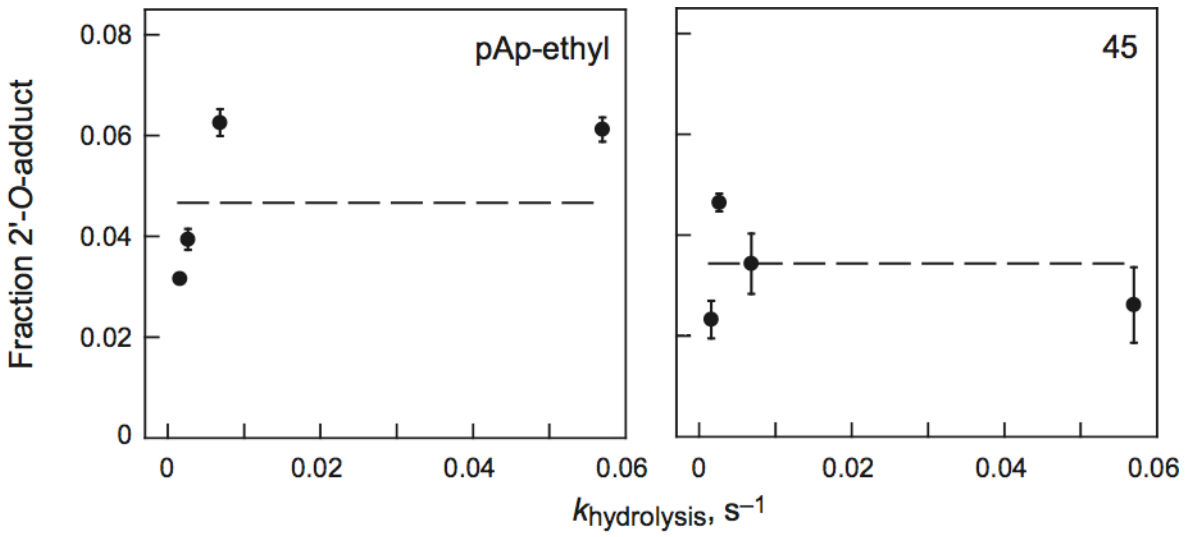


Figure 3.5. Absence of a dependence of fraction adduct formed as a function of $k_{hydrolysis}$ for reaction of the (unconstrained) model nucleotide pAp-ethyl and for nucleotide 45 in the loop of the C2′-endo construct.

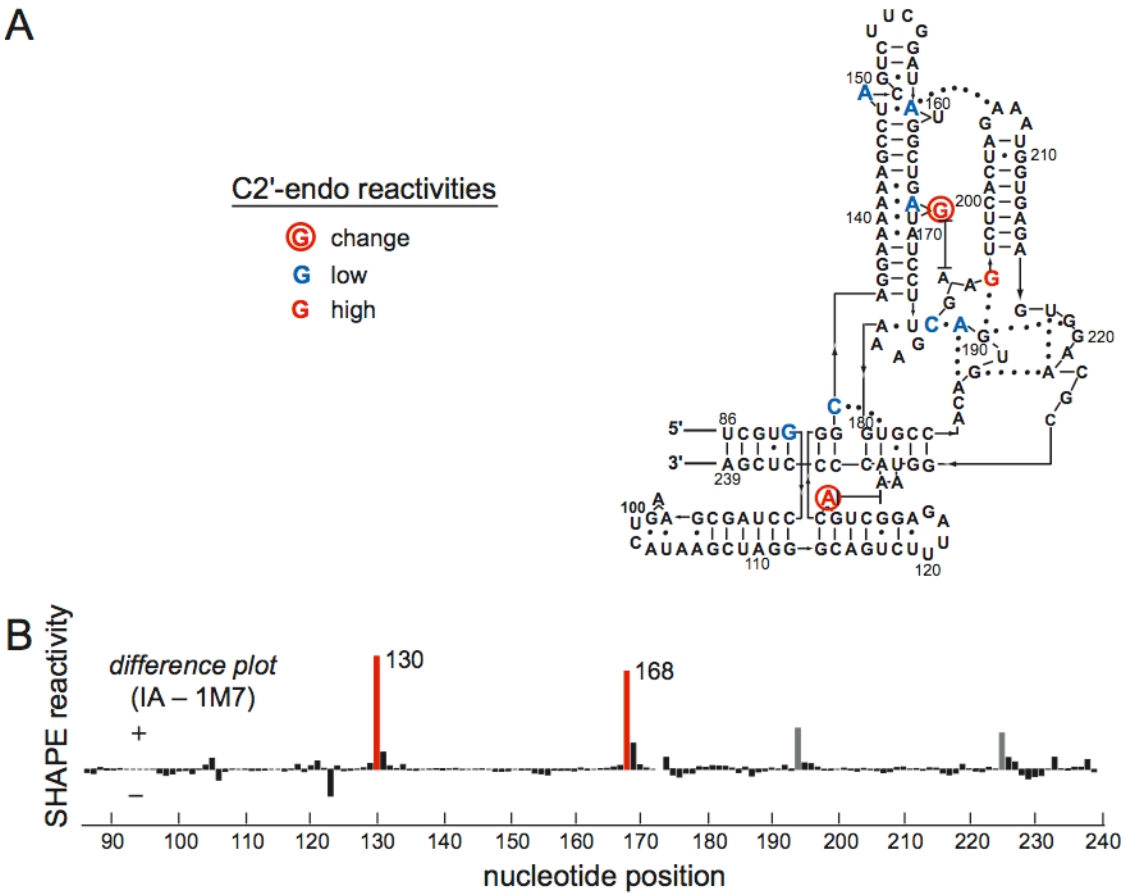


Figure 3.6. SHAPE reactivities at C2´-endo nucleotides in the specificity domain of RNase P. (A) Secondary structure showing SHAPE reactivities at the 10 well-defined C2´-endo nucleotides. (B) Differential SHAPE reactivities for slow (IA) versus fast (1M7) reagents.

3.2.2. Identification of slow-moving nucleotides in RNase P. We next explored whether the differential reactivity between the 1M7 and IA can be used to identify nucleotides that undergo slow conformational dynamics in an RNA with a complex structure, the specificity domain of the *B. subtilis* ribonuclease P enzyme (RNase P) [12]. After excluding nucleotides where the electron density was not well defined or that participate in crystal contacts, we identified 10 C2'-endo nucleotides in the RNase P RNA (in color, Figure 3.6).

For the vast majority of RNase P nucleotides, including all positions with C3'-endo conformations, SHAPE reactivities were identical for both the fast (1M7) and slow (IA) reagents (Figure 3.6). These nucleotides reflect limit 1. The 10 well defined C2'-endonucleotides fell into three categories: (i) most C2'-endo nucleotides are highly constrained and, as expected [3], unreactive towards both reagents (blue nucleotides, Figure 3.6A), (ii) one nucleotide is not constrained and is reactive towards both electrophiles (red nucleotide, Figure 3.6A), and (iii) two C2'-endo positions show large changes in reactivity (nts A130 and G168, circled nucleotides and red bars, Figure 3.6). Two other nucleotides showed smaller changes in reactivity but were in regions of the structure where experimental electron density was poorly defined (gray columns, Figure 3.6B). A similar distribution of reactive and unreactive C2'-endo nucleotides occurs in the *Tetrahymena* P5-P4-P6 domain using NMIA [13].

3.3. Discussion

While the C2'-endo conformation by itself clearly does not govern SHAPE reactivity, a distinct class of C2'-endo nucleotides experiences extraordinarily slow local

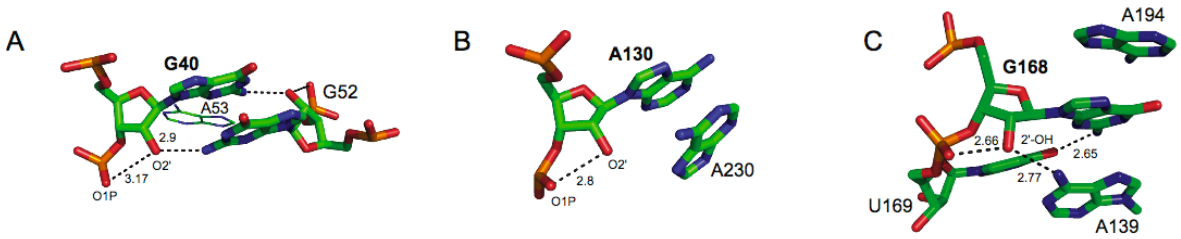


Figure 3.7. Base stacking and hydrogen bonding interactions at C2′-endo nucleotides that undergo slow conformational changes. (A) G40 from the C2′-endo RNA construct and (B) A130 and (C) G168 in the RNase P RNA. Distances are in angstroms.

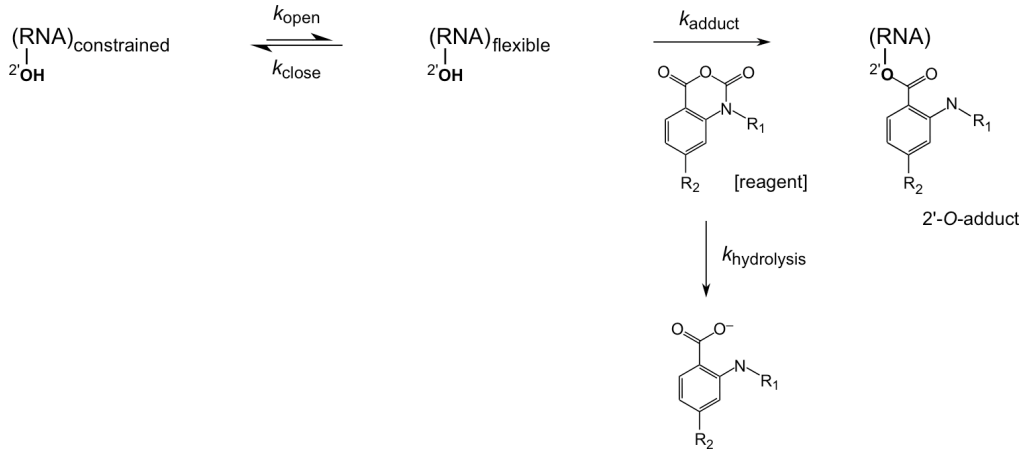
dynamics. These nucleotides in both the simple C2'-endo RNA (Figure 3.2B) and in the RNase P RNA (Figure 3.6B) share key characteristics: (1) the ribose group has the C2'-endo conformation and (2) the nucleotide conformation is partially constrained by base stacking and hydrogen bonding interactions (Figure 3.7). These C2'-endo dynamics are orders of magnitude slower than for other local RNA conformational changes like base opening reactions [14, 15] and are also slower than folding processes that involve assembly of whole domains in large RNAs [2].

In our view, the best precedent for slow conformational changes at a single residue in a biopolymer is the cis-trans isomerization of prolyl residues in proteins. Proline cis-trans conformations interconvert on the order of 10^{-2}–10^{-5} s^{-1} [16-18] and are thus comparable to the rates measured here for local dynamics at some C2'-endo nucleotides in RNA. Cis-trans isomerization can function as a molecular switch in biology [16].

We postulate that slow conformational dynamics at C2'-endo nucleotides also have the potential to function as rate-determining molecular switches and will play important, but currently unexplored, roles in RNA folding, ligand recognition, and catalysis.

3.4 Experimental

3.4.1. Derivation of Eqn. 1. The electrophile-dependent reaction of RNA to form a 2'-*O*-adduct involves the following mechanism and four relevant rate constants:



Scheme 1

The observed rate of 2'-*O*-adduct formation is given by:

$$k_{\text{obs}} = \frac{k_{\text{open}} k_{\text{adduct}} [reagent]}{k_{\text{open}} + k_{\text{close}} + k_{\text{adduct}} [reagent]}$$

then [3], $$f \approx 1 - e^{-\frac{k_{\text{obs}}}{k_{\text{hydrolysis}}}(1 - e^{-k_{\text{hydrolysis}} t})}$$

where f is the fraction 2'-*O*-adduct formed at any given nucleotide.

If the reaction is allowed to proceed until reagent hydrolysis is complete ($t \rightarrow \infty$), this equation simplifies to:

$$f \approx 1 - e^{-\frac{k_{\text{obs}}}{k_{\text{hydrolysis}}}}$$

as given in Eqn. 1.

3.4.2. IA, NMIA, 4NIA and 1M7 hydrolysis. Hydrolysis was followed by adding reagent (2.0 mM IA, 1.5 mM NMIA, 2.5 mM 4NIA, or 2.0 mM 1M7 in 300 μL DMSO) to 1.1× buffer [2.7 mL, 6.7 mM $MgCl_2$, 111 mM NaCl, 111 mM HEPES (pH 8.0)] equilibrated at 37 °C in a cuvette. Pseudo-first-order rates were obtained by monitoring the

absorbance of the hydrolysis product (at 345 nm for 2-aminobenzoate, 360 nm for 2-methylaminobenzoate, 440 nm for 2-amino-4-nitrobenzoate, and 430 nm for 2-methylamino-4-nitrobenzoate).

3.4.3. RNA constructs. The C2'-endo hairpin RNA (Figure 3.2A) and the specificity domain of the RNase P RNA were synthesized by *in vitro* transcription using a single stranded DNA (IDT) or a PCR-generated template [6], respectively. In both cases, the RNAs were embedded in the context of 5' and 3' structure cassette [5] sequences. RNAs were purified by denaturing polyacrylamide gel electrophoresis, excised from the gel, and recovered by electroelution and ethanol precipitation. Purified RNAs were resuspended in TE [10 mM Tris (pH 8.0), 1 mM EDTA] at concentrations of about 30 μM and stored at -20 °C.

3.4.4. SHAPE analysis. pAp-ethyl was 5'-end radiolabeled using γ-[^{32}P]-ATP, purified by denaturing gel electrophoresis, and resuspended in 1/2× TE. The pAp-ethyl (1 μL, 10000 cpm) was heated to 95 °C for 2 min, cooled on ice, mixed with 3 μL of 3.3× folding buffer [264 mM NaCl, 66 mM Hepes (pH 8.0), 16.5 mM $MgCl_2$], and incubated at 37 °C for 20 min. The pAp-ethyl solution was treated with reagent (1 μL; 100 mM; in anhydrous DMSO), allowed to react for 36 min (equal to five IA hydrolysis half-lives[3]). The no-reagent control contained 1 μL neat DMSO. The C2'-endo RNA (4 pmol) SHAPE experiments were performed as described for pAp-ethyl with the addition of a primer extension step. Modified RNA was recovered by ethanol precipitation [90 μL sterile H_2O, 5 μL NaCl (4 M), 1 μL glycogen (20 mg/mL), 400 μL ethanol; 30 min at -80 °C] and resuspended in 10 μL of TE. Analysis of the RNase P RNA was performed similarly except that the 3.3× folding buffer contained 333 mM Hepes (pH 8.0), 333 mM NaCl, and

33.3 mM $MgCl_2$.

3.4.5. Primer Extension. All the C2'-endo containing hairpin RNA experiments were performed by Costing Gherghe. The primer extension reaction was performed using a 5'-[^{32}P]-label primer as described [5], with the exception that the extension reaction was incubated at 45 °C for 1 min, 52 °C for 30 min, and 65 °C for 5 min. For the RNase P RNA, primer extension experiments were performed exactly as described [6], with the exception that the DNA primers (5'-GAA CCG GAC CGA AGC CCG-3') were labeled with either VIC or NED for the (–) and (+) reagent experiments; dideoxy sequencing markers were generated using unmodified RNA and primers labeled with 6-FAM or PET. cDNA extension products were separated by capillary electrophoresis using an Applied Biosystems 3130 Genetic Analyzer capillary electrophoresis instrument.

3.4.6. Data Analysis. For pAp-ethyl experiments, band intensities were quantified by phosphorimaging (Molecular Dynamics Storm instrument). For the C2'-endo RNA, individual band intensities for the (+) and (–) reagent reactions were integrated using SAFA [19], as described. Fraction adduct formation at each position was calculated as the band intensity divided by the full-length band. Reactivities were normalized to the intensity at position 80, a flexible nucleotide outside the RNA of interest. k_{open} was determined by fitting the fraction 2'-*O*-adduct using Eqn. 5 (given in the main text) with the addition of a plateau term:

$$f = 1 - e^{-k_{\text{open}}/k_{\text{hydrolysis}}} + A$$

where A accounts for the small amount of misfolded RNA in which the C2'-endo nucleotide exists in a uniformly reactive conformation. The addition of the A term

improves the quality of the fit, but has no material effect on the estimate for k_{open}, within error. Least squares fitting was performed with Kaleidagraph (ver. 4.01). For the RNase P RNA, raw traces from the ABI 3130 were processed using ShapeFinder [20] as described [6]. Overall reactivities for the IA and 1M7 experiments were normalized to intensities at positions 103 and 122; negative intensities were set to zero. Structure figures were composed with Pymol [21].

3.4.7. Refinement of the RNase P Structure. The 3.15 Å resolution structure of RNase P was refined from the deposited model and structure factors (PDB accession 1NBS) [12] by Joseph M. Krahn (NIEHS, Research Triangle Park, NC). The new model includes several regions characterized by ambiguous experimental density in order to reduce overall map noise. Electron density maps from the original model contained large difference peaks in the |Fo|–|Fc| maps. Many of these sites had been modeled as Mg^{2+}, which do not adequately account for the density. Forty-seven sites were modeled as Sr^{2+} ions, which were a component of the crystallization solution. In addition, bulk solvent parameters were adjusted empirically, which helps to account for the relatively high level of strongly diffracting solution ions. Refinement with CNS 1.1 [22] resulted in a significant improvement in both the crystallographic R-factors and map interpretability. Some of the ion sites may in fact reflect partial occupancy of Pb^{2+}, or a mixture of ions, but strontium is a good fit to the density in most cases. The final model includes 11 chloride ions, 16 magnesium ions, and 15 waters based on charge and density considerations. At this point, the model could be extended to include 16 of the 40 missing nucleotides in poorly ordered regions. Inclusion of these nucleotides reduces overall noise in the electron density map. This model was further improved by

including non-crystallographic symmetry and geometric restraints to idealize ribose geometries in regions of the structure that formed regular, well-defined, A-form RNA helices [23]. Refinement restraints for other regions were initially modified to allow unrestrained pucker conformations, and later restrained to either C3'-endo or C2'-endo based on electron density, geometric considerations, and analysis of the distance between the plane of the nucleobase and the phosphate using MolProbity. Improvement to the free R-factor was used to guide the steps outlined above, using the original test set from the deposited structure factors in all R-free calculations. The working and free R-factors of 28.0 and 30.7% from the original model were improved to 20.3 and 24.1%, respectively. Nucleotides categorized as C2'-endo (in well-defined regions of the experimental electron density map and excluding those involved in crystal contacts) are 90, 130, 134, 148, 160, 167, 168, 191, 192, 196 (see Figure 3.6A).

3.5 References

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

2. Buchmueller, K.L., et al., *RNA-tethered phenyl azide photocrosslinking via a short-lived indiscriminant electrophile.* J. Am. Chem. Soc., 2003. **125**(36): p. 10850-10861.

3. Merino, E.J., et al., *RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE).* J. Am. Chem. Soc., 2005. **127**: p. 4223-4231.

4. 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**: p. 4659-4667.

5. 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. 1610-1616.

6. Mortimer, S.A. and Weeks, K.M., *A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry.* Journal of the American Chemical Society, 2007. **129**(14): p. 4144-4145.

7. Fresht, A., *Enzyme structure and mechanism*. 1985, W.H. Freeman & Company. p. 117-118.

8. Richardson, J.S., et al., *RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution).* RNA, 2008. **14**(3): p. 465-81.

9. SantaLucia, J., Jr. and Turner, D.H., *Structure of (rGGCGAGCC)2 in solution from NMR and restrained molecular dynamics.* Biochemistry, 1993. **32**(47): p. 12612-23.

10. Heus, H.A., et al., *The detailed structure of tandem G.A mismatched base-pair motifs in RNA duplexes is context dependent.* J. Mol. Biol., 1997. **271**(1): p. 147-58.

11. Tolbert, B.S., et al., *NMR structures of (rGCUGAGGCU)2 and (rGCGGAUGCU)2: probing the structural features that shape the thermodynamic stability of GA pairs.* Biochemistry, 2007. **46**(6): p. 1511-22.

12. Krasilnikov, A.S., et al., *Crystal structure of the specificity domain of ribonuclease P.* Nature, 2003. **421**(760-764).

13. Vicens, Q., et al., *Local RNA structural changes induced by crystallization are revealed by SHAPE.* RNA, 2007. **13**(4): p. 536-48.

14. Leroy, J.L., et al., *Internal motions of transfer RNA: a study of exchanging protons by magnetic resonance.* J. Biomol. Struct. Dyn., 1985. **2**(5): p. 915-39.

15. Spies, M.A. and Schowen, R.L., *The trapping of a spontaneously "flipped-out" base from double helical nucleic acids by host-guest complexation with beta-cyclodextrin: the intrinsic base-flipping rate constant for DNA and RNA.* J. Am. Chem. Soc., 2002. **124**(47): p. 14049-53.

16. Lu, K.P., et al., *Prolyl cis-trans isomerization as a molecular timer.* Nat. Chem. Biol., 2007. **3**(10): p. 619-29.

17. Grathwohl, C. and Wuthrich, K., *Nmr studies of the molecular conformations in the linear oligopeptides H-(L-Ala)n-L-Pro-OH.* Biopolymers, 1976. **15**(10): p. 2043-57.

18. Schoetz, G., Trapp, O., and Schurig, V., *Determination of the cis-trans isomerization barrier of several L-peptidyl-L-proline dipeptides by dynamic capillary electrophoresis and computer simulation.* Electrophoresis, 2001. **22**(12): p. 2409-15.

19. Das, R., et al., *SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments.* RNA, 2005. **11**(3): p. 344-54.

20. Vasa, S.M., et al., *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.

21. DeLano, W.L. *The Pymol Molecular Graphics System.* 2005; Available from: http://www.pymol.org.

22. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr., Sect D: Biol. Crystallogr., 1998. **54**: p. 905-921.

23. Murray, L.W., et al., *RNA backbone is rotameric.* Proc. Natl. Acad. Sci. U. S. A., 2003. **100**: p. 13904-13909.

CHAPTER 4

Time-Resolved RNA SHAPE Chemistry

4.1 Introduction

SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) chemistry is emerging as a powerful approach for quantitative analysis of the equilibrium structures of diverse biological RNAs [1-8]. SHAPE chemistry exploits the discovery that the nucleophilic reactivity of the ribose 2'-hydroxyl position is strongly gated by the underlying nucleotide flexibility (Figure 4.1). Flexible nucleotides preferentially adopt conformations that react with a hydroxyl-selective electrophile to form a 2'-*O*-adduct, while base paired or otherwise conformationally constrained nucleotides are unreactive. The electrophile is concurrently consumed by hydrolysis making a specific quench step unnecessary, provided the adduct-forming reaction with RNA is allowed to continue until all reagent is consumed [1-3]. SHAPE works extremely well for interrogating the equilibrium structure for RNA motifs from a few nucleotides to large catalytic RNAs and viral genomes spanning hundreds to thousands of nucleotides [1-10].

To date, SHAPE chemistry has been based on electrophilic reagents derived from an isatoic anhydride (IA) scaffold. Reactivity can be tuned by varying the electron withdrawing potential of the functional groups on the IA scaffold. We have created a suite of reagents with hydrolysis half lives from 14 to 430 s [1-4, 7]. In our view, the most reactive of these reagents, 1-methyl-7-nitroisatoic anhydride (1M7) [4], represents a near-ideal probe for nucleotide-resolution analysis of RNA structures at equilibrium. The intrinsic reactivity of 1M7 is insensitive to solution conditions like Mg^{2+} concentration and temperature. Complete degradation of the reagent requires 70 s: this moderate level of reactivity makes the reagent easy to handle in the laboratory

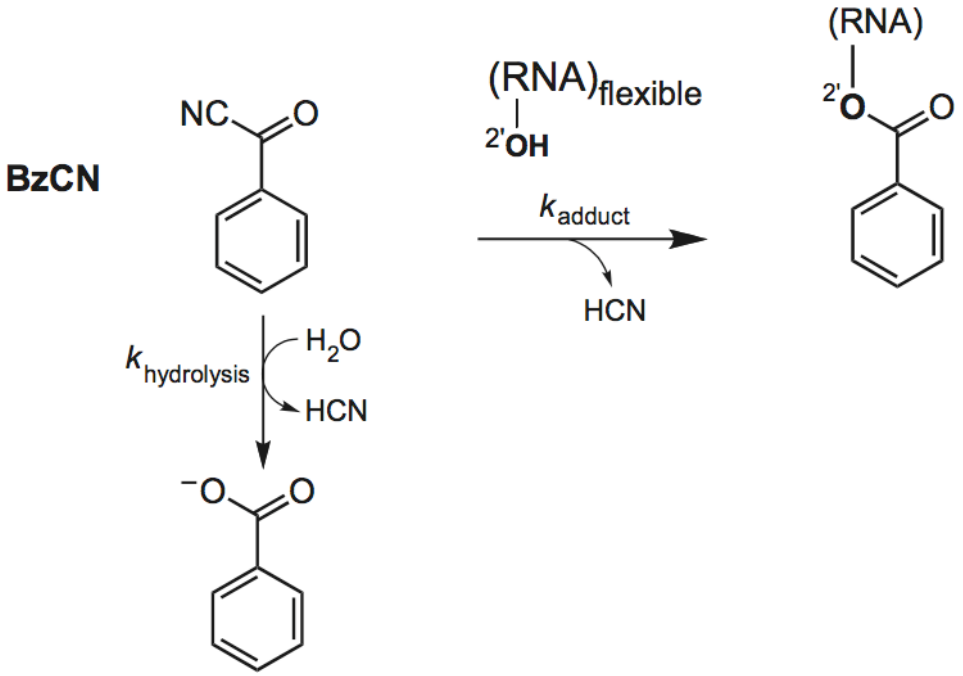


Figure 4.1. Mechanism of RNA SHAPE chemistry with BzCN. BzCN reacts with 2′-hydroxyl groups at flexible positions in RNA to form a 2′-*O*-adduct and also undergoes inactivation by hydrolysis.

but means that RNA structures are averaged on the minute timescale [4].

A full understanding of structure-function relationships in RNA biology requires a nucleotide-resolution view of the time-resolved mechanisms by which RNA molecules fold, interconvert between distinct states, and function in ribonucleoprotein complexes [11-14]. The reactivity of IA derivatives is circumscribed by the electron withdrawing potential of the substituents that can be introduced into this reagent. The limits of this useful scaffold have likely been reached.

We therefore sought to develop a new chemical framework to interrogate RNA structure on the second timescale. Benzoyl cyanide (BzCN) reacts with hydroxyl functional groups to yield a stable ester. Formation of the stable cyanide ion leaving group renders the reaction irreversible (Figure 4.1). As with all SHAPE electrophiles, this scaffold should satisfy two important criteria: (1) react with RNA 2′-hydroxyl groups in a structure-selective manner and (2) undergo rapid inactivation by hydrolysis with water (Figure 4.1).

4.2 Results

4.2.1 BzCN reacts with 2′-hydroxyl groups in a structure-selective manner.

We used BzCN to assess the structure of the specificity domain of *Bacillus subtilis* ribonuclease P (RNase P), whose structure has been well characterized using IA-based reagents [4, 7]. The RNase P RNA was treated with BzCN under conditions that stabilize the native tertiary fold (10 mM $MgCl_2$, 100 mM NaCl, pH 8.0) and also in the absence of Mg^{2+} where the RNA forms only its secondary structure. Sites of 2′-*O*-adduct formation were identified as stops to primer extension using fluorescently labeled DNA primers,

resolved by capillary electrophoresis [4, 9]. Absolute SHAPE reactivities observed for the RNase P RNA using BzCN are very similar to those obtained with 1M7. Both reagents react preferentially with loops and, in both cases, loops that participate in tertiary interactions are protected from forming 2'-*O*-adducts in the presence of Mg^{2+} (Figure 4.2). There is a very good linear correlation (r = 0.81) between 1M7 and BzCN reactivities, both in the presence and absence of Mg^{2+} (Figure 4.3) Thus, like 1M7 [4], BzCN reactivity measure local nucleotide flexibility and is selective for the 2'-OH group. BzCN thus meets the first of the two criteria for time-resolved SHAPE because it accurately reports the secondary and tertiary structure of the RNase P specificity domain, both in the presence and absence of Mg^{2+}.

4.2.2. BzCN undergoes rapid inactivation by hydrolysis with water. We next determined the rate of BzCN hydrolysis by monitoring the competition between hydrolysis versus 2'-*O*-adduct formation using the model nucleotide, 3′-phosphoethyl-5′-adenosine monophosphate (pAp-ethyl) (Figure 4.4A). Hydrolysis is rapid and was quenched at defined time points with DTT (see filled circle, Figure 4.4B). The rate constant for BzCN hydrolysis was determined by fitting the fraction adduct formed as a function of reaction time (Figure 4.4B). BzCN undergoes hydrolysis with a half-life of 0.25 s at 37 °C (Figure 4.4A).

The reaction between BzCN and RNA is complete in ~1 and 2.5 s at 37 and 25 °C (Figure 4.4). BzCN is thus an ideal reagent for following RNA folding on the second timescale.

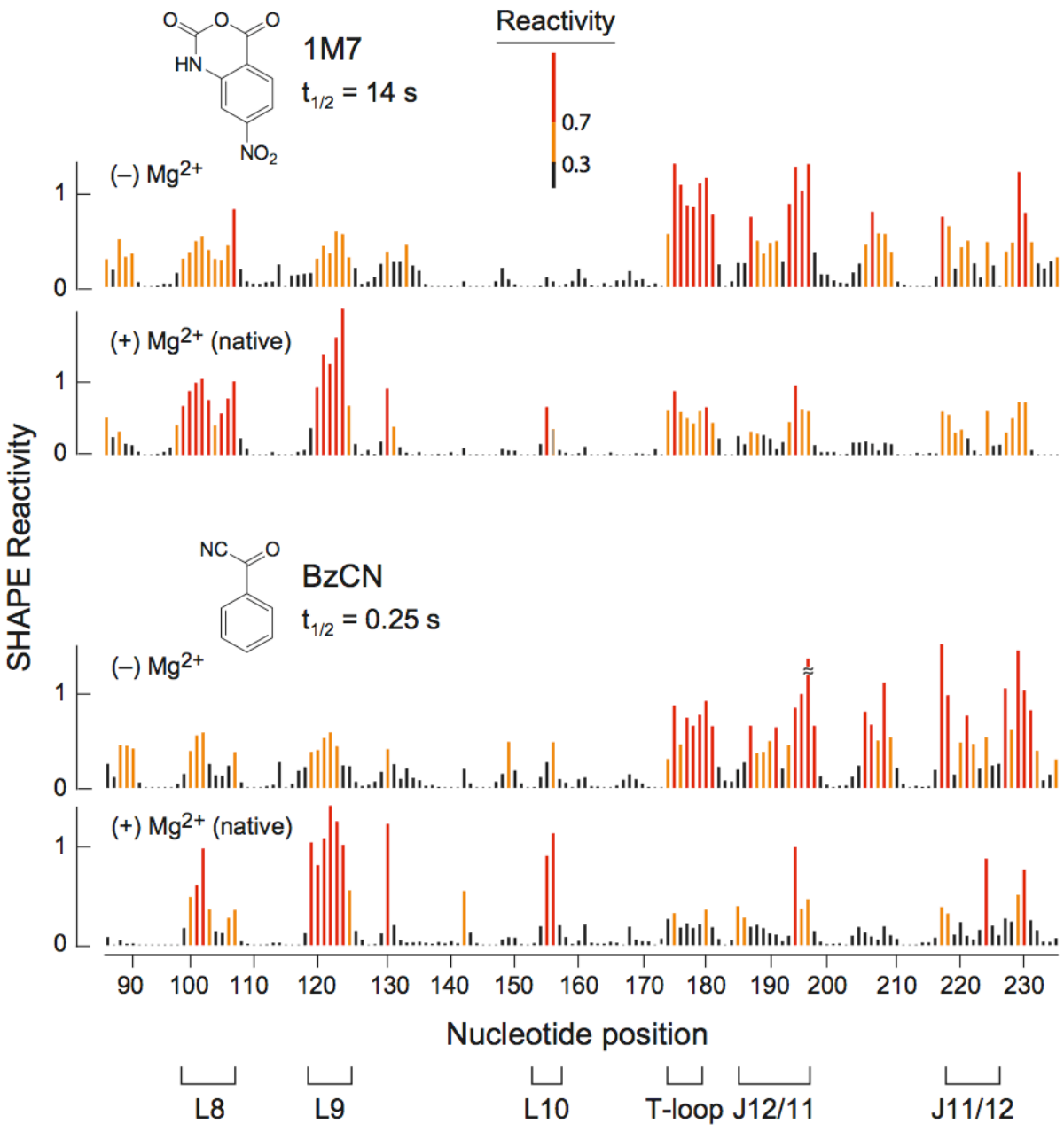


Figure 4.2. Histograms comparing absolute SHAPE reactivities for the RNase P specificity domain in the presence and absence of Mg^{2+} obtained using 1M7 (top) and BzCN (bottom). Loop (L*x*) and joining structure (J*x*) landmarks are emphasized.

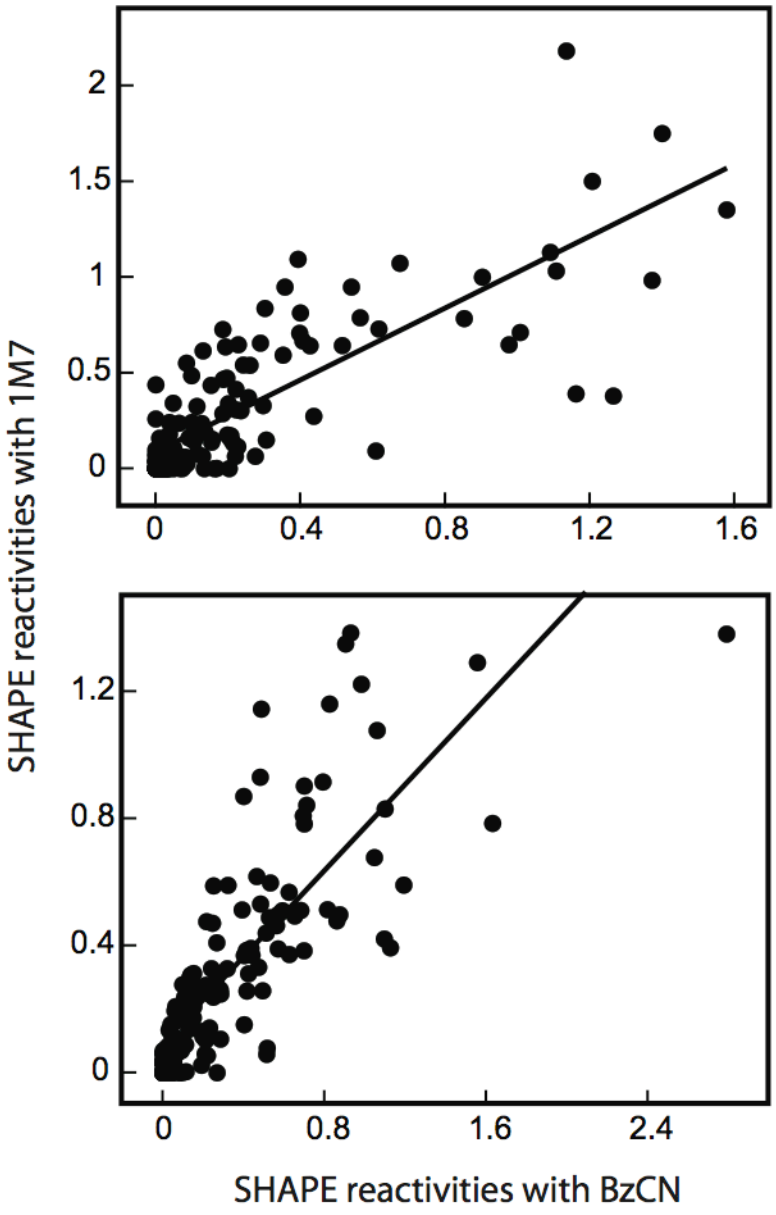


Figure 4.3. Correlation between SHAPE reactivities obtained with 1M7 and BzCN in the presence (top) and absence (bottom) of Mg^{2+}. Pearson's linear *r*-values are 0.81 for both sets of data.

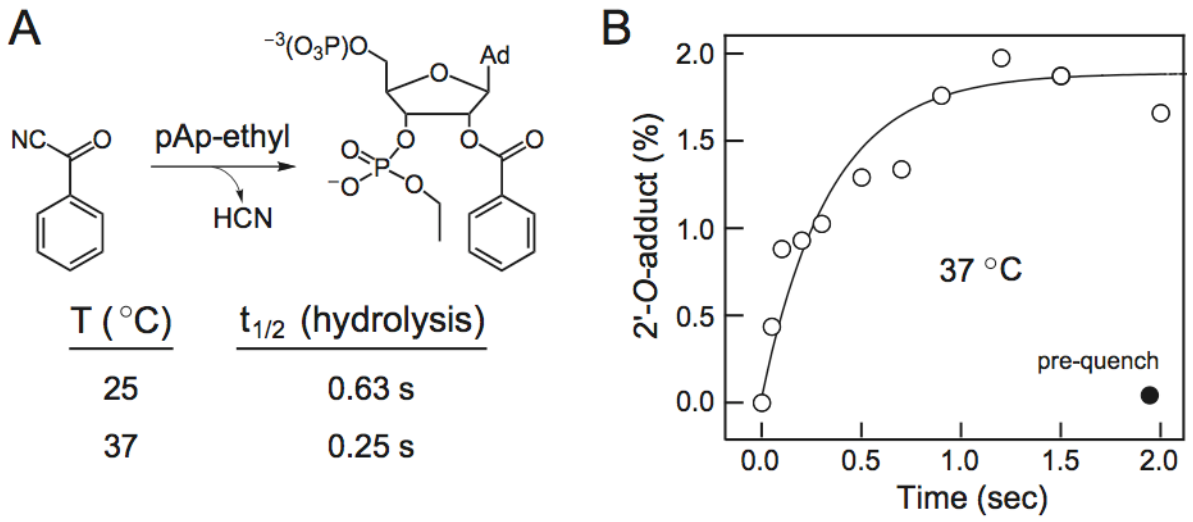


Figure 4.4. Reaction half-life of BzCN. (A) Reaction of BzCN with pAp-ethyl. (B) Time-dependent reaction of BzCN with pAp-ethyl at 37 °C. Half-life of BzCN in water was determined by fitting the fraction 2′-*O*-adduct formed to 1 – *exp*[(0.02)(*exp*(-$k_{\text{hydrolysis}}t$) – 1)].

4.2.3. Time-resolved SHAPE with BzCN yields a nucleotide-resolution view of RNase P tertiary folding. We then used BzCN to analyze the folding behavior of the RNase P RNA in a time-resolved SHAPE experiment. The RNase P RNA (154 nt) RNA forms several classes of tertiary structure modules: a U-shaped T-loop motif that docks into the P7-P10 helices through an A-minor interaction, numerous stacking interactions (including A130 on A230 in P9), two large internal loops (J11/12 and J12/11) which form extensive noncanonical interactions, and a tetraloop-receptor interaction (involving L12 and P10.1) [15].

Tertiary folding of the RNase P RNA was initiated by adding $MgCl_2$ (to 10 mM) to a solution of RNA (0.5 μM) preincubated at 37 °C in 100 mM NaCl (at pH 8.0). Aliquots were removed at time intervals between 5 and 1200 s and added directly to BzCN. This experiment provides 1 s snapshots of RNase P structure over 20 minutes (Figure 4.5A). Within 5 s after addition of Mg^{2+}, there is a rapid ordering of the RNA secondary structure, observed as a rapid decrease in SHAPE reactivity throughout the RNA. We focused our analysis on the tertiary folding that occurs after this step. After data normalization, significant changes in SHAPE reactivity occur at 16 nucleotides in the RNA. These nucleotides fell into one of two kinetically distinct categories (Figure 4.5B). The first category is a fast phase that folds with a rate constant of 0.06 s^{-1}; the second category is a slower phase that folds with a rate constant of 0.004 s^{-1}. The nucleotides that fall into the first, fast, phase also exhibit conformational changes consistent with the slow phase, suggesting that the slow phase involves a whole-scale tightening of the RNA structure (Table 4.1.).

The time-resolved SHAPE data make possible modeling of kinetic intermediates

in the folding pathway for the RNase P RNA. Consistent with the two categories of reactivity changes, we observe two distinct folding steps. The first tertiary folding step is characterized by significant protections in the structural module involving the J11/12 and J12/11 loops (residues 185-196 and 217-225) and also in the interaction of A229 with C134 and C232 (in red, Figure 4.6). The second, slower, folding step involves the formation of the core of the specificity domain: docking of the T-loop (U175-A179) at P7-P10, stacking of A230 on A130 in P9, and the docking of the GAAA tetraloop (P12) into its receptor in P10.1 (in blue, Figure 4.6).

There are several critical features of this folding mechanism. Although it is becoming clear that RNA folding is not obligatorily hierarchical [2, 6, 16], in the case of the RNase P RNA, tertiary structure folding does clearly follow the near-complete formation of the secondary structure. Second, tertiary interactions involving local interactions between the J11/12 and J12/11 loops and involving a base stacking interaction at A229 form first (in red, Figure 4.6). Third, the rate-limiting folding step, despite being on the order of minutes, does not involve the breaking of non-native contacts. Instead, the rate-determining step involves formation of tertiary interactions located 55 Å apart (in blue, Figure 4.6).

A three-state folding mechanism for the RNase P specificity domain was also developed using an equilibrium and kinetic approaches [17-19]. Time-resolved SHAPE yielded a highly detailed, nucleotide resolution, view of folding in this RNA in a single, very concise, set of experiments (Figure 4.5A).

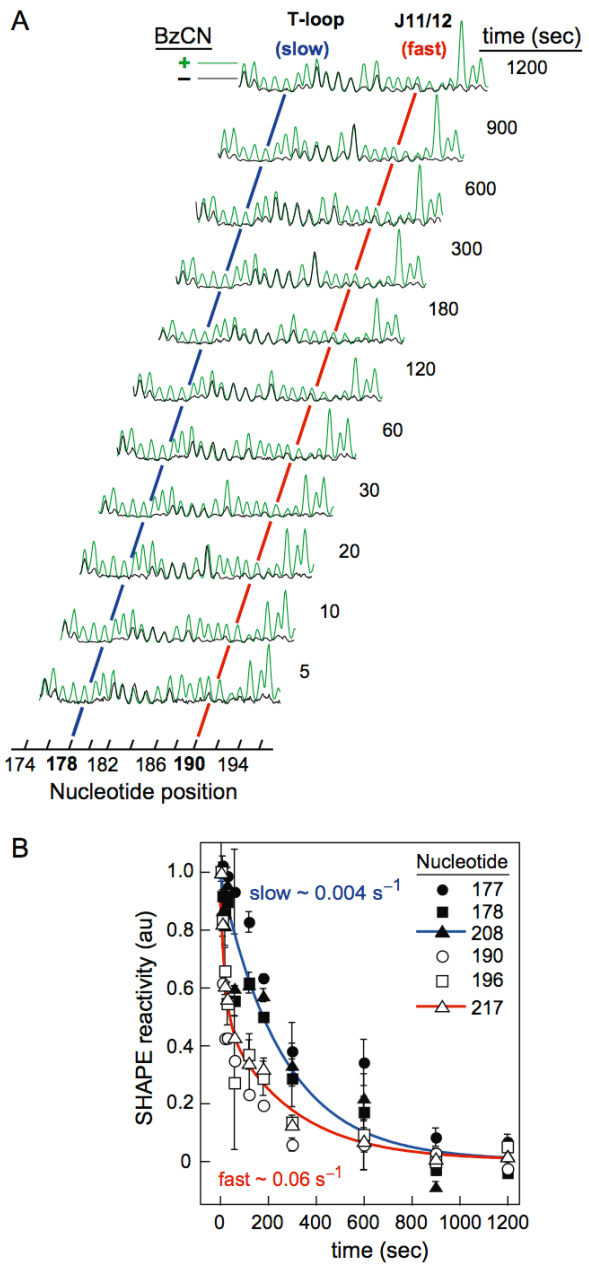


Figure 4.5. Formation of tertiary interactions in the RNase P specificity domain. (A) Processed capillary electrophoresis traces obtained for the reaction of BzCN with the RNase P specificity domain from 5 s to 20 min. The difference in peak intensity between the (+) and (–) BzCN channels reports the extent of 2′-*O*-adduct formation at each point. Nucleotides involved in tertiary interactions show a decrease in reactivity over time. Nucleotide 178 undergoes this transition slowly, while nucleotide 190 undergoes this transition more rapidly (blue and red lines, respectively). (B) Folding rate constants monitored by time-resolved SHAPE. All nucleotides show a decrease in SHAPE reactivity as a function of time, indicating an increase in structural constraints as the RNA forms tertiary interactions. Nucleotides folding in the fast and slow phases are indicated by open and closed symbols, respectively.

Nucleotide	fast k_1 (s^{-1})	slow k_2 (s^{-1})	Folding domain
A130	—	0.006 ± 0.002	
U175	—	0.006 ± 0.001	T-loop
G176	—	0.004 ± 0.002	
A177	—	0.002 ± 0.0005	
A178	—	0.004 ± 0.0008	
A179	—	0.004 ± 0.001	
G180	—	0.0030 ± 0.001	
U189	0.06 ± 0.03	0.002 ± 0.003	J11/12
G190	0.07 ± 0.03	0.004 ± 0.004	
A195	0.05 ± 0.04	0.002 ± 0.006	
G196	0.04 ± 0.01	0.003 ± 0.003	
A206	—	0.002 ± 0.001	GAAA Tetraloop-Receptor
A208	—	0.004 ± 0.0009	
G217	0.05 ± 0.02	0.003 ± 0.002	J12/11
A221	0.08 ± 0.07	0.005 ± 0.005	
A229	0.02 ± 0.04	0.001 ± 0.001	

Table 4.1. Folding rate constants for nucleotides in the RNase P specificity domain involved in forming tertiary interactions.

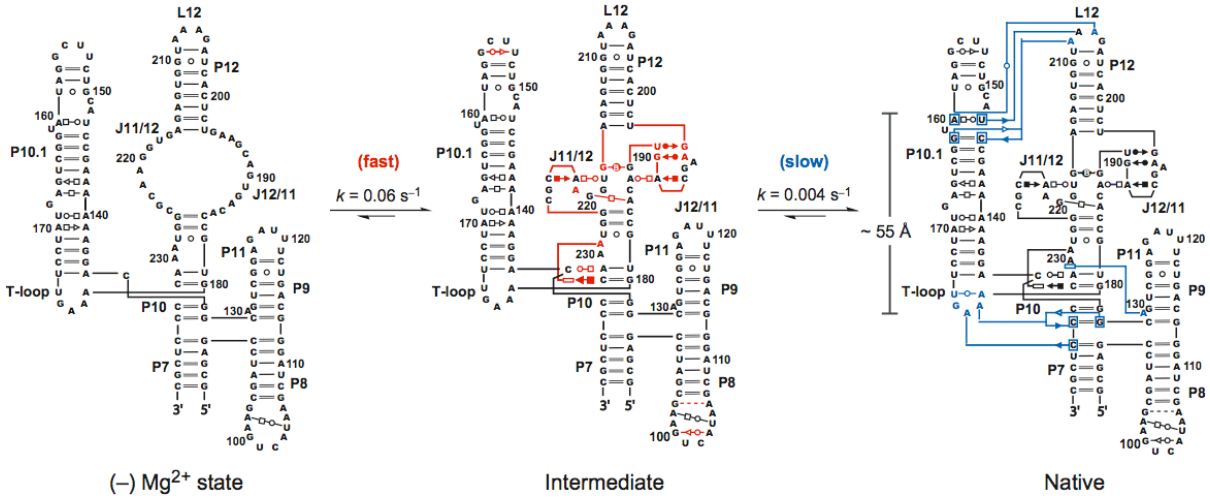


Figure 4.6. Mechanism for folding of the RNase P specificity domain. Fast-forming interactions found in the intermediate are red, interactions that form in the rate-determining step are in blue. Structure is drawn to approximate the arrangement of helices in three-dimensional space [20].

4.3 Discussion

Time-resolved SHAPE with BzCN makes it possible to monitor the pathways for formation of individual sets of tertiary interactions in complex RNAs on timescales as short as a few seconds. Because BzCN is self-quenching, yields nucleotide resolution information at every position simultaneously, and affords 1 s snapshots of RNA structure, time-resolved SHAPE should find wide application for developing detailed models for time-dependent processes involving RNA.

4.4. Experimental

4.4.1. Benzoyl cyanide 2′-*O*-adduct formation and hydrolysis. Adduct formation between [^{32}P]-labeled 3′-phosphoethyl-5′-adenosine (pAp-ethyl) [4] and benzoyl cyanide (BzCN, Sigma-Aldrich) was followed by adding 10% (vol/vol) BzCN (60 mM final, in DMSO; 35.2 μL) to 1.1× reaction buffer [235.4 μL, 11.1 mM $MgCl_2$, 111 mM NaCl, 111 mM Hepes (pH 8.0)] using a chemical quench-flow instrument [21] (KinTek, Model RQF-3; drive syringes were 1.0 and 10 mL for the BzCN reagent and buffer reaction loops, respectively) and quenched with 1 vol 250 mM dithiothreitol [at the following reaction times (in sec): 0.02, 0.07, 0.15, 0.25, 0.40, 0.60, 0.80, 1.0, 1.3, 1.7, 2.5]. Each quenched reaction was precipitated by addition of 10 vol isobutanol, resolved by gel electrophoresis (30% polyacrylamide; 29:1 acrylamide:bisacrylamide; 0.4 mm × 28.5 cm × 23 cm; 30 W, 45 min), and quantified by phosphorimaging. We observed a single adduct band under all conditions, consistent with a mechanism in which BzCN reacts exclusively with the 2′-OH groups and not with other positions in the ribose or base moieties.

Rate constants for BzCN hydrolysis were obtained using an equation that accounts for parallel reaction of BzCN by 2′-*O*-adduct formation (k_{adduct}) and hydrolysis ($k_{\text{hydrolysis}}$) [1]:

$$\text{Fraction product} = 1 - exp[(k_{\text{adduct}}/k_{\text{hydrolysis}})(e^{-k_{\text{hydrolysis}}t} - 1)]$$

(see Figure 3 in main text). The term $k_{\text{adduct}}/k_{\text{hydrolysis}}$ was obtained from the fraction adduct formed at long time points, where (as $t \rightarrow \infty$) $k_{\text{adduct}}/k_{\text{hydrolysis}}$ = –ln(1 – fraction product) = 0.02 for all experiments.

4.4.2. Structure-Selective RNA Modification. The RNase P specificity domain RNA [15] was synthesized as described [4]. RNA (60 pmol) in 60 μL sterile water was heated at 95 °C for 2 min, cooled on ice, treated with 36 μL of 3× no-Mg^{2+} folding buffer [333 mM NaCl, 333 mM Hepes (pH 8.0)], and incubated at 37 °C for 5 min. Tertiary structure folding was initiated by adding 12 μL of 10× $MgCl_2$ (100 mM). After mixing, 9 μL of this solution was removed (at 5, 10, 15, 20, 30, 60, 120, 180, 300, 600, 900, and 1200 s) and added directly to 1 μL 10× BzCN (600 mM in DMSO). No-reagent control reactions were added to 1 μL neat DMSO. Modified RNA was recovered by ethanol precipitation [90 μL sterile water, 5 μL NaCl (5 M), 1 μL glycogen (20 mg/mL), 400 μL ethanol; 30 min at –80 °C] and resuspended in 10 μL of TE [10 mM Tris (pH 8.0), 1 mM EDTA].

4.4.3. Primer Extension. The general procedure was that outlined previously [3, 4, 9]. Briefly, a fluorescently labeled DNA primer (5′ VIC- or NED-labeled GAA CCG GAC CGA AGC CCG; 3 μL, 0.3 μM) was annealed to the RNA (10 μL, from the previous step) by heating at 65 °C (6 min) and 35 °C (5 min). Reverse transcription buffer and

Superscript III were added and the reactions incubated at 52 °C for 30 min. Primer extension reactions were quenched by adding an equal volume of a mixture of sodium acetate (1.5 M, pH 5.2) and EDTA (40 mM, pH 8.0) and the resulting cDNAs were recovered by ethanol precipitation, washed twice with 70% ethanol, dried by vacuum for 10 min, and resuspended in 10 μL de-ionized formamide. Dideoxy sequencing markers were generated using unmodified RNA and primers labeled with unique fluorophores (6-FAM or PET, 0.6 μM), and by adding 1 μL of 2′,3′-dideoxycytosine (10 mM) or 2′,3′-dideoxyguanosine (0.25 mM) triphosphate after addition of reverse transcription buffer. cDNA extension products were separated by capillary electrophoresis using an Applied Biosystems 3130 DNA sequencing instrument.

4.4.4. Data Analysis. Raw traces from the ABI 3130 were processed using ShapeFinder [22]. Data sets were normalized by excluding the 2% most reactive nucleotides (3 total) and dividing by the average intensity of the next 8% most reactive nucleotides (12 total). After data normalization, time-dependent changes in reactivity were judged to be significant if they were 0.2 SHAPE units or greater. Sixteen nucleotides exhibited significant reactivity changes by this criterion; the average change in SHAPE reactivity was 0.4. Folding rates for individual nucleotides were obtained by normalizing intensities (I) to the first time point and fitting to either a single

$$I = A + (1 - A)e^{-k_2 t}$$

or double

$$I = A + (1 - A - B)e^{-k_1 t} + Be^{-k_2 t}$$

exponential. Nucleotides 130, 175, 176, 177, 178, 179, 180, 206 and 208 were fit as a single exponential and nucleotides 189, 190, 195, 196, 217, 220, and 229 were fit to the double exponential (Table 4.1). Secondary structure for the RNase P specificity domain was adapted from Lescoute and Westhof [20].

4.5 References

1. Merino, E.J., et al., *RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE).* J. Am. Chem. Soc., 2005. **127**: p. 4223-4231.

2. 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**: p. 4659-4667.

3. 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**: p. 1610-1616.

4. 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.

5. Badorrek, C.S., Gherghe, C.M., and Weeks, K.M., *Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization.* Proc. Natl. Acad. Sci. U. S. A., 2006. **103**(37): p. 13640-5.

6. 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**: p. 3454-3461.

7. Gherghe, C.M., et al., *Slow Conformational Dynamics at C2'-endo Nucleotides in RNA.* J. Am. Chem. Soc., 2008. **130**: p. 8884-8885.

8. Gherghe, C.M., et al., *Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA.* J. Am. Chem. Soc., 2008. **130**(37): p. 12244-5.

9. Wilkinson, K.A., et al., *High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states.* PLoS Biol., 2008. **6**(4): p. e96.

10. Duncan, C.D.S. 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**: p. 8504-8513.

11. Hennelly, S.P., et al., *A Time-resolved Investigation of Ribosomal Subunit Association.* J. Mol. Biol., 2005. **346**(5): p. 1243-1258.

12. Tijerina, P., Mohr, S., and Russell, R., *DMS footprinting of structured RNAs and RNA-protein complexes.* Nat. Protoc., 2007. **2**(10): p. 2608-2623.

13. Sclavi, B., et al., *RNA Folding at Millisecond Intervals by Synchrotron Hydroxyl Radical Footprinting.* Science, 1998. **279**: p. 1940-1943.

14. Shcherbakova, I. and Brenowitz, M., *Monitoring structural changes in nucleic acids with single residue spatial and millisecond time resolution by quantitative hydroxyl radical footprinting.* Nat. Protoc., 2008. **3**: p. 288-302.

15. Krasilnikov, A.S., et al., *Crystal structure of the specificity domain of ribonuclease P.* Nature, 2003. **421**: p. 760-764.

16. Wu, M. and Tinoco, I., *RNA folding causes secondary structure rearrangement.* Proc. Natl. Acad. Sci. U.S.A., 1998. **95**: p. 11555-11560.

17. Qin, H., Sosnick, T.R., and Pan, T., *Modular Construction of a Tertiary RNA Structure: The Specificity Domain of the Bacillus subtilis RNase P RNA.* Biochemistry, 2001. **40**: p. 11202-11210.

18. Baird, N.J., et al., *Structure of a Folding Intermediate Reveals the Interplay Between Core and Peripheral Elements in RNA folding.* J. Mol. Biol., 2005. **352**: p. 712-722.

19. Baird, N.J., et al., *Folding of a universal ribozyme:the ribonuclease P RNA.* Q. Rev. Biophys., 2007. **40**(2): p. 113-161.

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

21. Holmes, S.F., Forster, E.J., and Erie, D.A., *Kinetics of multisubunit RNA polymerases: experimental methods and data analysis.* Methods Enzymol., 2003. **371**: p. 71-81.

22. Vasa, S.M., et al., *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.

CHAPTER 5

C2′-Endo Nucleotides as Molecular Timers in RNA Folding

5.1 Introduction

In order to function properly inside the cell, RNA molecules must form specific three-dimensional structures [1-3]. These essential structures involve both local base pairing and also complex higher-order tertiary interactions. RNAs achieve their functional tertiary interactions over a wide range of time scales spanning microseconds to minutes. A persistent, and incompletely, explained observation is that many RNA folding reactions involve a rate-limiting step that occurs very slowly, on the minute timescale. Such slow folding reactions likely have important biological consequences either because slow folding governs the rate at which an RNA can perform its biological function or because it potentially requires that a cellular mechanism exists to overcome slow folding.

In some cases, slow folding results from formation of incorrectly folded and kinetically trapped states [4, 5]. Such misfolding is remedied, in part, by cellular chaperone activities [6-8]. In many other cases, the physical basis of slow folding is unknown [9, 10].

The intricate three-dimensional structures formed by RNA molecules are ultimately due to the underlying conformations of individual nucleotides. For example, most nucleotides in an RNA molecule exist in the C3′-endo conformation: however, the less frequent C2′-endo conformation is highly overrepresented in catalytic active sites and in critical tertiary structures [11-13]. The C2′-endo conformation in RNA nucleotides induces greater helical twist in an A-form helix [14, 15] and is capable of spanning a much longer 5′-to-3′ distance [11]. Recently, we discovered that C2′-endo

nucleotides can exhibit extremely slow conformational dynamics with half-lives on the 10-100 second timescale [16].

Here we show that the slow conformational dynamics that characterize a single C2′-endo nucleotide can ultimately govern the folding of a large ribozyme with numerous constituent tertiary interactions. This dramatic effect on folding can be attributed to a single nucleotide because deletion of one C2′-endo nucleotide accelerates folding of the RNase P specificity domain by an order of magnitude. We propose that slow local dynamics at C2′-endo nucleotides function to time many RNA folding reactions and envision that many rate-limiting steps in RNA biology reflect slow conformational dynamics at C2′-endo nucleotides.

5.2 Results

5.2.1 A130 has distinct local dynamics and a critical role in RNase P function. RNase P is a conserved ribonucleoprotein enzyme found in all kingdoms of life. This ribozyme catalyzes cleavage of the 5′ ends of tRNAs to achieve tRNA maturation [17, 18]. Bacterial RNase P RNAs are composed of two independently folding domains, a catalytic and a specificity domain [19]. One critical and conserved tertiary interaction in the specificity domain is the stacking interaction between A130 and A230 (in the *Bacillus subtilis* numbering) [20, 21]. A130 forms the C2′-endo conformation [16] and, together with its stacking partner at A230, are the only two nucleotides in the B-type RNase P specificity domain shown to be directly involved in recognizing and binding the tRNA substrate [22-24] (shown in red, Figure 5.1A). Both nucleotides experience a significant decrease in local nucleotide flexibility when tRNA

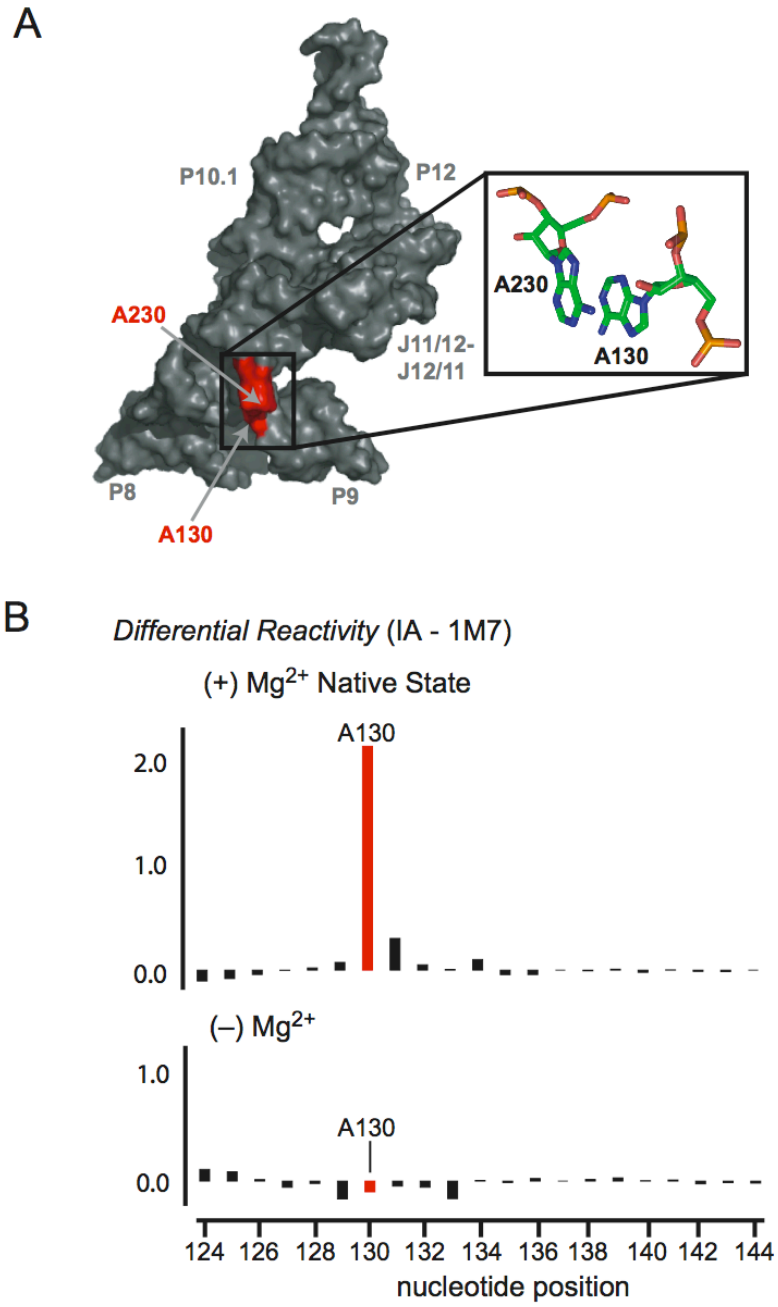


Figure 5.1. Structural context and local nucleotide dynamics of A130. (A) Three dimensional surface representation of the *Bacillus subtilis* specificity domain [20] illustrating the location, in red, and conformation, in stick models, of A130 and its stacking partner A230. (B) Differential SHAPE reactivities between slow (IA) and fast (1M7) 2′-hyhdroxyl selective reagents for the A130 region in the RNase P specificity domain in the presence (top) and absence (bottom) of Mg^{2+}. A large difference in reactivity indicates that a nucleotide experiences conformational dynamics [16].

is bound (Figure 5.2). Mutation of A130 severely compromises RNase P function [24].

In the context of the folded RNase P specificity domain, A130 shows low reactivity towards fast (1M7, $t_{1/2}$ = 14 s) SHAPE reagents but is highly reactive towards a slow (IA, $t_{1/2}$ = 430 s) reagent (red bar in top panel, Figure 5.1B). This strong differential reactivity is consistent with this C2′-endo nucleotide exhibiting extraordinarily slow conformational changes ($t_{1/2}$ = 10 – 100 s) [16]. When a similar experiment is performed in the absence of Mg^{2+}, where the RNA does not form a tight tertiary structure, fast and slow SHAPE reagents show equivalent reactivities at A130 (red bar in bottom panel, Figure 1B). Thus, the slow conformational dynamics at A130 require the native tertiary structure. We performed an identical experiment in the context of the intact RNase P ribozyme. A130 retains its strong differential reactivity towards fast and slow SHAPE reagents and therefore exhibits slow conformational dynamics in the full length RNA (Figure 5.3). In sum, A130 is critical to RNase P function [23, 24], forms the C2′-endo conformation [20], and exhibits highly unusual and very slow local nucleotide dynamics.

The RNase P specificity domain folds slowly (0.004 s^{-1}) and in two kinetically distinct steps [25]. The first phase folds with a rate constant of 0.06 s^{-1} (in red, Figure 5.4); the second, even slower, phase folds with a rate constant of 0.004 s^{-1} (in blue, Figure 5.4). The rate-limiting step involves formation of tertiary interactions that are ~ 55 Å distant in the RNA. One of the interactions that occurs in this very slow rate-determining step is the stacking of A130 in A230, which raises the possibility that the slow conformational dynamics at a single nucleotide, A130, slows folding of the entire

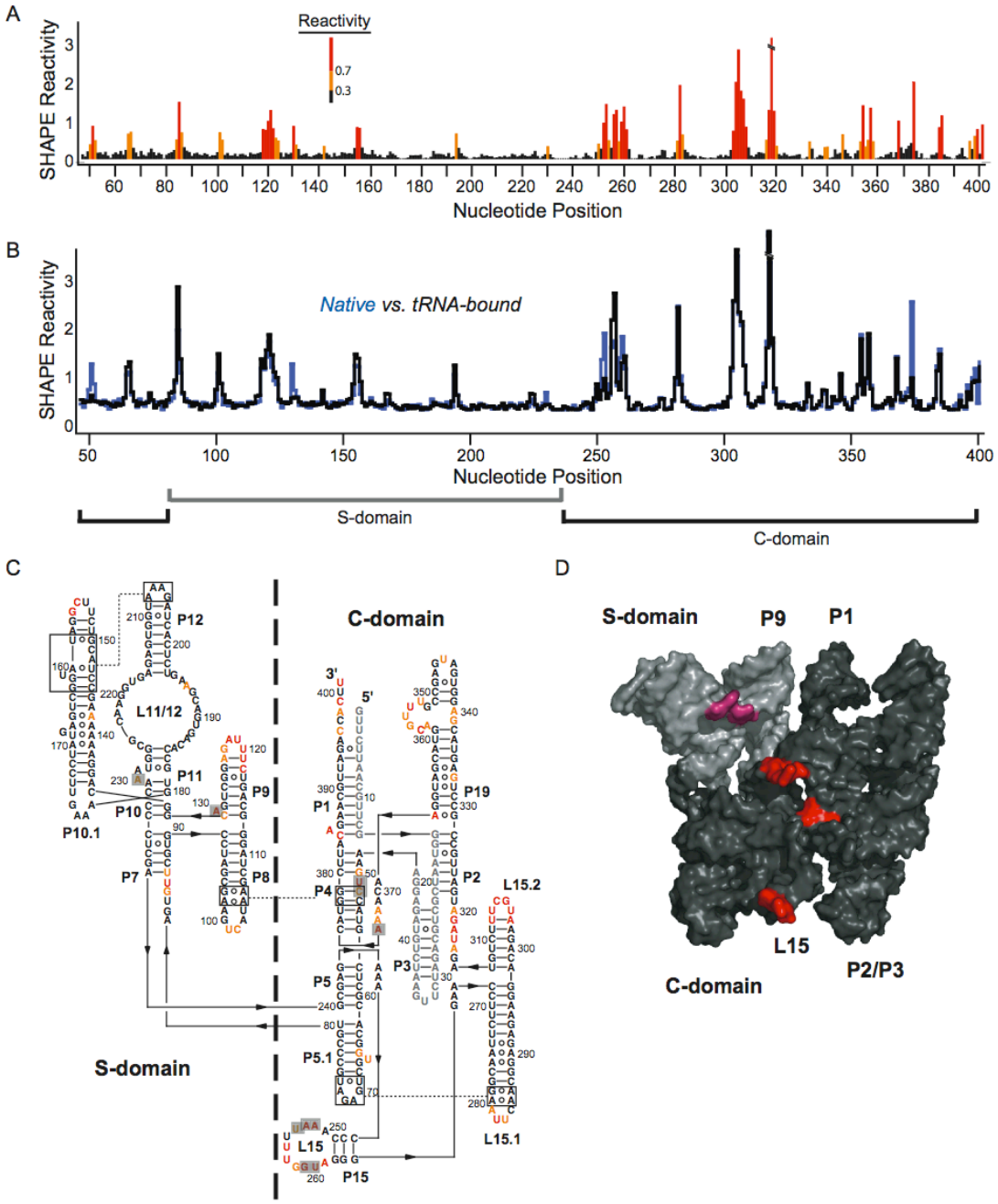


Figure 5.2. Binding of mature tRNA to the *Bacillus subtilis* ribonuclase P enzyme. (A) Absolute SHAPE reactivities in the absence of tRNA. (B) Comparison of absolute SHAPE reactivities in the presence and absence of tRNA. Specificity and Catalytic domains are indicated explicitly. (C) Superposition of SHAPE reactivities (in color) and tRNA protection sites (grey shaded boxes) on the secondary structure for RNase P. (D) Superposition of SHAPE-detected tRNA protection sites on the three dimensional structure of the B-type RNase P (PDB 2A64). All sites of protection are solvent accessible and occur on the same face of the RNA.

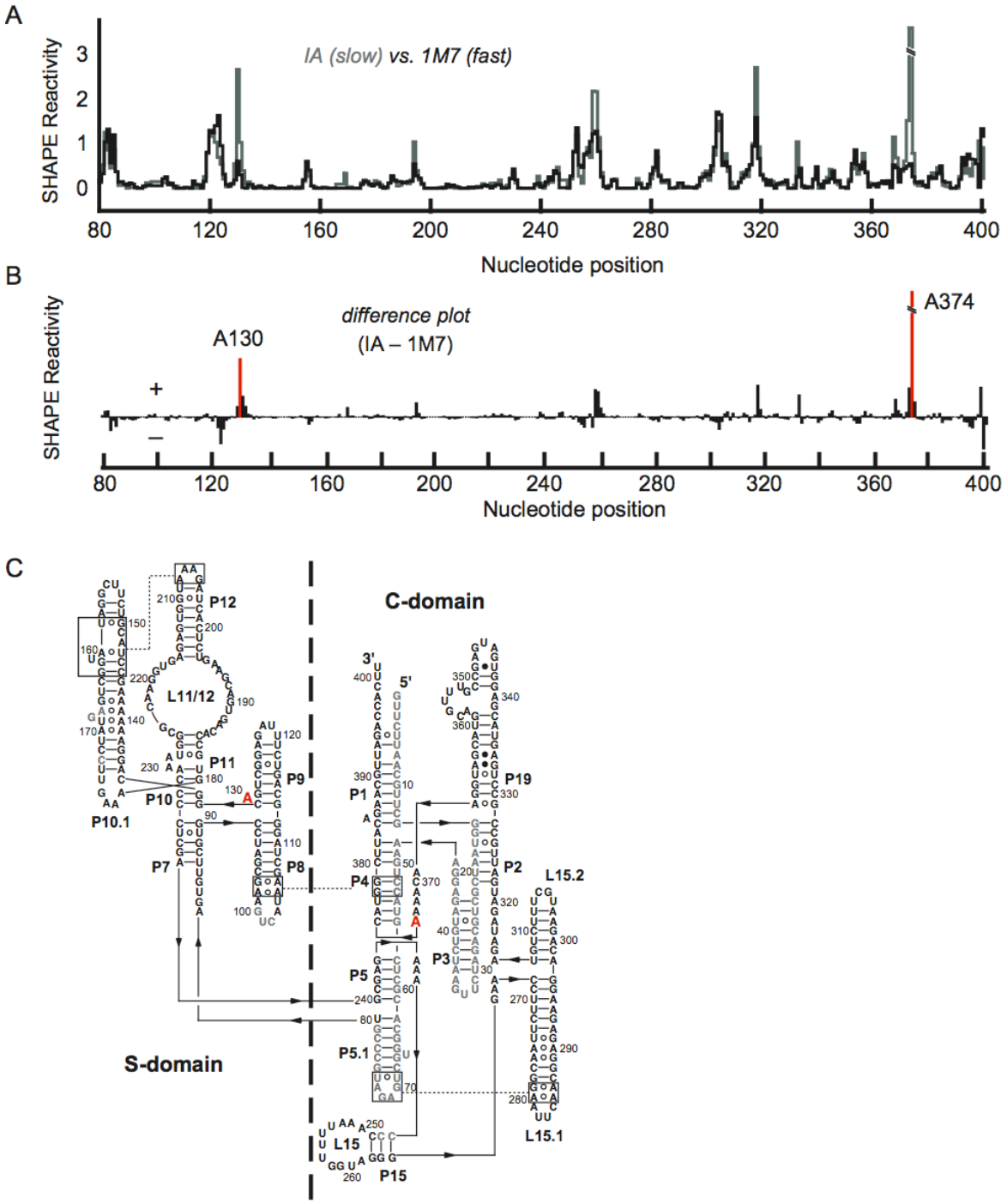


Figure 5.3. A130 and A374 experience slow conformational dynamics in the complete *Bacillus subtilis* RNase P RNA. (A) Comparison of absolute SHAPE reactivities obtained with slow (IA) and fast (1M7) SHAPE reagents. (B) Difference plot of SHAPE reactivities obtained with IA and 1M7. Red bars emphasize nucleotides showing the largest difference in SHAPE reactivity as a function of reagent. (C) RNase P secondary structure diagram showing the location of A130 and A374.

RNA. To test the hypothesis that slow conformational dynamics at A130 gate folding of the RNase P RNA, we evaluated an RNA in which this nucleotide was deleted (ΔA130 RNA).

5.2.2 The ΔA130 RNA has the same global fold as the native RNA. We first confirmed that the equilibrium structures of the native and ΔA130 RNAs were the same despite the disruption of the A130-A230 stacking interaction. We first compared the structures of the two RNAs using SHAPE chemistry [26, 27]. The native and ΔA130 RNAs were treated with the 2′-hydroxyl-selective reagent benzoyl cyanide (BzCN) under conditions that stabilize the native tertiary fold (10 mM $MgCl_2$, 100 mM NaCl, pH 8.0). BzCN selectively forms 2′-*O*-adducts at flexible nucleotides in RNA [25]. Sites of 2′-*O*-adduct formation are identified as stops to primer extension using fluorescently labeled DNA primers, resolved by capillary electrophoresis. Absolute SHAPE reactivities are virtually identical for both RNAs. RNA loops are reactive while the sites of unreactive nucleotides, which are constrained by base pairing and tertiary interactions, are the same in both RNAs (Figure 5.5A). The single exception is that the bulged A130 is reactive in the native sequence but, as expected, this reactivity disappears when this nucleotide is deleted.

We assessed the tertiary structure of the native and mutant RNase P RNAs in a second way using hydroxyl radical footprinting. Hydroxyl radicals react preferentially with solvent accessible regions of the RNA backbone [28]. The native and ΔA130 RNA were exposed to hydroxyl radicals, generated from H_2O_2 in the presence of Fe(II)-EDTA, under conditions that stabilize the tertiary fold (10 mM $MgCl_2$, pH 7.0). Cleavage patterns are very similar for both RNAs (Figure 5.5B). Nucleotides with low cleavage

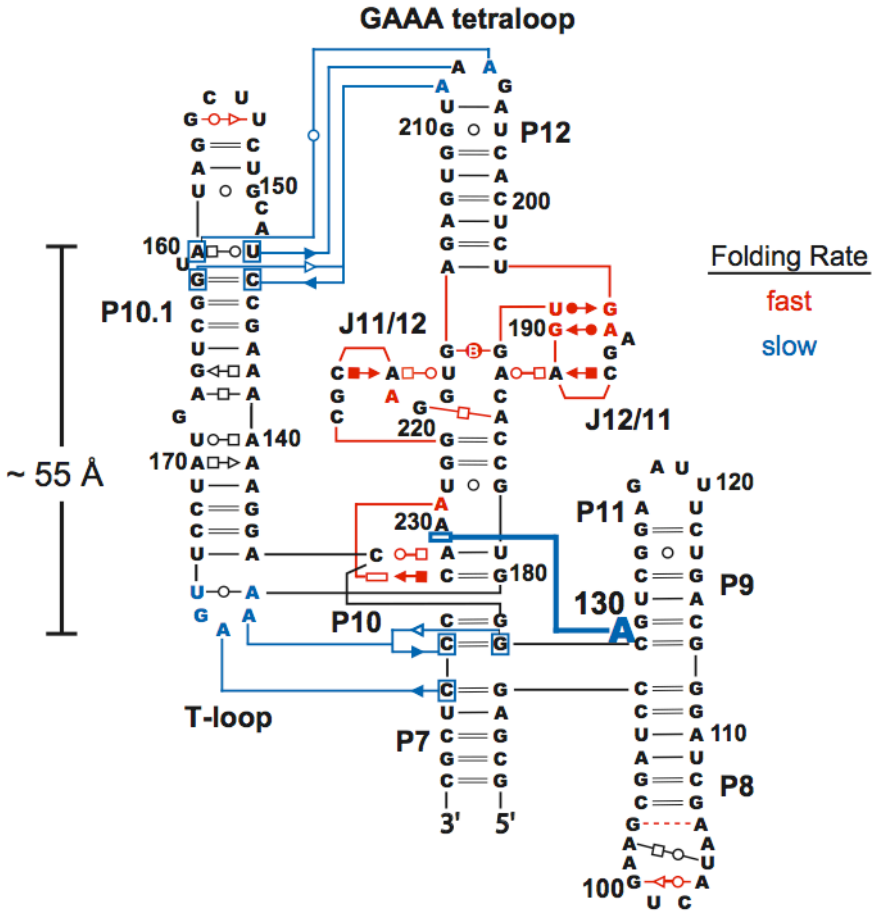


Figure 5.4. Two-step mechanism for tertiary folding in the RNase P specificity domain [25]. Tertiary interactions that form in the fast versus slow steps are illustrated in red and blue, respectively. A130 is emphasized in blue and a bold blue line indicates its stacking interaction with A230.

intensities are protected from solvent to the same extent in both RNAs, indicating the global fold is the same. Critically, the two sets of tertiary interactions that form in the rate-limiting folding step of the native RNA, docking of the T-loop into stacked helices and formation of the GAAA tetraloop-receptor interaction (in blue, Figure 5.4) have identical low cleavage intensities in both RNAs (dashed boxes, Figure 5.5B).

The SHAPE and hydroxyl radical probing experiments (Figure 5.5) indicate that the native and ΔA130 mutant RNase P RNAs form the same secondary and tertiary structures and thus posses the same global fold. Any differences in folding rates observed for the mutant RNA must then reflect the pathway for structural biogenesis, rather than differences in the final folded state.

5.2.3 Fast folding kinetics of the ΔA130 RNA. The folding behavior of the native sequence RNase P specificity domain was analyzed by time-resolved SHAPE [25]. Time-resolved SHAPE takes advantage of the discovery that BzCN degrades completely in water in ~1 sec. Thus, addition of an RNA to this reagent makes it possible to observe RNA folding at single nucleotide resolution in 1 sec snapshots. RNA folding was initiated by addition of Mg^{2+} to a pre-equilibrated RNA using the same conditions used to analyze RNase P structure at equilibrium. Tertiary folding of the RNase P RNA occurs in two kinetically distinct steps: the first step is a fast phase that folds with a rate constant of 0.06 s^{-1}, and the second step is a slower phase that folds with a rate constant of 0.004 s^{-1} (in red and blue, Figure 5.6A).

We then analyzed the folding behavior of the ΔA130 RNase P mutant, also using time-resolved SHAPE. Unlike the native RNA, the mutant RNA folds in a single step that

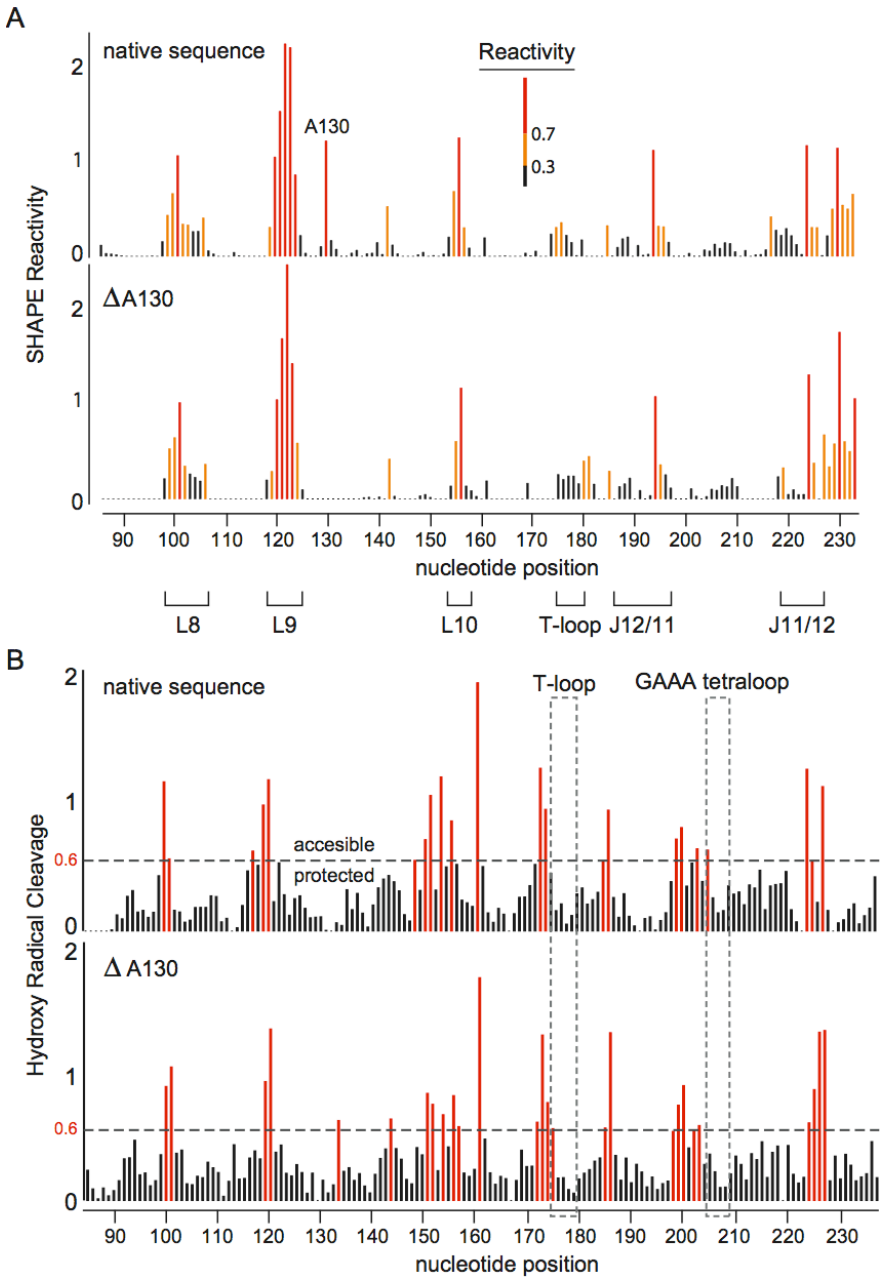


Figure 5.5. Native sequence and the ΔA130 mutant RNase P specificity domain RNAs form similar tertiary structures. (A) SHAPE reactivity histograms for the native (top) and ΔA130 (bottom) RNase P RNAs. Highly reactive nucleotides are indicated in red (≥ 0.7); moderately reactive nucleotides are orange (<0.7, >0.3); and nucleotides with low to near-zero reactivities are black. (B) Hydroxyl radical cleavage intensities, which provide a measure of nucleotide solvent accessibility, for native (top) and ΔA130 (bottom) RNase P RNAs. Solvent accessibility nucleotides (normalized reactivity ≥0.6) are red and protected, solvent inaccessible, nucleotides are black. Protections due to tertiary interactions involving the T-loop and the GAAA tetraloop that form in the slow step for the native sequence are emphasized with gray dashed boxes.

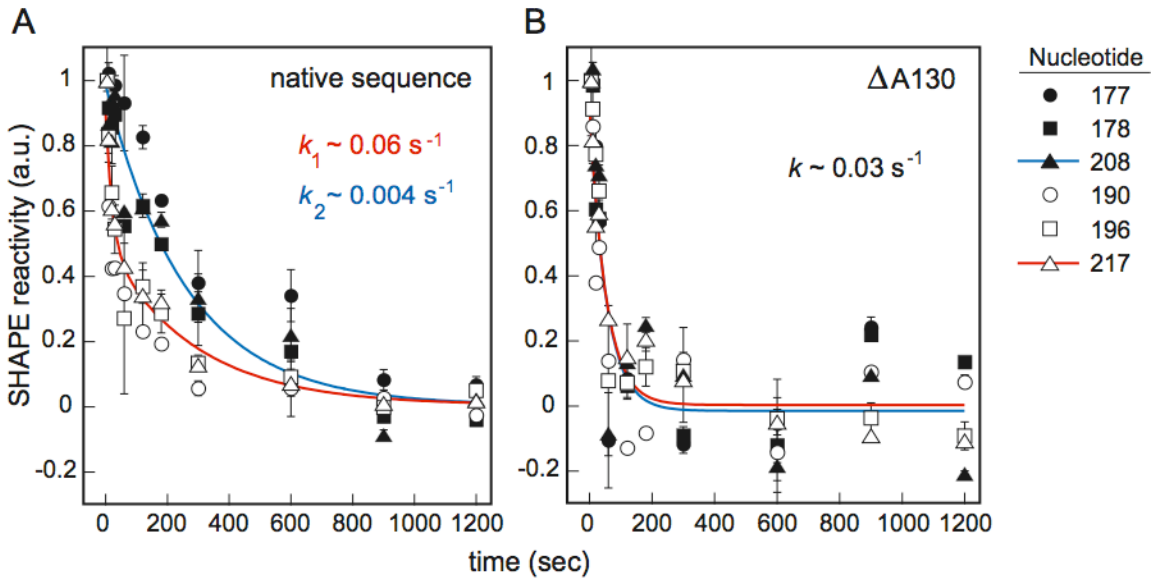


Figure 5.6. Deletion of A130 accelerates RNA folding. Folding rate constants for the RNase P specificity domain were monitored by time-resolved SHAPE. (A) The native sequence folds in two kinetically distinct steps. Nucleotides folding in the fast and slow steps are indicated by open and closed symbols, respectively. (B) All nucleotides in the ΔA130 RNA fold at the same rate, compare open and closed symbols.

is characterized by a rate constant of 0.03 s^{-1} (Figure 5.6B). The ΔA130 mutant thus folds 10-fold faster than the rate-limiting step that governs folding of the native RNA.

We confirmed this striking result using a second independent method, time-resolved fluorescence spectroscopy. The native sequence and ΔA130 RNase P RNAs were labeled at their 5′ ends with an environmentally sensitive fluorophore (Oregon green). The fluorescence emission of Oregon green is strongly dependent on the folded state of the RNA. The fluorescence emission of the 5′-labeled RNAs folded in the absence of Mg^{2+} is quenched by ~40% upon the addition of Mg^{2+} and subsequent formation of the native tertiary fold (Figure 5.7). Control reactions indicate that Mg^{2+}-induced changes in fluorescence reflect RNA conformational changes because the free fluorophore shows no change in fluorescence emission upon addition of Mg^{2+} (Figure 5.7).

Tertiary folding pathway of the native and mutant RNase P RNAs was followed as the decrease in fluorescence emission (at 515 nm) as a function of time after addition of Mg^{2+}. Folding in the native RNA exhibits a clear double exponential, indicative of two folding rates (Figure 5.8A), with rate constants of 0.07 s^{-1} and 0.004 s^{-1}. These rates agree exactly with the rates established by time-resolved SHAPE.

When otherwise identical experiments were performed with the ΔA130 point mutant, the fluorescently detected profile changes significantly. The mutant RNA folds in a single step (Figure 5.8B), with a rate constant of 0.03 s^{-1}, and 10-fold faster than the native RNA. An observed rate constant of 0.03 s^{-1} is in exact agreement with the time-resolved SHAPE experiment.

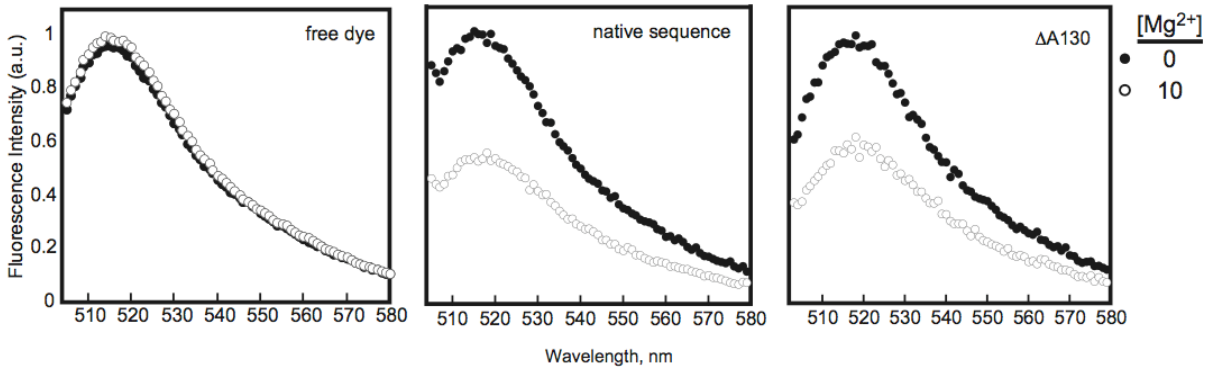


Figure 5.7. Comparison of emission spectra of free Oregon green versus dye-labeled native and mutant RNAs in the presence (open circles) and absence (closed circles) of Mg^{2+}.

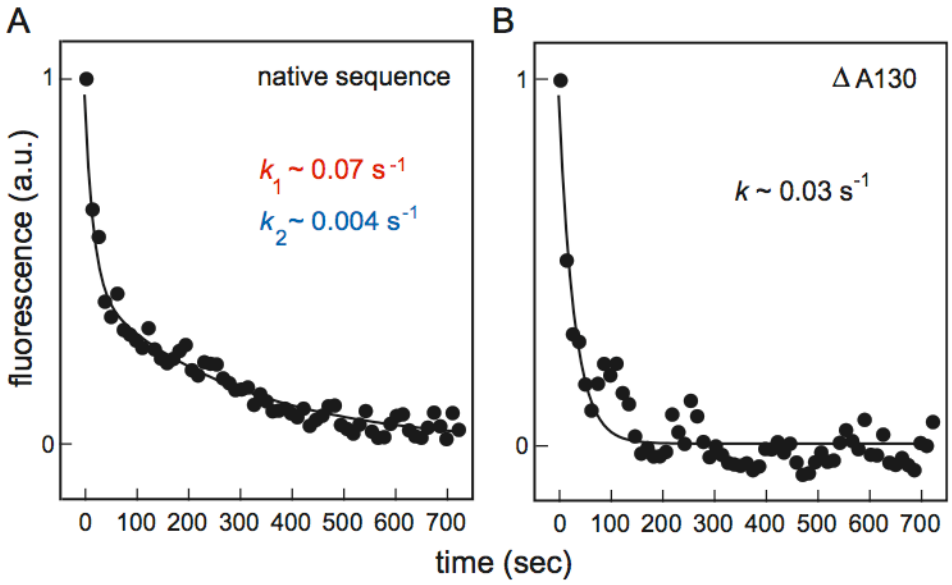


Figure 5.8. Folding rate constants for the RNase P specificity domain monitored by fluorescence spectroscopy. (A) The native sequence folds in two kinetically distinct steps. (B) The ΔA130 mutant RNA folds in a single kinetic step.

Time-resolved SHAPE and fluorescence experiments both reveal that the ΔA130 RNA folds an order of magnitude faster than the native RNA despite the fact that both RNAs fold to the same final tertiary structure. Deletion of a single C2′-endo nucleotide that exhibits slow conformational dynamics has a dramatic effect on tertiary folding in a large RNA.

5.3 Discussion

A surprisingly large number of RNAs appear to fold slowly, on the minute time-scale or even more slowly. Slow folding can reflect the requirement to escape from a non-functional, kinetically trapped state [4, 5] and to complete an extensive conformational search for correct tertiary interactions. For many RNAs, the structural basis for slow folding is unknown. Here we show that the structural requirement to form a single C2′-endo conformation governs folding of a large ribozyme. Deletion of the C2′-endo nucleotide at A130 accelerates RNA folding by an order of magnitude but does not detectably change the global folding of this RNA. We can also rule out that the A130 nucleotide contributes to the stability of the RNA: the native sequence and mutant RNAs possess almost identical thermal melting profiles (Figure 5.9).

Nucleotides in the C2′-endo conformation are relatively rare in RNA structures but occur overwhelmingly in specialized, functionally important, structures in an RNA. The A130 nucleotide in RNase P is typical in this respect because its interaction with A230 creates the local structure necessary for recognizing and binding the pre-tRNA substrate [22, 24, 29].

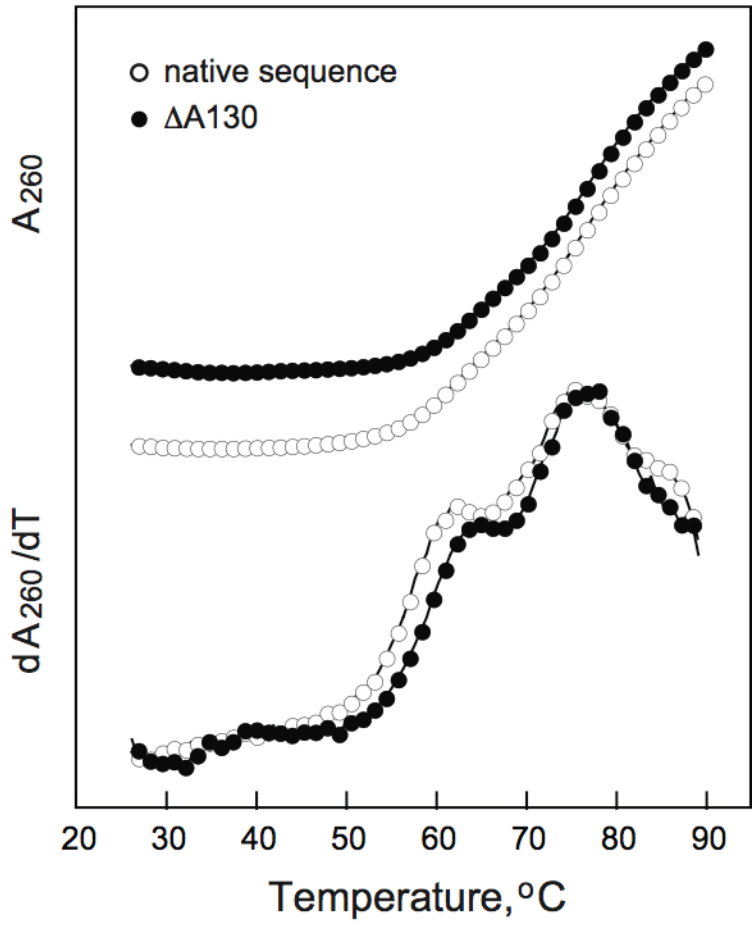


Figure 5.9. Thermally induced unfolding of RNase P, monitored by absorbance at 260 nm.

Although other simple RNA motifs that undergo slow local conformational changes may be discovered, the C2′-endo nucleotide motif is likely to be especially important. First, the C2′-endo conformation is an especially simple variation on RNA structure. Second, C2′-endo nucleotides that experience slow conformational dynamics occur in a wide variety of structural contexts [16]. Third, as shown in this work, slow folding at a single C2′-endo nucleotide is sufficient to function as a rate-determining step for folding a large RNA with a complex tertiary structure.

We envision that C2′-endo nucleotides characterized by slow local conformational dynamics have three potential roles in RNase P and in the function of large RNAs in general. First, a bulged C2′-endo nucleotide appears to be well suited for bridging a helix with another motif in RNA. For example, in the context of the specificity domain, both A130 and A168 are C2′-endo nucleotides characterized by slow nucleotide dynamics and both function to stabilize tertiary interactions between a helix and other RNA elements [16, 20]. In the complete RNase P ribozyme, two extra helical C2′-endo nucleotides, A130 and A374, make direct interactions with the tRNA substrate. A critical corollary of the role of C2′-endo nucleotides in stabilizing RNA tertiary structures is that slow dynamics are an “artifact” of this conformation.

Second, we also think it is likely that slow conformational dynamics might be exploited directly as a molecular timer in RNA folding reactions. This molecular timer function could play a structural role on both long-range and local levels. On a long-range level, A130 gates folding in the specificity domain and thus likely causes this domain to fold more slowly than other structural domains in the RNase P RNA. In this

way, slow conformational dynamics would order folding of the specificity domain relative to the rest of the RNA.

Third, on a local level, we find it striking that two C2′-endo nucleotides, A130 and A374 (Figure 5.3), both appear to contact tRNA (Figure 5.2). These nucleotides may function as molecular timers to facilitate selection of tRNA structures or to coordinate binding with cleavage of the tRNA precursors.

Because the motif is so simple and so highly overrepresented in functionally critical RNA motifs, we infer that slow dynamics at C2′-endo nucleotides is routinely exploited in the RNA world to time RNA folding, ligand recognition, and ribonucleoprotein assembly reactions. We encourage the RNA community to evaluate their favorite RNAs for the phenomenon, and consequences, of slow conformational dynmics at C2′-endo nucleotides.

5.4 Experimental

5.4.1 Synthesis of *Bacillus subtilis* RNase P RNAs. All RNAs were synthesized by *in vitro* transcription using a PCR-generated template [30]. RNAs analyzed by SHAPE and hydroxyl radical footprinting were embedded in 5′ and 3′ structure cassette sequences [31]. For fluorescence-detected folding, 5′-monophosphorothioate wild type and ΔA130 RNase P specificity domains were transcribed in the presence of 10 mM 5′-GMPS (BioLog). The flanking 5′ and 3′ structure cassette sequences were omitted and the RNAs contained an initial G86-C240 base pair instead of the U86-A240 base pair. RNAs were purified by denaturing polyacrylamide gel electrophoresis, excised from the gel, and recovered by passive elution and ethanol precipitation. Purified RNAs were

resuspended in TE [10 mM Tris (pH 8.0, 1 mM EDTA] at concentrations of about 40 μM and stored at –20 °C.

5.4.2 SHAPE Analysis. For equilibrium experiments, native and mutant RNase P specificity domain RNAs (5 pmol in 6 μL H_2O) were heated at 95 °C for 2 min, cooled on ice, treated with 3 μL of 3.3× folding buffer [333 mM NaCl, 333 mM HEPES (pH 8.0), 33.3 mM $MgCl_2$], and incubated at 37 °C for 20 min. The RNA solution was added directly to 1 μL 10× BzCN (600 mM in anhydrous DMSO). For time-resolved experiments, the procedure was identical to that described previously [25]. tRNA binding to the full length RNase P was measured using a published procedure [29]. Briefly, the full length RNA (20 pmol) and mature $tRNA^{Phe}$ (60 pmol, if present) were heated in buffer [200 mM HEPES (pH 8.0] at 85 °C for 2 min, incubated at room temperature for 3 min, supplemented with 1 μL 10× $MgCl_2$ (1 M)) and 3 μL 6.6× KCl (4 M). The RNase P RNA and the $tRNA^{Phe}$ were incubated at 50 °C and 37 °C, respectively, for 10 min. The RNase P solution was divided in half and added to $tRNA^{Phe}$ or a no-tRNA solution and incubated at 37 °C for 15 min. RNA solutions were then added directly to 1 μL 10× BzCN (200 mM in anhydrous DMSO). The differential reactivity experiment, used to assess local conformational dynamics was performed similarly except the 10× [$MgCl_2$] was 100 mM and no tRNA or salt was added. The RNA solution was treated with 1M7 or IA (1 μL, 50 mM in anhydrous DMSO), and allowed to react for 70 s (equal to five 1M7 hydrolysis half-lives) or 36 min (equal to five IA hydrolysis half-lives). For all of the above, no-reagent controls contained 1 μL neat DMSO. Modified RNA was recovered by ethanol precipitation [90 μL sterile H_2O, 5 μL NaCl (4 M), 1 μL glycogen (20 mg/mL), 400 μL ethanol; 30 min at –80 °C] and resuspended in 10 μL TE.

5.4.3 Hydroxy Radical Probing. RNA (5 pmol) in 4 μL sterile water was heated at 95 °C for 2 min, cooled on ice, treated with 3 μL of 3.3× folding buffer [66 mM HEPES (pH 7.0), 33 mM $MgCl_2$], and incubated at 37 °C for 20 min. were made fresh prior to each experiment. A 1:1 mixture of the sodium ascorbate (50 mM) and Fe-EDTA (20 mM EDTA pH 8.0; 10 mM ferrous ammonium sulfate) solutions were prepared and 2 μL spotted in the lid of each reaction. A 0.03% H_2O_2 solution (1 μL) was spotted on a separate area of each lid. Control reactions contained 3 μL H_2O. The reactions were initiated by briefly centrifuging the tubes and further incubated at 37 °C for 2 min. Reactions were stopped by ethanol precipitation and resuspended in 10 μL TE.

5.4.4 Primer Extension and Data Analysis of Modified RNA. The general procedure was that outlined previously [30]. Briefly, a fluorescently labeled DNA primer was annealed to the RNA (10 μL, from the previous modification steps) and extended with Superscript III. Primer extension reactions were quenched with 4 μL of a 1:1 mixture containing EDTA (100 mM , pH 8.0) and sodium acetate (3 M, pH 5.2); the resulting cDNAs were recovered by ethanol precipitation, washed twice with 70% ethanol, vacuum dried for 10 min, and resuspended in 10 μL deionized formamide. cDNA extension products, along with dideoxy sequencing markers [30] were separated by capillary electrophoresis using an Applied Biosystems 3130 Genetic Analyzer. Raw traces from the ABI 3130 were processed using ShapeFinder [32]. All data sets were normalized by excluding the 2% most reactive nucleotides and dividing by the average intensity of the next 8% most reactive nucleotides. Folding rates for individual nucleotides were calculated as described [25] by fitting to either a single ($A + Be^{-kt}$) or a double exponential ($A + Be^{-k_1 t} + Ce^{-k_2 t}$).

5.4.5 Folding Kinetics Monitored by Fluorescence Spectroscopy. A fluorescently labeled RNase P RNA was created by incubating the 5′-phosphorothioate RNA (6.28 nmol, from above) in 60 μL of buffer (10mM HEPES, pH 7.0) with Oregon green 488 maleimide (2 μL, 20 mg/mL, Invitrogen) in the dark at room temperature for 3 hours. Labeled RNA was recovered by ethanol precipitation and resuspended in 6 μL TE and 18 μL formamide (90%, 1X TBE). The labeled RNA was purified by denaturing polyacrylamide gel electrophoresis, excised from the gel, and recovered by passive elution [0.5 M potassium acetate (pH 6.5), 1 mM EDTA, ~12 hrs, 4 °C) and ethanol precipitation [1/10 vol NaCl (3M), 1 vol 2-propanol, 2.5 vol ethanol, 20 min at –80 °C]. The purified labeled RNA (~1500 pmol) was resuspended in TE and stored at –20 °C in the dark. Folding was initiated by adding a solution of RNA (400 pmol in 522 μL; 100 mM NaCl, pH 8.0) at 37 °C to 58 μL of $MgCl_2$ (100 mM) equilibrated in a 400 μL quartz cuvette (at 37 °C for 15 min) in a Varian Cary Eclipse Fluorescence Spectrophotometer. Measurements were averaged over 3 sec every 12 sec for 30 min at λ_{ex} = 491 ± 10 nm and λ_{em} = 515 ± 10 nm.

5.5 References

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

2. Leontis, N.B. and Westhof, E., *Analysis of RNA motifs.* Curr. Opin. Struct. Biol., 2003. **13**(3): p. 300-8.

3. Gesteland, R.F., Cech, T.R., and Atkins, J.F., *The RNA World*. 3rd ed. 2004, New York: Cold Spring Harbor.

4. Pan, T. and Sosnick, T.R., *Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity.* Nature, 1997. **4**: p. 931-938.

5. Treiber, D.K. and Williamson, J.R., *Exposing the kinetic traps in RNA folding.* Curr. Opin. Struct. Biol., 1999. **9**: p. 339-345.

6. Herschlag, D., *RNA Chaperones and the RNA Folding Problem.* J. Biol. Chem., 1995. **270**(36): p. 20871-20874.

7. Weeks, K.M., *Protein-facilitated RNA folding.* Curr. Opin. Struct. Biol., 1997. **7**(3): p. 336-342.

8. Schroeder, R., Barta, A., and Semrad, K., *Strategies for RNA folding and assembly.* Nat. Rev. Mol. Cell Biol., 2004. **5**(11): p. 908-919.

9. Treiber, D.K. and Williamson, J.R., *Beyond Kinetic Traps in RNA Folding.* Curr. Opin. Struct. Biol., 2001. **11**: p. 309-314.

10. Waldsich, C. and Pyle, A.M., *A Kinetic Intermediate that Regulates Proper Folding of a Group II Intron RNA.* J. Mol. Biol., 2008. **375**: p. 572-580.

11. Richardson, J.S., et al., *RNA backbone: Consensus all-anhle conformers and modular string nomenclature (an RNA Ontology Consortium contribution).* RNA, 2008. **14**: p. 465-481.

12. Correll, C.C., et al., *The common and the distinctive features of the bulged-G motif based on a 1.04 A resolution RNA structure.* Nucleic Acids Res., 2003. **31**(23): p. 6806-18.

13. Wadley, L.M. and Pyle, A.M., *The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery.* Nucleic Acids Res., 2004. **32**(22): p. 6650-9.

14. Heus, H.A., et al., *The Detailed Structure of Tandem GA Mismatched Base-Pair Motifs in RNA Duplexes is Context Dependent.* J. Mol. Biol., 1997. **271**: p. 147-158.

15. SantaLucia, J.J. and Turner, D.H., *Structure of (rGGCGAGCC)$_2$ in solution from NMR and Restrained Molecular Dynamics.* Biochemistry, 1993. **32**: p. 12612-12623.

16. Gherghe, C.M., et al., *Slow Conformational Dynamics at C2'-Endo Nucleotides in RNA.* J. Am. Chem. Soc., 2008. **130**: p. 8884-8885.

17. Frank, D.N. and Pace, N.R., *Ribonuclease P: Unity and diversity in a tRNA processing ribozyme.* Annu. Rev. Biochem., 1998. **67**: p. 153-180.

18. Walker, S.C. and Engelke, D.R., *Ribonuclease P: the evolution of an ancient RNA enzyme.* Crit. Rev. Biochem. Mol. Biol., 2006. **41**(2): p. 77-102.

19. Pan, T., *Higher Order Folding and Domain Analysis of the Ribozyme from Bacillus subtilis Ribonuclease P.* Biochemistry, 1995. **34**: p. 902-909.

20. Krasilnikov, A.S., et al., *Crystal structure of the specificity domain of ribonuclease P.* Nature, 2003. **421**: p. 760-764.

21. Krasilnikov, A.S., et al., *Basis for Structural Diversity in Homologous RNAs.* Science, 2004. **306**: p. 104-107.

22. LaGrandeur, T.E., et al., *Phylogenetic comparative chemical footprint analysis of the interaction between ribonuclease P RNA and tRNA.* EMBO J., 1994. **13**: p. 3945-3952.

23. Pan, T., Loria, A., and Zhong, K., *Probing of tertiary interactions in RNA: 2'-Hydroxyl-base contacts between the RNase P RNA and pre-tRNA.* Proc. Natl. Acad. Sci. U.S.A., 1995. **92**: p. 12510-12514.

24. Loria, A. and Pan, T., *Recognition of the T Stem-Loop of a Pre-tRNA Substrate by the Ribozyme from Bacillus subtilis Ribonuclease P.* Biochemistry, 1997. **36**: p. 6317-6325.

25. Mortimer, S.A. and Weeks, K.M., *Time-Resolved RNA SHAPE Chemistry.* J. Am. Chem. Soc., 2008. **In press**.

26. Merino, E.J., et al., *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).* J. Am. Chem. Soc., 2005. **127**: p. 4223-4231.

27. Gherghe, C.M., et al., *Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA.* J. Am. Chem. Soc., 2008. **130**: p. 12244-12245.

28. Tullius, T.D. and Greenbaum, J.A., *Mapping Nucleic Acid Structure by Hydroxy Radical Cleavage.* Curr. Opin. Struct. Biol., 2005. **9**: p. 127-134.

29. Odell, L., et al., *Interaction of structural modules in substrate binding by the ribozyme from Bacillus subtilis RNase P.* Nucleic Acids Res., 1998. **26**: p. 3717-3723.

30. 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**: p. 4144-4145.

31. 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.

32. Vasa, S.M., et al., *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.

CHAPTER 6

Structure Selective N-Silylation of Guanosine Residues in RNA

6.1 Introduction

Higher order RNA structures are built up from a base-paired secondary structure augmented by local and long-range tertiary interactions [1-3]. The ability to analyze structure in solution at both secondary and tertiary structure levels is necessary to understand the structure-function relationships that govern RNA folding, catalysis, and interactions with small molecules and protein ligands.

A valuable approach for analyzing RNA structure has been the use of small molecules that react to form covalent adducts with specific functional groups in the nucleobases [4-6]. For example, guanosine reacts with dimethylsulfate (DMS) to form an adduct at the N7 position and with kethoxal to form a cyclic adduct at the N1 and N2 positions [6]. A significant recent advance has been the development of SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension), which uses hydroxyl-selective electrophiles to react at the 2′-ribose positions at conformationally flexible positions in RNA [7-10]. Because almost all RNA nucleotides have a free 2′-hydroxyl, all nucleotides in an RNA can be interrogated in a single experiment, regardless of base identity.

An important, but incompletely met, experimental objective is the ability to map solvent accessibility in RNA, also at single nucleotide resolution. As Lavery and Pullman pointed out in 1984, none of the widely used nucleobase-selective reagents, including DMS, kethoxal or carbodiimides, measure solvent accessibility exclusively. Instead, these carbon-based electrophiles are sensitive to a mixture of steric accessibility and electrostatic factors [11]. This dependence on electrostatic factors for classic reagents, DMS and kethoxal, likely reflects (1) that nucleophilic attack must originate from the

direction opposite of the leaving group and at an angle of 107° to the plane of the carbonyl and (2) that the resulting transition state is highly polar.

Solvent accessibility at the RNA backbone can be estimated using hydroxyl radical footprinting [12], which makes use of an Fe(II)-EDTA^{2-} catalyst and an oxygen source to induce cleavage in the RNA backbone. Hydroxyl radical footprinting has been widely and very advantageously used to map RNA tertiary structure [12-15]. The hydroxyl radical is an aggressive reagent which reacts at multiple backbone positions [16] to induce strand cleavage; in addition; the Fe(II)-EDTA moiety carries a net –2 charge. Thus, both steric and electrostatic factors contribute to reactivity of this reagent as well. Absolute correlations between hydroxyl radical reactivity and solvent accessibility are good, but imperfect [17, 18]. We therefore sought to develop new chemical probes of RNA structure whose reactivity is strictly governed by solvent accessibility.

Our initial step in this direction was to explore silyl chlorides. Silicon-based electrophiles are species capable of reacting with both oxygen and nitrogen nucleophiles [19], including those present in RNA. Attack of a nucleophile on the silicon d-orbital leads to formation of a discrete, negatively charged, pentacoordinate intermediate that breaks down with loss of a leaving group [20]. In this work, we explore the reactivity of *N,N*-(dimethylamino)dimethylchlorosilane (DMAS-Cl) and show that this reagent reacts at the N2 position of guanosine to quantitatively report solvent accessibility in a folded RNA (Figure 6.1).

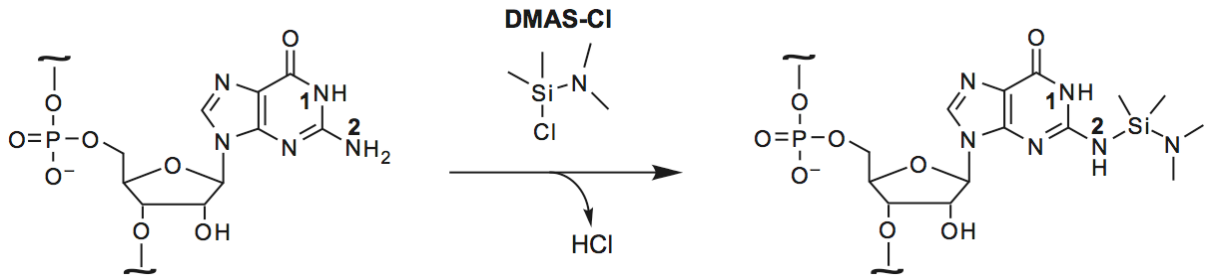


Figure 6.1 Reaction of *N,N*-(dimethylamino)dimethylchlorosilane (DMAS-Cl) with the guanosine N2 position in RNA.

6.2. Results

6.2.1. Silylation at the Guanosine N2 Position. We explored the ability of DMAS-Cl to form stable covalent adducts with RNA nucleotides by treating the *B. subtilis mgtE* aptamer domain (termed the M-box RNA) with DMAS-Cl. The M-box RNA was chosen because it is a relatively large RNA (154 nt) with a known three-dimensional structure solved to good resolution (2.6 Å) [21]. This RNA contains numerous typical base pairing, base stacking, and higher-order tertiary interactions.

We initially assessed whether DMAS-Cl reacts to form stable, covalent adducts with RNA. The M-box RNA was treated with DMAS-Cl under denaturing conditions (no added ions, 70 °C) where all functional groups in the RNA should be accessible to the reagent. Sites of potential adduct formation were detected by their ability to inhibit primer extension by reverse transcriptase. Extension products, visualized using a fluorescently labeled primer, were resolved to single nucleotide resolution by capillary electrophoresis. Potential adducts at any base functional group are reported as strong peaks in the resulting electropherogram (Figure 6.2).

The strongest sites of reactivity occurred at guanosine nucleotides in the RNA (upper panel, Figure 6.2). To assess whether DMAS-Cl forms a covalent adduct at the guanosine N2 position, we tested an M-box RNA in which inosine replaced guanosine (termed the inosine RNA). Inosine is iso-structural with guanosine except that it lacks the exocyclic N2 amine (structures, Figure 6.2). The ITP RNA was treated with DMAS-Cl under conditions identical to those used for the M-box RNA experiment; no reactivity above background was observed at any position (middle traces, Figure 6.2). These

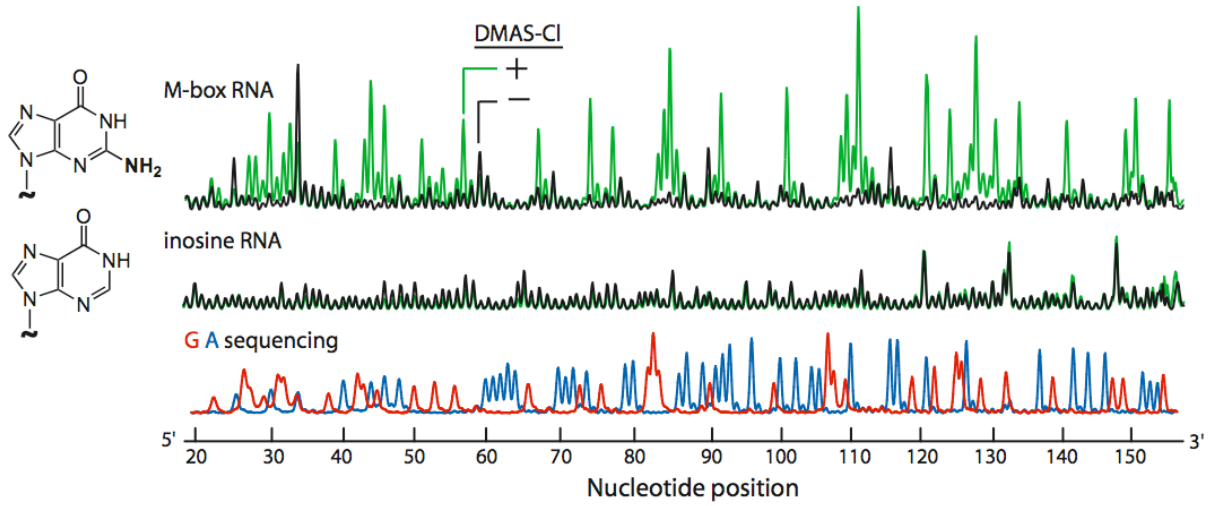


Figure 6.2 Visualizing RNA–DMAS-Cl adducts by primer extension, resolved by capillary electrophoresis. Experiments were performed using the native sequence M-box RNA (upper panel) and with an M-box RNA containing inosine in place of guanosine (middle panel). Sequencing ladders (G & U) were used to assign peak positions.

experiments strongly support the interpretation that DMAS-Cl reacts selectively at the guanosine N2 position in RNA.

We further characterized the product formed between guanosine and DMAS-Cl by forming an adduct between 2′-deoxyguanosine 5′-monophosphate and DMAS-Cl under conditions similar to those used for the experiments performed with the M-box RNA. The structure of the resulting adduct was established by NMR (see Experimental Procedures). The NMR data confirms that the dimethyl amino group remains bound to silicon, consistent with the adduct postulated in Figure 6.1.

6.2.2. Structure-Selective Reaction of DMAS-Cl with RNA. The M-box RNA was treated with DMAS-Cl at 37 °C both under conditions where the RNA is expected to form only its secondary structure (100 mM NaCl, pH 8.0) and also under conditions (+ Mg^{2+}) that stabilize the native tertiary fold (10 mM $MgCl_2$, 100 mM NaCl, pH 8.0). For both experiments, reactivity at every position was quantified by primer extension followed by resolution using capillary electrophoresis.

Again, reaction occurred at guanosine residues in the RNA. Overall, guanosine residues in the M-box RNA are more reactive in the absence of Mg^{2+}, where the RNA lacks most tertiary structure, and are substantially less reactive in the presence of Mg^{2+}, where the RNA forms extensive higher-order tertiary interactions (compare upper and lower histograms; Figure 6.3A). The reactivity of each guanosine residue in the RNA was calculated as a percentage of the most reactive residue. Guanosine residues could then be divided into three categories based on their reactivity patterns: those that are unreactive (<20%) under both conditions; those that are reactive only in the absence of Mg^{2+}; and those that remain reactive in the presence of Mg^{2+} (Figure 6.3B).

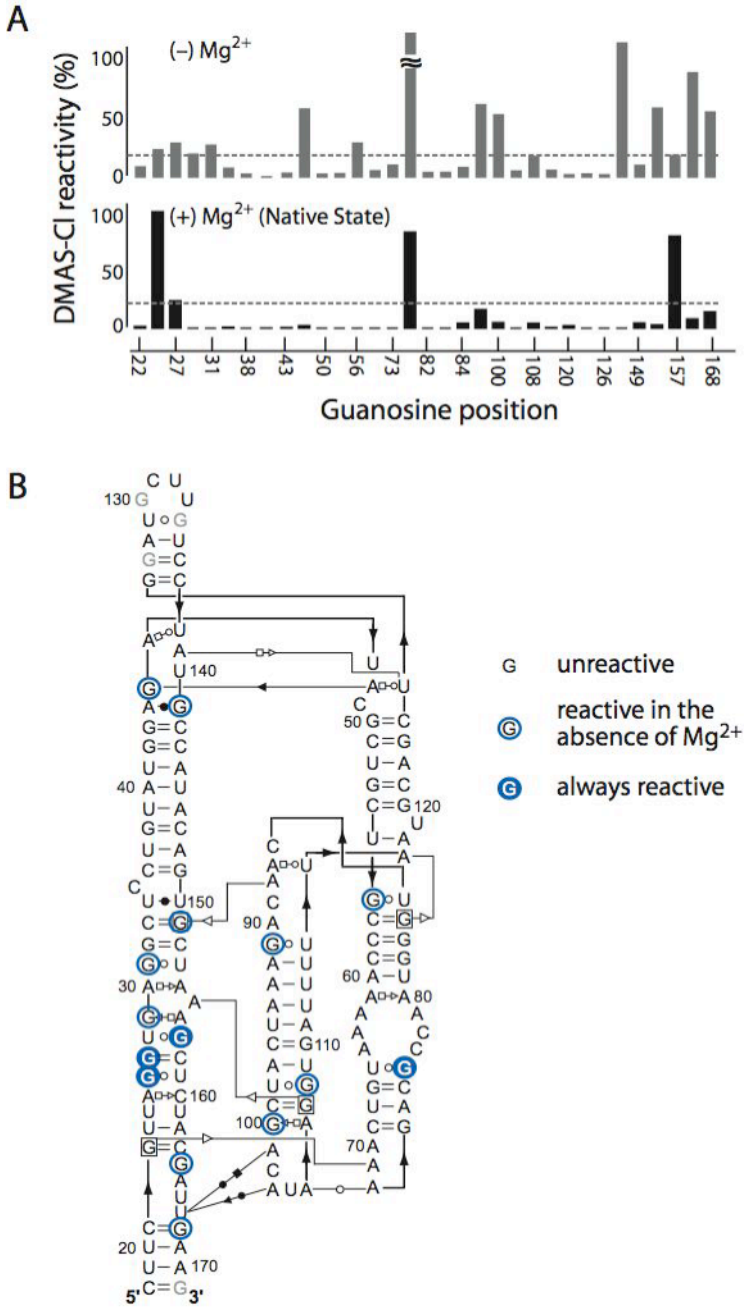


Figure 6.3. Structure-selective reaction of DMAS-Cl with RNA. (A) Guanine reactivity towards DMAS-Cl in the absence and presence of Mg^{2+}. (B) Guanosine reactivity superimposed in the secondary structure of the M-box RNA [21]. Hydrogen bonding interactions involving the bases [22] are illustrated explicitly. Guanosines with >20% reactivity towards DMAS-Cl are taken to be solvent accessible.

We compared DMAS-Cl reactivities with the solvent accessibility of the guanosine N2 position in the native RNA as a function of spherical probe sizes ranging from 1.0–7.0 Å [23]. There is a strong correlation between reactivity and solvent accessibility that reaches a maximum at a 5.0 Å probe size (open circles, Figure 6.4A). The correlation coefficient, *r*, is 0.82 for all residues and 0.98 if position G76 (indicated with an asterisk) is excluded. Excluding G76 is justified because this nucleotide forms a G-U base pair, analogous to those at several other reactive positions in the RNA. The low calculated accessibility at this one position may reflect constraints in the crystal that do not occur in solution. An effective probe size of 5.0 Å is consistent with the overall molecular dimensions of the DMAS-Cl molecule [24] (Figure 6.4B).

Furthermore, guanosine residues that become reactive in the absence of Mg^{2+} all either lie in single stranded regions or form G-A or G-U base pairs. These are exactly the nucleotides most likely to preferentially sample solvent accessible states in the partially denatured RNA. Thus, the differential reactivity patterns we observe correlate with the expectation that DMAS-Cl maps solvent accessibility at the guanosine N2 position in RNA, both in the presence and absence of Mg^{2+}.

6.2.3. Comparison of DMAS-Cl and Kethoxal Reactivity. An alternate, classic, chemical probe for RNA structure that is also selective for the guanosine N2 position is kethoxal (β-ethoxy-α-keto-butyraldehyde) [6]. Kethoxal is thought to react initially at the N2 position in guanosine via its aldehyde group followed by reaction at N1 to form a cyclic adduct [25]. We evaluated the reactivity of kethoxal by treating the M-box RNA with this reagent, both in the presence and absence of $MgCl_2$, under conditions similar to those used for DMAS-Cl (Figure 6.5A). These reactivities were then superimposed on

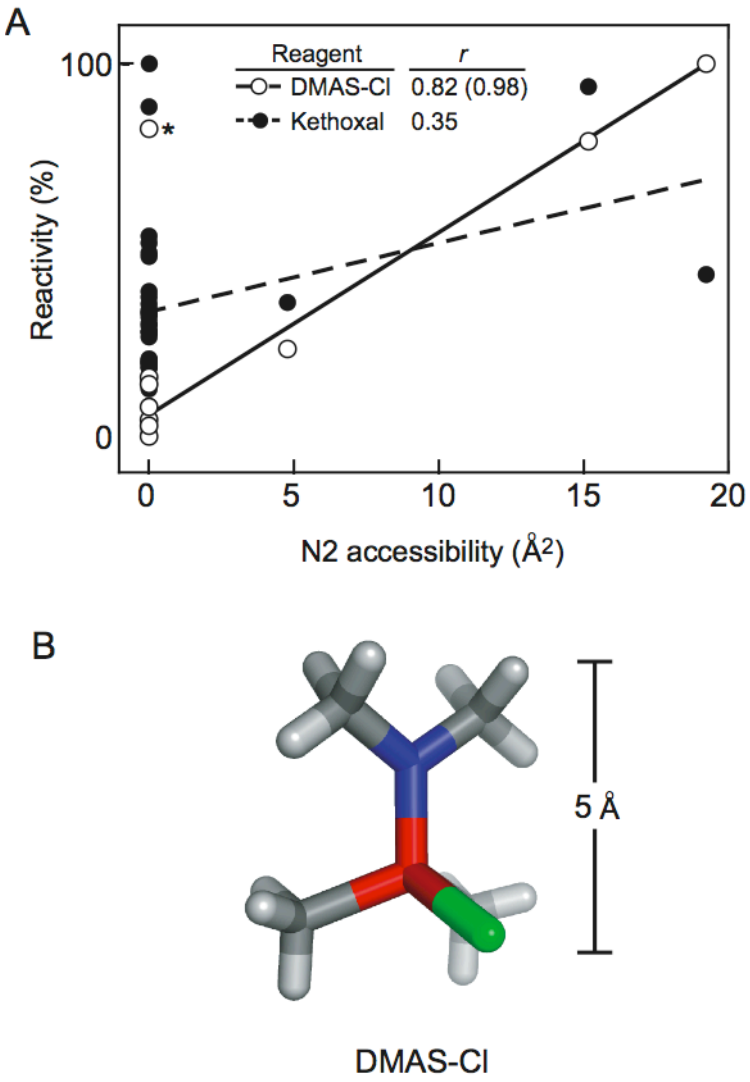


Figure 6.4. Correlation between solvent accessibility and guanosine reactivity with DMAS-Cl and with kethoxal at the N2 position. (A) Solvent accessibility correlation using a 5 Å probe size. Nucleotide G76 (asterisk) was not included for the *r*-value indicated in parenthesis. (B) Molecular dimensions of DMAS-Cl.

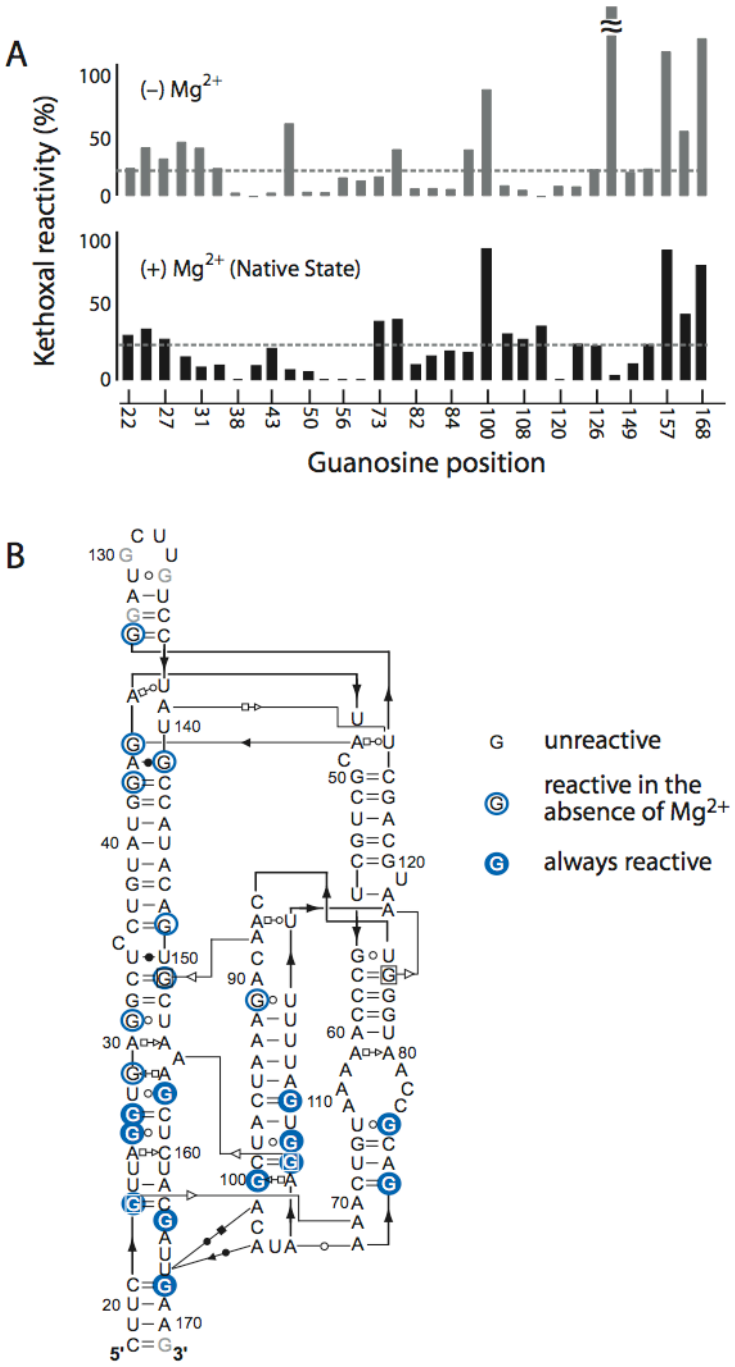


Figure 6.5. Reaction of kethoxal with RNA. (A) Guanine reactivitiy towards kethoxal in the absence and presence of Mg^{2+}. (B) Superposition of kethoxal reactivities on the secondary structure of the M-box RNA [21].

the secondary structure of the M-box RNA (Figure 6.5B).

Despite the fact that both kethoxal and DMAS-Cl form covalent adducts with the same N2 functional group in guanosine, the observed reactivity patterns differ (compare Figures 6.3B & 6.5B). The determinants that govern the reactivities of these two reagents must therefore also be different.

Correlation between kethoxal reactivity and solvent accessibility at the guanosine N2 position reaches a maximum at probes sizes >3 Å but is very poor in all cases ($r \leq 0.35$, closed circles, Figure 6.4A). The lack of a good correlation between kethoxal reactivity and solvent accessibility using the M-box RNA is consistent with the poor correlation seen previously with tRNA [11].

6.3. Discussion

At present, there are very few ways to probe RNA solvent accessibility in a direct and quantitative way. The correlation between solvent accessibility and hydroxyl radical–induced cleavage appears to be modest, with *r*-values of only ~0.6 (unpublished data). This work shows that, with the right chemistry, it is possible to create chemical probes of RNA structure that yield near-perfect measures of solvent accessibility (Figure 6.4A).

DMAS-Cl demonstrates very strong selectivity for reaction with the guanosine exocyclic N2 position. Adenosine and cytidine nucleotides also possess exocyclic amine groups (N6 in adenosine and N4 in cytidine). The feature that distinguishes the guanosine N2 versus the N6 and N4 positions in adenosine and cytidine is that, while

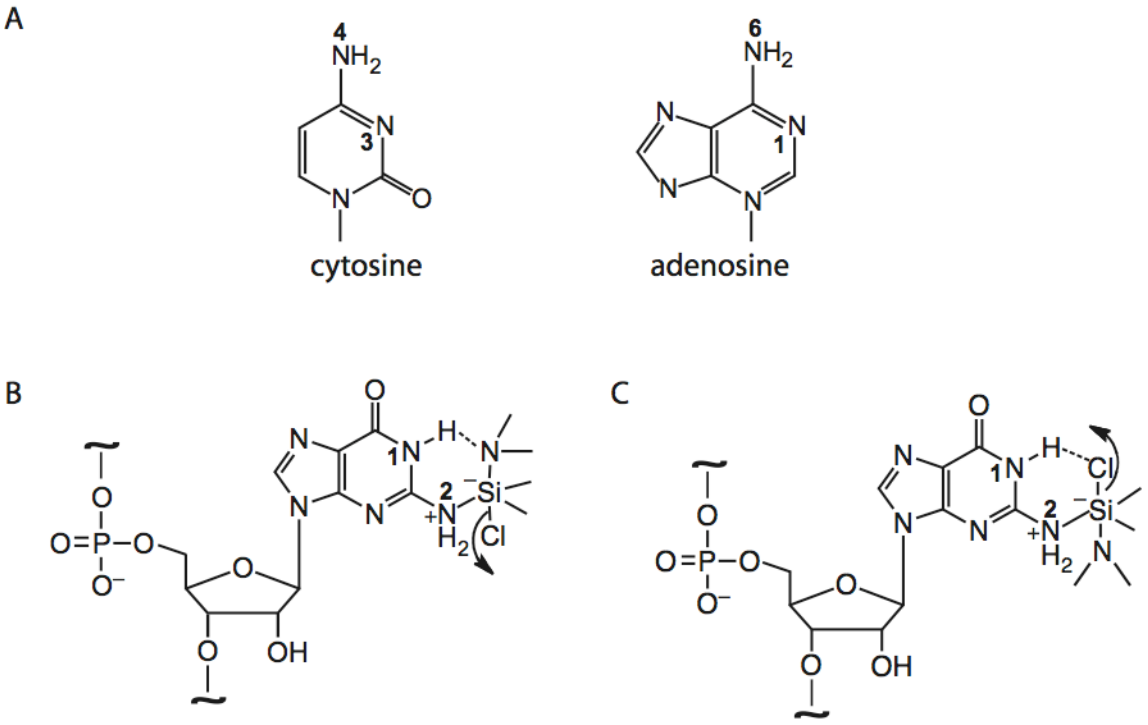


Figure 6.6. Mechanisms consistent with guanosine selectivity of DMAS-Cl reactivity. (A) Structures of adenosine and cytidine nucleobases highlighting the exocyclic amine and adjacent hydrogen bond acceptor. (B,C) Models for selective reaction with the guanosine N2 position, requiring an adjacent hydrogen bond donor.

there are hydrogen bond *acceptors* adjacent to the exocyclic amines in adenosine and cytidine (N1 and N3, respectively), only guanosine has a hydrogen bond *donor* adjacent to N2 (at N1) (Figure 6.6A). Two plausible mechanisms follow that can account for the guanosine selectivity of DMAS-Cl. The nitrogen atom in DMAS-Cl may form a transient hydrogen bond with the N1 hydrogen bond donor in guanosine. This interaction would then function to position silicon for nucleophilic attack by the N2 group and also decrease the electron density on silicon to enhance its electrophilicity (Figure 6.6B). Alternatively, the N1 hydrogen may facilitate ionization of the chlorine leaving group and thereby enhance the electrophilicity of the silicon center (Figure 6.6C) [26]. Both models posit formation of a transient six-membered ring in conjunction with the pentacoordinate intermediate of the silyl S_N2 reaction (Figure 6.6).

Reaction at the guanosine N2 position by DMAS-Cl and kethoxal is clearly sensitive to different local features in RNA as evidenced by the distinct reactivity patterns for these two electrophiles. These reactivity differences can be rationalized by the distinctive features of carbon- versus silicon-based electrophilic chemistry. A carbonyl-based electrophile like kethoxal reacts via a strongly polar transition state and requires that nucleophilic attack occur from a precise orientation of 107° either above or below the plane of the carbonyl moiety [27]. The lack of correlation between reactivity and solvent accessibility for kethoxal is readily rationalized by considering that the rate-determining step in kethoxal chemistry, the nucleophilic attack of the N2 on the aldehyde carbonyl to form a tetrahedral transition state [28, 29] is sensitive not only to the accessibility of the N2 position but also on orientation of the carbon electrophile relative to the nucleophilic N2. In contrast, nucleophilic attack at silicon

occurs via bonding between lone pairs in the nucleophile and an empty d-orbital in silicon to form a stable pentacoordinate intermediate [20]. Silicon-based electrophiles can attract a nucleophile from many spatial orientations due to the ionic nature of the bonding interactions and the participation by silyl d-orbitals [30]. Thus, solvent accessibility plays a much larger role in governing the reactivity of DMAS-Cl (silicon electrophile) than for kethoxal (carbonyl electrophile).

These differences assume a critical importance in the practical analysis of RNA structure in solution. Protection from DMAS-Cl reactivity precisely reports those guanosine residues that are involved in tertiary interactions or are in the interior of the molecule. In contrast, kethoxal reactivity is complex: some guanosine residues that are involved in tertiary interactions remain reactive; whereas, other guanosine nucleotides that are not involved in higher-order interactions are unreactive (compare blue circles, Figures 6.3 and 6.5). We believe that DMAS-Cl may now be the best reagent choice, and advantageously replaces kethoxal, for analyzing RNA interactions involving the base pairing face of guanosine.

6.4 Experimental

6.4.1. Synthesis of Wild Type *Bacillus subtilis mgtE* aptamer domain (M-box) and inosine M-box RNAs. A DNA template for transcription of the *B. subtilis mgtE* M-box, inserted in the context of 5′ and 3′ flanking structure cassette sequences [7], was generated by PCR [1 mL; containing 20 mM Tris (pH 8.4), 50 mM KCl, 2 mM $MgCl_2$, 75 μM each dNTP, 115 nM each forward and reverse primer, 10 pM template, and 0.025 units/μL Taq polymerase; denaturation at 95 °C, 30 s, annealing at 55 °C, 30 s, and

elongation at 72 °C, 45 s; 30 cycles]. The PCR product was recovered by ethanol precipitation and resuspended 20 μL sterile H_2O. Transcription reactions for the native RNA (1 mL, 37 °C, 6 h) contained 50 mM HEPES (pH 8.0), 20 mM $MgCl_2$, 40 mM DTT, 2 mM spermidine, 0.01% Triton X-100, 2 mM each NTP, 10 μL PCR-generated template, 20 μL SuperRNaseIN (Ambion) and 0.1 mg/mL of T7 RNA polymerase. For the inosine M-box RNA, transcription reactions (1.5 mL, 37 °C, 4 h) contained 40 mM Tris (pH 8.0), 10 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 0.01% (v/v) Triton X-100, 4% (w/v) poly(ethylene) glycol 8000, 2 mM ATP, CTP, UTP, and ITP (Trilink), 10 μL PCR-generated template, and 0.1 mg/mL of T7 RNA polymerase. RNA products were purified by denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 7 M urea, 29:1 acrylamide:bisacrylamide), excised from the gel, and recovered by passive elution [0.5 M potassium acetate (pH 6.5), 1 mM EDTA, ~12 hrs, 4 °C)] and ethanol precipitation. The purified RNA (~3000 pmol) was resuspended in water (40 μM) and stored at –80 °C.

6.4.2. Structure selective N-silylation of RNA. RNA (5 pmol) in 6 μL sterile water was heated at 95 °C for 2 min, cooled on ice, treated with 3 μL of 3× folding buffer [333 mM NaCl, 333 mM HEPES (pH 8.0), 33.3 mM $MgCl_2$ (or no $MgCl_2$)], and incubated at 37 °C for 20 min. For reactions performed under denaturing conditions, the RNA was instead treated with 3 μL of 3× buffer (333 mM HEPES, pH 8.0) and incubated at 70 °C for 5 min. The RNA solution was treated with N,N-(dimethylamino)dimethylchlorosilane (DMAS-Cl, from Gelest; 1 μL, 600 mM in anhydrous DMSO) and allowed to react for 1 hour at 37 °C or 0.5 hour at 70 °C. No-reagent control reactions contained 1 μL DMSO. The reaction was quenched with an

equal volume (10 μL) of buffered methoxylamine (100 mM, pH 6.0; Sigma-Aldrich). Modified RNA was recovered by ethanol precipitation [90 μL sterile H_2O, 5 μL NaCl (5 M), 1 μL glycogen (20 mg/mL), 400 μL ethanol; 30 min at –80 °C] and resuspended in 10 μL sterile water.

6.4.3. RNA modification with kethoxal. RNA was refolded in buffer either containing or omitting $MgCl_2$, as described for the DMAS-Cl reaction. The RNA solution was then treated with kethoxal (1 μL, 20 mM in sterile water; from USB) and allowed to react for 5 min at 37 °C. No-reagent control reactions contained 1 μL H_2O. The reaction was quenched with an equal volume of boric acid (10 mM) followed by ethanol precipitation and resuspended in 10 μL sterile water.

6.4.4. Primer extension of modified RNA. The general procedure was that outlined previously [10]. Briefly, a fluorescently labeled DNA primer (5′ VIC- or NED-labeled GAA CCG GAC CGA AGC CCG; 3 μL, 0.3 μM) was annealed to the RNA (10 μL, from the previous step) by heating to 65 °C (6 min) and placing on ice. Reverse transcription buffer and Superscript III were added and reactions were incubated at 45 °C for 1 min; 52 °C for 20 min; and 65 °C for 5 min. Primer extension reactions were quenched with 4 μL of a 1:1 mixture containing EDTA (100 mM , pH 8.0) and sodium acetate (3 M, pH 5.2); the resulting cDNAs were recovered by ethanol precipitation, washed twice with 70% ethanol, vacuum dried for 10 min, and resuspended in 10 μL deionized formamide. Dideoxy sequencing markers were generated using unmodified RNA and primers labeled with unique fluorophores (6-FAM or PET, 0.6 μM), and by adding 1 μL 2′,3′-deoxythymidine (10 mM) or 2′,3′-dideoxyadenosine (10 mM) triphosphate after addition of reverse transcription buffer. cDNA extension products

were separated by capillary electrophoresis using an Applied Biosystems 3130 Genetic Analyzer.

6.4.5. Reaction of DMAS-Cl with 2´-deoxyguanosinemonophosphate (2´-dGMP). Equal volumes of DMAS-Cl (800 mM, DMSO) and 2′-dGMP (400 mM, H_2O) were combined in a 0.6 mL Eppendorf tube and allowed to react at 37 °C for ~3 hours. Solvent and excess DMAS-Cl were removed under reduced pressure at elevated temperatures using a SpeedVac Concentrator. The isolated white powder was resuspended in D_2O (Cambridge Isotopes Laboratories). NMR spectra were recorded with a Bruker 400 MHz DRX spectrometer.

^{1}H NMR (D_2O, 400 MHz, 298 K): δ = 7.94 (s, 1H, *H*CN_2), 6.25 (t, 1H, *H*CCNO), 4.43 (s, 1H, *H*CC_2O), 4.07 (s, 1H, *H*CC_2O), 3.84 (t, 2H, *H*CHCO), 2.74 (m, 1H, *H*CHC_2), 2.44 (m, 1H, *H*CHC_2), 2.06 (s, 6H, *H*CH_2N), -0.069 (s, 6H, *H*CH_2Si).

6.4.6. Data Analysis. Raw traces from the AB 3130 were processed using ShapeFinder [31]; reactivities are reported as a percentage of the most reactive nucleotide after subtracting background from the (–) reagent trace. Solvent accessibility calculations for the M-box RNA (PDB ID 2QBZ) [21] were performed with CNS [23] using probe sizes of 1.0, 1.2, 1.4, 1.6, 1.8, 2.4, 2.8, 3.5, 4.2, 5.0, 5.5, 6.0, 6.5, and 7.0 Å. M-box nucleotides 127 and 171 were excluded because they are adjacent to an undefined loop in the crystal structure and at the end of the RNA, respectively.

6.5 References

1. Tinoco, I. and Bustamante, C., *How RNA folds.* J. Mol. Biol., 1999. **293**: p. 271-281.

2. Gesteland, R.F., Cech, T.R., and Atkins, J.F., *The RNA World*. 3rd ed. 2004, Cold Spring Harbor, New York.

3. Leontis, N.B. and Westhof, E., *Analysis of RNA motifs.* Curr. Opin. Struct. Biol., 2003. **13**: p. 300-8.

4. Peattie, D.A. and Gilbert, W., *Chemical Probes for Higher-Order Structure in RNA.* Proc. Natl. Acad. Sci. U.S.A., 1980. **77**: p. 4679-4682.

5. Stern, S., Moazed, D., and Noller, H.F., *Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension.* Methods Enzymol., 1988. **164**: p. 481-9.

6. Ehresmann, C., et al., *Probing the Structure of RNAs in Solution.* Nucleic Acids Res., 1987. **15**: p. 9109-9128.

7. Merino, E.J., et al., *RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE).* J. Am. Chem. Soc., 2005. **127**: p. 4223-4231.

8. Wilkinson, K.A., Merino, E.J., and Weeks, K.M., *RNA SHAPE chemistry reveals non-hierarchical interactions dominate equilibrium structural transitions in tRNA^{Asp} transcripts.* J. Am. Chem. Soc., 2005. **127**: p. 4659-4667.

9. 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**: p. 1610-6.

10. 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**: p. 4144-4145.

11. Lavery, R. and Pullman, A., *A New Theorectical Index of Biochemical Reactivity Combining Steric and Electrostatic Factors.* Biophys. Chem., 1984. **19**: p. 171-181.

12. Tullius, T.D. and Dombroski, B.A., *Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein.* Proc. Natl. Acad. Sci. U.S.A., 1986. **83**: p. 5469-73.

13. Latham, J.A. and Cech, T.R., *Defining the Inside and Outside of a Catalytic RNA Molecule.* Science, 1989. **245**: p. 276-280.

14. Brenowitz, M., et al., *Probing the Structural Dynamics of Nucleic Acids by Quantitative Time-Resolved and Equilibirium Hydroxy Radical 'Footprinting'.* Curr. Opin. Struct. Biol., 2002. **12**: p. 648-653.

15. Tullius, T.D. and Greenbaum, J.A., *Mapping Nucleic Acid Structure by Hydroxy Radical Cleavage.* Curr. Opin. Struct. Biol., 2005. **9**: p. 127-134.

16. Balasubramanian, B., Pogozelski, W.K., and Tullius, T.D., *DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone.* Proc. Natl. Acad. Sci. U.S.A., 1998. **95**: p. 9738-43.

17. Cate, J.H., et al., *Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing.* Science, 1996. **273**: p. 1678-1685.

18. Adams, P.L., et al., *Crystal Structure of a Group I Intron Splicing Intermediate.* RNA, 2004. **10**: p. 1867-1887.

19. Wannagat, U. and Schreiner, G., *Aminochlorosilane und ihre Umsetzung mit Ammoniak und Methylamin.* Monatsh. Chem., 1965. **96**: p. 1889-1894.

20. Bassindale, A.R., Glynn, S.J., and Taylor, P.G., *Reaction mechanisms of nucleophilic attack at silicon*, in *The chemistry of organic silicon compounds*, Rappoport, Z. and Apeloig, Y., Editors. 1998, John Wiley & Sons Ltd: New York. p. 495-511.

21. Dann, C., et al., *Structure and Mechanism of a Metal-Sensing Regulatory RNA.* Cell, 2007: p. 878-892.

22. Leontis, N.B. and Westhof, E., *Geometric nomenclature and classification of RNA base pairs.* RNA, 2001. **7**: p. 499-512.

23. Brunger, A.T., et al., *Crystallography & NMR System (CNS), A New Software Suite for Macromolecular Structure Determination.* Acta Cryst.D, 1998. **54**: p. 905-921.

24. Mitzel, N.W. and Voijinovic, K., *The Crystal Structures of Chlorodimethyl(dimethylamino)silane and Dimethyl-bis-(dimethlyamino)silane.* Z. Naturforsch., B: Chem. Sci., 2003. **58b**: p. 708-710.

25. Staehelin, M., *Inactivationm of Virus Nucleic Acid with Glyoxal Derivatives.* Biochim. Biophys. Acta, 1959. **31**: p. 448-454.

26. Raheem, I.T., et al., *Enantioselective Pictet-Spengler-Type Cyclizations of Hydroxylactactams: H-bond Donor Catalysis by Anion Binding.* J. Am. Chem. Soc., 2007. **129**: p. 13404-13405.

27. Burgi, H.B., Dunitz, J.D., and Shefter, E.J., *Geometrical Reaction Coordinates. II. Nucleophilic Addition to a Carbonyl Group.* J. Am. Chem. Soc., 1973. **95**: p. 5065-5067.

28. Shapiro, R. and Hachmann, J., *The Reaction of Guanine Derivatives with 1,2-Dicarbonyl Compounds.* Biochemistry, 1966. **5**: p. 2799-2807.

29. Shapiro, R., et al., *On the Reaction of Guanine with Glyoxal, Pyruvaldehyde, and Kethoxal, and the Structure of the Acylguanines. A New Synthesis of N^2-Alkyguanines.* Biochemistry, 1969. **8**: p. 238-245.

30. Shi, Z. and Boyd, R.J., *The Laplacian of the charge density as a probe of reaction paths and reactivity: a comparison of S_N2 reactions at C and Si.* J. Phys. Chem., 1991. **95**: p. 4698-4701.

31. Vasa, S.M., et al., *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.