From Primary Sequence to Static and Dynamic RNA Tertiary Structures

Costin M. Gherghe

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

Advisor: Kevin M. Weeks

Reader: Linda Spremulli

Reader: Howard Fried

Abstract:

Costin M. Gherghe

From Primary Sequence to Static and

Dynamic RNA Tertiary Structures

(Under the direction of Kevin M. Weeks)

RNA molecules fold back on themselves to form complex secondary and tertiary structures. Knowledge of the three-dimensional fold as well as the underlying dynamics is essential to understand sequence-function relationships in RNA. A novel technique, called SHAPE chemistry, has been created for determining RNA secondary structure. In this work, I expand the applications of SHAPE chemistry in two directions: calculating opening rates for slow moving nucleotides and analysis of RNA dynamics at nucleotide resolution. First, I use SHAPE to identify a novel class of slow-moving nucleotides and calculate their opening rates. The major mechanism is that certain C2'-endo nucleotides have slow, seconds long, opening rates. Second, I find a strong correlation between SHAPE chemistry and molecular order as measured by ^{13}C NMR relaxation experiments. This finding validates SHAPE as an adequate substitute for ^{13}C NMR relaxation

experiments, expanding nucleotide-resolution dynamics analysis to large

RNAs, inaccessible to NMR methods. Finally, I develop a fast, fully

automated modeling method to determine three-dimensional structures of

RNAs. This method blends (1) SHAPE-derived secondary structure

information; (2) biochemical experiments that yield high quality tertiary

structure information with (3) a coarse-grained molecular dynamics

refinement. It therefore does not require any prior assumptions about

secondary or tertiary structure, establishing the foundation for modeling and

refining large RNAs inaccessible to crystallography and NMR techniques.

"The most exciting phrase to hear in science, the one that heralds new

discoveries, is not 'Eureka!' but 'That's funny...' "

Isaac Asimov

ACKNOLEDGEMENT

I want to thank my wife, Cristina, for always being there for me and her continuous support. I want to thank my daughter, Nadia, for always wanting me to be there for her. She is a continuous source of happiness. My family in Romania, thank you for supporting me. It seems that having an ocean between us got us closer than we ever were.

I am greatly honored to have met Kevin Weeks. He trusted me from the very beginning. He guided my career with extraordinary scientific insight and offered me the greatest gift of all: knowledge. Thank you!

I appreciate all the good time that I spent in the lab. That would have never been possible without a great group. I enjoyed playing pool and the happy hours on top of Kenan building as much as talking science with you guys. Thank you for everything!

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
Asp	aspartic acid
C	cytosine
Ci	curie
^{o}C	degree Celsius
DMD	discrete molecular dynamics
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiotreitol
EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol
G	guanosine
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
L	liter
Mg^{2+}	magnesium ion
$MgCl_2$	magnesium chloride
MiDAS	minimal dimerization active sequence
min	minute
mg	milligram
MPE	methidium propyl EDTA
mRNA	messenger RNA
ms	millisecond
MuSV	Moloney murine sarcoma virus
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
^{32}P	phosphorus, isotope
PAGE	polyacrylamide gel electrophoresis

PAL   palindrome

PCR   polymerase chain reaction

Phe   phenylalanine

pmol   picomole

ps   picosecond

RNA   ribonucleic acid

RNase   ribonuclease

s   second

SHAPE   selective 2'-hydroxyl acylation analyzed by primer extension

SL   stem loop

TBE   90 mM Tris-borate, 2 mM EDTA

TE   10 mM Tris (pH 7.5), 1 mM EDTA

TEMED   N, N, N', N', -tetramethylethylenediamine

Tris   tris(hydroxymethyl)aminomethane

tRNA   transfer RNA

U   uridine

v   volume

V   volt

w   weight

W   watt

Chapter 1

Towards detailed analysis of RNA dynamics

and three-dimensional RNA structure

1. Introduction

1.1.1 RNA in biology. RNAs are known to play key roles in the cellular life cycle [1]. RNAs carry amino acids (tRNAs) to RNA-based complex protein synthesizing machinery (ribosomes), serve as enzymes (ribozymes), or are involved in gene regulation (riboswitches or microRNAs). In order for RNAs to fulfill their biological roles, RNA forms secondary and tertiary interactions that intercommunicate to form complex networks [2], rendering a precise three-dimensional architecture. Following the initial transcription event, most RNAs first form a complex secondary structure, followed by achieving a three-dimensional, functional state. In order to elucidate RNA sequence-function relationships, a detailed analysis of the folding pathways, as well as interactions that define the three-dimensional functional state are required.

1.1.2 SHAPE chemistry and RNA secondary structure determination. One of the first steps towards a comprehensive understanding of the roles of any RNA in biology is elucidating the RNA secondary structure. Several biochemical methods have been developed in the past few decades that provide data necessary to develop empirically supported structural models. Biochemical footprinting techniques [3] using base modifying reagents, such as dimethyl sulfate [4], kethoxal [5], diethyl pyrocarbonate [4] or selective enzymatic degradation [6] were central to RNA secondary structure interrogation just a few years ago. However, these methods often require the use of multiple reagents to probe different bases in

RNA or are not completely selective towards single stranded nucleotides. A novel method for RNA secondary structure probing, called Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) addresses many of the drawbacks of the traditional RNA secondary structure determination approaches [7]. In contrast with the classic methods for RNA secondary structure determination, SHAPE is a comprehensive method that interrogates, in a single experiment, RNA flexibility at nucleotide resolution, independent of the residue identity [7].

SHAPE chemistry is based on the observation that flexible, unpaired nucleotides preferentially react at the generic 2'-hydroxyl position with selected electrophilic reagents; whereas, nucleotides involved in base pairing or tertiary interactions are unreactive (Figure 1.1). SHAPE chemistry thus offers an impressive amount of information about nucleotide flexibility at every position in RNA in a single experiment.

In order to be able to infer the secondary structure from raw SHAPE data, SHAPE reactivity information is converted to a pseudo-energy function in an RNA secondary structure prediction program, called RNAstructure [8]. SHAPE information embedded in RNAstructure software allows a significant improvement in the accuracy of secondary structure prediction, especially for large RNAs. For example, secondary structure prediction of 16S rRNA using SHAPE chemistry was over 90% accurate when compared to the predicted phylogenetic structure [9]. If SHAPE chemistry is excluded from the secondary structure prediction algorithm, the accuracy drops to 50%.
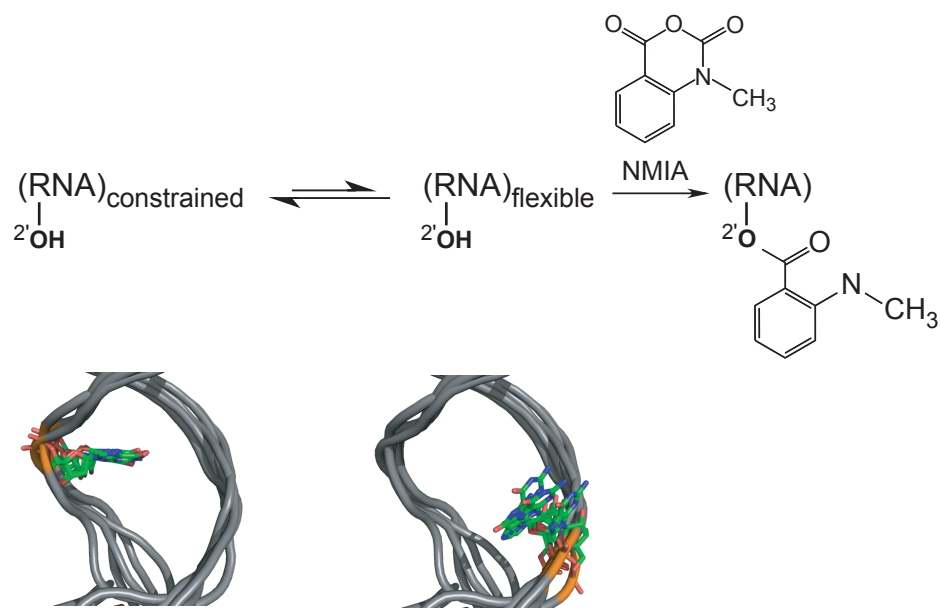
Figure 1.1. SHAPE chemistry scheme. SHAPE chemistry reagents, such as NMIA (N-methyl isatoic anhydride), selectively react with RNA nucleotides that are flexible. Nucleotides that are constrained are not able to achieve conformations reactive towards electrophiles.

Melding SHAPE chemistry with RNA secondary structure determination programs has been used to accurately map the secondary structure on a wide variety of RNAs, such as tRNA[Asp] [10], RNAse P specificity domain [11], dimerization domains from several murine retroviruses [12, 13] and even the entire 9000+ nucleotide HIV genomic RNA [14].

1.1.3 RNA tertiary structure determination. In order to gain insight into three-dimensional molecular structural details, the current gold standard methods are X-ray crystallography and NMR spectroscopy. Compared to proteins, RNA molecules are often difficult to crystallize. This is due to the inherent flexibility of RNA and the negatively charged backbone that hampers the formation of RNA-RNA crystal contacts [15]. Similar to crystallography, NMR methods offer impressive structural details, but are limited to small RNAs, generally 50 nucleotides or less. In order to overcome these limits, computational three-dimensional modeling approaches have facilitated RNA tertiary structure determination for small and large RNAs [16].

Several computational algorithms for building three-dimensional RNA structures have been developed and evaluated in recent years [16]. Some difficult algorithms take on an enormous challenge and predict *de novo* three-dimensional structures based on the primary sequence only [16, 17]. Importantly, these approaches require protocols for secondary structure determination prior to tertiary fold modeling. Success is generally limited to short RNA sequences that are composed of a small number of stems and loops. *De novo* three-dimensional structure prediction of large RNAs still has

5

important barriers to overcome, especially because these algorithms must parse high numbers of possible secondary as well as tertiary solutions. Faster, more facile, approaches can use as primary input one pre-determined RNA secondary structure, obtained either via phylogenetic analyses or biochemical experiments. In addition to secondary structure information, experimentally-derived instructive pair-wise tertiary constraints can be employed as long-range RNA-RNA interactions.

One of the classic ways to build and refine RNA structures relies on all atom modeling. This is a complex and computationally intensive process. All atom modeling is highly successful [18] in modeling small RNAs, generally less than 100 nucleotides. To overcome the intensive computational requirements that all atom modeling requires, especially for large RNAs, novel, coarse-grained approaches have been devised [19, 20]. Central to the coarse-grained methods is replacing all atoms from RNA bases either with one large sphere [19], or three spheres that replace the phosphate, sugar and base positions [17]. This approximation greatly reduces the computation requirements, thus allowing refinement of large RNAs over long time periods otherwise inaccessible to classic all atom modeling.

1.1.4 RNA dynamics at nucleotide resolution. Besides RNA secondary or tertiary structural determination, I am interested in elucidating the dynamics of nucleotides in RNA. Knowledge of RNA dynamics is central in exploring RNA-RNA and RNA-protein interactions and local and global structural rearrangements that occur while RNAs perform their biological

functions. Currently, the only method able to achieve insight into RNA dynamics at nucleotide resolution is NMR spectroscopy. Information on RNA base opening events, on ns-ps time scales, are obtained from measurements of exchange rates of imino protons using NMR [21]. Additionally, nucleotide dynamics on a longer, ms time are achieved via ^{13}C relaxation experiments [22]. Such methods were successfully applied to describe RNA dynamics in biology for several RNAs, such as the sarcin-ricin domain RNA from the large ribosomal subunit [21], U1A protein binding RNA [23, 24] or TAR RNA region from the 5' untranslated end of the HIV [25].

    1.1.5 Research overview. SHAPE chemistry has been extensively described and characterized in previous work in Weeks laboratory. It is an essential tool for RNA secondary structure prediction that appears to be poised to replace most of the traditional techniques. The focus of my research is to develop novel applications of SHAPE chemistry (Figure 1.2) to: (1) directly calculate RNA base opening rates; (2) describe RNA dynamics at nucleotide resolution and (3) create experimental and computational algorithms for three-dimensional RNA structure determination. In Chapter 2, I show that by using SHAPE chemistry I was able to identify, for the first time, a distinct class of C2'-endo nucleotides that have opening rates on the time scale of seconds. The discovery of slow RNA nucleotide motions directly challenges the commonly accepted belief that all RNA nucleotides display fast dynamics over ps-ms timescales. In Chapter 3, I find a strong correlation between SHAPE reactivity and NMR relaxation parameters that describe

7

5'-GAG ACU AUC GAC AUU UGA UAC ACU
AUU UAU CAA UGG AUG UCU C-3'

primary sequence

SHAPE

```
              U        G A          U          C A
5' G A G A C   A U C      C A U U  G A U A         C
   | | | | |   | | |      | | | |  | | | |          U
3' C U G U C   U A G      G U A A  C U A U          A
                                          U U
```
secondary structure

(3) SHAPE
biochemically-derived pair-wise information

(1) identify
slow moving
nucleotides

SHAPE

nucleotide-resolution
dynamics
(2) analysis

tertiary
structure
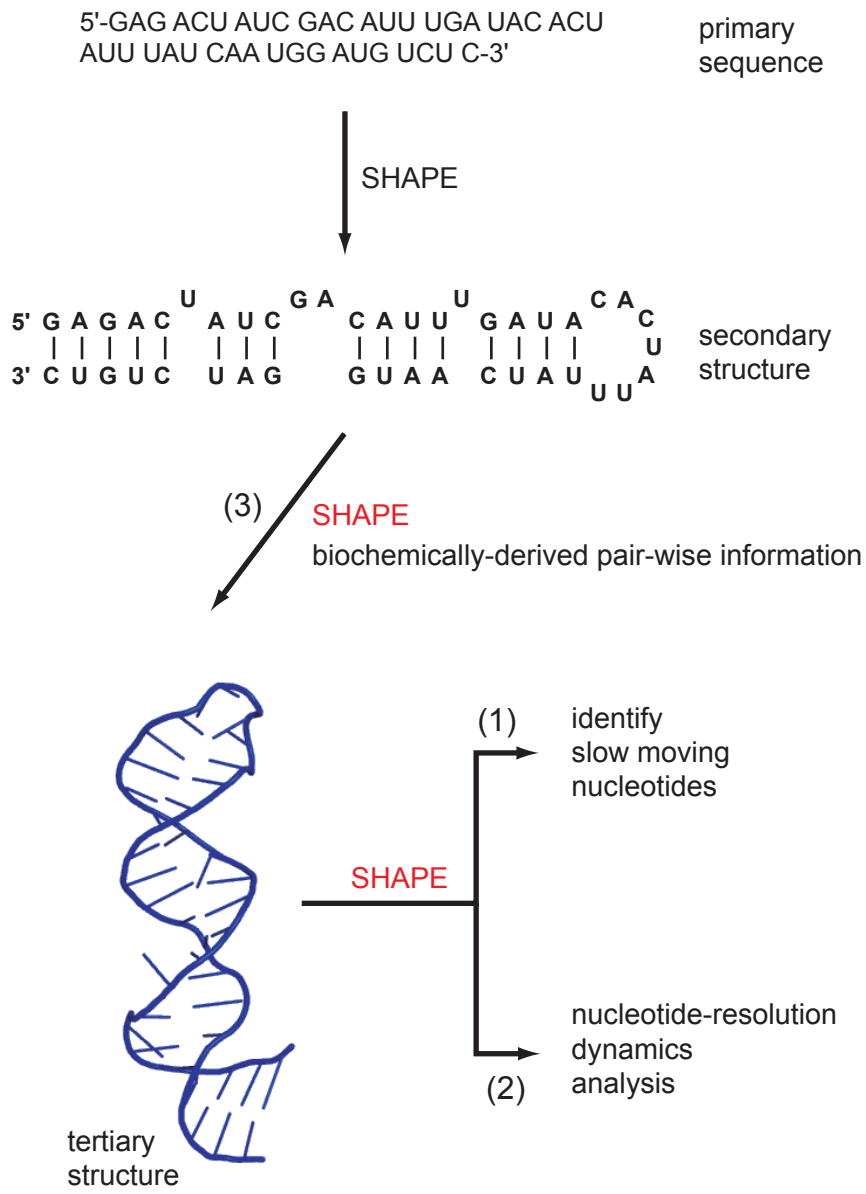
Figure 1.2. Outline of SHAPE chemistry applications. Besides classic secondary structure determinations, SHAPE chemistry can now be utilized to: (1) identify slow moving nucleotides in RNA, (2) analyze RNA dynamics at nucleotide resolution and (3) generate 3D structural information in concert with experimentally-derived tertiary information.

local nucleotide motions. I, thus, validate SHAPE chemistry as a tool for accurate description of RNA motion at nucleotide resolution.

Finally, in Chapters 4 and 5, via the use of two different methods, I blend biochemical methods that obtain long-range tertiary information of folded RNAs with three-dimensional modeling programs. I first refine the SL1-SL2 domain of the minimal dimerization active structure from Moloney Murine Sarcoma virus using all atom modeling. I then develop a much faster approach that embeds experimentally-derived, pair-wise long range constraints with a coarse-grained three-dimensional modeling program, DMD (Discrete Molecular Dynamics) [17]. I validate this approach by refining the *Saccharomyces cerevisiae* tRNA^{Asp}, to a 3.5 Å RMSD as compared to the crystal structure.

1.1.6 Perspective. In this work I explore the interface between SHAPE chemistry and nucleotide-resolution RNA dynamics analysis. I show that SHAPE can be used to describe nucleotide motions, especially for RNAs that are inaccessible to NMR relaxation methods. SHAPE chemistry has the advantage of rapidly and accurately analyzing RNAs of arbitrary length and complexity in diverse chemical and physical environments. I have also created a general approach to rapid and accurate three-dimensional RNA structure determination. I used this approach to refine tRNA^{Asp} to 3.5 Å RMSD as compared to the native crystal structure. I postulate that this novel computational method can be further expanded to refine large, flexible RNAs, unsuitable for X-ray crystallography or NMR techniques.

1.2 References

1.   Gesteland, R.F., T.R. Cech, and J.F. Atkins, *The RNA world.* 2006, Cold Spring Harbor: Cold spring harbor laboratory press.

2.   Moore, P.B., *Structural Motifs in RNA.* Annu. Rev. Biochem., 1999. **68**(1): p. 287-300.

3.   Ehresmann, C., et al., *Probing the structure of RNAs in solution.* Nucl. Acids Res., 1987. **15**: p. 9109-28.

4.   Peattie, D.A. and W. Gilbert, *Chemical probes for higher-order structure in RNA.* Proc. Natl. Acad. Sci. USA, 1980. **77**(8): p. 4679-82.

5.   Shapiro, R., et al., *On the reaction of guanine with glyoxal, pyruvaldehyde, and kethoxal, and the structure of the acylguanines. A new synthesis of N2-alkylguanines.* Biochem., 1969. **8**(1): p. 238-45.

6.   Knapp, G., Meth. Enzymol., 1989. **180**: p. 192-212.

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**(12): p. 4223-4231.

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

9.   Deigan, K.E., et al., *Accurate SHAPE-constrained RNA structure prediction.* in preparation, 2008.

10.   Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transition in tRNA(Asp) transcripts.* J. Am. Chem. Soc., 2005. **127**(13): p. 4659-4667.

11.   Mortimer, S.A. and K.M. Weeks, *A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry.* J. Am. Chem. Soc., 2007. **129**(14): p. 4144-4145.

12.   Gherghe, C. and K.M. Weeks, *The SL1-SL2 (Stem-Loop) domain is the primary determinant for stability of the gamma retroviral genomic RNA dimer.* J. Biol. Chem., 2006. **281**(49): p. 37952-37961.

13. Badorrek, C.S., C.M. Gherghe, and K.M. Weeks, *Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization.* Proc. Natl. Acad. Sci. USA, 2006. **103**(37): p. 13640-13645.

14. Watts, J.M., et al., *The structure of an entire HIV-1 genome.* in preparation, 2008.

15. Ke, A. and J.A. Doudna, *Crystallization of RNA and RNA-protein complexes.* Methods, 2004. **34**(3): p. 408-414.

16. Shapiro, B.A., et al., *Bridging the gap in RNA structure prediction.* Curr. Opin. Chem. Biol., 2007. **17**(2): p. 157-165.

17. Ding, F. and N.V. Dokholyan, *Large scale simulations of 3D RNA folding by Discrete Molecular Dymanics: from structure prediction to folding mechanisms.* RNA, *in press*, 2008.

18. Major, F., D. Gautheret, and R. Cedergren, *Reproducing the three-dimensional structure of a tRNA molecule from structural constraints.* Biochem., 1993. **90**: p. 9408-9412.

19. Wang, R., et al., *Three-dimensional placement of the conserved 530 loop of 16 S rRNAand of its neighboring components in the 30 S subunit.* J. Mol. Biol., 1999. **286**(2): p. 521-40.

20. Ding, F. and N.V. Dokholyan, *Simple but predictive protein models.* Trends Biotech., 2005. **23**(9): p. 450-455.

21. Chen, C., et al., *Structural energetics and base-pair opening dynamics in sarcin-ricin domain RNA.* Biochem., 2006. **45**(45): p. 13606-13613.

22. Shajani, Z. and G. Varani, *NMR studies of dynamics in RNA and DNA by ^{13}C relaxation.* Biopolymers, 2007. **86**(5-6): p. 348-359.

23. Shajani, Z., G. Drobny, and G. Varani, *Binding of U1A protein changes RNA dynamics as observed by 13C NMR relaxation studies.* Biochem., 2007. **46**(20): p. 5875-5883.

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

25. Shajani, Z., P. Deka, and G. Varani, *Decoding RNA motional codes.* Trends Bioch. Sci., 2006. **31**(8): p. 421-424.

Chapter 2

Slow conformational dynamics at C2'-endo nucleotides in RNA

2.1 Introduction

Local and global dynamics in folded RNAs occur over broad timescales spanning picoseconds to minutes [1, 2]. Slow motions may play important roles in governing RNA folding and ribonucleoprotein assembly reactions. Currently, slow local motions are extremely difficult to detect.

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 reacts with electrophilic reagents to form 2'-*O*-adducts (Figure 2.1).

The chemical reactivity at the 2'-ribose position is strongly influenced by the adjacent 3'-phosphodiester anion. Attack of the 2'-OH position by an electrophilic reagent requires the formation of a significantly more reactive 2'-oxyanion. Formation of the 2'-oxyanion is severely impaired if the negatively charged 3'-phosphodiester is adjacent to the 2' position, such as in constrained nucleotides. Flexible nucleotides more readily achieve states where the 2'-oxyanion is formed, therefore reacting easily towards electrophilic reagents, forming bulky, 2'-*O*-adducts. SHAPE reagents undergo a parallel self-quenching hydrolysis with water, thus SHAPE reactions are considered finished after consumption of more than 99% of the reagent, or 5 hydrolysis half lives ($t_{1/2}$).

Positions that form 2'-*O*-adducts are detected by primer extension [3-

$$\text{(RNA)}_{\text{constrained}} \underset{k_{\text{close}}}{\overset{k_{\text{open}}}{\rightleftharpoons}} \text{(RNA)}_{\text{flexible}} \xrightarrow{k_{\text{adduct}}} \text{(RNA)}$$

2'OH   2'OH   2'O
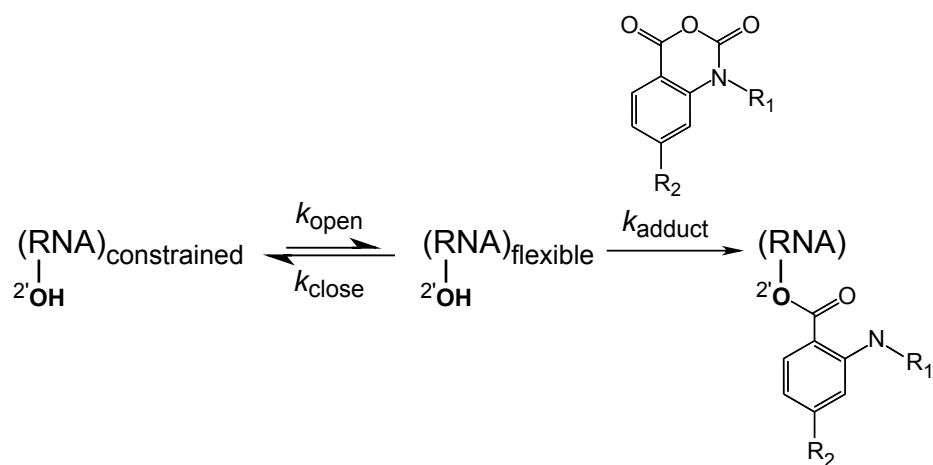
Reagent	hydrolysis $t_{1/2}$ (s)	R1	R2	
1M7	14	CH_3	NO_2	fast
4NIA	101	H	NO_2	
NMIA	260	CH_3	H	
IA	433	H	H	slow

Figure 2.1 Mechanistic framework for RNA SHAPE chemistry

14

5]. A DNA primer is annealed to the 3' end of the RNA, allowing cDNA synthesis using a reverse transcription reaction. When sites of adduct formation are encountered during the primer extension step, reverse transcriptase stops and dissociates from the RNA, ending DNA synthesis. This process allows formation of a cDNA library that accurately reports positions in RNA where an electrophile reacted with flexible nucleotides as well as the degree of flexibility. In order to facilitate the analysis of the entire RNA of interest and offer a stable binding site for the DNA primer, a small structure cassette is placed at the 3'-end of the RNA (Figure 2.2).

SHAPE experiments work well using electrophiles based on the isatoic anhydride (IA) scaffold [3, 6]. IA derivatives both react with the RNA 2'-OH group and also undergo concurrent degradation by hydrolysis (Figure 2.1). The reactivity of the 2'-hydroxyl 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, as described in the experimental section:

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

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

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

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

These relationships lead to two limits. In Limit 1, $k_{close} \gg k_{adduct}[reagent]$,
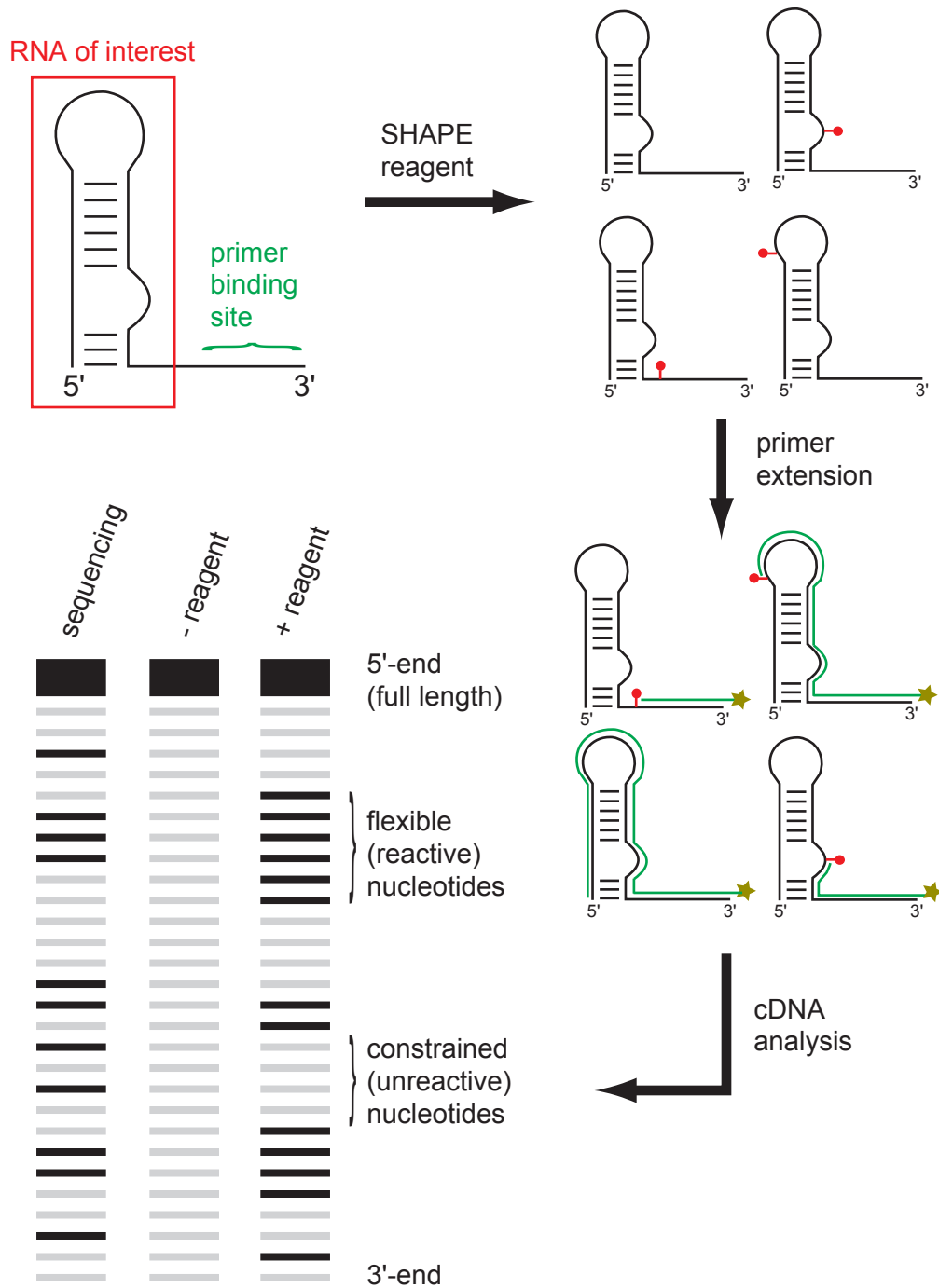
Figure 2.2. SHAPE chemistry experimental overview. The RNA of interest, containing the 3'-end structure cassette, is treated first with the SHAPE reagent, modifying preferentially flexible nucleotides. Sites of modification are analyzed using primer extension based on a radiolabeled DNA primer. The generated cDNA library is then separated using denaturing gel electrophoresis.

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

For Limit 2, $k_{\text{close}} << k_{\text{adduct}}[reagent]$,

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

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

It should, therefore, be possible to monitor local nucleotide dynamics in RNA under conditions where Limit 2 applies by varying the reactivity (or $k_{\text{hydrolysis}}$) of the hydroxyl-selective electrophile. IA has a hydrolysis half-life ($t_{1/2}$) of 430 s at 37 °C (table, Figure 2.1). 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-nitroisatioc anhydride (4NIA) and 1-methyl 7-nitroisatoic anhydride (1M7) [6] have progressively shorter hydrolysis half-lives (Figure 2.1).

2.2 Results

2.2.1 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 catalytically 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]. We 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 C2'-endo nucleotides (in red, Figure 2.3A).

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 2.3B). 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 to the 1M7 experiment, the four G nucleotides that adopt the C2'-endo conformation were highly reactive, even more so than some nucleotides in the flexible loop (Figure 2.3B, 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 2.3B). Nucleotides constrained in these C2'-endo conformations are thus unreactive towards fast-reacting electrophiles but are highly reactive towards the slower reagents (Figure 2.3B and Figure 2.4).

Figure 2.3 (A) The C2'-endo RNA construct containing tandem GA basepairs that adopt either C2'- (red) or C3'- (black) endo ribose conformation. (B). Absolute SHAPE reactivities as a function of reagent electrophilicity. Red bars represent SHAPE reactivities at the C2'-endo positions.

19
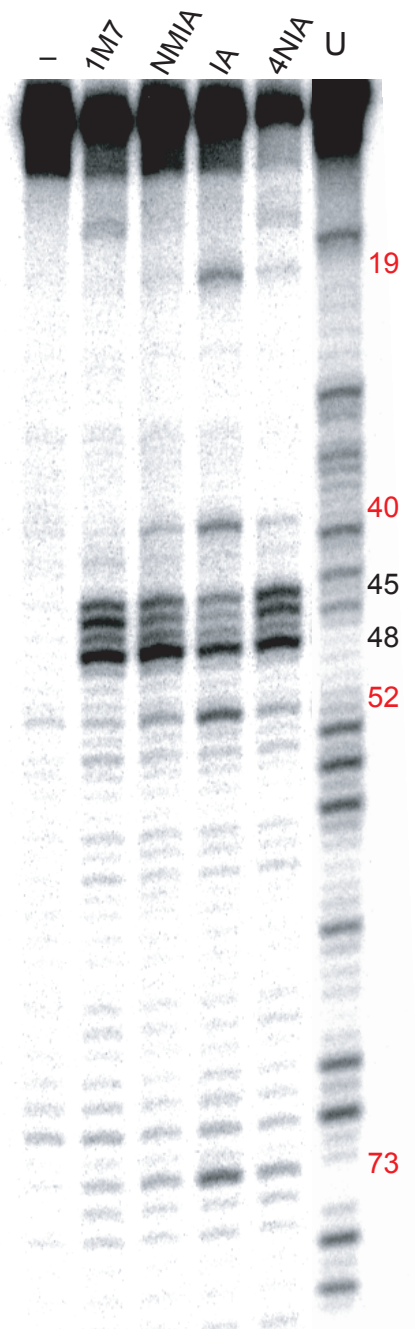
Figure 2.4. SHAPE experiment for the C2´-endo RNA as a function of reagent electrophilicity. (-) is the no reagent control. U, dideoxy sequencing marker. Positions marked with red correspond to the C2'-endo nucleotides.

Limit 1 (Eqn. 4) predicts that adduct formation is a function of the equilibrium constant for formation of the open state (k_{open}/k_{close}) and is independent of the reagent hydrolysis rate, as has been observed for most nucleotides thus far analyzed by SHAPE, including in tRNA [3-5] and in an RNase P RNA [6].

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 2.3A. 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). We 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 2.5).

We therefore 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 2.6). 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 2.7). NMR experiments show that the C2'-endo conformations at the G•A pairs illustrated by positions 19/73 and 40/52 are stable [9-11], indicating that k_{close} is comparable to or smaller than k_{open}.
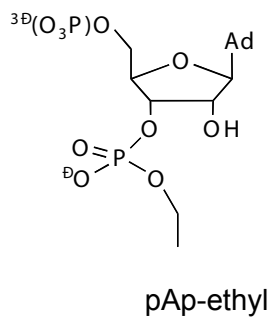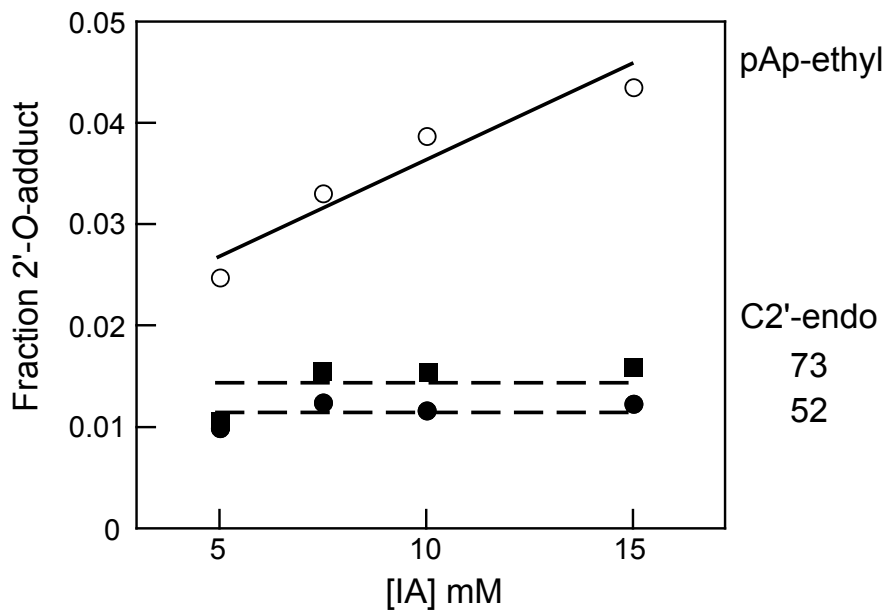
Figure 2.5. Concentration dependence for reaction at positions 52 and 73 in the C2'-endo RNA construct and for the (unconstrained) model nucleotide pAp-ethyl.
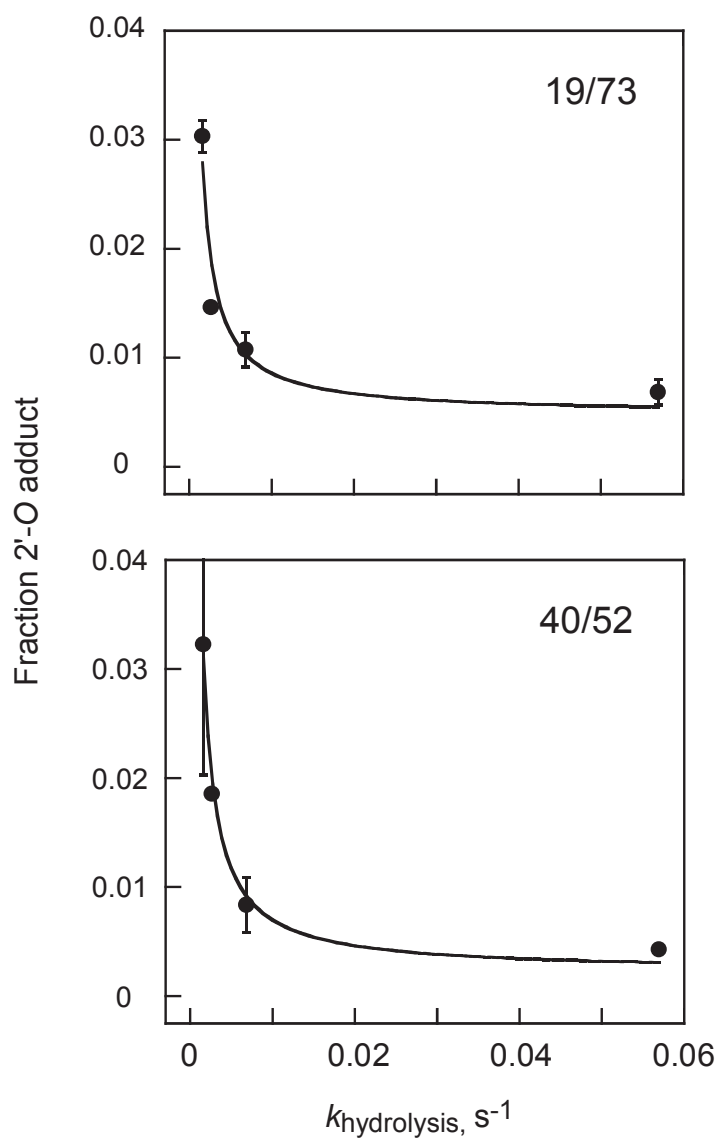
Figure 2.6. Determination of k_{open} for C2'-endo position 19/73 and 40/52. Lines represent a fit to Eqn. 5; rate constants are $\sim 4*10^{-5}$ s^{-1}. Error bars indicate standard deviations from two independent measurements.
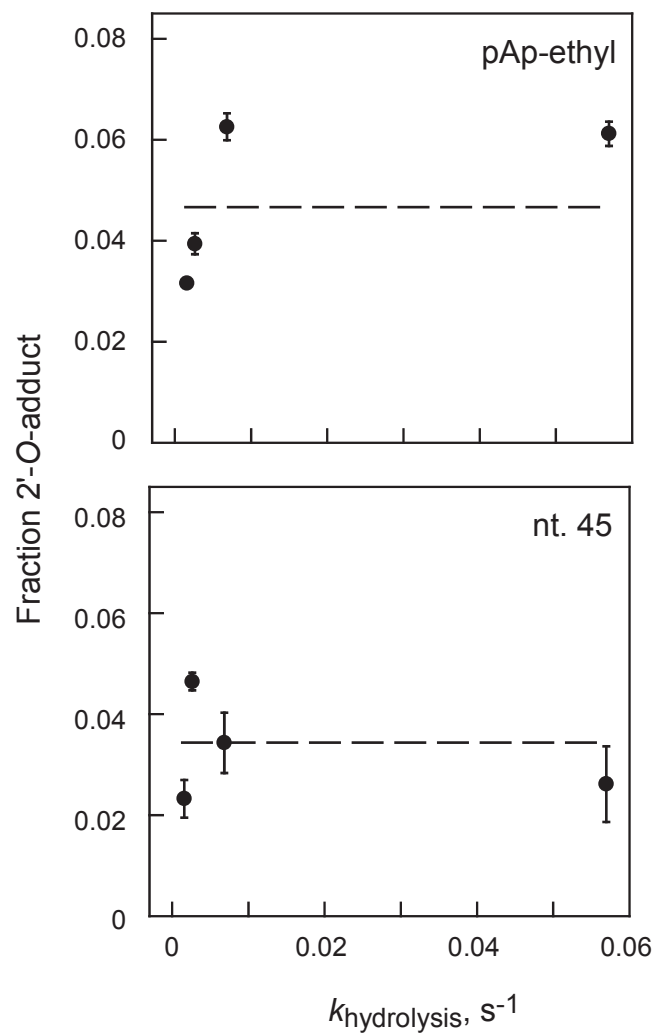
Figure 2.7 Absence of a dependence of fraction adduct formed as a function of $k_{hydrolysis}$ for reaction of the (unconstrained) model nucleotide pAp-ethyl and for nucleotide 45 in the loop of the C2'-endo construct.

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

2.2.2 Identification of slow-moving nucleotides in RNase P. We next explored whether the differential reactivity between 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 2.8).

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 2.8). These nucleotides reflect Limit 1. The 10 well defined C2'-endo nucleotides fell into three categories: (i) most C2'-endo nucleotides are highly constrained and, as expected,[3-5] unreactive towards both reagents (blue nucleotides, Figure 2.8A), (ii) one nucleotide is not constrained and is reactive towards both electrophiles (red nucleotide, Figure 2.8A), and (iii) two C2'-endo positions show large changes in reactivity (nts 130 and 168, circled nucleotides and red bars, Figure 2.8). 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 2.8). A similar distribution of reactive and
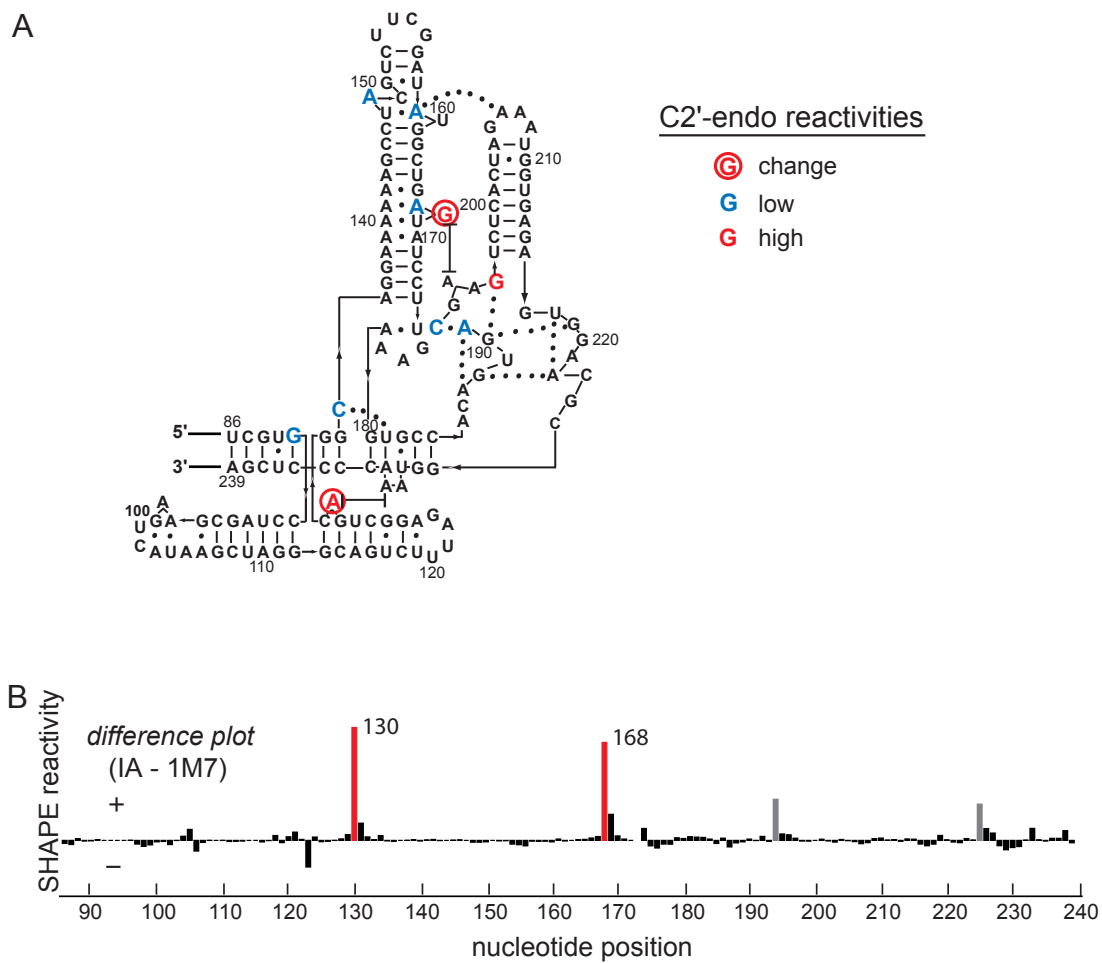
Figure 2.8 SHAPE reactivities at C2'-endo nucleotides in the specificity domain of RNase P. (A) Secondary structure showing SHAPE reactivities at the 10 well defined C2'-endo nucleotides. (B) Differential SHAPE reactivity for the slow (IA) versus fast (1M7) reagents.
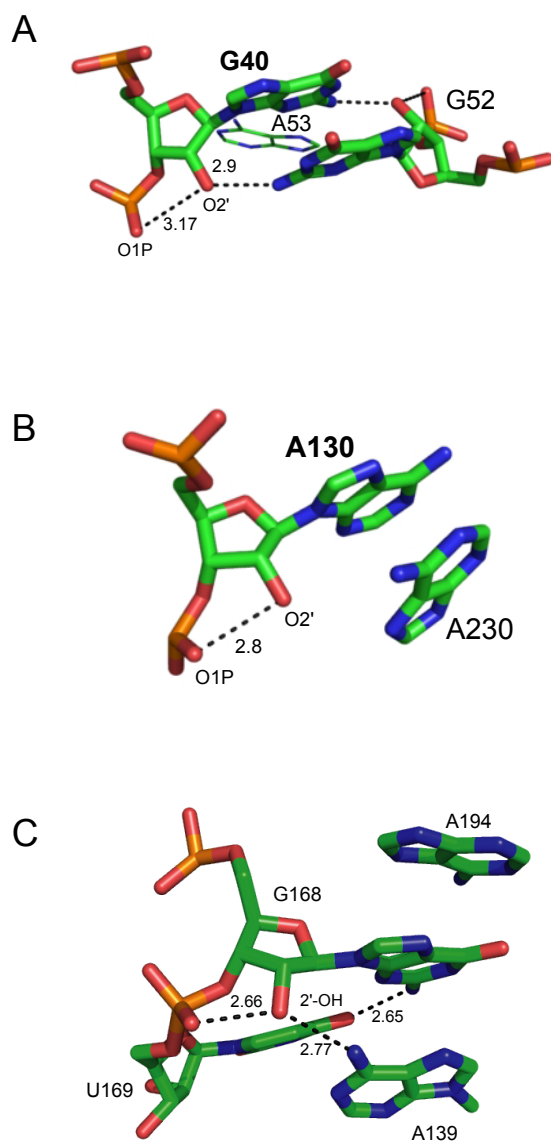
Figure 2.9 Base stacking and hydrogen bonding interactions at C2'-endo nucleotides that undergo slow conformational changes. (A) G40 from the C2'-endo RNA construct and (B) A130 and (C) G168 in the RNase P RNA. Distances are in angstroms.

unreactive C2'-endo nucleotides occurs in the *Tetrahymena* P5-P4-P6 domain using NMIA [13].

2.3 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 dynamics. These nucleotides in both the simple C2'-endo RNA (Figure 2.3B) and in the RNase P RNA (Figure 2.8) share key characteristics: (1) the ribose group has the C2'-endo conformation and (2) the nucleotide participates in base stacking and additional hydrogen bonding interactions (Figure 2.9). These C2'-endo dynamics are orders of magnitude slower than for other local conformational changes like base opening reactions[14, 15] and are slower than RNA 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 10^{-5} s^{-1} 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 molecular switches, they may govern certain specific conformational changes in RNA, and may play

important, but currently unexplored, roles in RNA folding, ligand recognition, and catalysis.

2.4 Experimental

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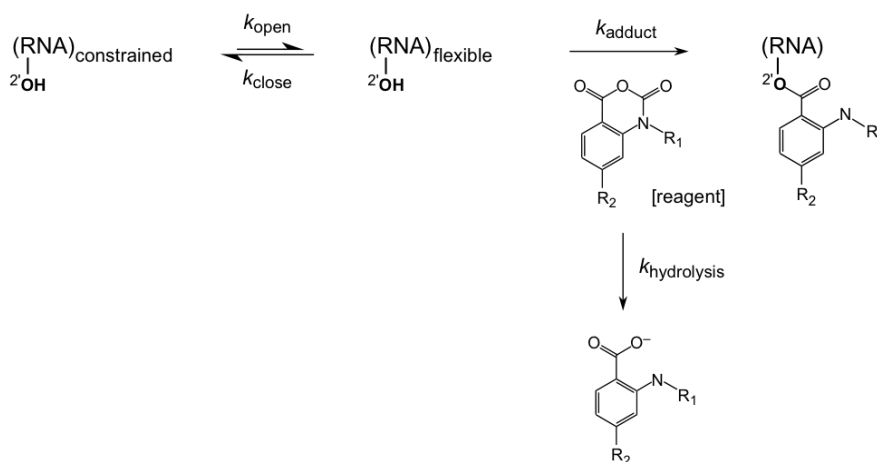
       Scheme 1

Taking

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

then,[3]

$$f = 1 - e^{-\frac{k_{obs}}{k_{hydrolysis}}(1 - e^{-k_{hydrolysis}t})}$$

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 ($t \to \infty$), this equation simplifies to:

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

as given in Eqn. 1.

2.4.2. IA, NMIA, 4NIA and 1M7 hydrolysis. Hydrolysis experiments were performed by Stefanie A. Mortimer. 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-NaOH (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).

2.4.3. RNA constructs. The C2'-endo hairpin RNA (Figure 1) and the specificity domain of the RNase P RNA[12] were synthesized by *in vitro* transcription using a single stranded DNA (Integrated DNA Technologies) 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-HCl (pH 8.0), 1 mM EDTA] at concentrations of about 30 μM and stored at -20 °C.

2.4.4. SHAPE analysis. pAp-ethyl was 5'-end radiolabeled using γ-[^{32}P]-ATP, purified by denaturing gel electrophoresis, and resuspended in 1/2× TE [TE is 10 mM Tris, pH 7.5, 1 mM EDTA]. The pAp-ethyl (1 μL, 10000 cpm) was heated to 95 °C for 2 min, cooled on ice, treated with 3 μL of 3.3× folding buffer [264 mM NaCl, 66 mM Hepes-NaOH (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 the primer extension step. Modified RNA was recovered by ethanol precipitation [90 μL sterile H$_2$O, 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-NaOH (pH 8.0), 333 mM NaCl, 33.3 mM MgCl$_2$.

2.4.5. Primer Extension. For the C2'-endo containing hairpin RNA, 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. All RNase P RNA experiments were performed by Stefanie Mortimer. 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 fluorophores for the (–) and (+) reagent

experiments; dideoxy sequencing markers were generated using unmodified RNA and primers labeled with 6-FAM or PET fluorophores. cDNA extension products were separated by capillary electrophoresis using an Applied Biosystems 3130 Genetic Analyzer capillary electrophoresis instrument.

2.4.6. 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 [5]. Fraction adduct formation at each position was calculated as the fraction of the band intensity divided by the full-length band. Reaction lanes were normalized to intensities at position 80, a flexible nucleotide outside the RNA of interest. Least square fitting to Eqn. 5 was performed with Kaleidagraph (ver. 4.01). For the RNase P RNA, raw traces from the ABI 3130 were processed 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 [20].

2.4.7. Refinement of RNase P Structure. RNase P was refined by Joseph M. Krahn (NIEHS, NIH, Research Triangle Park, NC). Electron density maps from the deposited 3.1 Å resolution RNase P structure (model and structure factors from PDB ID 1NBS) [12] contained large difference peaks in the |Fo|–|Fc| maps, consistent with ordered Sr^{2+} ions, which were a component of the crystallization solution. Many of these sites had been modeled as water. Replacing these waters with strontium significantly

improved both the crystallographic R-factors and the map interpretability. Four chloride atoms were also added to the model based on charge considerations. The bulk solvent mask parameters were then adjusted to account for the high concentrations of strongly diffracting ions. 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 the overall noise in the structure. To determine the ribose conformation for each nucleotide in the structure, the new model was refined using the annealing function in CNS 1.1 [21] with the dihedral restraints related to sugar pucker geometries disabled. The revised map was used to make initial estimates for the sugar pucker geometry at each nucleotide, with two observations for most positions (one for each of the two copies of RNase P in the asymmetric unit). 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. Analysis of RNA rotamers indicates that RNA nucleotides exist almost exclusively in either the C2'-endo or C3'-endo conformations [22]. Although ribose geometries are not discernable directly in electron density maps at this resolution, the combined information from the geometries of ring substituents and the restricted choice of sugar puckers was sufficient to assign most nucleotides. After several rounds of iterative refinement, most nucleotides could be assigned to the C2'-endo or C3'-endo geometry with confidence. In a final step to improve model geometries in regions with van der Waals

clashes, the model was refined with all hydrogen atoms included using REFMAC5 [23]. Improvement to the free R-factor was used to guide the steps outlined above; the original test set from the deposited structure factors was retained in all R-free [21] calculations. The working and free R-factors of 28.0 and 30.7% from the original model were improved to working and free R-factors of 17.8 and 24.1%, respectively.

2.5 References

1. Shajani, Z. and G. Varani, *NMR studies of dynamics in RNA and DNA by ^{13}C relaxation.* Biopol., 2007. **86**: p. in press.

2. Buchmueller, K.L., et al., *RNA-Tethered Phenyl Azide Photocrosslinking via a Short-Lived Indiscriminant Electrophile.* J. Am. Chem. Soc., 2003. **125**: 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., E.J. Merino, and K.M. Weeks, *RNA SHAPE chemistry reveals non-hierarchical interactions dominate equilibrium structural transitions in tRNA^{Asp} transcripts.* J. Am. Chem. Soc., 2005. **127**: p. 4659-4667.

5. Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution.* Nature Protocols, 2006. **1**: p. 1610-1616.

6. Mortimer, S.A. and K.M. Weeks, *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.

7. Fersht, 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. 465-481.

9. Santalucia, J.J. and D.H. Turner, *Structure of (rGGCGAGCC)$_2$ in solution from NMR and restrained molecular dynamics.* Biochemistry, 1993. **32**: p. 12612-12623.

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. 147-158.

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.* Biochem., 2007. **46**(6): p. 1511-1522.

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

13. Vicens, Q., et al., *Local RNA structural changes induced by crystallization are revealed by SHAPE.* RNA, 2007. **13**: 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**: p. 915-39.

15. Spies, M.A. and R.L. Schowen, *The trapping of a spontaneously "flipped-out" base from double helical nucleic acids by host-guest complexation with b-cyclodextrin: The intrinsic base-flipping rate constant for DNA and RNA.* J. Am. Chem. Soc., 2001. **124**: p. 14049 - 14053.

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

17. Grathwohl, C. and K. Wuthrich, *NMR studies of the rates of proline cis-trans isomerization in oligopeptides.* Biopolymers, 1981. **20**(2623-33).

18. Schoetz, G., O. Trapp, and V. Schurig, *Determination of the cis-trans isomerization barrier of several L-peptidyl-L-proline dipeptides by dynamic capillary electrophoresis and computer simulation.* Electrophoresis, 2001. **22**: 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**: p. 344-354.

20. DeLano, W.L., *The Pymol Molecular Graphics System.* DeLano Scientific, South San Francisco, CA, USA. http://www.pymol.org, 2005.

21. Brunger, A.T., et al., *Crystallography and NMR system (CNS): A new software system for macromolecular structure determination.* Acta. Cryst., 1998. **54**: p. 905-921.

22. Murray, L.J., et al., *RNA Backbone is rotameric.* Proc. Natl. Acad. Sci. USA, 2003. **100**: p. 13904-13909.

23. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Cryst., 1997. **D53**: p. 240-255.

Chapter 3

Strong correlation between SHAPE chemistry and the generalized order parameter (S^2) in RNA

3.1 Introduction.

RNA molecules perform important cellular functions that depend on the ability to form both rigid and dynamic structural elements and that often require large changes in conformation and motion. Critical examples include ribosomal protein synthesis, ribonucleoprotein assembly, and riboswitches [1-5]. The intrinsic RNA motions that underlie these processes occur over a wide range of time scales, from very fast ps motions to global conformational changes that require minutes [6, 7].

A number of NMR methods have been devised to measure RNA dynamics at single nucleotide resolution [6, 8]. One of the most useful is the model-free framework, originally developed for protein motions [9, 10], that interprets T1, T2 and NOE measurements in terms of two physically intuitive parameters: an effective correlation time for global molecular tumbling, and the generalized order parameter, S^2, which describes the spatial restriction of motion on a per-residue basis. S^2 can adopt values ranging from 0 (completely disordered) to 1 (fully ordered) (Figure 3.1) [9, 10].

Detailed motions at nucleotide resolution have been described for paradigmatic RNAs, including the TAR element from HIV [11-13], a regulatory element that binds the U1A protein [14], a stem loop derived from the U6 RNA [15], the lead-dependent ribozyme [16], and tetraloop-containing RNAs [17-19]. These studies have yielded important information regarding RNA dynamics both alone and in complex with proteins [20] and small molecule ligands [11-13]. However, analysis of RNA dynamics by NMR is

disordered
nucleotide
(low S^2)

ordered
nucleotide
(high S^2)
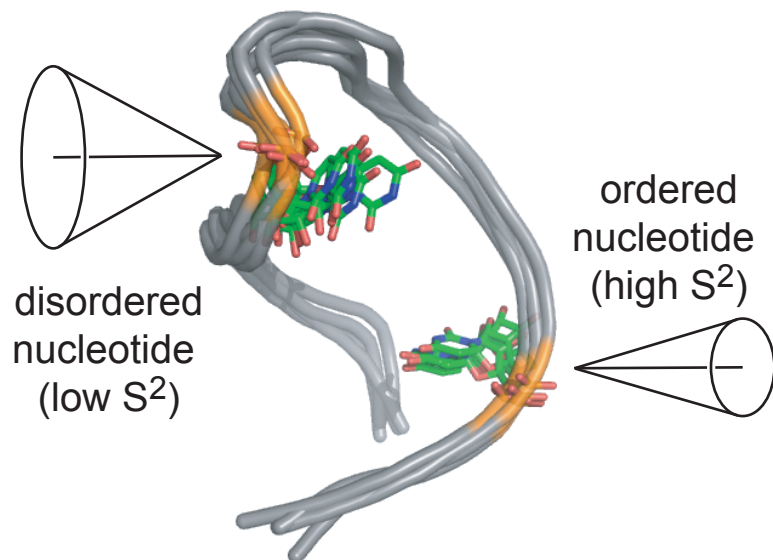
Figure 3.1. Schemes for understanding the generalized order parameter, S^2.
The cones represent the conformational space available for individual
nucleotides.

limited to small and spectroscopically well-behaved RNAs, of ~45 nucleotides or less in size. In order to analyze single-nucleotide resolution dynamics for large RNAs alone, as part of ribonucleoproteins, or *in vivo*, a different and more versatile approach is required.

Local motion in RNA can be measured using very simple and rapidly performed chemistry in a SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) experiment [21-23] SHAPE chemistry is based on the discovery that conformationally flexible nucleotides react preferentially towards acylating agents such as 1-methyl-7-nitroisatoic anhydride (1M7) to form a 2'-*O*-adduct. In contrast, nucleotides that are constrained by base pairing or tertiary interactions are unreactive. Sites of 2'-*O*-adduct formation are then detected as stops to primer extension [21-23].

SHAPE is proving to be a powerful approach for addressing a wide variety of structure-function relationships in RNA, from short oligonucleotides to RNAs thousands of nucleotides long [21-28]. To date, the correlation between SHAPE reactivity and local RNA motion, while clearly plausible, has not been rigorously proven. In contrast, S^2 is derived from a well-understood quantitative framework [9, 10] and is strongly correlated with local molecular motions and structure [29, 30].

We therefore sought to test whether SHAPE chemistry captures local nucleotide dynamics in a way that correlates with NMR observables with clear physical meaning, specifically the generalized order parameter S^2. SHAPE and NMR analyses were conducted in parallel for three RNAs: (1)

40

the TAR RNA from the 5' untranslated region of HIV-1 which activates transcriptional elongation in concert with the Tat protein; (2) the U1A protein binding site RNA that autoregulates mRNA processing; and (3) the *Tetrahymena* telomerase stem loop 4 RNA (T-SL4) which promotes folding of a pseudoknot essential for enzyme activity (Figure 3.2A) [14, 31-34].

3.2 Results

3.2.1. S2 measurements. S^2 values based on ^{13}C relaxation at the C1' ribose position have been reported for the U1A target RNA [14]. For this work, we have additionally measured relaxation parameters and calculated S^2 values for HIV-1 TAR and for most C and A nucleotides in the T-SL4 RNA (the results are represented as black and red nucleotides in Figure 3.2 and blue spheres, Figure 3.3).

3.2.2. SHAPE chemistry for TAR, U1A and T-SL4. We then measured SHAPE reactivities for these RNAs under conditions similar to those used in the NMR experiments. As expected, SHAPE reactivity profiles recapitulate the properties of the secondary structure. Base paired nucleotides are unreactive towards 2'-*O*-adduct formation, whereas bulges and loops are reactive (bars, Figure 3.3).

We then compared SHAPE reactivities with the generalized NMR order parameter, S^2, for all positions where both measurements could be made (compare blue spheres with columns, Figure 3.3). Because high SHAPE reactivities and low S^2 values both correspond to a disordered site,

41
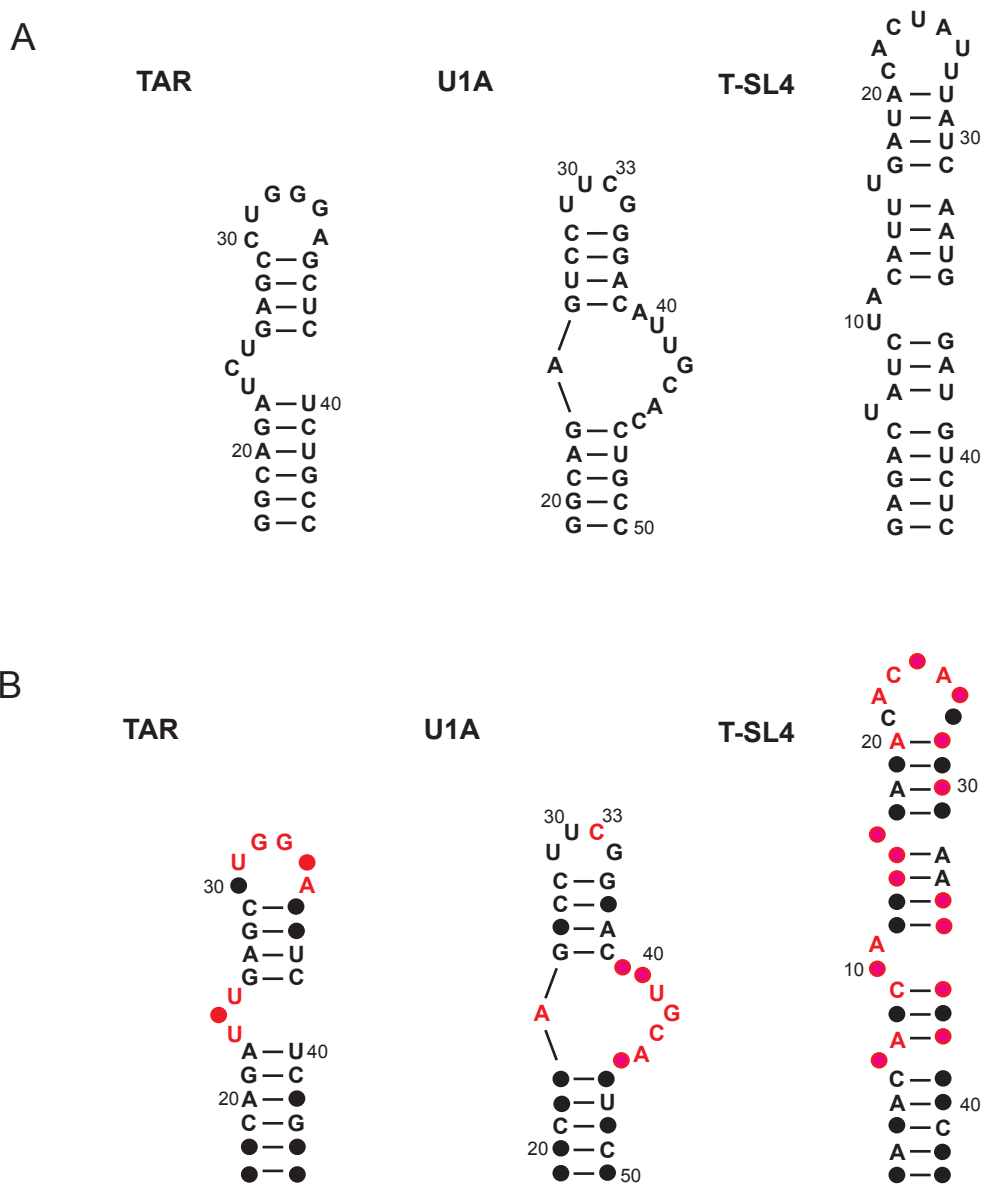
Figure 3.2. Secondary structures for TAR, U1A and T-SL4 RNAs. (A) Complete sequence for these RNAs. (B) Positions where both SHAPE and S2 data could be measured are shown with letters; nucleotides for which S2 was not obtained are shown as spheres. Nucleotides with SHAPE reactivities greater than 0.5 for TAR and U1A, or greater than 0.6 for T-SL4 are red.

42
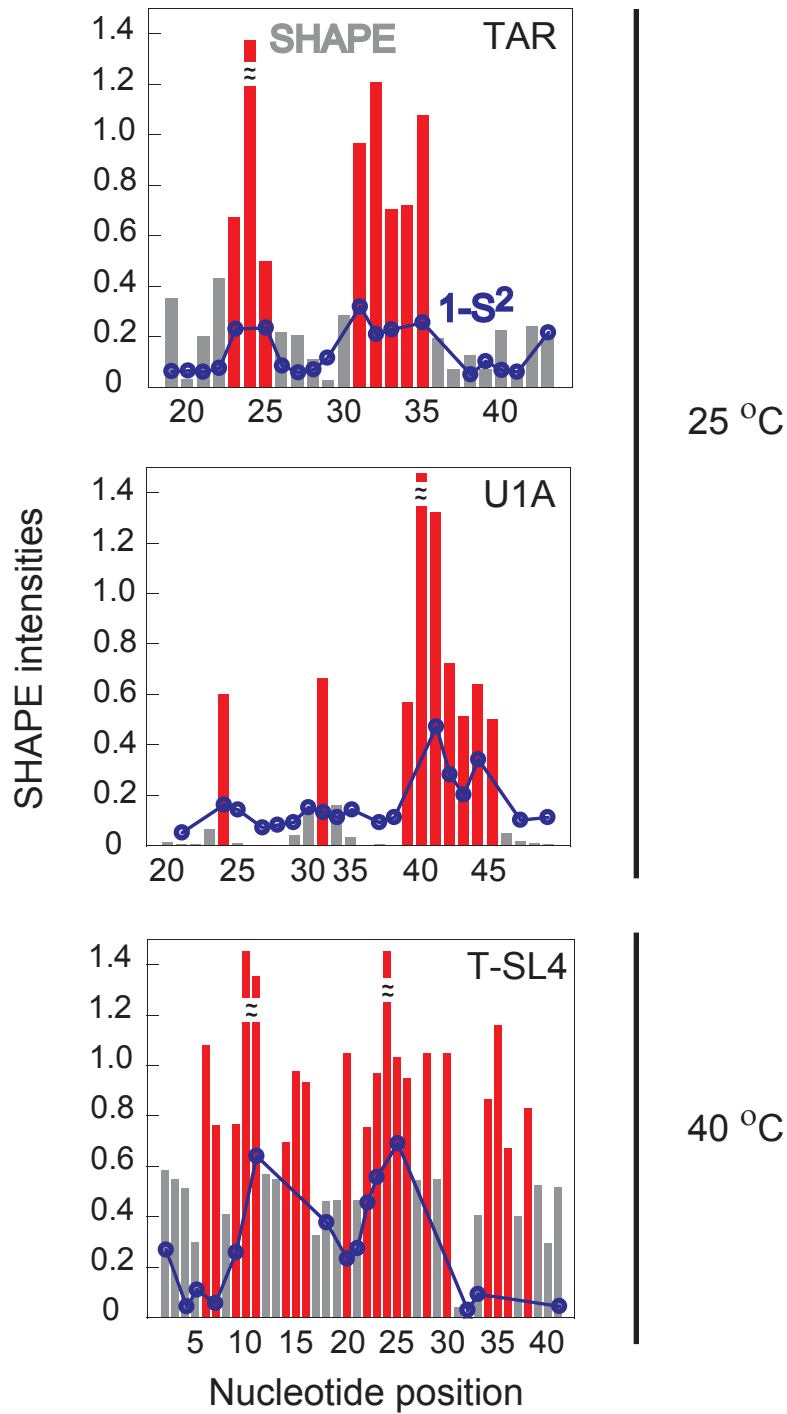
Figure 3.3. Histograms of SHAPE reactivities as a function of nucleotide position (columns) compared to 1-S2 (spheres). Nucleotides with SHAPE reactivities greater than 0.5 for TAR and U1A, or greater than 0.6 for T-SL4 are red.

we plot the generalized order parameter as $1-S^2$. For the TAR RNA, the three-nucleotide UCU bulge and the apical loop are both reactive by SHAPE and also disordered, as indicated by high $1-S^2$ values. For the U1A RNA, the 39-45 loop is clearly identifiable by both SHAPE and S^2 values. Importantly, three of four nucleotides in the 29-34 loop are both unreactive by SHAPE and have S^2 values that reflect a high level of order, in agreement with previous observations that UUCG loops are unusually stable, reflecting intramolecular base stacking and hydrogen bonding interactions [35]. SHAPE experiments with T-SL4 were performed at 40 °C, as required by the NMR analysis of this RNA. Nucleotides in the T-SL4 RNA are more reactive than in the TAR and U1A target RNAs, consistent with a temperature-induced increase in RNA dynamics (compare panels, Figure 3.3). For T-SL4, both SHAPE and S^2 detect increased nucleotide dynamics in the apical loop and at each of the three small bulges in this RNA.

There are two nucleotides where SHAPE and $1-S^2$ do not correlate well: C33 of U1A and A20 of T-SL4. Both occur at or near single stranded nucleotides, suggesting that SHAPE better represents the local structure in these regions.

3.2.3. Strong correlation between SHAPE chemistry and the generalized order parameter, S^2. We quantified the correlation between SHAPE reactivities and S^2 by plotting these values on a linear scale and calculating Pearson correlation coefficients, R (Figure 3.4). We analyzed correlation coefficients in two ways. First, we determined R using all

44
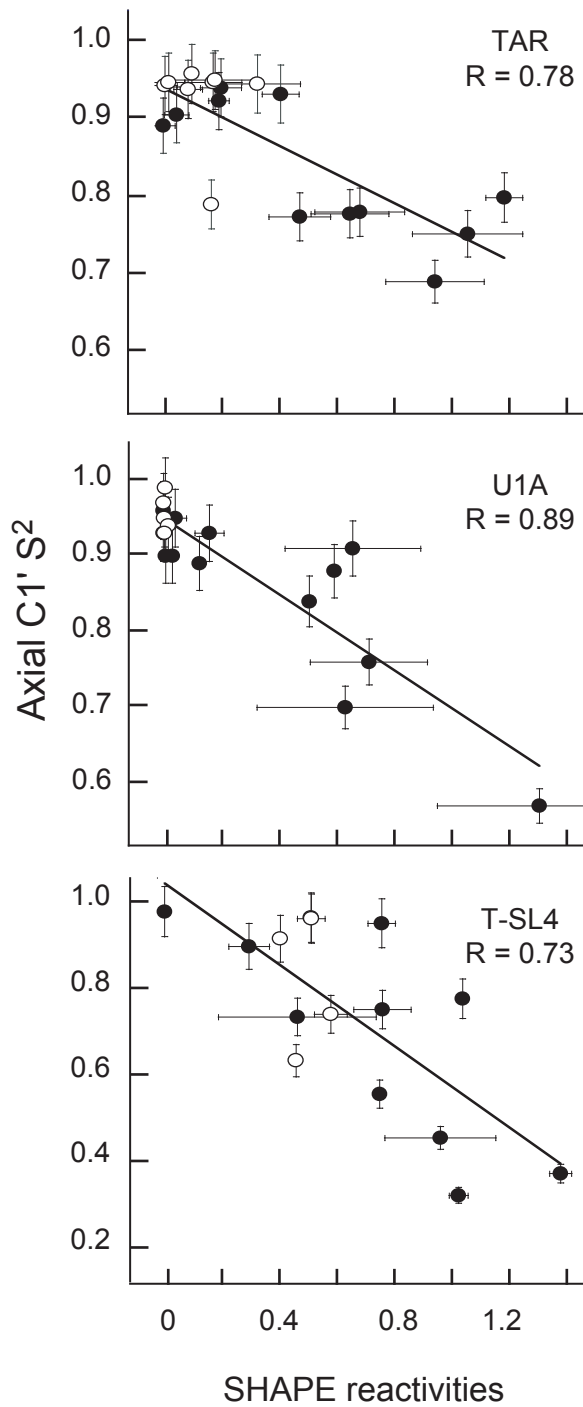
Figure 3.4. Correlation between SHAPE reactivities and S^2, measured at C1'
Pearson's linear R-values are shown. Nucleotides that do not form canonical,
internal, base pairs are emphasized with filled symbols.

measured values and observed a strong correlation between SHAPE reactivities and S^2 in all cases: R values are 0.78, 0.89 and 0.73 for the TAR, U1A and T-SL4 RNAs, respectively (all points, Figure 3.4). We also calculated R values after excluding nucleotides that form canonical base pairs, which always have near-zero SHAPE reactivities and high S^2 values. Correlation coefficients from this more stringent calculation (filled spheres, Figure 3.4) had similar correlations: R values are 0.79, 0.86 and 0.75 for the TAR, U1A and T-SL4 RNAs, respectively.

3.2.4 SHAPE chemistry is independent of solvent accessibility. Finally, we assessed whether SHAPE chemistry is influenced by the solvent accessibility of the 2'-hydroxyl position. This is a critical control to establish whether SHAPE might also report a reactive spatial orientation of the 2'-OH group in addition to measuring RNA dynamics. There is essentially no correlation between SHAPE reactivity and solvent accessibility at the 2'-hydroxyl group for any of the three RNAs (Figure 3.5).

3.3 Discussion

We conclude that the dynamics of individual RNA nucleotides as quantified by S^2 correlate strongly with SHAPE reactivities but not with solvent accessibility. Thus, SHAPE reliably measures spatial disorder and structural dynamics at single nucleotide resolution in RNA. We anticipate that SHAPE chemistry will create many opportunities for understanding the roles of individual nucleotide dynamics in the structure of large RNAs, during
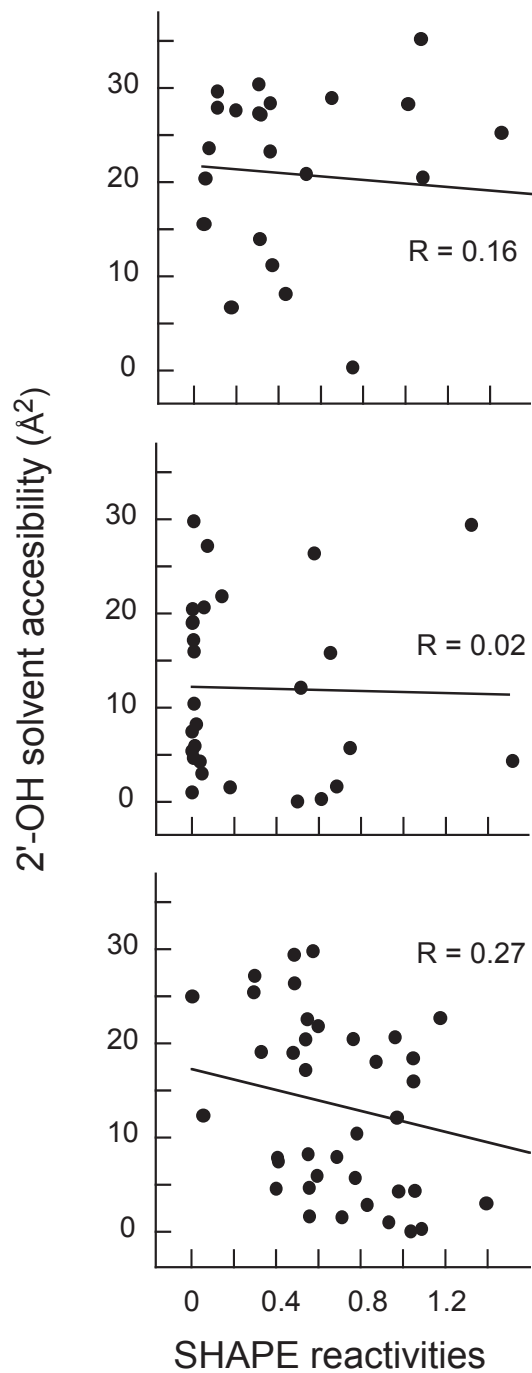
Figure 3.5. Lack of correlation between SHAPE reactivities and the solvent accessibility of the 2'-hydroxyl group. Solvent accessibilities were calculated using a 1.4 Å spherical probe, as described in the experimental section [41].

ribonucleoprotein assembly reactions, and upon RNA binding by proteins and other ligands and drugs, both *in vitro* and *in vivo*.

3.4 Experimental

3.4.1. RNA constructs. RNAs (TAR: 5'-GGCAG AUCUG AGCCU GGGAG CUCUC UGCC-3'; U1A: 5'-GGCAG AGUCC UUCGG GACAU UGCAC CUGCC-3'; T-SL4: 5'-GAGAC UAUCG ACAUU UGAUA CACUA UUUAU CAAUG GAUGU CUC-3') were synthesized by T7 RNA polymerase-mediated *in vitro* transcription using a PCR-generated template (SHAPE experiments) or a single stranded template with a double stranded promoter region (NMR measurements). SHAPE experiments for the TAR RNA were performed using a 976 nt RNA transcript [25]. For SHAPE experiments, U1A and T-SL4 RNAs were embedded within 5' and 3' structure cassette [36] sequences. RNAs were purified by denaturing polyacrylamide gel electrophoresis, excised from the gel, and recovered by electroelution and precipitation with ethanol. Purified RNAs were resuspended in either TE [10 mM Tris (pH 8.0), 1 mM EDTA] or sterile water and stored at –20 °C.

3.4.2 SHAPE analysis. SHAPE analysis of TAR RNA was performed by Kevin Wilkinson. RNAs were heated to 95 °C for 2 min, cooled on ice, and incubated at 37 °C for 10 min in SHAPE buffer [50 mM Hepes (pH 8.0), 50 mM KCl]. TAR and U1A RNAs were allowed to slowly cool to room temperature over 15 min; the T-SL4 RNA was incubated at 40 °C for 15 min. These conditions mimic those used for the NMR experiments, with the

exception that the pH was 8.0. RNAs were then treated with 1-methyl-7-nitro-isatoic anhydride (1M7) [24] reagent (1 μL; 100 mM; in anhydrous DMSO) and allowed to react for 5 min. No-reagent control reactions contained 1 μL neat DMSO. Modified RNAs were recovered by ethanol precipitation [80 μL water, 10 μL NaCl (5 M), 1 μL glycogen (20 mg/mL), 400 μL ethanol; 30 min at -80 °C; centrifugation at 10,000 g for 10 min] and resuspended in 5 μL TE. For the U1A and T-SL4 RNA, sites of 2'-O-adduct formation were identified by primer extension using a 5'-[^{32}P]-label primer as described [23], with the exception that the extension reaction was incubated at 52 °C for 7 min. Dideoxy sequencing markers were generated using unmodified RNA. cDNA extension products were separated by gel electrophoresis and visualized using phosphorimaging. cDNA band intensities for the (+) and (−) reagent reactions were integrated using SAFA [37]. Primer extension products for the TAR RNA were analyzed by capillary electrophoresis, as described [25]. SHAPE reactivity profiles were obtained by subtracting the (−) reagent control from the (+) reaction intensities. SHAPE reactivities were normalized to a scale in which 1.0 is defined as the average intensity of highly reactive positions, after excluding outliers using a box plot analysis [38]. Two independent SHAPE experiments were performed for each RNA. Errors in reactivities are reported as the absolute differences between experiments, divided by the maximum expected reactivity value, 1.

3.4.3. NMR relaxation experiments. NMR relaxation experiments were performed by Zahra Shajani, University of Washington, Seattle. RNAs

used for the NMR experiments were dialyzed into 10 mM potassium phosphate buffer containing 0.1 mM EDTA at pH 6.0; the final concentration of potassium ion (as the obligate RNA counterion) was ~50 mM. Data collection was executed on a Bruker Avance-500 spectrometer using a TXI triple resonance HCN probe in 99.9% D_2O at 25 °C. T_1, $T_{1\rho}$ and Het-NOE experiments were recorded as a series of 2D NMR spectra, in which the relaxation delay (τ) was parametrically increased. Experiments were performed in constant-time mode, essentially as described [14]. Relaxation studies with ^{13}C are complicated by the fact that dipolar couplings between adjacent carbons contribute significantly to the observed relaxation behavior in uniformly labeled samples. This interference increases with the square of the correlation time and becomes substantial for molecules that tumble slowly and can introduce large errors in the model-free analysis. Therefore, to obtain accurate data for the T-SL4 RNA, NMR data were collected at 40 °C to increase the rate of rotational diffusion.

3.4.4. Generalized order parameter calculations. Quantitative analysis of the relaxation data was conducted using the model-free approximation [9, 10], using ModelFree 4.15 [39]. In the analysis, we assumed that the chemical shift tensors are axially symmetric ($\eta=0$) and that the symmetry axis of this tensor is collinear with the C–H bond. Model selection was as described [39] with some modifications [40]. Specifically, when neither model 2 nor 3 could be applied satisfactorily, models 4 and 5 were introduced. ModelFree parameters were fit to one of five models, in

which the following parameters are varied: (1) S^2; (2) S^2 and an effective internal correlation time for fast motions (τ_e); (3) S^2 and the transverse relaxation exchange parameter (R_{ex}); (4) S^2, τ_e and R_{ex}; (5) the order parameters for shorter and longer time scale motion (S_f^2 and S_s^2, respectively). For all analyses, a CSA of 30 ppm was used. Errors in S^2 calculations were estimated to be ±5%.

3.4.5. Solvent Accessibility Calculations were performed using a 1.4 Å radius spherical probe [41].

Figure 3.1 was composed using Pymol [42].

3.5. References

1.   Kim, J.N. and R.R. Breaker, *Purine sensing by riboswitches.* Biol. Cell., 2008. **100**(1): p. 1-11.

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

3.   Buchmueller, K.L., et al., *A collapsed non-native RNA folding state.* Nature Struct. Biol., 2000. **7**(5): p. 362-366.

4.   Maity, T.S. and K.M. Weeks, *A three-fold RNA-protein interface in the signal recognition particle gates native complex assembly.* J. Mol. Biol., 2007. **369**: p. 512-24.

5.   Bokinsky, G., et al., *Two distinct binding modes of a protein cofactor with its target RNA.* J. Mol. Biol., 2006. **361**(4): p. 771-84.

6.   Shajani, Z. and G. Varani, *NMR studies of dynamics in RNA and DNA by ^{13}C relaxation.* Biopolymers, 2007. **86**(5-6): p. 348-359.

7.   Buchmueller, K.L., et al., *RNA-tethered phenyl azide photocrosslinking via a short-lived indiscriminant electrophile.* J. Am. Chem. Soc., 2003. **125**(36): p. 10850-61.

8.   Gueron, M. and J.L. Leroy, *Studies of base pair kinetics by NMR measurements of proton exchange.* Methods Enzymol., 1995. **261**: p. 383-413.

9.   Lipari, G. and A. Szabo, *Model free approach to the interpretation of nuclear magnetic resonance relaxation in macrololecules. 1. Theory and range of validity.* J. Am. Chem. Soc., 1982. **104**: p. 4546-59.

10.   Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results.* J. Am. Chem. Soc., 1982. **104**: p. 4559-70.

11.   Dayie, K.T., A.S. Brodsky, and J.R. Williamson, *Base flexibility in HIV-2 TAR RNA mapped by solution (15)N, (13)C NMR relaxation.* J. Mol. Biol., 2002. **317**: p. 263-78.

12.   Vallurupalli, P., et al., *New RNA labeling methods offer dramatic sensitivity enhancements in ^{2}H NMR relaxation spectra.* J. Am. Chem. Soc., 2006. **128**: p. 9346-7.

13. Hansen, A.L. and H.M. Al-Hashimi, *Dynamics of large elongated RNA by NMR carbon relaxation.* J. Am. Chem. Soc., 2007. **129**: p. 16072-82.

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

15. Blad, H., et al., *Dynamics and metal ion binding in the U6 RNA intramolecular stem-aoop as analyzed by NMR.* J. Mol. Biol., 2005. **353**(3): p. 540-555.

16. Hoogstraten, C.G., J.R. Wank, and A. Pardi, *Active site dynamics in the lead-dependent ribozyme.* Biochemistry, 2000. **39**(32): p. 9951-9958.

17. Hall, K.B. and C. Tang, *^{13}C relaxation and dynamics of the purine bases in the iron responsive element RNA hairpin.* Biochemistry, 1998. **37**: p. 9323-32.

18. Duchardt, E. and H. Schwalbe, *Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by MNMR 13C relaxation.* J. Biomol. NMR, 2005. **32**: p. 295-308.

19. Ferner, J., et al., *NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops.* Nucl. Acids Res., 2008. **36**: p. 1928-1940.

20. Shajani, Z., G. Drobny, and G. Varani, *Binding of U1A protein changes RNA dynamics as observed by 13C NMR relaxation studies.* Biochemistry, 2007. **46**(20): p. 5875-5883.

21. 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**(12): p. 4223-4231.

22. Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transition in tRNA(Asp) transcripts.* J. Am. Chem. Soc., 2005. **127**(13): p. 4659-4667.

23. Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.* Nat. Protocols, 2006. **1**(3): p. 1610-1616.

24.   Mortimer, S.A. and K.M. Weeks, *A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry.* J. Am. Chem. Soc., 2007. **129**(14): p. 4144-4145.

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

26.   Wang, B., K.A. Wilkinson, and K.M. Weeks, *Complex ligand-induced conformational changes in tRNA^{Asp} revealed by single nucleotide resolution SHAPE chemistry.* Biochemistry, 2008. **47**: p. 3454-3461.

27.   Duncan, C.D.S. and K.M. Weeks, *SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA.* Biochemistry, 2008. **47**: p. in press.

28.   Gherghe, C.M., et al., *Slow conformational dynamics at C2'-endo nucleotides in RNA.* J. Am. Chem. Soc., 2008. **130**: p. in press.

29.   Palmer, A.G., *NMR characterization of the dynamics of biomacromolecules.* Chem. Rev., 2004. **104**(8): p. 3623-3640.

30.   Jarymowycz, V.A. and M.J. Stone, *Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences.* Chem. Rev., 2006. **106**(5): p. 1624-1671.

31.   Sharp, P.A. and R.A. Marciniak, *HIV TAR: an RNA enhancer?* Cell, 1989. **59**(2): p. 229-30.

32.   Hall, T.M., *Poly(A) tail synthesis and regulation: recent structural insights.* Curr. Opin. Struct. Biol., 2002. **12**: p. 82-88.

33.   Sperger, J.M. and T.R. Cech, *A stem-loop of tetrahymena telomerase RNA distant from the template potentiates RNA folding and telomerase activity.* Biochemistry, 2001. **40**(24): p. 7005-7016.

34.   Chen, Y., et al., *Structure of stem-loop IV of Tetrahymena telomerase RNA.* EMBO J., 2006. **25**: p. 3156-3166.

35.   Cheong, C., G. Varani, and I. Tinoco, *Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC.* Nature, 1990. **346**: p. 680-682.

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

37. Das, R., et al., *SAFA: Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments.* RNA, 2005. **11**: p. 344-354.

38. Deigan, K.E., et al., *Accurate SHAPE-constrained RNA structure prediction.* in preparation, 2008.

39. Mandel, A.M., M. Akke, and A.G. Palmer, *Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme.* J. Mol. Biol., 1995. **246**: p. 144-63.

40. Chen, J., C.L. Brooks, and P.E. Wright, *Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation.* J. Biomol. NMR, 2004. **29**: p. 243-57.

41. Brunger, A.T., et al., *Crystallography & NMR System.* Acta. Cryst., 1998. **D54**: p. 905-921.

42. DeLano, W.L., *The Pymol molecular graphics system, DeLano Scientific, South San Francisco, CA, USA.* 2002.

Chapter 4

Three-dimensional structure of the dimeric

SL1-SL2 RNA

4.1 Introduction

Retroviral genomes consist of two sense-strand RNAs that are noncovalently linked near their 5' ends to form a dimeric structure [1-3]. This dimeric state contains exactly two RNA genomes. The RNA genomes are specifically recognized and packaged into each nascent virion [2, 3] against a large background of cellular RNA, thus representing a striking example of molecular recognition in biology.

Our laboratory has recently identified a minimal dimerization active sequence (MiDAS) [4] for a representative gamma retrovirus, the Moloney murine sarcoma virus (Figure 4.1). The MiDAS domain correlates closely with retroviral genomic sequences sufficient to package heterologous RNAs into virions [5-8], as dimers [7]. The MiDAS domain also includes conserved sequence elements previously proposed to specify the noncovalent interactions that mediate RNA dimerization.

Conserved sequence elements include self-complementary (palindromic) sequences (PAL1 and PAL2) and stem–loop structures 1 and 2 (SL1 and SL2) [9-13]. SL1 and SL2 contain GACG tetraloops that form stable loop–loop interactions with a second RNA molecule. Loop–loop interactions are mediated by canonical intermolecular C-G base pairing and additional stacking and intra- and intermolecular hydrogen bonds [14] (Figure 4.2). The self-complementary PAL1 and PAL2 sequences form extended heteroduplexes involving both strands in the dimer [9, 10, 12, 13].
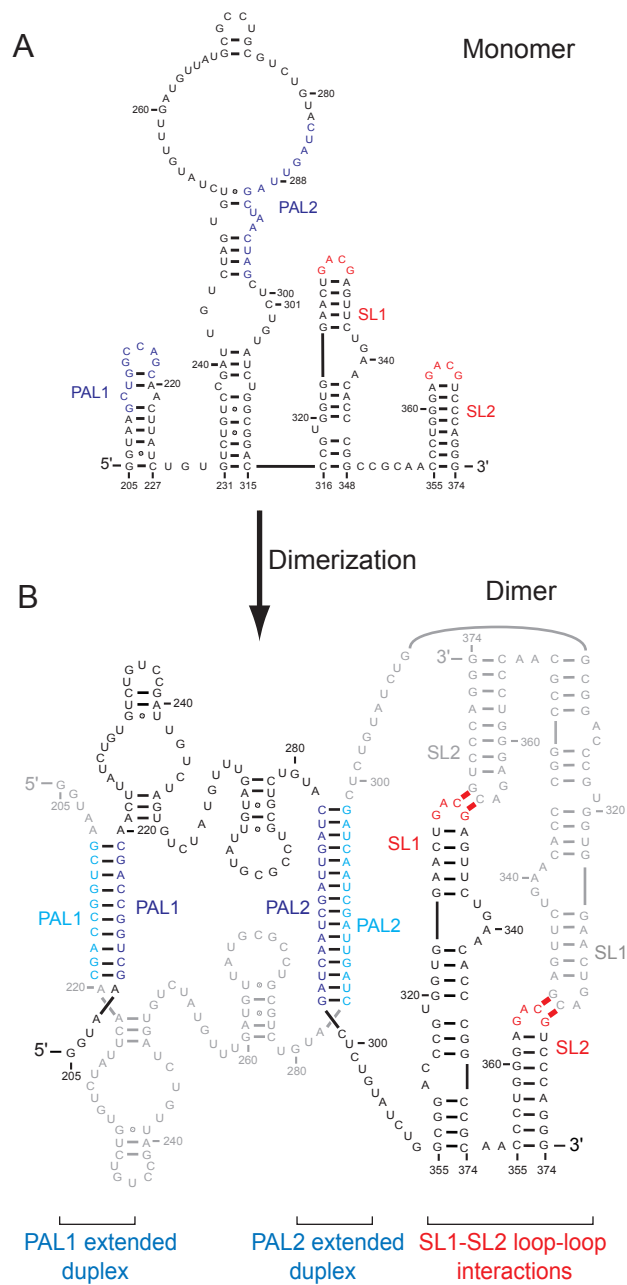
Figure 4.1. Monomeric and dimeric MiDAS RNA secondary structures. (A) Structure of the MiDAS domain for the Moloney murine sarcoma virus in the monomeric starting state. Upon dimerization (B) PAL 1 and PAL2 form intermolecular extended duplexes. SL1-SL2 domains interact via apical GACG interactions.
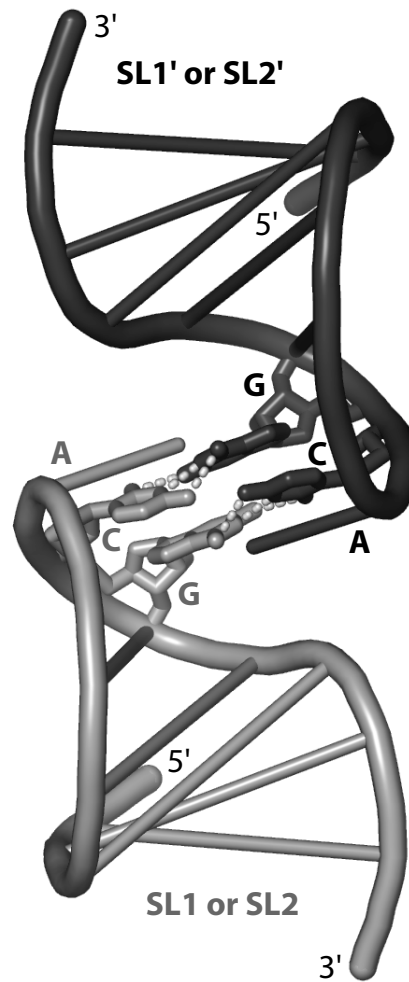
Figure 4.2. Structure of the loop–loop interaction formed between GACG sequences. Nucleotides in loop positions 3 and 4 that form canonical C-G base pairs are shown explicitly with their intermolecular hydrogen bonds; these base pairs also are stabilized by flanking stacked adenosine bases. Other nucleotides are represented as cylinders.

MiDAS conformational changes encountered during dimerization (Figure 4.1) have been proposed to expose single-stranded RNA regions that are recognized by the retroviral Gag protein [15, 16]. Selective packaging of retroviral genomes is likely to involve a mechanism significantly more complex than simple protein recognition of short RNA sequences. The RNA conformation itself is likely to present protein-binding elements in the context of a specific three-dimensional architecture that enables selective recognition by the viral Gag protein. Therefore, knowledge of the three-dimensional architecture of the MiDAS can set the foundations for understanding the genomic dimer recognition by the viral proteins.

We decided to focus our three-dimensional structural analysis on the SL1-SL2 domain. The secondary structure model of the SL1-SL2 has been previously determined using SHAPE chemistry [4]. Because the SL1-SL2 domain contains two GACG RNA tetraloops, one SL1 can interact with either an SL1 or SL2 from another monomer. Therefore, additional experiments are required to establish the interaction partners for SL1 and SL2 in one RNA with their tetraloop partners from the second RNA in the dimer (Figure 4.1B).

4.2 Results

4.2.1 SL1-SL2 three-dimensional orientation. We used site-directed hydroxyl radical cleavage experiments [17] (i) to identify the orientation and pairing partners for the SL1 and SL2 loops and (ii) to obtain constraints for the global architecture of the SL1-SL2 domain in the final dimer state. We

engineered two 65-nt RNAs (spanning nucleotides 310–374), confined to be in the dimer state, and derivatized these RNAs with Fe(II)-EDTA groups at defined positions (Figure 4.3). The 310-bromoacetamidobenzyl-EDTA-Fe(II) (BABE) construct contained a BABE adduct at the 5' end of the RNA. The 336-isothiocyanobenzyl-EDTA-Fe(II) (ITE) construct was derivatized with ITE via a 2' amino nucleotide at position 336. In the presence of hydrogen peroxide and ascorbic acid, highly reactive hydroxyl radicals capable of cleaving the RNA backbone are generated at the Fe(II)-EDTA moiety.

The 310-BABE or 336-ITE constructs were allowed to form loop–loop dimers with a second 5' ^{32}P-labeled SL1-SL2 domain RNA (radiolabels are indicated with asterisks in Figure 4.3). By this scheme, only intermolecular cross-strand RNA cleavages are detected. Cross-strand RNA cleavage products were resolved on sequencing gels. Cleavages greater than three or greater than five times background were judged to be moderate and strong, respectively (Figure 4.3)

For the 310-BABE construct, strong cleavage was obtained only at the apex of SL1' (Figure 4.3A, red symbols). With the 336-ITE construct, strong cleavage was observed in both SL1' and SL2' (Figure 4.3B). These cleavage data strongly support a secondary structure arrangement in which SL1 forms a loop–loop interaction with SL2' and SL2 interacts with SL1' (Figure 4.3).

4.2.2. Distance constraints refinement of the three-dimensional structure for the SL1-SL2 domain. RNA SHAPE chemistry and site-directed cleavage experiments provide extensive information that strongly

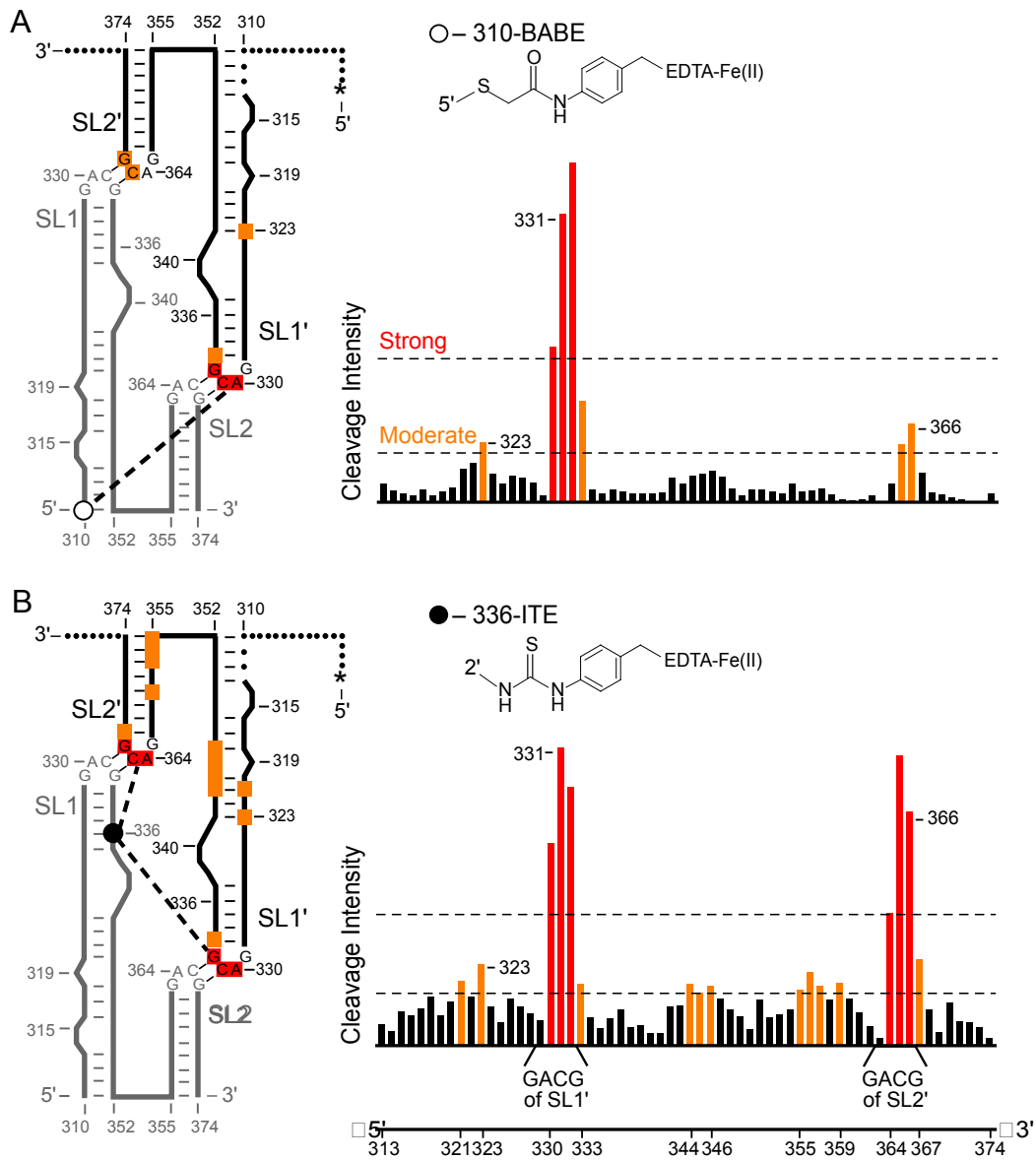Figure 4.3. Architecture of the SL1-SL2 interaction in the final dimer conformation mapped by site-directed hydroxyl radical footprinting. (A) Fe(II)-BABE (open circle) mediated cleavage from nucleotide 310. (B) Fe(II)-ITE (filled circle) cleavage from nucleotide 336. Small spheres indicate RNA regions that were not monitored; the position of 5' radiolabel on second RNA strand is indicated by an asterisk.

constrain allowed three-dimensional structures for the SL1-SL2 domain in the dimer state. We used this experimental information to refine structures for the dimer state by using distance constraints-based algorithms [18-20] . Canonical hydrogen bonding and planarity constraints were imposed for predicted [16] base pairs, as constrained by experimental SHAPE reactivities. Base pairing also was enforced between traditional Watson-Crick C-G pairs in the GACG tetraloops, as demonstrated in an independent NMR study [14]. Experiments with the 310-BABE construct yielded 3 strong and 4 moderate constraints; 336-ITE yielded 6 strong and 11 moderate constraints (Figure 4.3). In addition, seven positions that were unreactive by site-directed cleavage were constrained to be distant from the tethered Fe(II)-EDTA groups (positions 312, 340 and 350 for the 310-BABE experiment; positions 313, 340, 353 and 374 for the 336-ITE experiment). Because the SL1-SL2 domain is a symmetrical dimer, the number of constraints used to constrain the refinement is doubled to 62 (Figure 4.4).

Two classes of constraints were used during refinement. (i) Intramolecular constraints. Planarity and hydrogen-bonding constraints were imposed at all base paired positions; constraints involving idealized phosphate-phosphate and intra- and interribose distances were used to enforce A-form geometry. For unpaired nucleotides, backbone dihedrals were restrained to ±20° from A-form geometry; phosphate-phosphates distances for nucleotides 323–324 and 338–342 were constrained to be 5–7 Å and 16–18 Å, respectively. (ii) Intermolecular constraints. Base pairing was
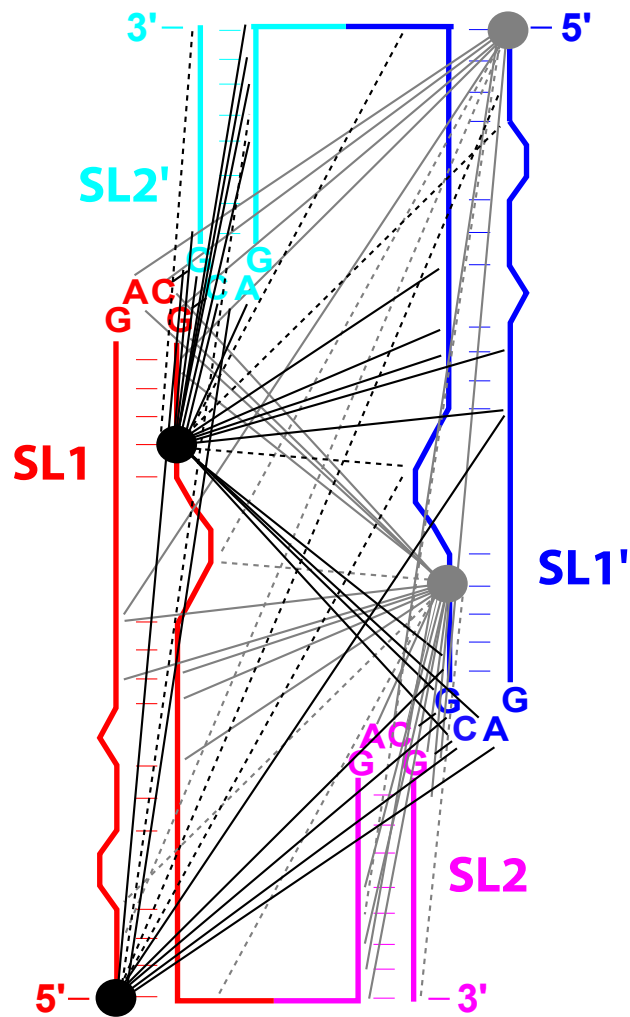
Figure 4.4 Summary of 62 long-range intermolecular distance constraints used for structure refinement. Adjacent and repelling constraints are shown with solid and dashed lines, respectively. One monomer is shown in red (SL1) and magenta (SL2), and a second monomer is shown in blue (SL1) and cyan (SL2).
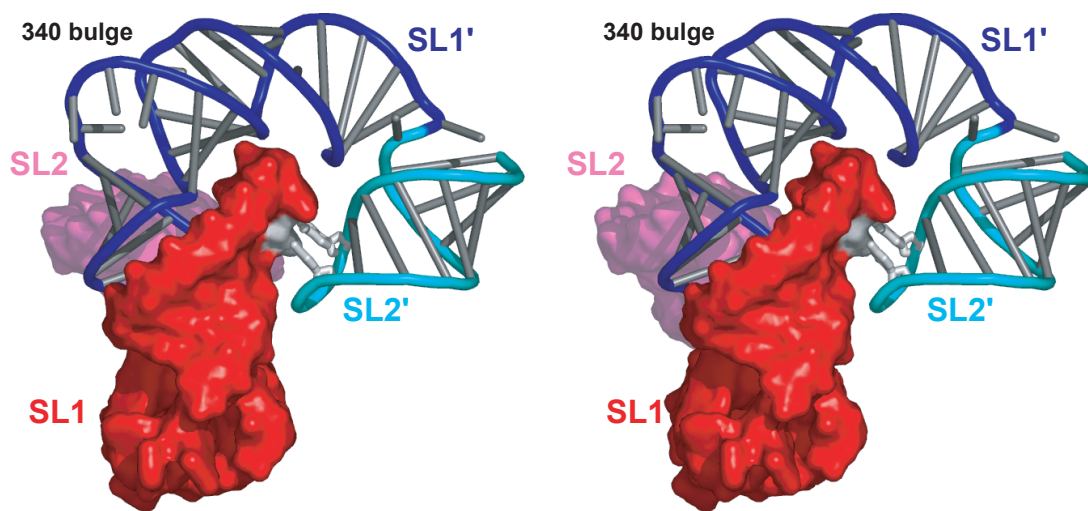
Figure 4.5 Stereo image of the SL1-SL2 domain in the final dimer state. One monomer (red and magenta) is shown in a surface representation. The second monomer (blue and cyan) is illustrated as a backbone cartoon; bases are shown as cylinders. Cross-strand G-C pairs in the tetraloops are white.

enforced at the G-C pairs in the tetraloops; strong and medium site-directed cleavages were refined to optimal distances of 0–25 and 0–35 Å (from the 5' OH or 2' OH position to the appropriate C4' atom), respectively, by using square-well potentials. These distance envelopes were established in ribosome-based work by Noller et al. [17, 21] and from our analysis of the site-directed cleavage data for tRNA by Han and Dervan [22]. Regions that were unreactive by site-directed cleavage were used as repelling interactions and were constrained to be ≥ 30Å from the tethered Fe(II)-EDTA group.

SL1-SL2 refinement was performed in two steps. (i) Starting from an initial structure in which both RNA monomers were in extended conformations, CNS 1.1 [19] was used initially to generate 75 independent, globally folded sets of monomers by using intramolecular constraints only; overall geometries then were improved by using Xplor-NIH [20]. (ii) These 75 structures then were folded to candidate dimer structures in CNS, using both intramolecular and intermolecular constraints; local interactions again were optimized by using Xplor-NIH. We evaluated multiple criteria for accepting the best refined structures, and the most persuasive criteria was found to be all-atom packing quality as defined by the clash score [23, 24]. Our analysis is based on the eight refined structures with clash scores <43. These eight structures superimpose with RMSD values of 3.5–5.0 (4.1 average) Å over 128 phosphate positions in the dimer. Refined models converged to an average RMSD of 4.1 Å over 128 phosphate positions (Figure 4.6 and Figure 4.7). The high degree of convergence reflects that the 62 constraints provide
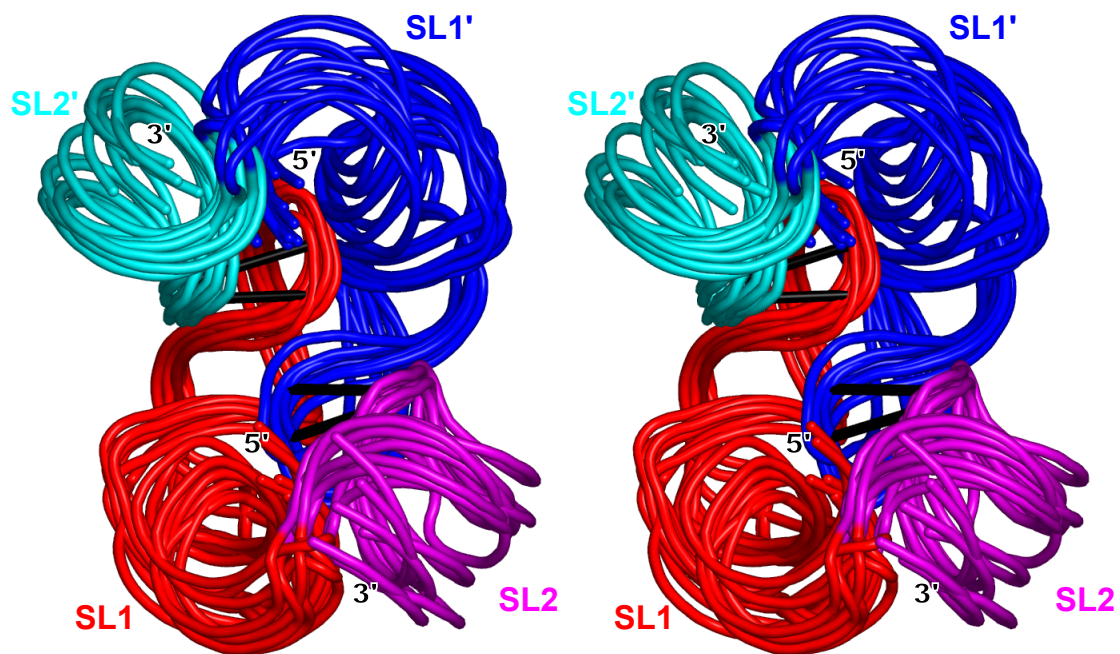
Figure 4.6. Stereo image of eight refined structures with the lowest all-atom clash scores. RNA backbones are shown as tubes; cross-strand G-C base pairs in the tetraloops are represented as black cylinders.
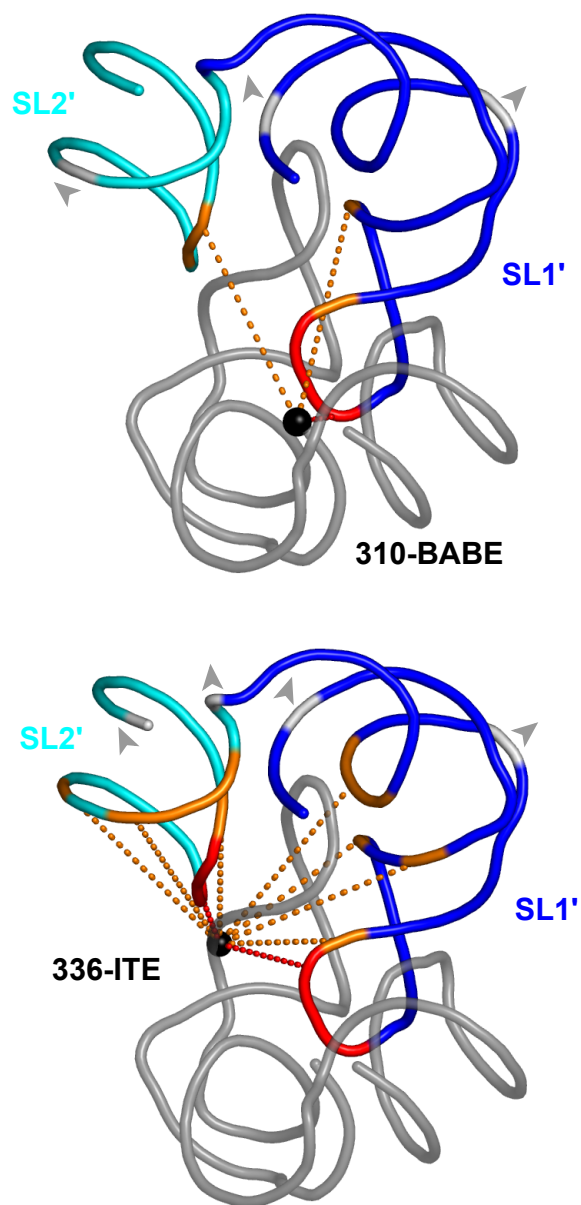
Figure 4.7. Distance constraints derived from the 310-BABE and 336-ITE RNAs superimposed on the mean refined model. Radiolabeled RNA is shown with colored backbone; red and orange indicate strong and moderate cross-strand cleavages, respectively. Positions used as repelling constraints are shown in white next to gray arrowheads. For clarity, only a single dotted line is illustrated for each group of multiple long-range constraints in a given region. RNAs derivatized with Fe(II)-EDTA groups are shown with gray backbones; attachment sites are emphasized with black spheres.

dense information with which to position the five independent helical elements in each SL1-SL2 domain monomer.

Finally, to evaluate the extent to which the final structures for the SL1-SL2 domain dimer were governed by the 62 long-range constraints obtained from the site-directed cleavage experiments, we also refined the SL1-SL2 dimer structure by using only local base pairing and helix geometry constraints. Structures refined without the use of long-range experimental constraints typically formed noncompact, pseudo-square-planar structures. Of 75 solutions, 27 had clash scores <43. The RMSD values of these 27 structures and of the eight structures with lowest clash scores were 13.9 and 14.7 Å, respectively. These values are much larger than the 4.1 Å deviation obtained when the long-range experimental constraints are included in the refinement.

4.3 Discussion

The SL1-SL2 domain forms a tightly packed tertiary structure in which SL1 is bent at the 340 bulge, such that each SL1-SL2 element is roughly U-shaped (Figures 4.5 and 4.6). Two U-shaped RNAs then form an interdigitated structure in which the apex of SL1 from one RNA forms extensive interfaces with both SL1' and SL2' from the second monomer. The minor groove side of SL1 faces the minor groove of SL2', whereas the major groove of SL1 fits snugly against the major groove side of SL1' in the second RNA.

Our experiments support a model for retroviral genomic dimerization in MuSV in which the SL1-SL2 domain plays a major role. This conclusion is strongly corroborated by recent experiments showing that the SL1-SL2 domain (in the absence of PAL2 and most of PAL1) is sufficient to direct modest packaging of a heterologous RNA, in a dimeric state, into virions [7].

4.4 Experimental

4.4.1. Retroviral RNA Constructs. The RNA constructs were synthesized by Christopher Badorrek. DNA templates for in vitro transcription of the MiDAS, SL1-SL2 (monomer), and SL1-SL2 (final dimer) RNA constructs were generated by PCR from the pLNBS plasmid or synthetic oligonucleotide (Midland Certified, Midland, TX) templates. RNA constructs were generated by T7 RNA polymerase-mediated transcription [500 µl, 37 °C, 5 h; containing 80 mM HEPES-NaOH (pH 7.4), 40 mM dithiothreitol (DTT), 0.01% (vol/vol) Triton X-100, 2 mM spermidine, 10 mM $MgCl_2$, 2 mM each nucleoside triphosphate, 25 µg PCR-generated template, 20 units SUPERase-In (Ambion, Austin, TX) and 0.1 mg/ml polymerase]. Internally labeled RNAs were generated by using 20 µCi (1 Ci = 37 GBq) γ-[^{32}P]ATP and 0.5 mM unlabeled ATP. RNAs were purified by denaturing gel electrophoresis [5 or 15% (wt/vol) polyacrylamide, 7 M urea; for MiDAS and SL1-SL2 RNAs, respectively], excised from the gel, eluted overnight into 0.5x TBE (45 mM Tris-borate/1 mM EDTA), and concentrated by ethanol

precipitation. RNAs were stored in 10 mM HEPES-NaOH (pH 7.5)/1 mM EDTA at –20 °C.

4.4.2. 310-BABE and 336-ITE RNAs. Site-directed hydroxyl radical footprinting experiments were performed by Christopher Badorrek. 310-BABE: A 65-nt RNA (nucleotides 310-374) was transcribed in vitro in the presence of 10 mM guanosine monophosphorothioate (GMPS, Dharmacon Research, Lafayette, CO); GMPS is incorporated only at the initiating 5' G during transcription. Fifty pmol of this RNA were treated with excess (50 µmol) bromoacetamidobenzyl-EDTA (BABE, Dojindo Labs, Gaithersburg, MD) in 10 µl of 0.15 M potassium phosphate (pH 8.0) at 37°C for 1 h, ethanol precipitated, and resuspended in 10 µl 0.15 M potassium phosphate (pH 8.0). This solution (8 µL) was incubated with 2 µl of 10 mM $(NH4)_2Fe(SO4)_2$ at 25°C for 10 min, ethanol precipitated, and resuspended in 0.15 M potassium phosphate (pH 8.0). 336-ITE: The precursor RNA was synthesized chemically (Dharmacon Research) with a 2' amino nucleotide at position U336. Thirty pmol of this RNA were treated with 300 µmol of isothiocyanobenzyl-EDTA (ITE; Dojindo Labs) in 10 µl buffer [0.05 M sodium borate (pH 8.0)/30% formamide/10 units SUPERase-In] at 37 °C overnight, ethanol precipitated, resuspended in 10 µl 0.15 M potassium phosphate (pH 8.0), and allowed to form a complex with Fe^{2+} as described above. Site-directed hydroxyl radical experiments were initiated by adding hydrogen peroxide and ascorbic acid (in 3 µl) to final concentrations of 0.1% and 2.5 mM, respectively. Cross-strand RNA cleavage required (i) 5' thiol or 2' amine

71

groups, (ii) addition of $(NH4)_2Fe(SO4)_2$, and (iii) derivatization with BABE or ITE.

4.4.3. RNA dimer structure refinement. Three-dimensional models were refined by using simulated annealing and molecular mechanics computations, analogous to routines used to develop macromolecular structure models from NMR information [19, 20]. Our general refinement routine is based on scripts generously provided to us by J. Davis and S. Butcher [18] and additionally incorporates the optimized force field parameters developed by Moore et al [25].

4.5 References

1.   D'Souza, V. and M.F. Summers, *How retroviruses select their genome.* Nat. Rev. Microbiol., 2005. **3**(8): p. 643-55.

2.   Berkowitz, R., J. Fisher, and S.P. Goff, *RNA packaging.* Curr. Top. Microbiol. Immunol., 1996. **214**: p. 177-218.

3.   Rein, A., *Take two.* Nature Struct. Mol. Biol., 2004. **11**: p. 1034-1035.

4.   Badorrek, C.S. and K.M. Weeks, *RNA flexibility in the dimerization domain of a gamma retrovirus.* Nat. Chem. Biol., 2005. **12**: p. 104-11.

5.   Sakuragi, J., T. Shioda, and A.T. Panganiban, *Duplication of the primary encapsidation and dimer linkage region of human immunodeficiency virus type 1 RNA results in the appearance of monomeric RNA in virions.* J. Virol., 2001. **75**(6): p. 2557-2565.

6.   Mougel, M. and E. Barklis, *A role for two hairpin structures as a core RNA encapsidation signal in murine leukemia virus virions.* J. Virol., 1997. **71**(10): p. 8061-8065.

7.   Hibbert, C.S., J. Mirro, and A. Rein, *mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers.* J. Virol., 2004. **78**(20): p. 10927-10938.

8.   Adam, M.A. and A.D. Miller, *Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions.* J. Virol., 1988. **62**(10): p. 3802-3806.

9.   Paillart, J.C., et al., *Dimerization of retroviral genomic RNAs: structural and functional implications.* Biochimie, 1996. **78**(7): p. 639-53.

10.   Ly, H. and T.G. Parslow, *Bipartite signal for genomic RNA dimerization in Moloney murine leukemia virus.* J. Virol., 2002. **76**(7): p. 3135-3144.

11.   Konings, D.A., et al., *Novel GACG-hairpin pair motif in the 5' untranslated region of type C retroviruses related to murine leukemia virus.* J. Virol., 1992. **66**(2): p. 632-640.

12.   Girard, P.M., et al., *A short complementary sequence in the 5' leader region is responsible for dimerization of MoMuLV genomic RNA.* Biochem., 1995. **34**(30): p. 9785-94.

13. Oroudjev, E.M., P.C. Kang, and L.A. Kohlstaedt, *An additional dimer linkage structure in Moloney murine leukemia virus RNA.* J. Mol. Biol., 1999. **291**(3): p. 603-13.

14. Kim, C.-H. and I. Tinoco, Jr., *A retroviral RNA kissing complex containing only two GC base pairs.* Proc. Natl. Acad. Sci. USA, 2000. **97**(17): p. 9396-9401.

15. Dey, A., et al., *Composition and sequence-dependent binding of RNA to the nucleocapsid protein of moloney murine leukemia virus.* Biochem., 2005. **44**(10): p. 3735-3744.

16. D'Souza, V. and M.F. Summers, *Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus.* 2004. **431**(7008): p. 586-590.

17. Joseph, S. and H.F. Noller, *Directed hydroxyl radical probing using iron (II) tethered to RNA.* Methods Enzymol., 2000. **318**: p. 175-90.

18. Davis, J.H., et al., *RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex.* J. Mol. Biol., 2005. **351**(2): p. 371-382.

19. Brunger, A.T., et al., *Crystallography & NMR System.* Acta. Cryst., 1998. **D54**: p. 905-921.

20. Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package.* J. Mag. Res., 2003. **160**(1): p. 65-73.

21. Lancaster, L., et al., *Orientation of ribosome recycling factor in the ribosome from directed hydroxyl radical probing.* Cell, 2002. **111**(1): p. 129-40.

22. Han, H. and P.B. Dervan, *Visualization of RNA tertiary structure by RNA-EDTA.Fe(II) autocleavage: analysis of tRNA(Phe) with uridine-EDTA.Fe(II) at position 47.* Proc. Natl. Acad. Sci. USA, 1994. **91**(11): p. 4955-9.

23. Davis, I.W., et al., *MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes.* Nucl. Acids Res., 2004. **32**p. 615-619.

24. Word, J.M., et al., *Visualizing and quantifying molecular goodness-of-fit: sall-probe contact dots with explicit hydrogen atoms.* J. Mol. Biol., 1999. **285**(4): p. 1711-33.

25.   Rife, J.P., et al., *Comparison of the crystal and solution structures of two RNA oligonucleotides.* Biophys. J., 1999. **76**(1): p. 65-75.

Chapter 5

Accurate RNA tertiary structures and

refinement by discrete molecular dynamics

using a sequence-encoded cleavage agent

5.1 Introduction

RNA molecules play a wide variety of functional roles inside the cell, from the tRNA intermediates that carry amino-acids [1], to complex ribozymes [2], and riboswitches [3]. Knowledge of the underlying structure of the RNA in these and many other elements is a fundamental prerequisite to a complete understanding of the function of RNA. Methods such as X-ray crystallography and NMR spectroscopy offer impressive insight into the details of the structure-function relationships of RNA. However, both methods require highly structured RNAs and NMR is generally limited to small RNAs. Moreover, many important motifs in RNA contain extensive flexible domains that are functionally essential, but render the molecule a poor candidate for structural analysis by these well established methods.

For many classes of RNA, three-dimensional structural model building offers a critically important approach [4-10]. A common theme among these programs is the ability to incorporate both secondary structure information and inter-residue distance constraints based on known tertiary interactions and various biochemical experiments. In general, existing programs either use computationally intensive all-atom reconstruction which limits use to small RNAs or employ overtly simplified models that omit key structural details, require high levels of expert user intervention, and rely extensively on chemical intuition derived from databases of known structures or derived from pre-existing information for tertiary interactions. We sought to create an approach for accurate *de novo* determination of the tertiary fold for large

77

RNAs that does not require expert user intervention nor impose heavy computational requirements and that is efficient for large RNAs.

To successfully reconstruct any RNA tertiary structure, information that defines both the secondary structure and long-range inter-residue distances is required. SHAPE chemistry [11, 12] (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension) is a powerful approach for analyzing the secondary structure, at single nucleotide resolution for RNAs of any length. SHAPE exploits the discovery that the 2'-OH group in unconstrained or flexible nucleotides reacts preferentially towards hydroxyl-selective electrophilic reagents; whereas, paired or constrained nucleotides are unreactive. The resulting reactivities can be used, in concert with a secondary structure prediction algorithm [13], to obtain a accurate secondary structures.

Information regarding the three-dimensional arrangements of RNA helices can be obtained from instructive sets of pair-wise inter-residue distances. The accuracy of any refinement is closely linked to the quality and interpretation of pair-wise constraint information. The most useful experimental information would come from a small, conformationally restricted probe that potentially yields a large number of high-quality constraints per experiment. These requirements are well met by site-directed hydroxyl radical footprinting. Generally, an Fe(II)-EDTA moiety is tethered to an RNA via a unique sulfur or 2'-amine group [14, 15]. However, these derivatization steps are experimentally challenging and time consuming. We

have developed a rapid approach for placing a tethered Fe(II)-EDTA moiety at well-defined sites in a large RNA via site-selective intercalation using the reagent methydium-propyl-EDTA (MPE) [16-18]. MPE binds selectively at a small, but well-defined motif, and thus, many regions in a large RNA can be probed simply by making small changes to the RNA sequence.

We then use discrete molecular dynamics (DMD) [19, 20] to convert this easily obtained secondary and tertiary structural information in terms of a three-dimensional structure. DMD is a special type of molecular dynamics simulation method where pair-wise interactions are approximated by step-wise functions. This approximation enables DMD to sample conformational space more efficiently than the traditional molecular dynamics simulations. We use a coarse-grained RNA model where each RNA nucleotide is represented as three pseudo-atoms representing the phosphate, sugar and base moieties. We show that DMD refinement is both computationally efficient yet yields highly useful representations of RNA. Because the DMD engine is closely tuned to the resolution of the biochemical information, we can probe thus RNA dynamics in a structurally instructive way over timescales largely inaccessible by all atom molecular dynamics simulations.

5.2 Results

The problem of refining accurate three-dimensional RNA structural models based on in-solution biochemical information can be conveniently divided into separate steps for determining the secondary and tertiary

structure. Here, we focus our analyses on refining the structure of yeast tRNA^{Asp}. Throughout this work, all constraints for both the base paired secondary structure and for tertiary interactions are derived from efficiently performed biochemical experiments. No prior assumptions are made based on pre-existing phylogenetic or crystallographic information.

5.2.1. SHAPE accurately determines tRNA^{Asp} secondary structure. The secondary structure of tRNA^{Asp} is readily determined using SHAPE chemistry [11, 21]. SHAPE reactivities are converted to pseudo-free energy charge terms and used to constrain a thermodynamically based minimum free energy prediction algorithm as implemented in the RNAstructure program [22, 23]. For tRNA^{Asp}, SHAPE-constrained prediction yielded a secondary structure that is exactly correct (Figure 5.1B).

5.2.2. Insertion of the MPE binding motif. We used the intercalating cleavage agent, MPE, to obtain comprehensive sets of pair-wise inter-residue distances (Figure 5.1A). MPE preferentially intercalates at CpG steps in RNA at sites adjacent to a single nucleotide bulge [24, 25]. In agreement with previous work [17], we find that MPE preferentially binds at a simple motif comprised of three base pairs and a bulged A. MPE intercalates with a preferred orientation with the Fe(II)-EDTA moiety pointing towards the bulged A. The intercalated MPE [26] occupies approximately the same space in a helix as a single RNA base pair. Thus, the CGAG/C(C/U)G motif, containing the intercalated methidium group, represents roughly four continuously replaced paired nucleotides in RNA (Figure 5.1). In principle, MPE can be
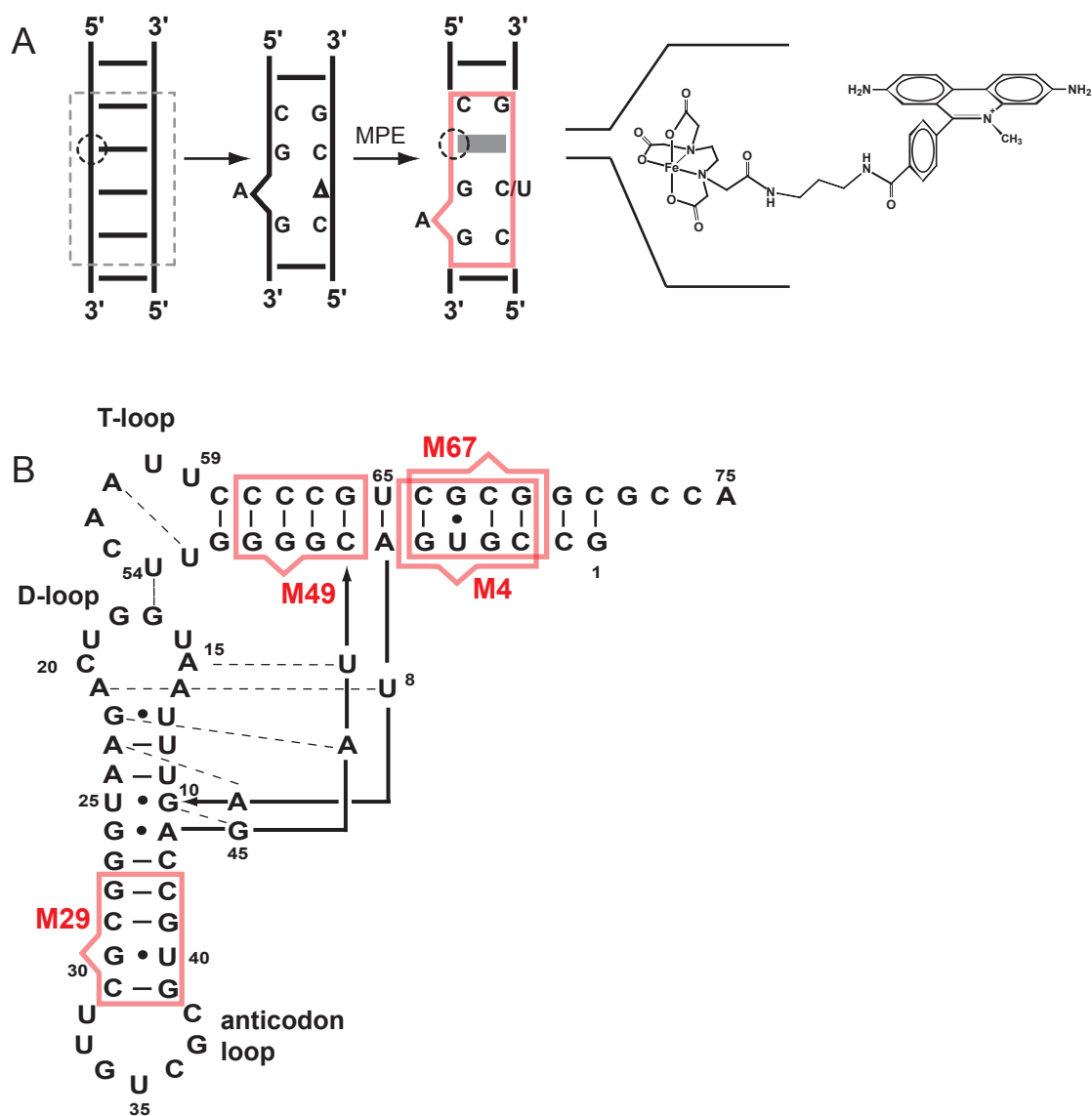
Figure 5.1. Analysis of RNA tertiary structure using a sequence-encoded cleavage agent. (A). Intercalation of MPE in a bulged, three base-pair helix that replaces 4 canonical base pairs (gray dashed square) in a helix (red square). MPE preferentially intercalates (solid gray box) such that the Fe(II)-EDTA moiety is oriented towards the bulged A nucleotide (dashed circle). (B). tRNAAsp secondary structure and design of MPE orientation. MPE binding sites are shown using the scheme from panel A. Mutants are numbered by the site occupied by the Fe(II)-EDTA group.

placed at many helical sites in RNA by replacing four base pairs with the MPE binding site (Figure 5.1A). We took the origin for residue-residue distance measurements to be the base replaced by MPE on the same side as the bulged A (dashed circle, Figure 5.1A). To obtain comprehensive three-dimensional structural constraints for tRNA^{Asp}, we designed four constructs that placed the MPE binding motif in each of the most stable, G-C rich helices in this RNA: in the acceptor stem (M4, M67), in the T stem (M49), and in the anticodon stem (M29) (where Mx indicates the nucleotide corresponding to where Fe(II)-EDTA is placed, Figure 5.1B).

The MPE motif is likely to be an isostructural replacement for many, but not all, four-base-pair sites in folded RNAs. We, therefore, directly tested whether each of these four MPE-binding intercalation sites are compatible with tRNA^{Asp} structure. We first measured the baseline SHAPE reactivity profile for the native sequence RNA. As expected [21], the T-, D- and anticodon loops are reactive, while nucleotides that participate in base pairing and tertiary interactions are unreactive (Figure 5.2, top). We then compared the native sequence SHAPE reactivity profile with the four mutant constructs. For M4, M49 and M67, reactivities were essentially identical with the native sequence. In addition, the position of the bulge A, introduced as part of the MPE binding site, is seen clearly as a single reactive position, exactly at the expected position (see Figure 5.2, central 3 panels). The M29 mutant, however, shows marked differences in SHAPE reactivity profile compared to the native sequence RNA (Figure 5.2, bottom) and clearly does
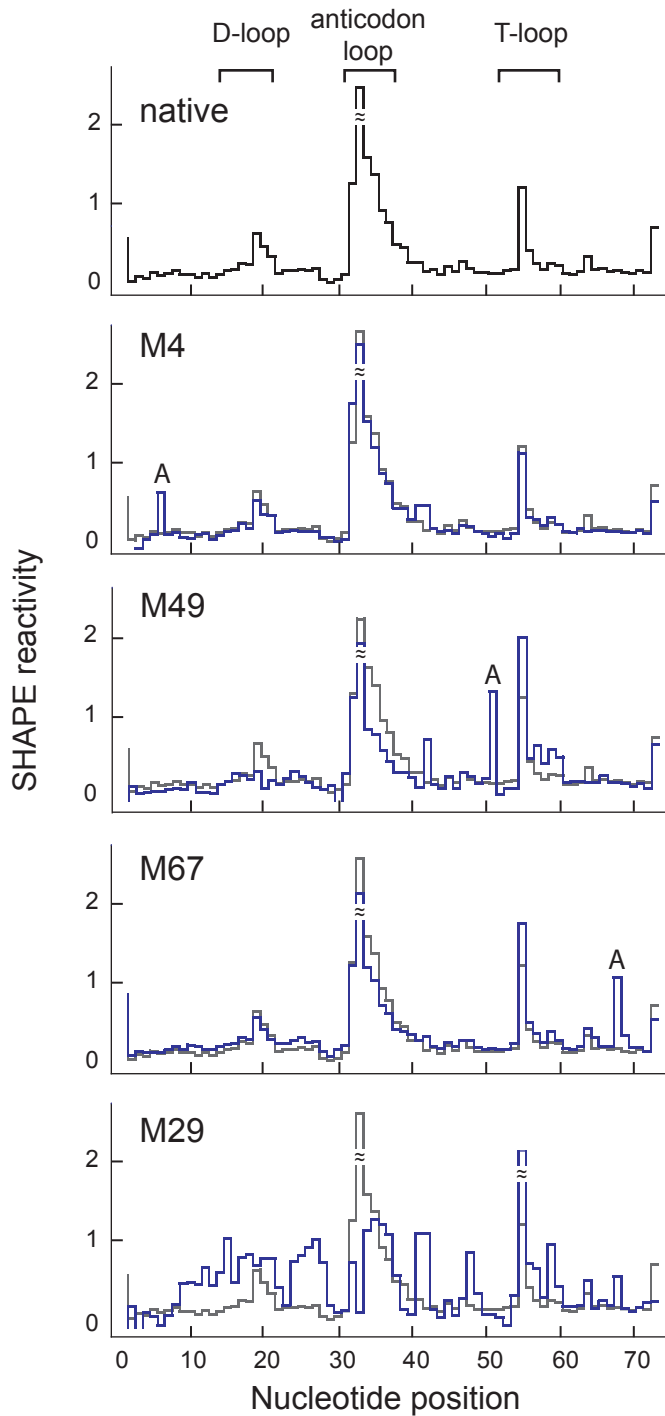
Figure 5.2. SHAPE reactivity profiles for the native sequence tRNAAsp and 4 mutant constructs. The native sequence SHAPE profile is shown as a black line in the top panel and as a gray trace when superimposed on the SHAPE profiles for the mutants (blue lines). The site of the bulged A nucleotide introduced as part of the MPE binding site is indicated explicitly.

not fold to a tRNA-like structure. Thus, the MPE binding site substitutes for many four-base-pair elements in a folded RNA and successful sites are readily identified by SHAPE. We focused our analysis of site-directed hydroxyl radical footprinting experiments on the three RNA constructs that fold to the correct tRNA secondary structure.

5.2.3. MPE-based site-directed hydroxyl radical footprinting for tRNA^{Asp}. In the presence of H_2O_2, the Fe(II)-EDTA moiety catalyses the formation of highly reactive hydroxyl radicals that will cleave proximal regions of the RNA backbone. As expected, the strongest backbone cleavages occur adjacent to the position at which the Fe(II)-EDTA is located in the MPE binding site (see arrows in the left-hand panels; in Figure 5.3). Cleavage intensity (I) at each position was calculated as a ratio relative to the mean value for all intensities (<I>). Cleavage ratios less than 1 were taken to be background. This approach eliminates the requirement for user interpretation of significant cleavages and the need to normalize between experiments.

Superposition of the backbone cleavages on the crystal structure clearly shows that the Fe(II)-EDTA moiety is selectively oriented within the MPE binding site (Figure 5.3, note the distinct pattern of cleavages in constructs M4 and M67, in which MPE is placed in different orientations in the acceptor stem). Critically, all strong backbone cleavages occur in a space around the MPE binding site. These experiments thus yield three sets of high quality, long-range, constraints that can be interpreted using DMD refinement.

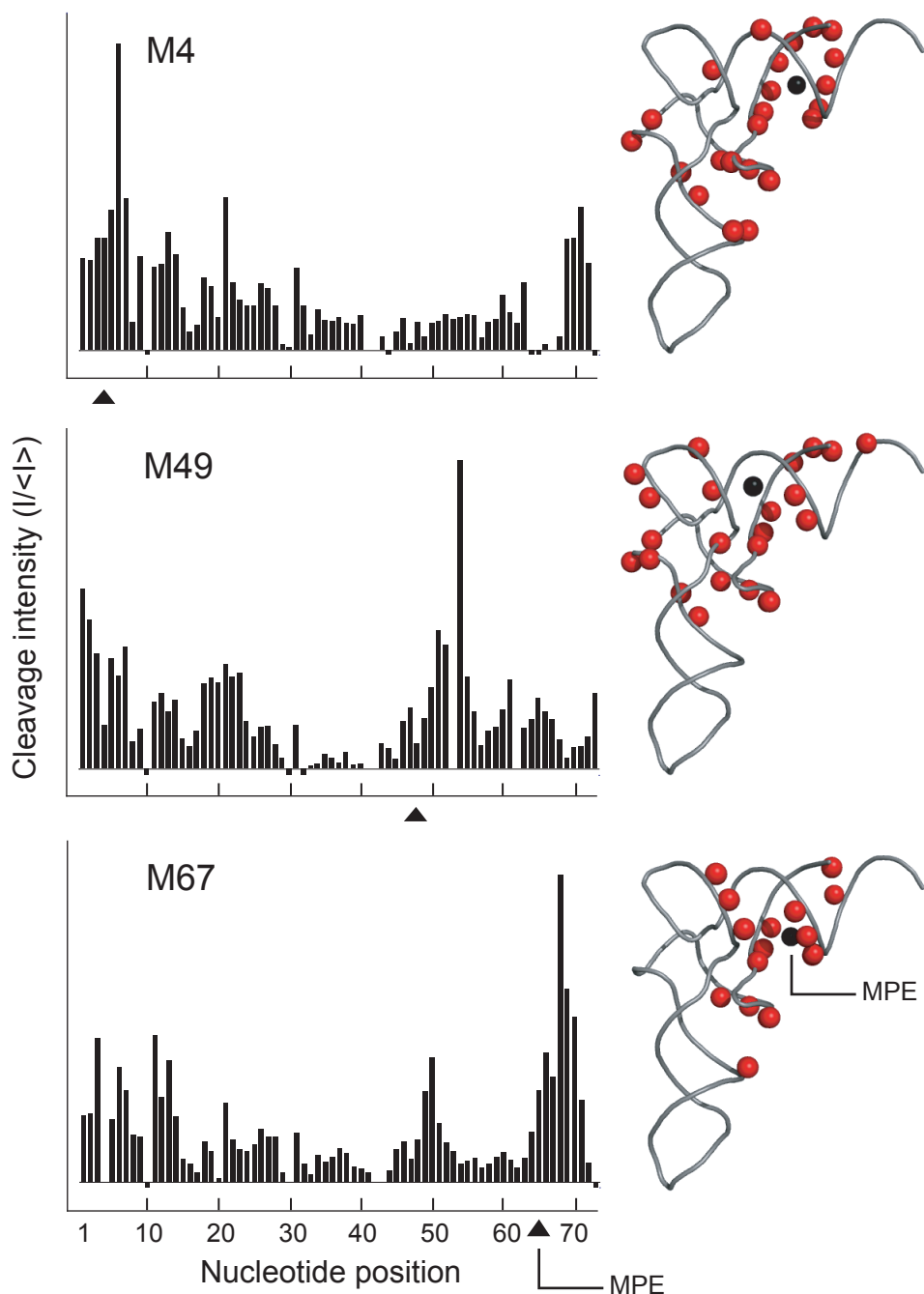Figure 5.3. Site-directed MPE cleavage experiments. MPE cleavage data are illustrated (1) as histograms of intensity versus nucleotide position, after subtracting the background observed for the native sequence RNA and (2) as backbone cleavages (red spheres) superimposed on the tRNAAsp tertiary structure. Positions of the Fe(II)-EDTA cleavage group and MPE insertion sites are indicated as black spheres and triangles, respectively.

Each observed cleavage event was defined as an energy potential well (ε) with a depth quantitatively proportional to the cleavage intensity [ε = ln (I/<I>)], using a fully automated procedure. The wells feature a "soft" energy wall at 25 Å, with smaller energy bonuses extending out to 35 Å (Figure 5.4B). The 25 Å barrier was derived from previous experiments with closely related cleavage agents [14, 27]. Pair-wise distances extending beyond the 25 Å barrier will be unlikely for strong cleavages and deep wells. In contrast, lower intensity cleavages, corresponding to shallower wells, can more readily achieve values greater than 25 Å without large overall energy losses (Figure 5.5). This approach for interpreting the cleavage data has the important advantages that (1) no user input is required to determine that a given cleavage is significant or not and (2) structure refinement is highly tolerant of the measurement errors inherent in any hydroxyl radical footprinting experiment.

5.2.4. Coarse-grained refinement of tRNA^{Asp}. Two classes of information were input into the DMD algorithm, base pairing information as determined by SHAPE chemistry [11, 21] (Figure 5.1B) and long-range constraints derived from site-directed hydroxyl radical footprints (Figure 5.6). Refinement is started using as input the RNA sequence and the base-pairing information. Once all base pairs have formed, the refinement adds weighted potentials for the experimentally-determined long-range constraints. The system is then allowed to cool slowly to a temperature of 0.15, corresponding roughly to room temperature, and finally equilibrated for 10^5 DMD time units
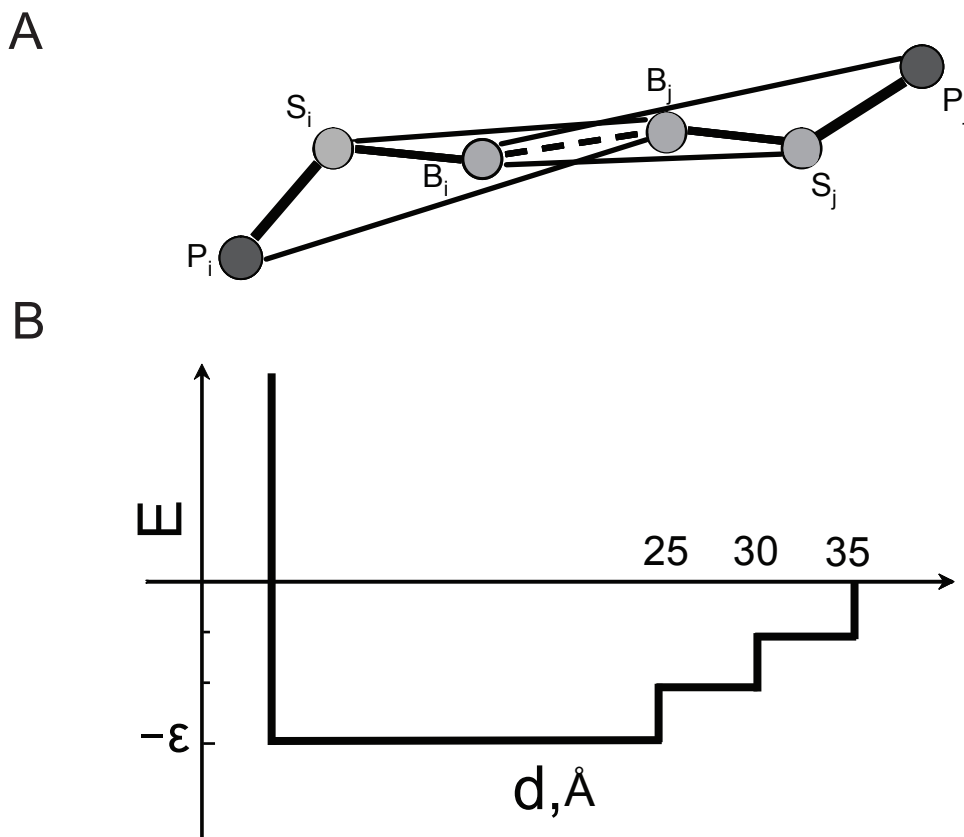
Figure 5.4. Schematic of three-bead construct and energy well design. (A) Three pseudo-atom model in which an RNA nucleotide is represented by base (S), ribose (C) and phosphate (P) groups. Lines indicate pair-wise potentials between atoms: Bi-Bj, base pairing interaction, Bi(j)-Sj(i) interaction energies for inter-nucleotide orientation dependences. The Bi(j)-Pj(i) interaction between base and phosphate14 reflects rigidity of double-helices observed in known RNA structures. (B) Energy diagram for the long-range constraint potential.
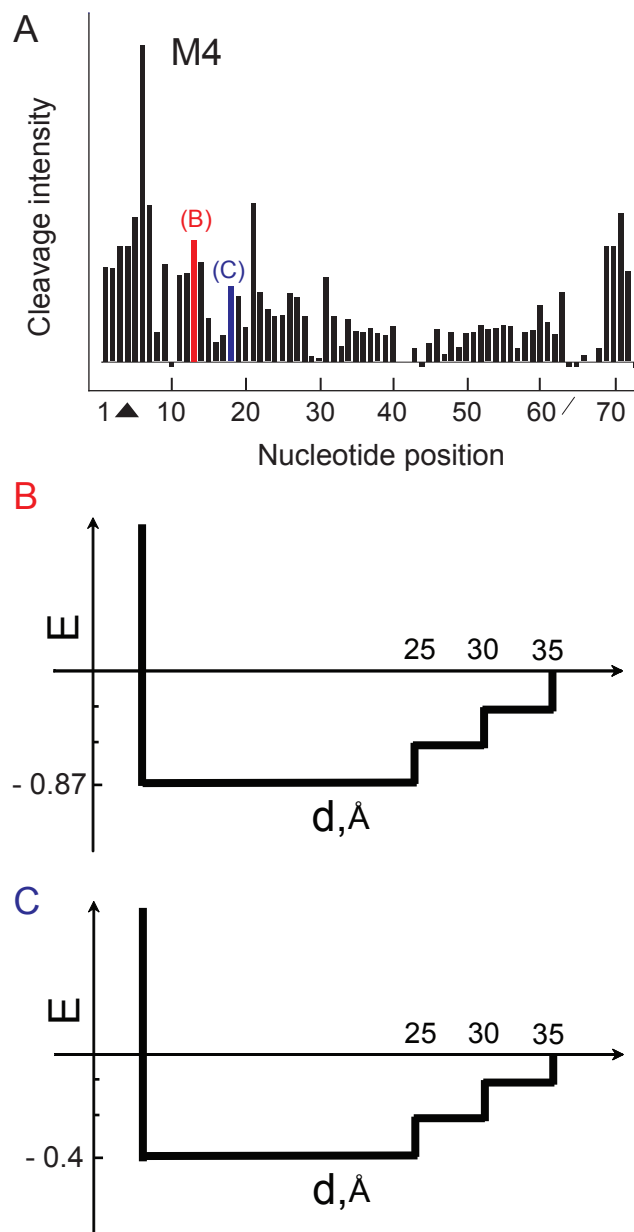
Figure 5.5. Two examples of cleavage reactivity differences translated into different potential wells. Positions 13 and 18 are indicated as red and blue, respectively in panel A. B. The cleavage at position 13, if satisfied during simulation will contribute to the overall energy with -0.87 kcals.mol. C. The cleavage at position 18, if satisfied during simulation will contribute to the overall energy with -0.4 kcals.mol.

Figure 5.6 Refinement algorithm for DMD simulations.

(tu). In DMD simulations, the temperature unit is kcal/(mol*k_B) [19]. We generated one tRNA^{Asp} model for every 10 DMD time units, for a total of 10^4 models per refinement.

Once the DMD simulation is complete, we selected representative structures using an algorithm that (1) preferentially selects low energy states and (2) requires selected structures to be at least 1000 tu apart to prevent analysis of 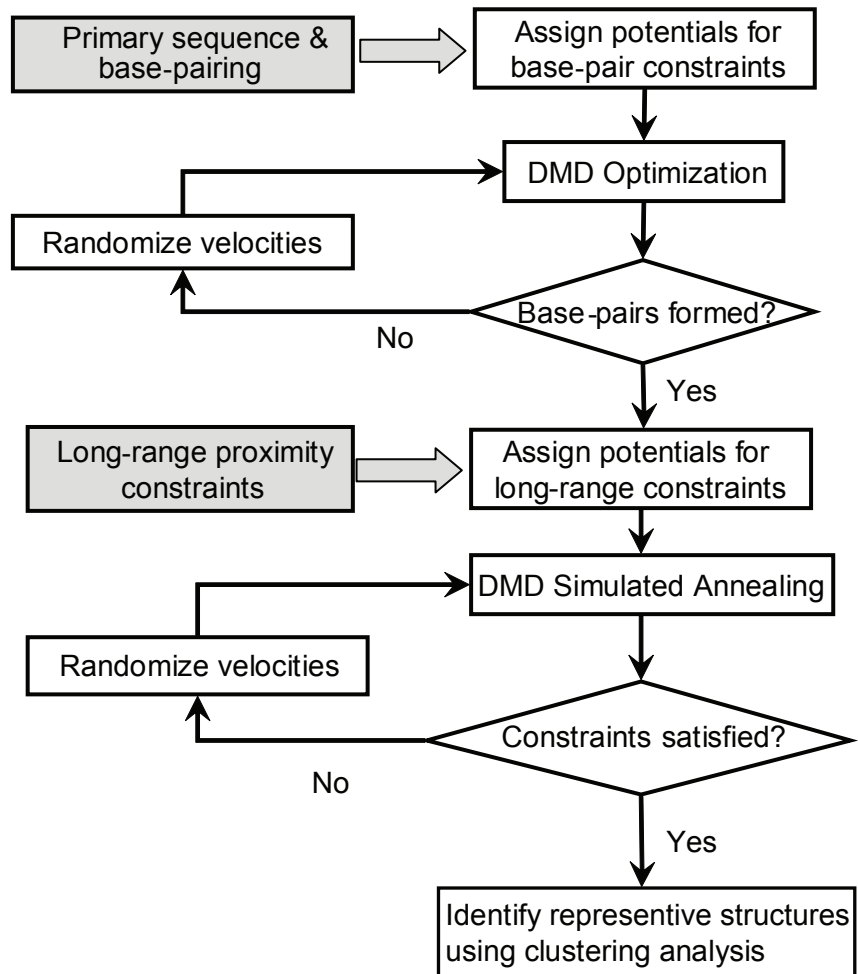consecutive, structurally similar models. This algorithm consistently yielded sets of about 95 representative structures. We identified predominant RNA conformations using distance-based hierarchical clustering using the OC [28] software. Final conformations were divided into 10 clusters, subject to the requirement that structures within a cluster agree to better than 5 Å RMSD (Table 1). From the 10 clusters, we focused on the top 1-2 ensembles. These ensembles always contained more than half of the total representative models. We focused our analysis on the most central structure in each of these 1-2 final clusters. These structures were then reconstructed into all-atom RNA models.

We first assessed the importance of adding long-range distance constraints to the DMD simulations. A baseline DMD simulation of tRNA^{Asp} structure, using only base pairing information and without any long range constraints, yields a structure in which the pair-wise stacking interactions between the acceptor and D-stems and between the anticodon and T-stem form correctly. This accurate reconstruction of stacking interactions reflects that the three pseudo-atom model for RNA and the inter-atom interactions

90

experiment	number of highly populated clusters	RMSD within the cluster (Å)	RMSD compared to the crystal (Å)
M4	1	3.7	4.2
M49	1	3.8	5.0
M67	1	3.7	4.33
M4-49	2	4.0	4.9 & 5.6
M4-67	2	3.3	3.8 & 5.2
M49-67	2	3.1	4.0 & 3.5
M4-49-67	2	3.2	4.0 & 5.4
no constraints	2	4.6	11.2 & 10.5

Table 1. DMD simulations using data from one, any two, or all three experiments.

implemented in our DMD algorithm robustly recapitulates local interactions in RNA. Whereas helix stacking interactions were correctly predicted, tertiary interactions in the T- and D-loops were not predicted and the L-shaped native fold of the tRNA[Asp] was not achieved. In the absence of tertiary constraints, tRNA[Asp] refines to an RMSD of 11.0 Å as compared to the crystal structure (Figure 5.7A). Thus, for an RNA the size of tRNA[Asp], an 11.0 Å RMSD reflects a structure that is not particularly helpful for generating biological hypotheses.

When we implemented the long range-hydroxyl radical footprinting-derived constraints in the DMD simulations, we consistently recapitulated the structure of tRNA[Asp] with RMSD values spanning 3.5 Å to 5.6 Å (Table 1). Since we performed three independent site-directed cleavage experiments, DMD simulations could be performed using long-range constraints obtained from a single experiment or from any combination of experiments, for seven possible total refinements (Table 1). Regardless of the number of experiments used, we obtained highly similar predicted structures. In all cases, both stacking interactions and the native L-shaped tRNA[Asp] molecule were predicted correctly. This striking convergence strongly suggests that, for a 75 nt long RNA, one site-directed hydroxyl radical experiment is sufficient for a high-quality analysis of the overall three-dimensional fold (Figure 5.7B).
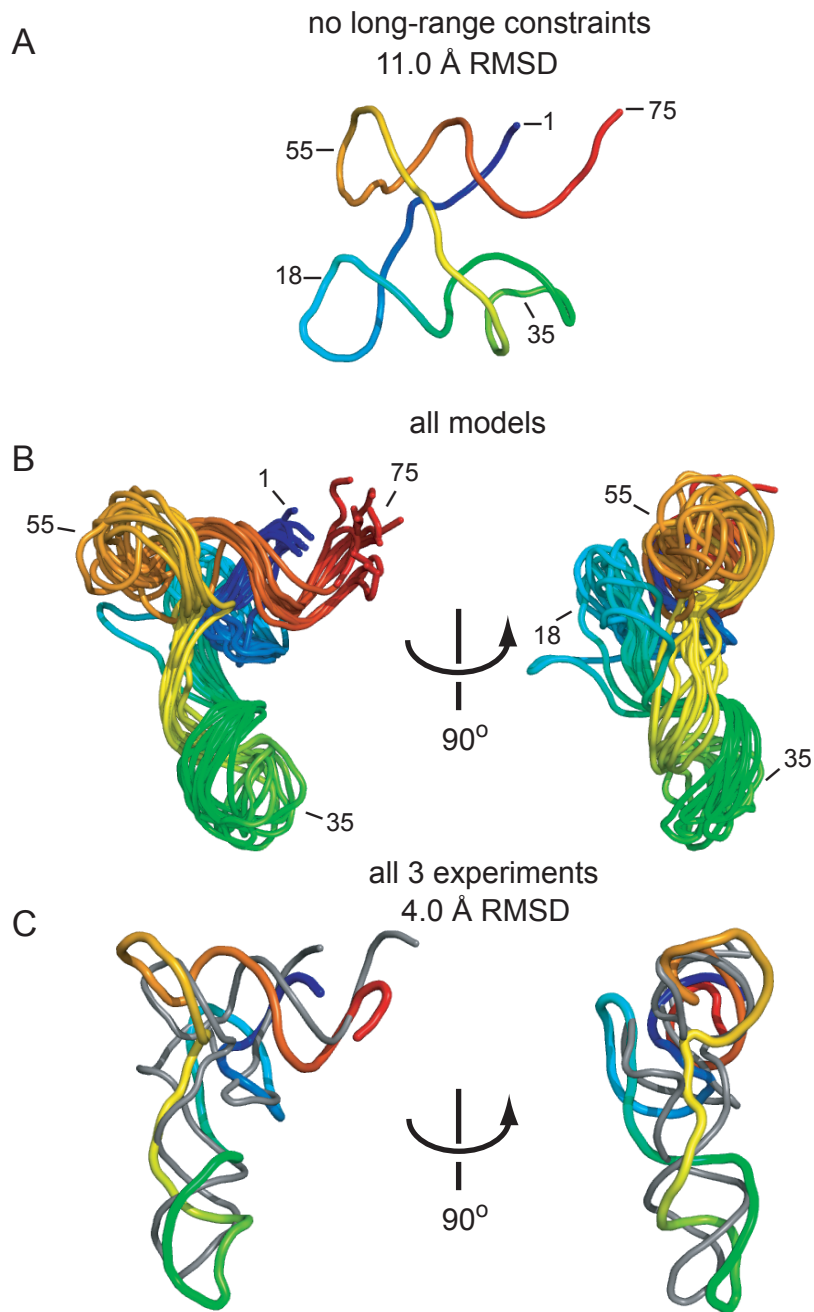
Figure 5.7. Models of the tRNAAsp obtained by blended experimental and computational refinement. (A) tRNAAsp structure predicted in the absence of inter-residue distance constraints. (B). Superposition of the six models refined using either one or two sets of constraints (see Table 1). All models yield self consistent families of structures that correctly recapitulate native-like folds for tRNAAsp. (C) tRNAAsp model using data from all three experiments (rainbow), superimposed on the crystallographic structure for tRNAAsp (in gray).

5.3. Discussion

Using a concise blended experimental and computational approach, we show that it is possible to obtain highly accurate three-dimensional RNA structures for our tRNA^{Asp} test case. No pre-existing or non-experimental information was required. The RNA secondary structure was established in a single experiment using SHAPE chemistry. Multiple high-quality, long-range constraints for the tertiary structure were obtained using site-directed cleavage with MPE. It appears that MPE can be directed to bind at many helical sites within RNA to yield true tertiary structure probing in any RNA.

These experimental data are interpreted using a coarse-grained DMD engine whose underlying model is tuned to the resolution at which biochemical experiments are performed. tRNA^{Asp} structure was consistently refined to an average RMSD of 4.0 Å or better, relative to the crystallographically determined structure (Table 1). Thus, refinement accuracy is consistent with the spatial resolution of a single nucleotide. We anticipate that our tightly melded biochemical and computational approach can be applied to a wide variety of RNAs, including large RNAs and in the absence of prior knowledge of tertiary interactions. If tertiary interactions are known, this information can be readily incorporated into our existing DMD refinement protocol by importing tertiary interaction constraints as either pseudo-base pairs or alternative long-range interactions. Reversal of the three atom coarse grain model to an all-atom model offers impressive local details and may establish a foundation for predicting tertiary interactions not

previously known. Blended biochemical and computational refinement also creates many new opportunities for establishing structure-function relationships for the large universe of RNAs that, due to their inherent flexibility, will never be suitable for analysis by crystallography or NMR.

5.4 Experimental

5.4.1. RNA constructs. RNAs (tRNA^{Asp} native sequence and 4 mutants) were synthesized using T7 RNA polymerase-mediated *in vitro* transcription using a single stranded DNA (IDT) with a double-stranded promoter region [29]. The RNAs were embedded in the context of 5' and 3' structure cassette [21] sequences. RNAs were purified by denaturing polyacrylamide gel electrophoresis (PAGE), excised from the gel, and recovered by electroelution and ethanol precipitation. Purified RNAs were resuspended in TE [10 mM Tris-HCl (pH 8.0), 1 mM EDTA] and stored at -20 ^{o}C.

5.4.2. SHAPE analysis. SHAPE experiments were performed by Christopher Leonard. RNAs were refolded by heating to 95 ^{o}C for 2 min, cooling on ice and incubation at 37 ^{o}C for 10 min in 1x reaction buffer [100 mM NaCl, 50 mM Hepes-NaOH (pH 7.2), 10 mM MgCl$_2$] and then allowed to slowly cool to 22 ^{o}C over 15 min. RNAs were then treated with 1-methyl-7-nitro-isatoic anhydride (1M7; 1 μL at 100 mM; in anhydrous DMSO) and allowed to react for 10 min at 22 ^{o}C. No-reagent control reactions contained 1 μL neat DMSO. Modified RNAs were recovered by ethanol precipitation

[80 μL sterile H$_2$O, 10 μL NaCl (5 M), 1 μL glycogen (20 mg/mL), 400 μL ethanol; 30 min at -80 °C] and resuspended in 5 μL of TE.

5.4.3. Primer Extension. Primer extension reactions were performed using a 5'- [^{32}P]-label primer as described [12], with the exception that the extension reaction was incubated at 52 °C for 7 min. Dideoxy sequencing markers were generated using unmodified RNA. cDNA extension products were separated by gel electrophoresis and visualized by phosphorimaging.

5.4.4. MPE cleavage experiments. MPE experiments were performed by Christopher Leonard. MPE was a generous gift from Peter A. Beal (University of California, Davis). tRNAs at 0.5 μM (either native sequence or mutants containing the MPE binding motif) were refolded as described above with the exception that the reaction buffer contained 100 mM bis-Tris-HCl (pH 7.0) in place of Hepes-NaOH. Unlabeled competitor tRNA^{Phe} (Sigma, R4018) at 7.2 μM was added and incubated with the refolded tRNA^{Asp} at room temperature for 10 min. MPE-Fe(II)-EDTA complex was pre-formed by incubating MPE-EDTA and Fe(II) (1.2:1 ratio) for 5 min. The MPE-Fe(II)-EDTA was then added to the RNA solution and allowed to intercalate for 10 min at 22 °C; control reactions omitted the MPE complex. The cleavage reaction was initiated by addition of DTT (5 mM) and H$_2$O$_2$ (0.1%) and incubated at room temperature for 3 min. Reactions were quenched by adding 80 μL water, 10 μL NaCl (5 M), 1 μg glycogen and 300 μL ethanol (100%). Following ethanol precipitation, RNAs were resuspended in 5 μL ½x TE and subjected to primer extension, as described above. The

resulting cDNAs were separated by denaturing PAGE (8% w/v, 72 W, 2 h: 40 min) and quantified by phosphorimaging.

5.4.5. Data Analysis. For all experiments, individual band intensities for the (+) and (–) reagent reactions were integrated using SAFA. [30] SHAPE reactivity profiles were obtained by subtracting the no-reagent background from the (+) reaction intensities. For the MPE experiments, baseline reactivity for the wild type and each mutant RNA was calculated by subtracting (–) MPE reaction from the (+) MPE reaction. MPE-specific cleavage in the MPE-binding mutants was calculated by subtracting the baseline reactivity for the native sequence RNA from that for each mutant.

5.4.6. DMD refinement. Sequence information and base pairs (as established by SHAPE [11, 21]) were subjected to one round of refinement by DMD [31] (T=0.3, 10^4 tu). After base pair formation was confirmed, long range interactions were added. The RNA was allowed to cool in 5 additional steps: (1) T=0.3, 10^4 tu; (2) T=0.25, 10^4 tu; (3) T=0.2, 10^4 tu; (4) T=0.15, 10^4 tu; and (5) T=0.15, 10^5 tu. One complete three-dimensional refinement of the 75 nt tRNA^{Asp} requires ~2 hrs on an IBM linux workstation (Intel Pentium 4 CPU at 3.2 GHz, Fedora Core 4 OS). Structures from the final trajectory were then filtered as described, yielding about 95 representative structures. These were then clustered and analyzed as described. Structure figures were composed with Pymol [32].

5.5. References

1.   Ibba, M. and D. Soll, *Aminoacyl-tRNAs: setting the limits of the genetic code.* Genes Dev., 2004. **18**(7): p. 731-738.

2.   Strobel, S.A. and J.C. Cochrane, *RNA catalysis: ribozymes, ribosomes, and riboswitches.* Curr. Opin. Chem. Biol., 2007. **11**(6): p. 636-643.

3.   Edwards, T.E., D.J. Klein, and A.R. Ferre-D'Amare, *Riboswitches: small-molecule recognition by gene regulatory RNAs.* Curr. Opin. Chem. Biol., 2007. **17**(3): p. 273-279.

4.   Massire, C. and E. Westhof, *MANIP: an interactive tool for modelling RNA.* J. Mol. Graph. Model, 1998. **16**: p. 197-205.

5.   Wang, R., et al., *Three-dimensional placement of the conserved 530 loop of 16 S rRNA and of its neighbouring components in the 30 S ribosomal subunit.* J. Mol. Biol., 1999. **286**: p. 521-40.

6.   Zwieb, C. and F. Muller, *Three-dimensional comparative modeling of RNA.* Nucleic Acids Symp. Ser., 1997. **36**: p. 69-71.

7.   Macke, T. and D. Case, *Modeling unusual nucleic acid structures*, in *Molecular Modeling of Nucleic Acids*, N. Leontes and J.J. SantaLucia, Editors. 1998, American Cemical Society: Washington, DC. p. 379-393.

8.   Jossinet, F. and E. Westhof, *Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure.* Bioinformatics, 2005. **21**: p. 3320-21.

9.   Major, F., *Building three-dimensional ribonucleic acid structures.* Comput. Sci. Eng., 2003. **5**: p. 44-53.

10.   Yingling, Y.G. and B.A. Shapiro, *The prediction of the wild-type telomerase RNA pseudoknot structure and the pivotal role of the bulge in its formation.* J. Mol. Graph. Model, 2006. **25**: p. 261-74.

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**(12): p. 4223-4231.

12. Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.,* 2006. **1**(3): p. 1610-1616.

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

14. Badorrek, C.S., C.M. Gherghe, and K.M. Weeks, *Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization.* Proc. Natl. Acad. Sci. USA, 2006. **103**(37): p. 13640-13645.

15. Joseph, S. and H.F. Noller, *Directed hydroxyl radical probing using iron(II) tethered to RNA*, in *Methods Enzymol.* 2000. p. 175-90.

16. Vary, C.P.H. and J.N. Vournakis, *RNA structure analysis using Methidiumpropyl-EDTA: Fe(II): A base-pair-specific RNA structure probe.* Proc. Natl. Acad. Sci. USA, 1984. **81**(22): p. 6978-6982.

17. Gooch, B.D., et al., *Binding of helix-threading peptides to E. coli 16S ribosomal RNA and inhibition of the S15-16S complex.* Chem. Bio. Chem., 2005. **6**(12): p. 2247-2254.

18. Dyke, M.W.V., R.P. Hertzberg, and P.B. Dervan, *Map of distamycin, netropsin, and actinomycin binding sites on heterogeneous DNA: DNA cleavage-inhibition patterns with methidiumpropyl-EDTA-Fe(II).* Proc. Natl. Acad. Sci. USA, 1982. **79**(18): p. 5470-5474.

19. Ding, F., et al., *Large scale simulations of 3D RNA folding by Discrete Molecular Dymanics: from structure prediction to folding mechanisms.* RNA, 2008. **14**.

20. Dokholyan, N.V., et al., *Molecular dynamics studies of folding of a protein-like model.* Folding and design, 1998. **3**: p. 577-87.

21. Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transition in tRNA(Asp) transcripts.* J. Am. Chem. Soc., 2005. **127**(13): p. 4659-4667.

22. Deigan, K.E., et al., *Accurate SHAPE-constrained RNA structure prediction.* in preparation, 2008.

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

24.   White, S.A. and D.E. Draper, *Effects of single-base bulges on intercalator binding to small RNA and DNA hairpins and a ribosomal RNA fragment.* Biochem., 1989. **28**(4): p. 1892-7.

25.   White, S.A. and D.E. Draper, *Single base bulges in small RNA hairpins enhance ethidium binding and promote an allosteric transition.* Nucl. Acids Res., 1987. **15**(10): p. 4049-4064.

26.   Jain, S.C. and H.M. Sobell, *Visualization of drug-nucleic acid interactions at atomic resolution.* J. Biomol. Struct. Dyn., 1984. **1**: p. 1179-94.

27.   Maity, T.S., et al., *Compartmentalization directs assembly of the signal recognition particle.* Biochemistry, 2006. **45**(50): p. 14955-14964.

28.   Barton, G.J., *OC-A cluster analysis program.* University of Dundee, Scotland, UK, 1993-2002.

29.   Milligan, J.F., et al., *Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates.* Nucl. Acids Res., 1987. **15**(21): p. 8783-8798.

30.   Das, R., et al., *SAFA: Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments.* RNA, 2005. **11**: p. 344-354.

31.   Ding, F. and N.V. Dokholyan, *Large scale simulations of 3D RNA folding by Discrete Molecular Dymanics: from structure prediction to folding mechanisms.* RNA, *in press*, 2008.

32.   DeLano, W.L., *The Pymol molecular graphics system, DeLano Scientific, South San Francisco, CA, USA. www.pymol.org*