Structural Implications of RNA Dimerization
in a Gamma Retrovirus

By
Christopher S. Badorrek

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
2005

Approved by

Advisor: Kevin M. Weeks
Reader: Linda Spremulli
Reader: Howard Fried
ABSTRACT

Christopher S. Badorrek

Structural Implications of RNA Dimerization in a Gamma Retrovirus  
(Under the direction of Kevin M. Weeks)

Retroviruses selectively package two sense-strand viral genomic RNAs that are non-covalently linked at their 5' ends to form a complex, but structurally unresolved dimer. Determining the molecular basis for dimerization is an initial step in understanding the role of this dimer in retroviral biology and also in devising novel anti-viral and retroviral vector therapies. Initially, I define a 170-nucleotide Minimal Dimerization Active Sequence (MiDAS) for a representative gamma retrovirus, the Moloney murine sarcoma virus, by stringent competitive dimerization. I then analyze, at single-nucleotide resolution, the secondary structure of the MiDAS monomer and final dimer states using quantitative Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE) chemistry. SHAPE analysis characterizes two novel structural features within the monomer and dimer states. (i) Both the monomer and dimer contain partially overlapping, but distinct flexible domains spanning 63 and 58 nucleotides, respectively. (ii) The well conserved SL1-SL2 domain that consists of two GACG-containing stem loop structures undergoes a modest conformational RNA switch in which SL1 is extended by four base-pairs. I then analyze the structural implications of both SL1-SL2 conformations and identify that this domain can mediate
specific interactions between RNA strands in the final dimer by a two-step mechanism. Finally, using site-directed and solvent-based hydroxyl radical footprinting, I show that the MiDAS final dimer adopts a compact three-dimensional structure containing multiple long-range tertiary interactions. Moreover, within the dimer, the SL1-SL2 domain folds autonomously into a tertiary U-shaped structure stabilized by interdigitated interactions.
ACKNOWLEDGEMENT

There are not enough words in the world to thank each and every one of the wonderful people who have supported me on this great adventure. To my beautiful and loving wife, Callie, thank you so much for your love, patience, understanding (especially on all those late night trips to the lab), and support during this trial of my life. I am so very blessed to have you in my life and would not have it any other way. To my Mom and Dad, I will always be thankful for all the drive and support you have provided me through the years of my life and education. It warms my heart greatly to know that I have put a big smile of pride on your faces. To Kevin Weeks, thank you for the opportunity to work on such a challenging and rewarding project. You showed me that I could go above and beyond what I initially thought of as my limits and for that I will always be eternally grateful. To all the great Weeks Lab members both past and present, thank you all for the great science talks, the laughs we have shared, the demise that we planned for the world, the racquet ball games, the trips to the pool, and who can forget the great lab outings, especially to He’s Not Here. Every one of you are wonderful individuals and I hope and wish the best for all of you. Lastly, I will leave you all on one final note and smile: “Victory is mine.”
# TABLE OF CONTENTS

List of Tables……………………………………………………………………………………...x

List of Figures……………………………………………………………………………………..xi

Abbreviations………………………………………………………………………………………xiii

Chapter 1  Introduction…………………………………………………………………………1

1.1. Introduction………………………………………………………………………………..2

1.1.1. Retroviral Life Cycle…………………………………………………………………2

1.1.2. Retroviral RNA Dimerization in the DLS………………………………………..4

1.1.3. Mapping Structural Consequences of Retroviral RNA
Dimerization in the MIDAS……………………………………………………………7

1.1.4. An RNA Switch Enforces Stringent Retroviral Genomic
RNA Dimerization………………………………………………………………………12

1.1.5. Summary of Chapters……………………………………………………………..13

1.1.6. Conclusion…………………………………………………………………………14

1.2. References…………………………………………………………………………………15

Chapter 2  RNA Flexibility in the Dimerization Domain of a Gamma
Retrovirus………………………………………………………………………………….18

2.1. Introduction……………………………………………………………………………..19

2.2. Results……………………………………………………………………………………21

2.2.1. Rigorous Definition of a MiDAS………………………………………………………21

2.2.2. A Minimal Sequence Active in Dimerization……………………………………24

2.2.3. Deletion of PAL2 Unmasks the Contribution of PAL1…………………27
2.2.4. A Minimal Dimerization Domain……………………………………..28

2.2.5. MiDAS Monomeric Structure Analyzed by RNA SHAPE Chemistry………………………………………………………………………………30

2.2.6. PAL2 is Unstructured in the MiDAS Monomer……………………………………..35

2.3. Discussion ……………………………………………………………..38

2.4. Methods and Materials…………………………………………………44

2.4.1. Retroviral RNA Transcripts………………………………………………44

2.4.2. Competitive Dimerization Assay……………………………………………….44

2.4.3. Time-resolved Dimerization Assay……………………………………45

2.4.4. SHAPE Analysis of Monomer MiDAS……………………………………45

2.4.5. Primer Extension………………………………………………………..46

2.5. References…………………………………………………………………47

Chapter 3 Quantitative Structure of a Retroviral RNA Dimer…………………..…...50

3.1. Introduction……………………………………………………………..…...51

3.2. Results…………………………………………………………………..…...54

3.2.1. SHAPE Analysis of the Retroviral Dimer……………………………54

3.2.2. Secondary Structure Model for the MiDAS Domain in The Final Dimer State……………………………………………………………………………….57

3.2.3. Long-range Interactions in The MiDAS Final Dimer…………………..60

3.2.4. Tertiary Structure in the MiDAS Dimer……………………………..62

3.3. Discussion…………………………………………………………………64

3.4. Methods and Materials…………………………………………………..68

3.4.1. Retroviral RNA Transcripts……………………………………………..68

3.4.2. SHAPE Analysis of Monomer and Dimer MiDAS……………………68
Chapter 4

An RNA Switch Enforces Stringent Retroviral Genomic RNA Dimerization

4.1. Introduction ................................................................. 74

4.2. Results ........................................................................ 76

4.2.1. The SL1-SL2 Domain Undergoes a Conformational Switch Upon Dimerization .............................................. 76

4.2.2. The SL1-SL2 Domain is an Autonomous Dimerization Motif ................................................................. 79

4.2.3. Architecture of SL1-SL2 Domain ........................................................................................................ 84

4.2.4. Tertiary Structure in the SL1-SL2 Domain ......................................................................................... 88

4.3. Discussion ....................................................................... 90

4.4. Methods and Materials ..................................................... 95

4.4.1. Retroviral RNA Constructs .............................................. 95

4.4.2. SHAPE Analysis of MiDAS and SL1-SL2 RNA ......... 95

4.4.3. Concentration-dependent Dimerization of SL1-SL2 ...... 96

4.4.4. 310-BABE and 336-ITE RNAs .................................... 96

4.4.5. Site-directed Hydroxyl Radical Cleavage ......................... 97

4.4.6. Solvent-based Hydroxyl Radical Cleavage ....................... 97

4.4.7. Model Building and Refinement .................................... 98

4.5. References ...................................................................... 100
Appendix A.................................................................................................................103

A.1. Using the Probability Drop Off Equation..............................................104

A.2. Derivation of Fraction RNA Dimer Equation........................................106
LIST OF TABLES

Table 2.1  Dimerization activity of 3’ and 5’ truncation mutants…………………………26
Table 4.1  Binding affinities (nM) for SL1-SL2 interactions in the initial monomer-like versus final dimer conformations…………………………83
LIST OF FIGURES

Figure 1.1. Retroviral life cycle..............................................................................3
Figure 1.2. Schematic of MuSV unspliced genomic RNA framework......................5
Figure 1.3. Conventional RNA secondary structure of a representative portion of the dimer linkage sequence (DLS).........................................................6
Figure 1.4. Exploratory Pb2+ cleavage probing data superimposed on the conventional secondary structure of the minimal dimerization active sequence (MiDAS, nt 205-374) in monomer form.................................8
Figure 1.5. Mechanism for RNA SHAPE chemistry....................................................10
Figure 1.6. Secondary structural models for monomer and final dimer MiDAS states.................................................................11
Figure 2.1. 5’Untranslated region of MuSV..............................................................22
Figure 2.2. Competitive dimerization assay for stringent definition of RNA structures essential for dimerization...............................................................23
Figure 2.3. The minimal dimerization active sequence for MuSV defined by competitive dimerization.................................................................25
Figure 2.4. Comparison of RNA Dimerization between two viral RNA constructs (nts 205-374 and 276-374)...............................................................29
Figure 2.5. SHAPE analysis of the MuSV MiDAS RNA and of the PALSTB and Δ289-300 mutants.................................................................32
Figure 2.6. Secondary structure model of the monomeric MuSV MiDAS RNA............34
Figure 2.7. Quantitative difference maps for the effects of the PALSTB and Δ289-300 mutations on MiDAS structure........................................................36
Figure 2.8. Proposed secondary structures for MuLV and HaSV dimerization domains.................................................................39
Figure 2.9. Structural model for overall flexibility in the 231-315 domain..................41
Figure 3.1. RNA dimerization by the MiDAS domain of a gamma retrovirus................52
Figure 3.2. SHAPE analysis of MiDAS RNA in monomer and dimer states..............55
Figure 3.3. Secondary structure model for MiDAS RNA in final dimer state..............58
Figure 3.4. Site-directed hydroxyl radical cleavage of the final dimer.........................61
Figure 3.5. Protection from solvent-based hydroxyl radical cleavages in the final dimer.................................................................63
Figure 3.6. Comparison of hydroxyl radical protection for the isolated SL1-SL2 domain (Chapter 4) and for this domain in the context of the intact MiDAS RNA in the final dimer state..............................................66
Figure 3.7. Architecture of a gamma retroviral RNA dimer..........................................67
Figure 4.1. Conformational switch in the SL1-SL2 domain during retroviral RNA dimerization, defined by RNA SHAPE chemistry.................................75
Figure 4.2. SHAPE analysis of the MiDAS RNA in starting monomer-like and final dimer conformations and of a simplified SL1-SL2 domain RNA in the final dimer state.................................................................78
Figure 4.3. Structure of the loop-loop interaction formed between GACG sequences .............................................................80
Figure 4.4. Dimerization specificity of the SL1-SL2 domain in monomer-like versus final dimer conformations .................................................................81
Figure 4.5. Architecture of the SL1-SL2 interaction in the final dimer conformation mapped by site-directed hydroxyl radical footprinting.....................................................85
Figure 4.6. Global architecture of the SL1-SL2 interaction in the monomer-like conformation mapped by Fe(II)-BABE mediated site-directed hydroxyl radical footprinting.........................................................87
Figure 4.7. Solvent-based hydroxyl radical footprinting of the SL1-SL2 domain in the final dimer conformation.....................................................89
Figure 4.8. Refined model of the SL1-SL2 domain in the final dimer state based on secondary structure and long-range site-directed cleavage restraints.................................................................91
Figure 4.9. Refined model and two-step assembly of the SL1-SL2 domain..............92
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-[32P]-ATP</td>
<td>ATP with $^{32}$P at alpha position</td>
</tr>
<tr>
<td>A</td>
<td>adenine</td>
</tr>
<tr>
<td>Å</td>
<td>angstrom (10^{-10} meters)</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>BABE</td>
<td>bromoacetamidobenzyl-EDTA</td>
</tr>
<tr>
<td>BME</td>
<td>2-mercaptoethanol</td>
</tr>
<tr>
<td>°C</td>
<td>degrees centigrade</td>
</tr>
<tr>
<td>C</td>
<td>cytidine</td>
</tr>
<tr>
<td>Ci</td>
<td>curie</td>
</tr>
<tr>
<td>cm</td>
<td>centimeter</td>
</tr>
<tr>
<td>CTP</td>
<td>cytidine triphosphate</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>delta (region deleted)</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>ds</td>
<td>double stranded</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>$\gamma$-[32P]-ATP</td>
<td>ATP with $^{32}$P at gamma position</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>G</td>
<td>guanosine</td>
</tr>
<tr>
<td>GMP</td>
<td>guanosine monophosphate</td>
</tr>
<tr>
<td>GMPS</td>
<td>guanosine monophosphorothioate</td>
</tr>
</tbody>
</table>
GTP  guanosine triphosphate
HEPES 4-(2-hydroxy-ethyl)-1-piperazine-ethanesulfonic acid
H₂O  water
hr  hour
kcal  kilocalorie
K₀  equilibrium binding constant
ITE  isothiocyanobenzyl-EDTA
µg  microgram
µl  microliter
µM  micromolar
µmol  micromole
Mg²⁺ magnesium ion
MgCl₂  magnesium chloride
mL  milliliter
mM  millimolar
mol  mole
NaCl  sodium chloride
NaOH  sodium hydroxide
nt  nucleotide
PCR polymerase chain reaction
PNK  polynucleotide kinase
RNA  ribonucleic acid
sec  second
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>thymidine</td>
</tr>
<tr>
<td>Taq</td>
<td><em>Thermus aquaticus</em></td>
</tr>
<tr>
<td>TBE</td>
<td>89 mM tris-borate, 2mM EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>10 mM Tris (pH 7.5), 1mM EDTA</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>U</td>
<td>uridine</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine triphosphate</td>
</tr>
<tr>
<td>V</td>
<td>voltage</td>
</tr>
<tr>
<td>W</td>
<td>watt</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
1.1 Introduction

1.1.1 Retroviral Life Cycle. During the early phases of the viral replication cycle, a mature viral particle introduces its genomic components, consisting of two sense-strand RNAs, into a host cell. The RNA strands are then reverse transcribed into viral DNA and integrated into the host genome where the virus can remain in a dormant state for long periods. The host is later stimulated to transcribe the virally encoded RNA. There are two distinct fates for the viral RNAs. They may be either spliced into mRNA coding for viral proteins or remain unspliced to function as new retroviral genomic RNA. Interestingly, two genomic RNA strands become linked at their 5' ends to form an RNA dimer (Fig. 1.1, highlighted in blue box) that has a significant structural role in various stages of the viral cycle, including recombination during reverse transcription and RNA packaging. In viral RNA packaging, only the RNA dimer structure (not the monomer!) is recognized by the nucleocapsid domain of the viral gag protein. Thus, RNA dimerization appears to be a primary mediator used in selectively packaging only the viral genome into nascent viral particles (Fig. 1.1). Notably, the viral RNA dimer is thought to be packaged in an “immature” initial conformational state. Then, in the newly released virion, many viral proteins, like the gag, are digested and the RNA dimer folds into a final “mature” conformation.

This thesis focuses on (i) extrapolating the minimal viral RNA sequence required to drive the dimerization process and (ii) solving the underlying monomeric and dimeric RNA structures formed within that sequence. Retroviruses are the causative agents of many diseases, such as acquired immunodeficiency syndrome and several cancers, and recently have become useful gene therapy vectors. Moreover, retroviral RNA dimerization is well
Figure 1.1. Retroviral life cycle. The viral diploid genome is introduced into the host cell (step 1). After reverse transcription (step 2), the viral DNA is integrated (step 3) into the host genome. Newly transcribed viral RNA (step 4) can then follow two pathways: (i) be spliced to encode for viral proteins or (ii) remain unspliced as genomic RNA. Two unspliced “genomic” RNAs then undergo RNA dimerization (step 5) at their 5’ ends to form a dimer. Viral proteins recognize and interact with the dimer and undergo vesicular transport to the cell surface where budding occurs (step 6). Finally, the newly formed virion matures (step 7) and can re-initiate the cycle.
conserved among retroviruses. Thus, understanding the structural basis for viral genomic RNA dimerization can yield novel opportunities for generating anti-viral therapies and enhancing vector function.

The Moloney murine sarcoma virus (MuSV), a gamma retrovirus, was chosen as a model since it contains key features conserved among retroviruses. The MuSV sequence is almost identical to that of the well-studied Moloney murine leukemia virus (MuLV), a retrovirus that can heterodimerize with the human immunodeficiency virus (HIV), indicating a possible structural similarity. The MuSV retroviral genomic framework is comprised of 2 non-coding regions: 5’ and 3’ untranslated regions (UTR) and 3 coding regions: gag, pol, and v-mos (Fig. 1.2). When the viral genomic RNA is spliced, the three coding regions are translated into the viral proteins required for the viral life cycle. However, if the RNA is not spliced, a 454 nt dimer linkage sequence (DLS) located in the 5’UTR (Fig. 1.2) interacts with another DLS of a second viral genomic RNA strand to form a dimeric complex. The mechanism that directs the genomic RNA to be spliced or remain in the unspliced genomic state is currently unresolved.

1.1.2 Retroviral RNA Dimerization in the DLS. At physiological temperatures, the nucleocapsid domain of the gag precursor protein is required to drive dimerization of the full length DLS. However, self-dimerization of retroviral RNA is possible for in vitro experiments at 60 °C in relevant ionic conditions, thus setting a precedent for facile analysis of retroviral RNA dimerization in vitro.
Figure 1.2. Schematic of MuSV unspliced genomic RNA framework. U5 and U3 are untranslated regions of the RNA. The region involved in retroviral RNA dimerization resides within the U5 and is termed the dimer linkage sequence (DLS). Viral protein encoding sites of the RNA include the gag, pol, and v-mos; gag for nucleocapsid proteins, pol for reverse transcriptase, integrase, and protease, and v-mos for a mos oncogene (serine/threonine kinase).
Figure 1.3. Conventional RNA secondary structure for a representative portion of the dimer linkage sequence. (a) Regions thought to be involved with RNA dimerization are highlighted with boxes. (b) Conventional PAL2 loop-loop interaction and intermolecular extended duplex formation.
Previous *in vitro* chemical mapping studies were used to propose what is now the conventional secondary structure for the DLS. The conventional RNA secondary structure of a representative portion for MuSV DLS is shown in Figure 1.3a. Multiple regions within the DLS have been implicated in dimerization (see Fig. 1.3a, highlighted in boxes). PAL and PAL2 are self-complementary sequences that are proposed to interact with PAL1 and PAL2 sequences of a second RNA and form intermolecular extended duplexes (Fig 1.3b). SL1 and SL2 contain well conserved GACG tetraloops that form loop-loop interactions and, along with PAL2, are proposed to form a structure recognized by the nucleocapsid domain of the gag precursor protein. G5 and G3 are purine rich sequences that can potentially form G-quartet motifs to stabilize the dimer structure.

Various regions throughout the DLS have been implicated in dimerization, but their specific involvement in the process was still unclear. Thus, my first goal was to resolve the minimal RNA sequence in the DLS that mediated retroviral RNA dimerization. In Chapter 2, I define a 170-nucleotide minimal dimerization active sequence (MiDAS, nts 205-374) for MuSV by stringent competitive dimerization with the full length DLS at 60 °C in near-physiological ionic conditions.

1.1.3 Mapping Structural Consequences of Retroviral RNA Dimerization in the MiDAS. In exploratory experiments, I used Pb2+-mediated cleavage (which targets single-stranded regions in RNA) to analyze the monomeric structure of the MiDAS RNA and found that large regions were relatively unstructured (Fig. 1.4). Surprisingly, these studies correlated poorly with current working models for this secondary structure, especially in the PAL2 hairpin. However, Pb2+ cleavage chemistry is well known to also target divalent metal
Figure 1.4. Exploratory Pb$^{2+}$ cleavage probing data superimposed on the conventional secondary structure of the minimal dimerization active sequence (MiDAS, nt 205-374) in monomer form. Denatured and natively folded monomeric MiDAS (Chapter 2) were treated with 2 mM Pb$^{2+}$ acetate at 37 °C for 5 min. All RNA sites cleaved by Pb$^{2+}$ in the denatured control are circled. Circles signify regions where cleavage was either identical (gray) or reduced (black), respectively, in the monomer as compared to the denatured control.
binding sites in RNA\textsuperscript{26} making it difficult to distinguish between a single-stranded region or divalent metal binding site. Therefore, no final conclusions were made, but this study hinted at the existence of a much more flexible RNA than previously proposed for the dimerization domain. Thus, I sought another chemical and/or enzymatic technique to corroborate these results.

At this time, Edward Merino, Ph.D., Kevin Wilkinson, and Jennifer Coughlan in the Weeks laboratory had been developing a novel chemical mapping method for probing RNA structure at single nucleotide resolution using quantitative Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE)\textsuperscript{27}. The reactivity of N-methylisatoic anhydride (NMIA) at the 2'-hydroxyl position of the RNA backbone is correlated directly to the local nucleotide flexibility. That is, NMIA reactivity is high in flexible (single-stranded) positions of the RNA and low in constrained (base-paired) positions (Fig. 1.5). Moreover, by targeting the 2' hydroxyl position of the ribose moiety in the backbone, this chemistry is not biased to any specific nucleotide as seen in many traditional chemical and/or enzymatic probing technologies. Thus, throughout this thesis, I applied this methodology to understand the secondary structural features of both the monomer and final dimer MiDAS conformations (Chapter 2 and Chapter 3).

In chapter 2, after defining the MiDAS, I analyzed the monomer structure at every nucleotide using SHAPE. Notably, SHAPE analysis confirmed the proposed structures for the PAL1 loop, top of SL1, and SL2. Strikingly, though, I observed that PAL2 does not exist in the conventional stem loop motif, but instead lies in a larger flexible domain (Fig. 1.6a, nts 242-304). RNA flexibility in this domain is quite extensive and is apparently maintained
**Figure 1.5.** Mechanism for RNA SHAPE chemistry. Formation of an oxyanion at the 2'-hydroxyl position is gated by its proximity to the 3'-phosphodiester anion. At constrained (base-paired) sites, the negatively charged 2'-oxyanion and 3'-phosphodiester positions would be near in proximity, thus the formation of the 2’-oxyanion would not be favorable. But, at flexible (single-stranded) sites, the nucleophilic oxyanion is readily available to react with NMIA and form the bulky 2’-ester adduct. Sites of 2’-O-adduct formation are then probed using primer extension.
Figure 1.6. Secondary structural models for (a) monomer and (b) final dimer states in the MiDAS. The self-complementary sequences PAL1 and PAL2 (colored red and blue, respectively) form extended heteroduplexes with the second RNA upon dimerization. Notably, as discussed in Chapter 4, an RNA conformational switch also occurs in which nts 310-313 and 349-352 (colored green) interact and extend the base of SL1 by four nucleotides.
through the interconversion of multiple structures, all of which may have distinct, yet unresolved, dimerization activities.

As discussed in Chapter 3, portions of the initial flexible domain become structured, while previously base-paired elements become incorporated into a new flexible domain in the final dimer state (Fig. 1.6b, nts 220-281). Much like the original, this new flexible domain also interconverts between structures and can form a two stem loop motif or even a pseudoknot. As expected, the self-complementary PAL1 and PAL2 sequences form extended duplexes with PAL1 and PAL2 sequences from the second RNA (Fig. 1.6b, colored red and blue, respectively) during dimerization. Interestingly, as the anchoring helix (nts 231-241, 305-315) breaks, the SL1 extends by four base-pairs (Figure 1.6b, colored green). Thus, the SL1-SL2 domain undergoes an RNA conformational switch.

1.1.4 An RNA switch enforces stringent retroviral genomic RNA dimerization.

The conventional model for retroviral dimerization suggests that (i) only the final dimer is recognized for RNA packaging in excess of a large background of cellular RNAs and (ii) self-complementary sequences, like PAL1 and PAL2, are the main contacts in that final dimer9,15-18. Thus, logic would dictate that heteroduplex formation at PAL1 and PAL2 should mediate the selective recognition and dimerization for viral genomic RNAs that are to be packaged. But, simple duplex formation, like at PAL1 and PAL2, is known to have poor sequence selectivity since mismatch and perfect duplexes can readily compete with one another28,29. Moreover, in vivo studies30,31 show that mutations in PAL1 and PAL2 do not hinder viral RNA packaging. Thus, it appears that another region of the MiDAS, rather than
PAL1 and PAL2, must mediate the selective interactions that are essential for viral RNA recognition.

In Chapter 4, I show that the SL1-SL2 domain undergoes a conformational RNA switch and extends the base of SL1 (Fig. 1.6b, green) upon dimerization. Therefore, this domain has two different conformations: an initial monomer-like and a final dimer. Both of these stem loop structures can form cross-strand loop-loop interactions with the SL1-SL2 domain of the second RNA. Strikingly, the initial monomer-like structure (see SL1-SL2 domain in Fig. 1.6a) forms highly selective cross-strand loop-loop interactions with the second viral RNA. But, in the final dimer structure (see Fig. 1.6b), these same interactions convert to a higher affinity, but less selective form.

Thus, viral RNA recognition in MuSV appears to be selectively regulated by a two-step mechanism that advantageously uses two conformations of the SL1-SL2 domain. In the initial stages of dimerization, the virus uses the stringent selectivity of the monomer-like structure to identify another viral RNA and form an interaction. Then, immediately upon recognition and initial dimerization, the SL1-SL2 domain switches to a higher affinity structure that would stabilize the interaction and prevent loss of the viral RNA strand.

1.1.5 Summary of Chapters. In Chapters 2 and 3, I apply two ideas that can be used in any RNA structure prediction. First, I used a competitively rigorous method to identify a minimal sequence of viral RNA that natively folds like the full length. I then interrogated the flexibility about every position of the RNA in both monomer and dimer state using a much higher resolution SHAPE chemistry. SHAPE chemistry provides quantitative, model-independent data that can be used to constrain secondary structure prediction algorithms, like
RNAstructure. In Chapters 3 and 4, I then show that retroviral RNA dimerization should no longer be regarded as a collection of linked stem-loop and helical structures, but instead as a complex RNA folding problem. Site-directed and solvent-based hydroxyl radical footprinting data indicate that the final dimer structure folds into a compact three-dimensional shape.

1.1.6 Conclusion. Retroviral dimerization is a well conserved event among retroviruses and thus is not only a potential target for disruption in anti-viral therapies, but is monumental for understanding the retroviral life cycle. However, when I first initiated this project, the RNA elements involved and the structural consequences associated with retroviral dimer formation were unresolved. Thus, in this thesis, I have created a generalizeable approach for characterizing the dimerization domain for any retroviral RNA model. Then, I further show that after resolving the dimerization domain, high throughput RNA SHAPE chemistry can be employed to authoritatively interrogate and define the structural features associated with both monomer and dimer states of the virus. My structural analysis experiments suggest that existing secondary structure models that have led the gamma retroviral field for the past ten years require careful reinterpretation. Thus, I have laid a new solid foundation that researchers can use to create useful theoretical proposals in regards to retroviral RNA dimerization and structure.
1.2 References


Chapter 2

RNA Flexibility in the Dimerization Domain of a Gamma Retrovirus
2.1 Introduction

Retroviruses selectively package two sense-strand genomic RNAs. These genomic RNAs are linked together near their 5' ends \(^1\) by a precise, but poorly understood, set of non-covalent interactions. The structure of this RNA "dimer" is implicated in multiple stages of the retroviral infectivity cycle, including RNA encapsidation into nascent viral particles \(^2-5\) and recombination during reverse transcription \(^3,6,7\). Retroviruses are both valuable biotechnology tools, as gene therapy vectors, and are the causative agents of serious diseases, including the acquired immunodeficiency syndromes and several cancers. Understanding the mechanism of retroviral dimerization thus represents an important opportunity for both enhancing vector function and for disrupting the infectivity cycle of pathogenic viruses.

Because the retroviral dimerization sequences are highly conserved \(^8-10\), a consensus secondary structure cannot be inferred by phylogenetic covariation analysis, which is the most robust method to determine the secondary structure for a large RNA \(^11-14\). Current secondary structure models for retroviral RNA dimerization domains are still provisional and likely only partially encompass the biological function of these RNAs. Determining the biologically relevant structure of the dimerization domain for any retroviral RNA is thus representative of a broad class of problem in biology. This challenge is to understand an RNA secondary structure in enough detail to be able to formulate hypotheses about biological function, even though only one or a few highly similar sequences are known.

Current algorithms for predicting an RNA secondary structure from a single sequence identify roughly 50-70% of known helices correctly \(^15,16\). However, prediction accuracy for a single RNA or for any helix within a larger structure is not known in advance. Incorrect
prediction of even a few helices in a functionally important region makes it difficult or impossible to develop robust biological models.

Among the gamma retroviruses, several sequences have been consistently proposed as important for dimerization of the RNA genome\(^8\)-\(^{10}\) (summarized in Fig. 2.1a). PAL1, also known as the 204-227 stem loop\(^{17}\), DIS1\(^{18}\), or SL-B\(^5\), and PAL2, also known as DIS2\(^{18}\), SL-B\(^{19}\) or H1\(^{20}\), are currently postulated to form hairpin loops (in green, Fig. 2.1a). PAL1 and PAL2 span self-complementary ('palindromic') sequences and are conventionally proposed to interact via loop-loop interactions with PAL1 and PAL2 sequences from a second RNA and eventually form extended duplexes in the dimer\(^{10,17-19}\). Highly conserved GACG tetraloops\(^9\) (Fig. 2.1a) in stem-loops 1 and 2 (SL1 and SL2, also known as SL-C\(^{19}\) or H2\(^{20}\) and SL-D\(^{19}\) or H3\(^{20}\)) have the potential to form stable loop-loop interactions via cross-loop G-C base pairs\(^{19,21}\) and appear to be important for packaging via interactions with the viral Gag protein\(^{22}\).

I develop a generalizable approach for obtaining a well constrained secondary structure for a retroviral dimerization domain and for many other classes of RNA. I first use competition experiments to define rigorously a contiguous minimal dimerization active sequence (MiDAS) for a representative gamma retrovirus, the Moloney murine sarcoma virus (MuSV). I then use a new chemical approach, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE)\(^{23,24}\), to obtain comprehensive, quantitative, nucleotide-resolution, and model-independent constraints for the secondary structure of the monomeric starting state of the retroviral dimerization domain. These experiments emphasize that existing structural models for the dimerization domain in gamma retroviruses require reinterpretation. Most strikingly, a large region in the dimerization domain is
conformationally flexible, which may facilitate retroviral RNA dimerization by decreasing the energetic cost of disrupting base pairing or other interactions in the monomer prior to forming functional structures specific to the dimer.

2.2 Results

2.2.1 Rigorous Definition of a MiDAS. *In vitro* studies using synthetic RNA transcripts have been essential for identifying candidate structures that contribute to dimerization in the gamma retroviruses\(^8,10,17-20,25-29\). A significant challenge in interpreting these experiments is that most RNAs containing a stem-loop structure will dimerize if dimerization reactions are performed at sufficiently high RNA concentration or ionic strength conditions\(^30,31\). Neither the minimal region required for dimerization nor the structure of the dimerization-active domain in the viral genome monomeric starting state is currently well defined.

In exploratory experiments, I identified a roughly physiological ion environment -- 50 mM Hepes, 200 mM potassium acetate, and 5 mM MgCl\(_2\) (pH 7.5) -- that yields well-behaved single conformation monomer and dimer complexes for an RNA spanning positions 135 to 599 in the MuSV 5' untranslated region (Figure 2.1a). This RNA spans all structures previously proposed to participate in dimerization in the gamma retroviruses. Thus, this RNA was used to impose a functional threshold in competitive dimerization experiments that, by design, strongly discriminate against promiscuous self-dimerization (Fig. 2.2).
Figure 2.1. 5'-Untranslated region of MuSV. Conserved sequences and conventionally proposed secondary structures are illustrated schematically. The 5' genomic RNA cap is position 1. (a) 5' and 3' truncation mutants are shown in blue and red, respectively. (b) ΔPAL2 mutant.
Figure 2.2. Competitive-dimerization assay for stringent definition of RNA structures essential for dimerization. Assay employs radiolabeled mutant RNA (mut; in red, with asterisk) and unlabeled full length RNA (FL; gray). Only species containing the mutant RNA are visualized in the non-denaturing gel.
I constructed an extensive series of viral sequences containing systematic truncations from their 5' and 3' ends (blue and red arrows, respectively, in Fig. 2.1a). Competitive dimerization experiments were performed at 60 °C in the presence of the full-length (FL) transcript and visualized by the selective detection of the radiolabeled, truncated RNA variants in non-denaturing gels (Fig. 2.2). Both the radiolabeled mutant-mutant homodimer (HOD) and mutant-full length heterodimer (HED) are visualized directly. Full length RNA homodimers also form but are not radiolabeled and thus are not observed.

To be scored as a structurally competent dimerization active sequence, the mutant RNA (mut) must quantitatively compete with homodimerization by the (unlabeled) full length RNA. RNAs that are only able to homodimerize or only heterodimerize (see middle two lanes in Fig. 2.2) are scored as structurally deficient. Experiments further used a range of full-length RNA concentrations in order to make the experiment structurally stringent (Fig. 2.3).

2.2.2 A Minimal Sequence Active in Dimerization. Truncation mutants are identified by the 5' or 3' nt at which the mutant sequence terminates (Fig. 2.1a). Competitive dimerization experiments were performed with ~ 1.5 nM radiolabeled, truncated RNA and 1, 5 or 15 nM full length (FL) RNA (Fig. 2.3). Markers for the mobilities of the mutant monomer (M) and for the mutant homodimer (HOD) were obtained by omitting the heating step or addition of FL RNA, respectively.

The progressively larger 3' truncations, 3'-479, 3'-419 and 3'-374 (Fig. 2.3a), yield native-like RNAs that both homo- and heterodimerize efficiently with the full-length RNA as visualized in a non-denaturing gel. In contrast, the 3'-339 truncation heterodimerizes
Figure 2.3. The minimal dimerization active sequence for MuSV defined by competitive dimerization. (a,b) 3'-end truncations in the native and ΔPAL2 contexts. (c,d) 5'-end truncations. (e) Schematic structure for the minimal dimerization active sequence (MiDAS) in the context of the conventional secondary structure. (f) Efficient dimerization of the MiDAS in both native and ΔPAL2 contexts. FL, full-length; M, mutant Monomer; HED, mutant-FL heterodimer; HOD, mutant-mutant homodimer
efficiently with the FL RNA, but forms almost no homodimer (3'-339 panel, Fig. 2.3a). Thus, the 3'-339 RNA is deficient in dimerization in a way that can be rescued by the full length RNA. This RNA also forms multiple monomeric (M) conformations. Similarly, the 3'-354 truncation forms multiple monomeric conformations and also forms heterodimers inefficiently (Table 2.1). The 3' boundary for the minimal dimerization active sequence lies between nucleotides 354 and 374; the largest tested deletion that is fully functional in the competitive dimerization assay spans position 374 (3'-374 panel, Fig. 2.3a).

### Table 2.1: Dimerization activity of 3' and 5' truncation mutants

<table>
<thead>
<tr>
<th>3' Truncations</th>
<th>native</th>
<th>ΔPAL2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HOD</td>
<td>HED</td>
</tr>
<tr>
<td>3'-569</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-539</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-509</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-479</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-449</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-419</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-407</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-381</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-374</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-354</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3'-339</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3'-324</td>
<td>-</td>
<td>+/-</td>
</tr>
<tr>
<td>3'-309</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>3'-303</td>
<td>+/-</td>
<td>+/-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5' Truncations</th>
<th>native</th>
<th>ΔPAL2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HOD</td>
<td>HED</td>
</tr>
<tr>
<td>5'-175</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5'-205</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5'-235</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5'-265</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5'-276</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5'-295</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>5'-325</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

+ native-like dimerization activity; +/- dimerization occurs but is substantially compromised. ✓ and X, overall native-like versus compromised dimerization. M*, forms multiple monomer conformations.
Truncations from the 5' end through position 276 yield RNAs that are fully functional in both homo-and heterodimerization (Fig. 2.3c). In contrast, truncation through position to 295 yields an RNA that homodimerizes well but is essentially incompetent to form heterodimers with the FL RNA (5'-295 panel in Fig. 2.3c). Further truncation through 5'-325 yields an RNA that forms neither homo- nor heterodimers. The 5' boundary for the minimum dimerization active sequence was thus set at position 276.

This minimal dimerization active sequence spans positions 276 to 374, and includes PAL2, SL1, and SL2. Importantly, the behavior of the 5'-295 mutant illustrates the stringency of the competitive dimerization assay. Although this RNA would have been scored as dimerization competent under less stringent conditions, it clearly lacks key elements required to dimerize competitively with the full length RNA.

### 2.2.3 Deletion of PAL2 Unmasks the Contribution of PAL1.

The minimal dimerization active region, defined by this initial analysis, includes the PAL2 sequence which several groups have proposed plays a role in dimerization. To explore whether any other accessory sequence contributes to dimerization, but is masked by PAL2, I compromised PAL2 by removing nts 283-294 (ΔPAL2 mutant, Fig. 2.1b) and re-tested the panel of 5' and 3' truncations by competition with a full-length RNA, also harboring the ΔPAL deletion. All of the 3' truncations in the ΔPAL context through nt 374 formed both homo- and heterodimers efficiently (Fig. 2.3b), indicating that the 3' boundary of the minimal dimerization active sequence remains at position 374.

When the 5' series was analyzed, truncations through nt 205 yielded fully functional RNAs. In contrast, both homo- and heterodimer formation was significantly impaired when
the RNA was truncated through 5'-235 in the ΔPAL2 context (5'-235 panel, Fig. 2.3d). In addition, when time resolved dimerization experiments are performed with a construct spanning the PAL2 through SL2 sequences, this RNA dimerizes only to about 80% (at concentrations up to 50 nM). In contrast, RNAs that also include the additional 5' 205-275 sequence dimerize essentially to completion (Fig. 2.3f and compare gels in Fig. 2.4). Thus, I concluded that the role of sequences that include the PAL1 region (at nts 205-217) is partially masked if PAL2 is present and assigned the 5' boundary of the minimal dimerization domain to be position 205 (Fig. 2.3d).

2.2.4 A Minimal Dimerization Domain. I constructed a minimal RNA spanning MuSV sequences from 205 to 374 and tested the ability of this RNA to function in the competitive dimerization assay, both in the native PAL2 and ΔPAL2 contexts (Fig. 2.3f). Both RNAs form single-conformation monomers, homodimers, and heterodimers identical to those observed for native-like truncations. I infer that nts 205-374 span the minimal dimerization domain for MuSV and, potentially, for most gamma retroviruses.

The minimal dimerization active sequence (MiDAS; Fig. 2.3e) incorporates RNA elements previously proposed to either contribute or be a primary determinant for retroviral dimerization\(^8,10,17-21,25,33\). The competitive dimerization assay supports proposals\(^17,18\) that the role of the PAL1 sequence may be especially important if the PAL2 sequence is compromised. Unique to the competitive dimerization assay is the strong inference that the MiDAS (Fig. 2.3e) spans all RNA sequences stringently required for dimerization. The dimerization sequences defined here in vitro correspond closely with analogous experiments designed to define a minimal dimerization active sequence in vivo: an RNA spanning
Figure 2.4. Comparison of time-resolved RNA dimerization between two viral RNA constructs (nts 276-374 and 205-374). Both RNA constructs (at 50 nM) were dimerized in dimer buffer at varying times over a course of 60 min. The samples were then resolved on non-denaturing gels and band intensities were integrated. Interestingly, the 276-374 construct which contained only PAL2, SL1, and SL2 could not completely form dimer resulting in a heterogeneous 20:80 mix between monomer (M) and dimer (D) species. But, when PAL1 is included, as demonstrated by the 205-374 construct, complete formation of dimer is reached.
positions 215 to 404 from the Moloney murine leukemia virus is sufficient to increase packaging of a nonviral RNA, in dimeric form, by 50-fold\(^5\). To the extent that other viral components like the nucleocapsid protein augment, but do not fundamentally alter, an RNA-centered process, the MiDAS represents a rigorously vetted minimal domain for retroviral dimerization.

2.2.5 MiDAS Monomeric Structure Analyzed by RNA SHAPE Chemistry. The Weeks laboratory has recently developed a single nucleotide resolution approach to interrogate, quantitatively, the local environment at every nucleotide in an RNA\(^{23,24}\). RNA ribose 2'-hydroxyl groups react with N-methylisatoic anhydride (NMIA) to form the nucleotide 2'-ester as shown in Equation 2.1:

\[
\text{(RNA)}_{\text{constrained}} \xrightleftharpoons[(\text{fast})]{k_2 [\text{NMIA}]} \text{(RNA)}_{\text{flexible}} \xrightarrow{\text{CO}_2} \text{RNA}^{2'-O-}\text{adduct}
\]

2'-Hydroxyl reactivity is gated by whether or not a given nucleotide is constrained by base pairing or tertiary interactions\(^{23,36}\). Flexible nucleotides react preferentially because they are better able to reach a conformation that facilitates nucleophilic attack of the 2'-hydroxyl on NMIA. Formation of the bulky 2'-O-adduct is readily detected as a stop to reverse
transcriptase-mediated primer extension: the complete experiment involves selective 2'-
hydroxyl acylation analyzed by primer extension (SHAPE).

SHAPE experiments were performed in the context of a MiDAS RNA with 30 and 5
nt native sequence extensions, respectively, at the 5' and 3' ends to facilitate analysis of the
entire domain by primer extension. I also appended an RNA cassette\textsuperscript{23} to the 3' end that
contains an efficient DNA primer binding site. This 3' cassette folds independently of
flanking RNA structures (as described below and in refs. 23,24).

Refolded MiDAS RNA was treated with NMIA under physiological-like conditions
(200 mM potassium acetate, 5 mM MgCl\textsubscript{2}, pH 8, 37 °C); the monomeric state of the RNA
was confirmed by native gel analysis. Sites of 2'-O-adduct formation were detected by
primer extension and resolved on sequencing gels (\textbf{Fig. 2.5a}). Comparison of reactions
performed in the presence of NMIA with reactions omitting the reagent reveals selective
formation of 2'-O-adducts at a subset of sites in the RNA (compare + and – NMIA lanes in
the MiDAS panel, \textbf{Fig. 2.5a}). Individual band intensities were integrated\textsuperscript{37} and absolute
reactivities were computed for every position in the MiDAS RNA construct.

The nucleotide resolution SHAPE experiment provides a large number of constraints
that must be accommodated in any secondary structure prediction for the MiDAS RNA. I
screened secondary structures for the MiDAS region (residues 205-374) by submitting
positions whose calculated reactivity was at least 25% of the strongest observed reactivities
(47 nts total) as chemical modification constraints to the RNAStructure 4.11 program\textsuperscript{15}. The
quantitative data is shown superimposed on a secondary structure consistent with the entire
body of SHAPE reactivity (\textbf{Fig. 2.6}).
Figure 2.5. SHAPE analysis of the MuSV MiDAS RNA and of the PALSTB and Δ289-300 mutants. (a) 2′-O-Adduct formation visualized in a sequencing gel. Reactions were performed in the presence (+) and absence (−) of NMIA. Sequencing lanes (seq) showing guanosine positions were generated by dideoxy nucleotide incorporation during primer extension. The dideoxy sequencing ladder is exactly 1 nt longer than the corresponding NMIA lane; nucleotide positions are labeled with respect to NMIA lanes. (b) Sequences of PAL2 and of the PALSTB and Δ289-300 mutants, drawn in the context of the conventional structure for PAL2.
The nucleotide resolution SHAPE experiment provides a large number of constraints that must be accommodated in any secondary structure prediction for the MiDAS RNA. I screened secondary structures for the MiDAS region (residues 205-374) by submitting positions whose calculated reactivity was at least 25% of the strongest observed reactivities (47 nts total) as chemical modification constraints to the RNAStructure 4.11 program\textsuperscript{15}. The quantitative data is shown superimposed on a secondary structure consistent with the entire body of SHAPE reactivity (Fig. 2.6).

Residues with high and moderate reactivity (red and orange, Fig. 2.6) towards NMIA are located in single-stranded loops and connecting structures. Positions with low or undetectable reactivity (blue and black, Fig. 2.6) lie largely in base paired helices. Because SHAPE is sensitive to any interaction, including non-canonical interactions, that constrain a nucleotide\textsuperscript{23}, reactive positions should fall cleanly in flexible RNA structures; whereas, some unreactive nucleotides, that lie in nominally single-stranded regions, may reflect tertiary structure constraints that remain to be defined at this stage of analysis. As expected\textsuperscript{23,24}, the 3’ RNA structure cassette (Fig. 2.6, top) has a reactivity pattern exactly consistent with its designed fold, indicating that this appended structure does not interfere with folding of the MiDAS RNA.

SHAPE analysis (Figs. 2.5 and 2.6) strongly supports secondary structures for the PAL1 loop, SL2, and the upper portion of SL1 that are consistent with prior proposals. Unexpectedly, the PAL2 sequence (positions 283-298, see Fig. 2.3e for comparison), which has been almost universally assumed to form a stable stem-loop structure, is highly reactive towards NMIA. Moreover, the reactive PAL2 sequence resides in the middle of a larger flexible domain in which most nts are reactive by SHAPE chemistry (Fig. 2.6).
Figure 2.6. Secondary structure model of the monomeric MuSV MiDAS RNA. Quantitative NMIA reactivities, minus background, are superimposed as columns at each nucleotide position. A structure cassette\textsuperscript{23}, containing a DNA primer binding site (gray circles), was appended to the 3' end of the viral sequence. Lower panel shows the same secondary structure as in the upper panel, but with residues labeled explicitly.
2.2.6 PAL2 is Unstructured in the MiDAS Monomer. Because the observed structure is significantly different from conventional models for the dimerization domain, I analyzed the structure of two MiDAS RNAs carrying instructive mutations in PAL2 (Fig. 2.5b). The first mutant, PALSTB (PAL stabilization), was designed to stabilize PAL2 in the conventional stem-loop structure by increasing the G-C base pair content at flanking helix positions (circled positions, Fig. 2.5b). Inspection of the SHAPE data shows that stabilizing the PAL2 duplex has the desired effect. Nucleotides located in the PAL2 loop are strongly reactive while base paired positions in the stem are now much less reactive than in the native sequence (compare MiDAS and PALSTB lanes, Fig. 2.5a).

The experimental SHAPE reactivity data for the PALSTB mutant was subtracted from that for the native MiDAS RNA to create a quantitative difference map for every position in the PALSTB RNA (Fig. 2.7a). In the difference map, residues that are more reactive or are more constrained in the mutant relative to the native MiDAS sequence are reported as positive and negative amplitudes, respectively (red and blue, Fig. 2.7). If PAL2 already existed as a hairpin in the monomeric native state, stabilizing this stem should have a minimal effect on global MiDAS RNA structure. In strong contrast to this expectation, stabilizing the PAL2 sequence as a stem-loop causes large changes to the SHAPE reactivity in the MiDAS domain.

Consistent with the design of this mutant, nucleotides in the loop at the apex of the PAL2 stem in the PALSTB mutant show an increase in reactivity and nucleotides in the stabilized stem are much less reactive than in the native MiDAS sequence (see PAL2, Fig. 2.7a). More significantly, the PALSTB mutant shows very large changes in global
Figure 2.7. Quantitative difference maps for the effects of the PALSTB and Δ289-300 mutations on MiDAS structure. Vertical bars report absolute NMIA reactivities at each base position for the mutant RNA minus reactivity of the native RNA. Positive (red) and negative (blue) differences indicate increased versus reduced reactivity, respectively, in the mutants relative to the native MiDAS. Vertical scales are the same in upper and lower panels; the shorter bars, overall, in the lower panel reflect the more modest structural perturbation introduced by the Δ289-300 mutant.
structure that extend almost the entire length of the RNA and up to 80 nts away (see especially nts 252-268 in the flexible domain and nts 374-382 between SL2 and the 3' end of the RNA, Fig. 2.7a). The peaks shown in the difference map are plotted on a scale comparable to that used in Fig. 2.6. Thus, the large positive peaks centered at positions 255, 312, 365 and 380 represent significant enhancements in absolute local nucleotide flexibility in these regions (for example, compare the 255 region in Fig. 2.5 with the difference map in Fig. 2.7a).

Our second mutant, Δ289-300, was designed to delete a large region within PAL2 that is flexible as judged by SHAPE chemistry (mutant is shown in Fig. 2.5b, flexible region is labeled in Fig. 2.6). If the conventional stem-loop model for PAL2 were correct, this 12 nt deletion should have a dramatic effect on global MiDAS structure. On the other hand, if PAL2 is unstructured in the native monomer state, as indicated by SHAPE chemistry, then this extensive deletion may have only a minimal effect on global MiDAS structure. A difference map for the Δ289-300 mutant shows that this deletion, in fact, introduces very modest changes to local nucleotide flexibility in the MiDAS RNA and induces virtually no significant structural change over large regions of the sequence (Fig. 2.7b). The most significant effect is an increase in SHAPE reactivity in the (already flexible) 267 region of the flexible domain and a decrease in reactivity in the 340 bulge in SL1. Thus, direct analysis of the MiDAS RNA (Fig. 2.6) and differential analysis of two mutants (Fig. 2.7) strongly support a new model for the dimerization domain of MuSV that contains an extensive flexible domain that spans the PAL2 sequence.
2.3 Discussion

The model for the minimal dimerization active structure of MuSV makes use of two innovations that are likely to be generalizable to any RNA structure prediction problem. First, I defined a minimal sequence for dimerization using an assay that requires the simplified RNA to functionally compete with a native-like sequence (Fig. 2.2). Second, SHAPE chemistry quantitatively interrogates every nucleotide in an RNA, which means that secondary structure models can be evaluated with much greater confidence than when using traditional chemical and enzymatic reagents.

Although the MiDAS secondary structure (Fig. 2.6) proposed here differs significantly from earlier models, this structure is consistent with the two existing sets of experimental information for dimerization domains in gamma retroviruses. The sequence of the murine leukemia virus (MuLV) is almost identical to that of MuSV. The nucleotide resolution SHAPE information strongly supports the original MuLV model for SL2 and the upper portion of SL1. In contrast, SHAPE does not support the earlier proposal that PAL2 forms a stable stem-loop structure. However, superposition of the chemical mapping information for MuLV on our secondary structure for MuSV shows that the prior information is exactly consistent with the current MiDAS proposal (Fig. 2.8). In particular, the PAL2 sequence is strongly reactive towards conventional single-strand-selective chemical reagents and thus consistent with the idea that PAL2 lies in a flexible domain.

Structure mapping studies on the Harvey sarcoma virus (HaSV) also emphasize the importance of SL1- and SL2-like structures in the dimerization domain. HaSV does not contain a PAL2 sequence but, strikingly, the HaSV RNA can be folded into a secondary structure that is both similar to that for MuSV and for which RNase-based cleavage data
Figure 2.8. Proposed secondary structures for MuLV and HaSV dimerization domains. Mapping data are from refs. 8 and 28. Solid circles indicate positions reactive towards single-strand selective reagents (DMS and CMCT; upper panel) or enzymes (RNases T1, T2, and A; lower panel). Open circles (lower panel) indicate positions cleaved by both single- and double-strand (RNase V1) selective enzymes. Alternate structures for the flexible domains are shown as inserts. Asterisks (top panel) indicate minor sequence differences between MuLV and MuSV.
strongly support formation of an internal flexible domain (Fig. 2.8). Many sites in the HaSV domain are cleaved by both single and double strand-selective RNases \textsuperscript{28} (Fig. 2.8): the HaSV RNA likely contains a flexible domain in which portions of the structure are alternately both paired and flexible in distinct conformations. That the MuLV and HaSV RNAs fold to similar monomeric starting structures provides a structural basis for the observation that these viruses readily heterodimerize\textsuperscript{28}, presumably via PAL1.

I folded the flexible domain, including its anchoring helix (spanning positions 231-315; Fig. 2.6), subject to the requirement that the 27 positions with high and moderate reactivities be single stranded. The lowest energy structure, which is compatible with all of the SHAPE information (Fig. 2.6) has a total calculated\textsuperscript{15} folding free energy of only –10.5 kcal/mol. This single low energy structure spans 84 nts and thus has a net stability comparable to a simple stem-loop structure containing roughly 3 base pairs. Moreover, although the entire flexible domain from nt 249 through 294 contains no instances in which there are more than two strongly constrained nts in a row (black positions, Fig. 2.6), individual nts vary significantly in their 2'-hydroxyl reactivity.

I therefore evaluated the alternate hypothesis that multiple, more stable, structures might be compatible with the SHAPE information and submitted the 12 most highly reactive sites (red, Fig. 2.6) as chemical modification constraints for RNA structure prediction\textsuperscript{15}. Four structures have calculated free energies within 10\% of the most stable structure. Three of these structures have distinctive folds (Fig. 2.9) while the fourth (not shown) contains elements of the other structures. Absolute SHAPE reactivities were superimposed on the three most distinctive structures (Fig. 2.9). These intensity data are colored by the fraction of
**Figure 2.9.** Structural model for overall flexibility in the 231-315 domain. Column heights indicate absolute NMIA reactivities (and are the same as reported in **Fig. 2.6**). Column colors illustrate the fraction in which a position is single-stranded in this ensemble of three representative structures.
structures in which they are single stranded (always paired and always single stranded are black and white, respectively). Each structure A, B and C is only partially consistent with the SHAPE data. Each of these structures, however, has a calculated folding free energy of approximately –17 kcal/mol and, thus is significantly more stable than the single consensus structure that incorporates all of the flexibility information. In this semi-quantitative analysis, in which the large universe of possible structures is approximated by three low energy structures (Fig. 2.9), there is a very strong correlation between SHAPE reactivity and the extent to which individual nucleotides are constrained, as averaged over these structures.

Models for the genomic RNA retroviral dimerization domain, in which PAL2 forms a stable stem-loop structure (see Figs. 2.1a and 2.5b), have guided the gamma retrovirus field for over a decade. However, the model-independent SHAPE intensity information (Figs. 2.5a and 2.6) emphasize that existing structural models merit careful reinterpretation.

For example, many mechanistic analyses of retroviral dimerization have used simplified RNAs or RNAs in which PAL2 was mutated to enforce the conventional structure (see Fig. 2.1a) for the RNA. The RNA structure in such mutations should be similar to the PALSTB mutant, which was also designed to artificially reinforce the conventional structure for PAL2. The PALSTB mutant yielded dramatic and extensive changes to the global structure of the MiDAS, including at residues up to 80 nts distant from PAL2 (Fig. 2.7a). Thus, the structure of the retroviral dimerization domain with a native sequence can be quite different from RNAs containing mutations in PAL2 or that are shorter than the MiDAS. Moreover, the significant global changes that occur in the MiDAS domain upon introducing mutations in PAL2 emphasize that the flexible domain communicates via long range interactions with other regions of the RNA.
Because PAL2 (and PAL1) sequences are self-complementary, an attractive model for the noncovalent interactions that stabilize the retroviral dimer is for these sequences to form an extended duplex in the dimer\textsuperscript{10,17-19}. Prior models that proposed that PAL2 initially exists as a stable stem-loop recognized that it might be energetically costly to disrupt the extensive pre-existing base pairing in this structure. These models thus generally proposed that dimerization proceeds stepwise, first, via base pairing between nucleotides in the loops of two PAL2 stem-loop structures, followed by helix extension.

The nucleotide-resolution SHAPE experiments demonstrate that the PAL2 sequence lies in an RNA domain in which, on average, most nucleotides are either unconstrained by base pairing or are transiently in a single stranded conformation (Fig. 2.9). SHAPE experiments support formation of extended duplexes in the dimer which is discussed in Chapter 3 of this thesis. Thus, formation of a flexible domain in the monomeric starting state has significant mechanistic implications for retroviral RNA genome dimerization. (i) Dimerization via PAL2, in the context of a flexible domain, is potentially much more thermodynamically favorable than previously thought because fewer base pair interactions in the monomer have to be disrupted to form extended duplexes between two PAL2 sequences in the dimer. (ii) Placing the PAL2 sequence in a flexible domain may also kinetically enhance retroviral RNA dimerization by lowering the activation barrier for extended duplex formation. (iii) The distinct conformations visualized for the flexible domain likely also have different dimerization activities. Retroviral dimerization could thus be modulated by interactions between these distinct conformations and other regions of the genomic RNA or retroviral proteins.
2.4 Materials and Methods

2.4.1 Retroviral RNA Transcripts. DNA templates for in vitro transcription of the full length RNA, 5' and 3' truncations, and MiDAS constructs were generated by PCR from the pLNBS\textsuperscript{26,27} plasmids generously gifted by A. Kaplan. PCR (1 mL; 1 cycle, 95 °C, 5 min, [32 cycles, 95 °C, 30 sec; 55 °C, 30 sec; 72 °C, 1 min], 1 cycle, 72 °C, 10 min) reactions contained 0.2 mM of each dNTP, 1x PCR buffer (0.01 M Tris (pH 7.4), 0.05 M KCl), 0.5 µM each of forward and reverse primer, 2.5 mM MgCl\textsubscript{2}, 0.1 ng/µl pLNBS template, and 5 µl of Taq polymerase (5 U/µl, Invitrogen). RNA constructs were generated using T7 RNA polymerase-mediated transcription (500 µL; 37 °C, 5 h) containing 80 mM Hepes (pH 7.4), 40 mM dithiothreitol (DTT), 0.01% (v/v) Triton X-100, 2 mM spermidine, 10 mM MgCl\textsubscript{2}, 2 mM each nucleoside triphosphate, ~ 25 µg of PCR-generated template, 20 U of SUPERase-In (Ambion), and 0.1 mg/mL polymerase. Internally labeled RNAs were synthesized using 20 µCi α-[\textsuperscript{32}P]-ATP and unlabeled ATP at 0.5 mM. RNAs were purified by denaturing gel electrophoresis (5% polyacrylamide, 7 M urea), excised from gel, eluted overnight into 1/2× TBE (45 mM Tris-borate, 1 mM EDTA), and concentrated by ethanol precipitation. RNAs were resuspended in HE [10 mM Hepes (pH 7.5), 1 mM EDTA] and stored at -20 °C.

2.4.2 Competitive Dimerization Assay. [\textsuperscript{32}P]-internally-labeled truncated RNA (~1.5 nM) was incubated with unlabeled full length RNA (at 1, 5 or 15 nM in 15 µL). Reactions were heated to 90 °C for 3 min to eliminate pre-existing dimers, rapidly cooled on ice, treated with 5 µL 4× dimerization buffer [25 °C; 200 mM Hepes (pH 7.5), 800 mM potassium acetate (pH 7.5), 20 mM MgCl\textsubscript{2}], incubated at 60 °C for 30 min, and placed on ice. Samples (3 µL) were mixed with 1 µL of 30% (v/v) glycerol (containing xylene cyanol and bromophenol blue) and resolved by non-denaturing electrophoresis at 4 °C. Gels (5%
polyacrylamide in 1x TBE) were pre-run for 15 min prior to sample loading and subsequently run for 2 hr at 20 W. Gel imaging was done on a phosphorimager (Molecular Dynamics) and analyzed using ImageQuant.

2.4.3 Time-resolved Dimerization Assay. [$^{32}$P]-internally-labeled viral RNA constructs (~1.5 nM) spanning nts 205-374 and 276-374, respectively, were incubated with their same unlabeled RNA (at 50 nM) in 15 µL. Reactions were then heated to 90 °C for 3 min, cooled on ice, and mixed with 5 µL 4× dimerization buffer [25 °C; 200 mM Hepes (pH 7.5), 800 mM potassium acetate (pH 7.5), 20 mM MgCl$_2$], and incubated at 60 °C for 0, 1, 2, 5, 10, 20, 40, and 60 min. Samples were separated at 4 °C as previously discussed above in Competitive Dimerization assay section (5% polyacrylamide in 1x TBE). ImageQuant software was then used to quantify the band intensities and fraction dimer was calculated by dividing intensity of dimer by the total sum of intensities for both monomer and dimer.

2.4.4 SHAPE Analysis of Monomer MiDAS. 2'-Hydroxyl acylation and primer extension (SHAPE) experiments were performed with a MiDAS RNA that contained flanking 5' and 3' extensions of viral sequence of 30 and 5 nucleotides, respectively and a 3' non-viral RNA cassette containing an efficient DNA primer binding site$^{23}$ was appended to the 3' end. The MiDAS RNA construct (10 pmol) was heated at 90 °C for 3 minutes in 7.2 µL water, cooled on ice, treated with 1.8 µL 5× dimerization buffer [250 mM Hepes (pH 8.0), 1 M potassium acetate (pH 7.5), 25 mM MgCl$_2$], incubated at room temperature (~ 25 °C) for 30 s, and returned to ice. The RNA solution was then equilibrated at 37 °C for 5 min, treated with NMIA (1 µL, 180 mM in anhydrous DMSO), allowed to react for 50 min (~ 5 half lives$^{23,24}$) at 37 °C, and placed on ice. Control reactions contained DMSO without NMIA.
2.4.5 Primer Extension. Two DNA primers were used to analyze the MiDAS RNA construct: primers were complementary to the 3' end of the RNA structure cassette (5'-GAA CCG GAC CGA AGC CCG) and to SL1 (5'-CAG AAC TCG TCA GTT CCA CCA). Primer extension reactions were performed by adding modified RNA (2 µL, 2 pmol) and 5'-[\textsuperscript{32}P]-DNA primer (1 µL, 1 pmol) to 9 µL water and annealing by incubation at 95 °C (30 sec), 60 °C (6 min), and 35 °C (10 min). Reverse transcription buffer [7 µL; 143 mM Tris (pH 8.3), 214 mM KCl, 7.14 mM MgCl\textsubscript{2}, 1.43 mM each dNTP, 14.3 mM DTT] was added. Superscript III (Invitrogen, 1 µL, 200 units) reverse transcription enzyme was added and reaction was heated at 48.5 °C for 5 min. Reactions were quenched by addition of 1 µL 4M NaOH, heated at 90 °C for 5 min, addition of 29 µL of gel loading solution (40 mM Tris-borate, 276 mM Tris-HCl, 5 mM EDTA, 0.01% xylene cyanol, 0.01% bromophenol blue, 73% (v/v) formamide), and finally heated again at 90 °C for 1 ½ min before loading. cDNA fragments were resolved on a series of 8% (w/v) denaturing polyacrylamide gels (in 1x TBE/7M Urea) to achieve nucleotide resolution throughout the analyzed region. After visualizing the gel with a phosphorimager (Molecular Dynamics), semi-automated footprinting analysis (SAFA)\textsuperscript{35} was employed to quantitatively analyze band intensities at each base position.
2.5 References


Chapter 3

Quantitative Structure of a Retroviral RNA Dimer
3.1 Introduction

Retroviruses are well-characterized diploid entities in biology. Retroviruses package exactly two copies of their genomes that are in a dimeric state into each nascent virion. Retroviral RNA genomes dimerize by conserved non-covalent interactions involving sequences near the 5’ end. Expectedly, this dimeric state appears to be a fundamental feature of retroviral biology; thus, it is important to understand the structural details of this domain at high resolution.

I have previously used competitive-dimerization experiments to define a minimal dimerization active sequence (MiDAS) for the Moloney murine sarcoma virus (MuSV). This domain, defined in vitro, spans nts 205-374. Strikingly, the MiDAS corresponds very closely to the minimal sequences required to direct packaging of heterologous non-viral RNAs, as dimers, into nascent virions.

The MiDAS domain includes sequence elements previously proposed to be important for forming the final dimer state including two self-complementary, or palindromic, sequences, PAL1 and PAL2, and two well conserved stem-loop motifs, SL1 and SL2 (Fig. 3.1a). PAL1 and PAL2 have the potential to form extended heteroduplexes with their complements in the dimer. As is typical for most RNAs, the MiDAS domain contains well-structured helical elements. In addition, in the monomer state, the MiDAS RNA also includes a large flexible domain that is linked by a stable anchoring helix (nts 231-241, 305-315; Fig. 3.1a).

The structures of retroviral dimerization domains, that have lead the field, have primarily been characterized in terms of their monomer starting states. But, in part, emphasis on the monomer state has made it difficult to analyze structures in the final dimer. For
**Figure 3.1.** RNA dimerization by the MiDAS domain of a gamma retrovirus. (a) Secondary structure of the MiDAS in the monomeric starting state. RNA structure is shown with grey lines. Key sequences involved in dimerization are emphasized in bold. (b) Well-resolved dimerization of MiDAS domain visualized on non-denaturing gel electrophoresis (8% polyacrylamide). Monomer and dimer species are labeled; O, gel origin.
example, conformational changes from an initial stem-loop structure to a heteroduplex will only yield a small number of local structural changes at certain nucleotide positions. Thus, monitoring the monomer-to-dimer conversion at single nucleotide resolution with conventional chemical footprinting methodologies has been difficult. However, the recently developed RNA SHAPE\textsuperscript{14} chemistry allows local structural analysis at every nucleotide in an RNA to be quantitatively assessed.

In this work, I use SHAPE chemistry to develop a single nucleotide resolution structure of the MuSV MiDAS domain in the final dimer state. Additionally, using site-directed and solvent-based hydroxyl radical footprinting, I also observe long range interactions in the dimer state that indicate retroviral RNA dimerization forms a compact three-dimensional architecture in the RNA genome.
3.2 Results

3.2.1 SHAPE Analysis of the Retroviral RNA Dimer. The MiDAS RNA forms well-defined, single conformation monomer and dimer species, as analyzed by non-denaturing gel electrophoresis (Fig. 3.1b). The monomer migrates as a single band and can be converted to the dimer state upon heating at 60 °C for 30 minutes (100 nM RNA; pH 7.5, 200 mM potassium acetate and 5 mM MgCl₂). Because the MiDAS RNA in the final dimer state migrates as a single species, it appears that my structural analysis reports on a single conformation without interfering contributions of alternatively folded RNAs.

I analyzed conformational changes that accompanied the monomer-to-dimer transition in the MiDAS domain using RNA SHAPE chemistry. In a SHAPE experiment, 2'-hydroxyl groups at flexible nucleotide positions will preferentially react with N-methylisatoic anhydride (NMIA) to form 2’-O-adducts. Thus, SHAPE chemistry quantitatively monitors local RNA flexibility at single nucleotide resolution.

MiDAS RNAs in the monomer or dimer states (Fig. 3.1b) were treated with NMIA and the sites of 2’-O-adduct formation were detected as stops to primer extension (Fig. 3.2a). Band intensities, corresponding to SHAPE reactivities, were then quantified for almost every nucleotide in the MiDAS domain for RNAs in both monomer and dimer states. Absolute reactivities were computed by subtracting background intensities observed in the absence of NMIA (compare (+) and (-) NMIA lanes, Fig. 3.2a; quantitative histograms are shown in Fig. 3.2b).

Absolute SHAPE reactivities for the monomer MiDAS are identical in terms of both overall reactivity and fine structure to studies in Chapter 2, even though these relevant experiments were performed a month apart (compare regions of monomer panel in
Figure 3.2. SHAPE analysis of MiDAS RNA in monomer (M) and dimer (D) states. (a) 2’-O-adduct formation detected by primer extension resolved in a sequencing gel. Experiments were performed in the presence (+) and absence (-) of NMIA. Sequencing lanes (SEQ) were generated by dideoxy cytosine nucleotide incorporation during the primer extension; nucleotide positions are labeled with respect to the NMIA lanes. (b) Histograms of SHAPE reactivities. Monomer and final dimer panel show absolute levels of 2’-O-adduct formation minus background. Difference plot was calculated by subtracting the monomer intensities from those of the final dimer; positive and negative (green and red) peaks indicate nucleotides that are more reactive or more structured in the final dimer state, respectively.
Fig. 3.2b with upper panel in Fig. 2.6 in Chapter 2). This similarity emphasizes the extreme reproducibility of a SHAPE experiment.

Upon forming the final dimer state, SHAPE chemistry reveals large structural changes throughout the MiDAS RNA (compare monomer and dimer experiments in Fig. 3.2a,b). I computed a monomer-to-dimer difference plot by subtracting absolute SHAPE reactivities in the monomer state from those observed in the final dimer conformation. By this definition, positions that are more reactive in the dimer are positive, while RNA regions that become structured in the dimer are negative (see green and red bars, respectively, in the difference plot, Fig. 3.2b).

SHAPE reactivities in both PAL1 and PAL2 self-complementary sequences (nts 210-219, 283-298) decrease significantly which is consistent with formation of extended heteroduplexes in these regions of the final dimer. In addition, I observe significant conformational changes in the SL1-SL2 region (compare nts 314-315, 319, and 338-343 in upper and middle panel of Fig. 3.2b). This domain has been analyzed and discussed in Chapter 4 and is known to undergo a conformational switch in which SL1 becomes extended by four base-pairs at its base (Badorrek et al., unpublished). Overall, a striking feature of the monomer-to-dimer conformational change is that, with exception of PAL1 and PAL2, local nucleotide flexibility increases in most regions of the dimer, relative to the monomer (see green bars in Fig. 3.2b).

An important feature in the secondary structure for the initial monomer state was an extensive flexible domain between the PAL1 and SL1 regions. This flexible domain spans positions 242-304 and has the ability to fold into multiple structures. Notably, the dimer state is also characterized by a new flexible domain that spans positions 221-278 (see final
dimer panel, Fig. 3.2b) and is constrained on each side by the PAL1 and PAL2 duplexes. Regions previously with no or lower reactivity in the monomer have now become more reactive and flexible in this domain (see nts 232, 243, 246-247, and 254-258, middle panel, Fig. 3.2b).

3.2.2 Secondary Structure Model for the MiDAS Domain in the Final Dimer State. The SHAPE information provides high resolution model-independent data about base-pairing and tertiary interactions in the final dimer. Therefore, I used SHAPE reactivities to constrain RNA secondary structure prediction programs\textsuperscript{16,17} to obtain specified models of the secondary structure of this domain. I currently constrain my structural predictions using RNAstructure v4.2\textsuperscript{16} by dividing absolute SHAPE reactivities into two classes. Positions with absolute reactivity between 50-100% of the maximum observed are required to be single-stranded (red in Fig. 3.3a). Positions with reactivities in the range of 25-50% of the maximum are required to be either single-stranded or adjacent to an unpaired nucleotide or a G-U base pair as implemented in the RNAstructure algorithm\textsuperscript{16}.

Model-independent SHAPE reactivities are superimposed, as columns, on a secondary structure model for the MiDAS domain in the final dimer state (Fig. 3.3a). Several features of this model are well-defined by the SHAPE data. PAL1 and PAL2 sequences form extended heteroduplexes of 10 and 16 base-pairs, respectively. Although the observed heteroduplex at PAL2 is consistent with the original 16mer extended duplex proposal\textsuperscript{12}, SHAPE- constrained predictions from RNAstructure show that PAL1 is only a 10mer extended duplex and is not extended by additional base-pairs as previously
Figure 3.3. Secondary structure model for MiDAS RNA in the final dimer state. Nucleotides are reported as color-coded columns (a) or letters (b) corresponding to their absolute SHAPE reactivities. For clarity, only one of the two strands in the dimer is highlighted; the second is shown in grey.
predicted\textsuperscript{10}. Instead, the region immediately 5’ to PAL1 (see Fig. 3.1a) does undergo a significant conformational change in which nucleotides 198-207 form a new stem loop containing a GNRA tetraloop at its apex. The SL1-SL2 forms an independent and compact tertiary structure domain (Badorrek et al., unpublished) as discussed in Chapter 4. SHAPE data indicates that the SL1-SL2 domain is linked to the PAL2 and the rest of the MiDAS via a highly flexible junction region (see nts 299-309, Fig. 3.3).

SHAPE reactivities do not support the formation of a single well-defined secondary structure in the region between the PAL1 and PAL2 extended duplexes in the final dimer. The thermodynamic model predicted by RNAstructure\textsuperscript{16} suggests a loose two stem loop motif (Fig. 3). RNAstructure, though, does not allow pseudoknot formation. I, therefore, searched for alternate secondary structures, consistent with SHAPE constraints, using the HotKnots program\textsuperscript{17}, which is a heuristic algorithm that allows for pseudoknot formations. Interestingly, the experimental constraints for this region indicate a reasonable pseudoknot (Fig. 3.3). Finally, I also note that another alternative structure mediated by cross-strand interactions at nts 268-271 is also consistent with the SHAPE information.

In summary, the SHAPE reactivity data emphasizes that the 221-278 flexible domain region is unlikely to form a single well-defined secondary structure in the context of the final RNA dimer. But, this is one of the additional advantages of the SHAPE approach because local flexibility can be interpreted for every nucleotide position. Thus, SHAPE constraint information provides a means to develop multiple secondary structural models for a single sequence in which a single structure is unlikely to form.
3.2.3 Long-range Interactions in the MiDAS Final Dimer. I performed site-directed hydroxyl radical cleavage\textsuperscript{18,19} on the MiDAS RNA in the final dimer state. A Fe(II)-EDTA moiety was attached at the 5' end of the MiDAS RNA via a phosphorothioate-mediated thiol linkage (see insert, Fig. 3.4a). This derivatized RNA was allowed to heterodimerize with the trace concentrations of 3'-\textsuperscript{[\textsuperscript{32}P]}-end labeled MiDAS RNA. Upon addition of hydrogen peroxide and ascorbic acid, hydroxyl radicals are selectively generated in the vicinity of the Fe(II)-EDTA group and will cleave nearby RNA structures. A key feature of this experiment is that, with cleavage observed only in the radiolabeled RNA, I can detect long-range and cross-strand proximal RNA structures.

I calculated the absolute site-directed cleavage intensities by subtracting background observed when using a mock-derivatized RNA lacking the phosphorothioate group (Fig. 3.1a). Cleavages are judged to be strong or moderate if they are 5- or 3- times greater than the average background, respectively. Cleavages are superimposed on our secondary structure model for the MiDAS domain in the final dimer state (Fig. 3.4b). For clarity, cleavage data is shown on one strand only, but is expected to be symmetrical over the two strands in the dimer.

Strong and moderate cleavages occur at nts 217, 219, and 220 in PAL1 consistent with the expected heteroduplex formation at this region of the RNA. Extensive strong and moderate cleavages occur in the flexible region between the PAL1 and PAL2 heteroduplexes. Interestingly, long-range cleavages were detected as far as PAL2 and show a sinusoidal pattern suggesting one turn of this extended heteroduplex is accessible to the 5'-tethered site of the Fe(II)-EDTA group. Additionally, I also observe other long-range cleavages in the junction linking PAL2 and the SL1-SL2 domain and in the SL1. Thus, the
Figure 3.4. Site-directed hydroxyl radical cleavage of the final dimer. The Fe(II)-BABE moiety was attached at the 5’end at position 205 of one RNA and cross-strand cleavage was selectively detected at positions in a 3’ end-labeled (*) second RNA. (a) Histogram showing absolute cleavage intensity. (b) Strong and moderate (red and orange boxes) site-directed cleavages are superimposed on the secondary structure for the MiDAS final dimer. Small spheres indicate regions not analyzed.
MiDAS domain appears to form a reasonably compact structure in which the PAL1 lies near the PAL2 and SL1-SL2 domain.

3.2.4 Tertiary Structure in the MiDAS Dimer. Site-directed hydroxyl radical cleavage supports the existence of long-range interactions in the MiDAS dimer such that the 5’ end of the RNA lies near the other conserved elements in the RNA. Thus, I probed for interactions in the MiDAS domain that could sufficiently protect the RNA from cleavages generated by solvent-based or untethered Fe(II)-EDTA reagent. MiDAS RNA was subjected to hydroxyl radical cleavage in conditions that either stabilized the formation of the native final dimer ([Fig. 3.1b]) or in the absence of mono and divalent ions (denaturing) where the RNA backbone is uniformly accessible to cleavage. A hydroxyl radical protection profile was generated by subtracting integrated band intensities for the RNA in the native final dimer state from those measured under denaturing conditions. Nucleotides that are protected or show enhanced cleavage in the final dimer in relation to the denatured state are indicated as positive and negative amplitudes, respectively ([Fig. 3.5a]).

Superposition of strong and moderate protections on the secondary structure of the MiDAS dimer ([Fig. 3.5b]) indicates extensive protections in the SL1-SL2 domain. The protection pattern in this domain, especially at the apexes of SL1 and SL2, coincides well with work in Chapter 4 that focused on an isolated SL1-SL2 domain ([Fig. 3.6]). These data support a model in which the SL1-SL2 domain does function autonomously and forms a compact tertiary structure in the final dimer. Moreover, reproducible hydroxyl radical cleavage protections are also observed in the 5’ GNRA tetraloop, PAL1, and PAL2 regions (see [Fig. 3.5]), while the flexible domain is completely solvent exposed.
Figure 3.5. Protection from solvent-based hydroxyl radical cleavages in the final dimer. (a) A quantitative histogram of hydroxyl radical protection was generated by subtracting the net cleavage intensities of the natively folded dimer from those observed at denaturing conditions. Strong and moderate cleavages are emphasized with black and gray bars, respectively. Asterisks indicate positions where RNA degradation prevented quantitative analysis. (b) Summary of observed protections superimposed on the secondary structure model for the final dimer. Small spheres indicate regions not analyzed.
3.3 Discussion

Using RNA SHAPE chemistry, I design a high resolution secondary structure model of the dimerization domain for the Moloney murine sarcoma virus. This model incorporates many features previously proposed to be important for retroviral genomic RNA dimerization. For example, since the SHAPE method provides robust single nucleotide resolution experimental information, I can now authoritatively define the termini of both the 10mer PAL1 and 16mer PAL2 heteroduplex. Previously, this was a structural detail that was unresolved in the literature.

In addition, SHAPE analysis also emphasizes the existence of flexible domains in retroviral dimerization sequences. Both MiDAS monomeric starting state and final dimer contain well-delimited unstructured, flexible domains that connect structured elements: the anchoring helix (nts 231-241/305-315) and PAL1 and PAL2 heteroduplexes, respectively (Figs. 3.1a, 3.3). Like in the monomer, flexibility in the dimer allows interconversion of multiple structures (see two stem loop and pseudoknot, Fig. 3.3). Thus, RNA flexibility appears to be a prominent feature in retroviral biology that is probably used to facilitate RNA dimerization by reducing the energetic costs for breaking and forming base-pairs during dimerization.

I used solvent-based hydroxyl radical footprinting to determine the solvent-accessibility of RNA elements within the dimerization domain. The flexible domain in the final dimer is completely solvent accessible as shown in solvent-based hydroxyl radical footprinting (see region 221-278, Fig. 3.5). Interestingly, I observe reproducible hydroxyl radical cleavage protection in PAL1, its flanking 5’ stem loop (nts 198-207), and in PAL2 (see Fig. 3.5b). The existence of protection patterns on both sides of the PAL2 duplex
suggests that this duplex is almost completely inaccessible to cleavage and must be buried in the MiDAS dimer. Thus, PAL2 must lie sandwiched between the other RNA elements.

As discussed in Chapter 4, a three-dimensional structure for the isolated SL1-SL2 (Fig. 3.6) domain was experimentally refined. Strikingly, a careful comparison of solvent inaccessible regions in the isolated SL1-SL2 domain versus the SL1-SL2 in the entire MiDAS RNA yields additional strong and moderate hydroxyl radical protections at the base of SL1 (nts 312-315, 351-353) and in the middle of the SL1 (nts 321-323, 343) (see Fig. 3.6). The conservation of the protection pattern for the SL1-SL2 domain in the context of MiDAS, as compared to its isolated version, suggests that this domain does fold independently as discussed in Chapter 4. Interestingly, the full length MiDAS dimer-specific extra protections lie on one face (side) of the SL1-SL2 RNA region (see asterisks and daggers, respectively; Fig. 3.7). The extensive protection pattern (both sides are protected!) of PAL2 and the additional protection sites (asterisks and daggers, respectively; Fig. 3.7) observed in the SL1-SL2 domain suggest that, after folding, the SL1-SL2 domain forms long-range tertiary structure interactions with PAL2.

Retroviral dimerization domains have almost universally been analyzed in terms of their constituent secondary structure elements. However, as judged by both site-directed (Fig. 3.4) and solvent-based (Fig. 3.5) hydroxyl radical cleavage experiments, the MiDAS RNA expectedly folds to a true three-dimensional dimeric shape. Thus, I conclude that the dimerization domain for this gamma retrovirus, and likely for many other retroviruses, functions not as a collection of linked stem-loop and helical structures as previously perceived, but instead as a three-dimensional entity.
Figure 3.6. Comparison of hydroxyl radical protection for the isolated SL1-SL2 domain (Chapter 4) and for this domain in the context of the intact MiDAS RNA in the final dimer state. Strong and moderate protections are emphasized in black and grey, respectively. Nucleotides not analyzed are shown as small spheres.
Figure 3.7. Architecture of a gamma retroviral genomic RNA dimer. The two RNA strands in the dimer are colored dark gray and white, respectively. PAL1 and PAL2 extended duplexes are shown as A-form helices; the refined structure of the SL1-SL2 domain is discussed in Chapter 4. Protected regions that are solvent-inaccessible to hydroxyl radicals are emphasized with a red backbone; * and † symbols indicate additional positions in the SL1-SL2 domain that are selectively protected in the context of the full length MiDAS RNA.
3.4 Materials and Methods

3.4.1 Retroviral RNA Transcripts. DNA templates for in vitro transcription of the MiDAS RNA construct that contained flanking 5' and 3' extensions of viral sequence of 30 and 5 nucleotides, respectively, were generated by PCR from the pLNBS\textsuperscript{20,21} plasmid using same conditions previously discussed in Chapter 2. RNA was generated using T7 RNA polymerase-mediated transcription (500 µL; 37 °C, 5 h) containing 80 mM Hepes (pH 7.4), 40 mM dithiothreitol (DTT), 0.01% (v/v) Triton X-100, 2 mM spermidine, 10 mM MgCl\textsubscript{2}, 2 mM each nucleoside triphosphate, ~ 25 µg of PCR-generated template, 20 U of SUPERase-In (Ambion), and 0.1 mg/mL polymerase. RNA was purified by denaturing gel electrophoresis (5% polyacrylamide, 7 M urea), excised from gel, eluted overnight into 1/2× TBE (45 mM Tris-borate, 1 mM EDTA), and concentrated by ethanol precipitation. RNA was stored in 10 mM Hepes (pH 7.5), 1 mM EDTA at -20 °C.

3.4.2 SHAPE Analysis of MiDAS Monomer and Dimer. NMIA modification, primer extension, and band quantification steps for both monomer and dimer were performed exactly as described previously in Chapter 2. The only added step was that after adding the dimer buffer and prior to adding NMIA, heat the dimer samples for 30 min at 60 °C to create the dimer species. The MiDAS RNA construct used in these experiments also contained a nonviral RNA cassette at its 3' end to facilitate analysis of the entire sequence by primer extension\textsuperscript{14}. The raw net NMIA SHAPE reactivity for both monomer and dimer was then fitted with the drop off equation (see Appendix A.1.) and the observed “drop off” in SHAPE reactivity was adjusted accordingly. The adjusted SHAPE reactivity data for both monomer and dimer were normalized at nt 393 in the UUCG tetraloop of the non-viral RNA cassette for comparative purposes (Fig. 3.2b).
3.4.3 205-BABE. A 170 nt viral RNA (nts 310-374) was transcribed in vitro in the presence of 10 mM guanosine monophosphorothioate (GMPS, Biolog Life Science Institute); GMPS is incorporated only at the initiating 5' G during transcription. 50 pmol of this RNA was treated with excess (50 µmol) bromoacetamidobenzyl-EDTA (BABE, Dojindo Labs) and (50 µmol) (NH₄)₂Fe(SO₄)₂ in 10 µl 0.15 M potassium phosphate (pH 8.0) at 37 °C for 1.5 h, ethanol precipitated, and resuspended in 10 µl 0.15 M potassium phosphate (pH 8.0).

3.4.4 Site-directed Hydroxyl Radical Cleavage. MiDAS RNA construct was initially 3'-end labeled using 600U poly(A) polymerase (600U/µl, USB) and [α-³²P] cordecypin-5'-triphosphate for 1 h at 25°C in 1x Poly A USB buffer. 3'-[³²P]-labeled MiDAS RNA construct (0.4 pmol) was then mixed with 2 pmol unlabeled 205-BABE in 5.25 µl water, heated for 3 min at 90 °C, snap-cooled on ice, mixed with 1.75 µl 4× dimer buffer [200 mM HEPES (pH 7.5), 800 mM potassium acetate (pH 7.5), 20 mM MgCl₂], and incubated for 30 minutes at 60 °C. Hydroxyl radical cleavage was initiated by adding hydrogen peroxide and ascorbic acid (in 3 µl) to final concentrations of 0.1% and 2.5 mM, respectively. Reactions (5 min at 25 °C) were quenched by adding 11 µl stop solution [0.1 M thiourea, 73% (v/v) formamide, 81 mM Tris-borate, 1.8 mM EDTA, and marker dyes]. Sites of cleavage were resolved by denaturing gel electrophoresis (8 and 12% polyacrylamide/7 M urea) and intensities for individual bands were obtained by integration. An important mock control, in which an RNA that did not contain a 5'-thiol group was also derivatized with BABE, showed that the observed cross-strand RNA cleavages were only due to the tethered Fe(II) moiety, rather than any free Fe(II) or Fe(II)-BABE complex that may have carried over from the ethanol precipitation.
3.4.5 **Solvent-based Hydroxyl Radical Cleavage.** 1 pmol of 5'-[\(^{32}\)P]-labeled SL1-SL2 RNA in 4.5 µl water was heated for 3 min at 90 °C, snap-cooled on ice, mixed with 1.5 µl 4× dimer buffer [200 mM HEPES (pH 7.5), 800 mM potassium acetate (pH 7.5), 20 mM MgCl\(_2\)], and incubated for 30 minutes at 60 °C to create the dimer species. Hydroxyl radical cleavage was then performed by adding (in 4 µl) final concentrations of 3 mM (NH\(_4\))\(_2\)Fe(SO\(_4\))\(_2\), 4.5 mM EDTA, 6 mM sodium ascorbate, and 15 mM DTT and reactions (1 hr at 25 °C) were quenched by addition of stop solution [0.1 M thiourea, 73% (v/v) formamide, 81 mM Tris-borate, 1.8 mM EDTA, and bromophenol blue and xylene cyanol dyes]. Cleavage products were resolved by denaturing electrophoresis (8 and 12% polyacrylamide/7 M urea) and protection at individual nucleotides in the MiDAS RNA was quantified\(^{15}\) and band intensity observed under native conditions was subtracted from the intensity obtained for the denatured control.

3.4.6 **Modeling and Refinement of Three-Dimensional Model.** The A-form helices of both PAL1 and PAL2 duplexes were built in Sybyl (Tripos, Inc). Construction and refinement of the SL1-SL2 domain is discussed in Chapter 4. The final model picture was created using PyMOL\(^{22}\).
3.5 References


71


Chapter 4

An RNA Switch Enforces Stringent Retroviral Genomic RNA Dimerization
4.1 Introduction

Retroviral genomes consist of two sense-strand unspliced RNAs that are non-covalently linked near their 5' ends to form a dimeric structure. Recognition of this dimeric state appears to be the mechanism by which retroviruses ensure that exactly two RNA genomes are encapsulated into each nascent virion. Mature retroviral virions contain almost exclusively retroviral genomic RNA plus a few select cellular RNAs. Specific recognition of retroviral genomic RNA against a background of cellular and spliced viral RNA thus represents a striking example of molecular recognition in biology.

In Chapter 2, I identified and discussed a minimal dimerization active sequence (MiDAS) for a representative gamma retrovirus, the Moloney murine sarcoma virus (MuSV; see Fig. 1a). The MiDAS domain includes conserved sequence and structural elements previously proposed to be important for RNA dimerization. These include self-complementary (palindromic) sequences (PAL1 and PAL2) and stem-loop structures 1 and 2 (SL1 and SL2). SL1 and SL2 contain GACG tetraloops that form loop-loop interactions with a second RNA molecule via canonical intermolecular C-G base pairing. The MiDAS domain correlates closely with gamma retroviral genomic sequences shown independently to be sufficient to package heterologous RNAs, as dimers, into virions. An unanticipated feature of the monomeric secondary structure for the MiDAS domain is that the PAL2 sequences lies in a larger structural element in which the PAL2 sequence is, on average, highly flexible (Fig. 4.1a). This flexible region extends from a stable anchoring helix (spanning nts 231-241 and 305-315, Fig. 4.1a).

The non-covalent interactions that specify the final dimer structure include base pairing interactions comprised of PAL1-PAL1' and PAL2-PAL2' extended
Figure 4.1. Conformational switch in the SL1-SL2 domain during retroviral RNA dimerization, defined by RNA SHAPE chemistry. (a) Structure of the minimum dimerization active sequence (MiDAS) for the Moloney murine sarcoma virus in the monomeric starting state. RNA sequences that contribute to the SL1-SL2 conformational change are shown explicitly; other MiDAS structures are represented with a gray line. (b) SL1-SL2 domain conformation in the final dimer state. Regions in the RNA that undergo significant conformational changes are emphasized with colored symbols or boxes. Dashed line (light blue) illustrates base pairing with PAL2 sequences in a second RNA in the dimer.
heteroduplexes between the two strands in the dimer (refs. 8-11 and Chapter 3). However, simple duplex formation typically exhibits poor sequence selectivity because duplexes containing mismatches readily compete with formation of thermodynamically more stable perfect duplexes\textsuperscript{17,18}. Consistent with this view, small changes to the sequences in PAL1 and PAL2 have almost no effect on RNA propagation in cell culture\textsuperscript{19,20}. What nucleic acid structures then allow retroviral RNA genomes to dimerize and be packaged into nascent virions with such exquisite selectivity?

High resolution RNA structure mapping using RNA SHAPE chemistry\textsuperscript{21,22} shows that dimerization is accompanied by a conformational change in the SL1-SL2 region during which SL1 extends by four base pairs at its base (boxed nts, \textbf{Fig. 4.1b}). I show that the SL1-SL2 region within the MiDAS domain operates as a functional unit and forms stable loop-loop interactions independently of other RNA elements. In the monomeric conformation, SL1-SL2 forms a lower affinity, but very stringent, complex with the SL1-SL2 element from a second RNA. This initial loop-loop interaction is converted to a high affinity, but less selective, interaction in the final dimer state. A refined three-dimensional model for the SL1-SL2 domain shows that high affinity interactions in the dimer are mediated by tertiary interactions between two interdigitated and U-shaped SL1-SL2 domains.

\textbf{4.2 Results}

\textbf{4.2.1 The SL1-SL2 Domain Undergoes a Conformational Switch Upon RNA Dimerization.} The structure of the monomeric RNA starting state for the MuSV MiDAS was determined previously using RNA SHAPE chemistry\textsuperscript{7} (\textbf{Fig. 4.1a}) and is discussed in Chapter 2. In an RNA SHAPE analysis, local nucleotide flexibility is monitored at every
position in an RNA via structure-selective reaction with an anhydride reagent (NMIA) at the ribose 2'-hydroxyl position. Flexible and single stranded nucleotides react efficiently; whereas, positions constrained by base pairing or tertiary interactions are unreactive\textsuperscript{21,22}.

I have also used SHAPE to monitor the structure of the final dimer state by allowing the monomeric MiDAS RNA to dimerize by incubating the RNA at 60 °C for 30 min. RNAs in the final dimer state and, for comparison, in the monomer-like starting state were subjected to SHAPE chemistry and the resulting 2'-O-adducts were identified by primer extension and sequencing gel analysis (Fig. 4.2a). Inspection of the (+) NMIA reactions for the initial monomer (M) and final dimer (D) states shows that multiple regions in the MiDAS domain undergo a conformational change upon dimerization (colored bars, Fig. 4.2a). Structural changes are highlighted using a consistent coloring scheme in Figs. 1 and 2.

Absolute 2'-hydroxyl SHAPE chemistry reactivities were quantified for every position in both the starting monomer-like and final dimer states (Fig. 4.2b). Nucleotides in PAL2 are reactive in the monomeric starting state (light blue, Figs. 4.1a, top panel of 4.2b). Upon dimerization, reactivity in PAL2 decreases dramatically, consistent with formation of an extended 16 base pair duplex in the dimer state (Figs. 4.1b and 4.2b).

In addition to the expected strong decrease in reactivity in the PAL2 sequence, SHAPE analysis identifies structural changes in several other MiDAS regions of the SL1-SL2 domain (green, orange, dark blue, and red symbols, Figs. 4.1 and 4.2). NMIA reactivity increases in the stem that anchors the flexible domain at nts 305-309, in the AC bulge (nts 314-315), and at the U319 bulge (compare green, orange and dark blue symbols, upper and middle panels of Fig. 4.2b). These changes in SHAPE reactivity are consistent with a
Figure 4.2. SHAPE analysis of the MiDAS RNA in starting monomer-like (M) and final dimer (D) conformations and of a simplified SL1-SL2 domain RNA in the final dimer state (D). (a) 2'-O-Adduct formation upon addition of NMIA (+) detected by primer extension; (−) control reactions omitting NMIA. Sequencing lanes (SEQ) were generated by dideoxy cytosine nucleotide incorporation; nucleotide positions are labeled with respect to NMIA lanes. (b) Quantitative histograms for NMIA reactivity. Columns are colored using the scheme shown in Fig. 1. Column heights report band intensities in the (+) NMIA reactions minus those for (−) NMIA.
conformational change such that the SL1 stem-loop structure becomes extended by four base pairs during dimerization (see boxed nts, Fig. 4.1).

To confirm this secondary structure assignment for the SL1-SL2 domain in the final dimer state, I analyzed the secondary structure of an RNA spanning nts 295-381. This simplified RNA is missing most of the flexible domain and is only capable of forming the SL1-SL2 domain structure found in the final dimer state (Fig. 4.1b). SHAPE reactivity in the simplified SL1-SL2 domain construct is essentially identical to this region in the complete MiDAS domain in the final dimer state (compare middle and lower panels, Fig. 4.2). Notably, the structural similarity includes the fine-scale reactivity characteristic of the 298-310, 314-319 and 328-361 regions. Thus, this construct accurately represents the conformation of the SL1-SL2 domain in the final dimer state. Moreover, the complete SHAPE analyses of the monomer, final dimer, and SL1-SL2 domain RNA indicate that the SL1-SL2 domain changes conformation upon dimerization (Fig. 4.1).

4.2.2 The SL1-SL2 Domain is an Autonomous Dimerization Motif. RNA stem-loops containing a GACG sequence at their apex form stable loop-loop interactions via canonical C-G base pairs mediated by the final two nts in each loop13 (Fig. 4.3). If the GACG tetraloops can form loop-loop interactions in the context of the SL1-SL2 domain, this implies that there are two distinct dimer states: a loop-loop dimer state in which the RNA is in the starting monomer-like conformation (Fig. 4.4a) and a loop-loop state that characterizes the final dimer (Fig. 4.4d).
Figure 4.3. Structure of the loop-loop interaction formed between GACG sequences\textsuperscript{13}. Nucleotides in loop positions 3 and 4 that form canonical C-G base pairs are shown explicitly with their intermolecular hydrogen bonds; other nucleotides are represented as cylinders.
Figure 4.4. Dimerization specificity of the SL1-SL2 domain in monomer-like versus final dimer conformations. (a, d) Monomer and dimer states for the SL1-SL2 domain. Monomer contains a single inverted base pair to facilitate transcription (open letters). Control experiments show this base pair change does not affect dimerization (data not shown). (b, e) Native gel analysis of RNA dimerization for wild type (CG/CG) and representative mutant sequences in monomer-like or final dimer conformations. (c, f) Binding curves for RNA dimerization in the monomer-like and final dimer states.
In order to understand the contribution of these loop-loop interactions for the monomer-to-dimer conformational change, I measured dimerization affinities for RNAs that were constrained to be in conformations observed in either the starting or final dimer states (Figs. 4.4a,d). I evaluated loop-loop dimerization affinities for native sequence RNAs and also for mutants containing changes in the CG sequence that mediates intermolecular base pairs in the loop-loop interaction. RNA constructs are identified by their sequence in the cross-strand base pairing positions. Thus, the native sequence RNAs are termed CG/C\textsubscript{G} and the mutants are GG/GG, GG/CC, CC/GG, CC/CC, AA/CG, CG/AA, and AA/AA, respectively. Dimers were formed in a roughly physiological environment (pH 7.5, 200 mM potassium acetate and 5 mM MgCl\textsubscript{2}) and resolved in non-denaturing gels supplemented with 5 mM MgCl\textsubscript{2}.

Strikingly, in the context of the monomer conformation, the native SL1-SL2 domain forms a high affinity complex, resolved as a well-defined species by non-denaturing gel electrophoresis ($K_{\text{dimer}} = 11$ nM; see CG/C\textsubscript{G} panel in Fig. 4.4b, red symbols in Fig. 4.4c). Of the tested mutants, only the variant in which the CG sequence in SL1 was changed to AA (to form a GNRA-type tetraloop\textsuperscript{23}) forms a detectable dimeric complex ($K_{\text{dimer}} = 51$ nM; Figs. 4.4b,c). This dimer appears to be relatively unstable because it partially dissociates during gel electrophoresis (AA/CG panel, Fig. 4.4b). Mutating the loops in SL1 and SL2 to all other tested sequences eliminates detectable dimerization, including for variants like CC/GG and GG/CC that, in principle, could maintain Watson-Crick-type base pairing interactions. Thus, in the monomer-like (initial) state, the SL1-SL2 domain dimerizes stringently and strongly discriminates against most non-native sequences.
I then performed comparable experiments using our SL1-SL2 domain construct that folds to the global conformation seen in the final dimer state (Figs. 4.2, 4.4d). The native (CG/CG) sequence dimerizes with high affinity ($K_{dimer} = 0.3$ nM) or ~40-fold more tightly than in the monomer state. Control experiments showed that removing the 5' and 3' flanking sequences (nts 295-309 and 375-381 had no effect on $K_{dimer}$ (data not shown); therefore, RNA mutants were constructed in the context of the RNA shown in Fig. 4.4d. Of the SL1-SL2 domain sequence variants tested, many dimerized with high affinities (Table 4.1; Figs. 4.4e,f). For example, both AA/CG and CG/AA variants have dimerization affinities comparable to wild type. Other sequence variants have lower, but still quite high, dimerization affinities (Fig. 4.4f; Table 4.1).

<table>
<thead>
<tr>
<th>SL1-SL2 sequence</th>
<th>Monomer-like</th>
<th>Final dimer</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG/CN (native)</td>
<td>11</td>
<td>0.3</td>
</tr>
<tr>
<td>AA/CG</td>
<td>51</td>
<td>0.1</td>
</tr>
<tr>
<td>CG/AA</td>
<td>&gt;&gt;1000</td>
<td>0.5</td>
</tr>
<tr>
<td>GG/CC</td>
<td>&gt;&gt;1000</td>
<td>1.2</td>
</tr>
<tr>
<td>CC/GG</td>
<td>&gt;&gt;1000</td>
<td>3.4</td>
</tr>
<tr>
<td>GG/GG</td>
<td>&gt;&gt;1000</td>
<td>12 (9.2)</td>
</tr>
<tr>
<td>CC/CC</td>
<td>&gt;&gt;1000</td>
<td>21 (13)</td>
</tr>
<tr>
<td>AA/AA</td>
<td>&gt;&gt;1000</td>
<td>16 (29)</td>
</tr>
</tbody>
</table>
Thus, the SL1-SL2 region functions as an autonomous domain that dimerizes in two distinct modes (Figs. 4.4c, f). In the monomer-like starting state conformation, loop-loop mediated dimerization is highly selective for the native sequence. Upon rearranging to the final dimer state conformation, dimerization affinity increases, but now shows much lower stringency (compare Figs. 4.4c with 4.4f).

4.2.3 Architecture of SL1-SL2 Domain. Because the SL1-SL2 domain contains two GACG RNA tetraloops, additional experiments are required to establish the interaction partners for SL1 and SL2 in one RNA with their tetraloop partners from a second RNA (see Fig. 4.3). I used site-directed hydroxyl radical cleavage experiments\textsuperscript{24,25} (i) to identify the orientation and pairing partners for loops in SL1 and SL2 and (ii) to obtain constraints for the global architecture of the SL1-SL2 domain in the monomer-like and final dimer states. I engineered two 65 nt RNAs (spanning nts 310-374), confined in the final dimer state and derivatized with Fe(II)-EDTA groups at defined positions (in gray, Fig. 5). The 310-BABE construct contained a bromoacetamidobenzyl-EDTA-Fe(II) (BABE) adduct at the 5' end of the RNA. The 336-ITE construct was derivatized with isothiocyanobenzyl-EDTA-Fe(II) (ITE) via a 2'-amino nucleotide at position 336. In the presence of hydrogen peroxide and ascorbic acid reagents, highly reactive hydroxyl radicals are generated at the Fe(II)-EDTA moiety that cleave proximal positions in the RNA backbone.

The 310-BABE or 336-ITE constructs were allowed to form loop-loop dimers with a second 5'-\textsuperscript{32P}-labeled SL1-SL2 domain RNA (radiolabels are indicated with asterisks in Fig. 5). By this scheme, only intermolecular cross-strand RNA cleavages are
Figure 4.5. 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 nt 310. (b) Fe(II)-ITE (closed circle) cleavage from nt 336. Small spheres indicate RNA regions that were not monitored; position of 5'-radiolabel on second RNA strand is indicated by an asterisk.
detected. Dimerization affinities for the Fe(II)-EDTA-containing complexes were ~ 0.4 nM or identical, within error, to underivatized complexes. Cross-strand RNA cleavage products were resolved on sequencing gels. Net band intensities over background were obtained for almost every position in each RNA construct by subtracting the background observed for mock conjugated RNAs lacking the 5'-thiol attachment group or omitting addition of the BABE or ITE reagents during the derivatization step. Cleavages greater than 3 or greater than 5 times background were judged to be moderate and strong, respectively (orange and red bars, Fig. 4.5).

For the 310-BABE construct, strong cleavage was obtained only at the apex of SL1' (red symbols, Fig. 4.5a). In contrast, with the 336-ITE construct, strong cleavage was observed in both SL1' and SL2' (Fig. 4.5b). These cleavage data strongly support models in which SL1 forms a loop-loop interaction with SL2' and SL2 interacts with SL1' (see secondary structure models, Fig. 4.5).

Finally, to evaluate structural changes in the SL1-SL2 domain that occur during dimerization, I performed site-directed cleavage experiments using our RNA construct that imitates the SL1-SL2 domain structure in the initial monomer-like starting state (see Fig. 4.4a). This RNA was derivatized with BABE at its 5' end, at nt 316 (Fig. 4.6). For this RNA, no strong cross-strand cleavages were detected; instead, moderate cleavages of approximately equal intensity were observed in both SL1' and SL2' apical loops (Fig. 4.6). These experiments support a model for the monomer-like starting state in which loop-loop interactions still form but that, overall, the structure is looser than in the final dimer state. A looser structure would also contribute to the 40-fold weaker dimerization affinities observed for the monomer-like state (Figs. 4.4c,f).
Figure 4.6. Global architecture of the SL1-SL2 interaction in the monomer-like conformation mapped by Fe(II)-BABE mediated (open circle) site-directed hydroxyl radical footprinting. Yellow and black bars in histogram indicate absolute Fe(II)-BABE reactivities in the monomer-like conformation. For comparison, cleavages observed for the analogous experiment performed with the SL1-SL2 domain in the final dimer state (from Fig. 4.5a) are shown with light gray bars. Position of 5' radiolabel is indicated with an asterisk.
4.2.4 Tertiary Structure in the SL1-SL2 Domain. I then used solvent-based hydroxyl radical footprinting to map tertiary contacts in the SL1-SL2 domain in the final dimer state. 5'-end-labeled SL1-SL2 domain RNA (Fig. 4.4d) was incubated under native conditions that yield loop-loop dimers and subjected to solvent-based hydroxyl radical cleavage using untethered Fe(II)-EDTA (Fig. 4.7a). Regions protected from cleavage in the final dimer state were identified by comparing cleavage patterns for the same RNA incubated in water (den, denaturing conditions with no mono or divalent cations). Under conditions that support loop-loop interactions, multiple regions in the SL1-SL2 domain are strongly protected from hydroxyl radical cleavage (black bars, Fig. 4.7a). Strong and moderate protections are emphasized in black and gray (Figs. 4.7b,c). Solvent inaccessible regions occur at the apex of the SL1 and SL2 stems and lie adjacent to one another in the secondary structure model for the SL1-SL2 domain in the final dimer state (Fig. 4.7c). Reassuringly, solvent accessible regions are compatible with the site-directed cleavage information obtained using the 310-BABE and 336-ITE constructs because strong and moderate site-directed cleavage occurs in regions that are accessible to the untethered Fe(II)-EDTA probe (compare Figs. 4.5 and 4.7c).
Figure 4.7. Solvent-based hydroxyl radical footprinting of the SL1-SL2 domain in the final dimer conformation. (a) Representative hydroxyl radical footprinting data visualized in a sequencing gel. D, final dimer state; den, denaturing control. Strong sites of protection are emphasized with black bars. Sequence lanes (G,U) were generated by iodine cleavage of phosphorothioate-substituted RNA. (b) Quantitative histogram of hydroxyl radical protection. Strong and moderate cleavages are emphasized with black and gray bars. Asterisks indicate positions where band compression prevented quantitative analysis. (c) Hydroxyl radical protection (black and gray) superimposed on an RNA secondary structure model for the SL1-SL2 domain in the final dimer state.
4.3 Discussion

RNA SHAPE chemistry (Fig. 4.2) and site-directed cleavage (Fig. 4.5) experiments provide a wealth of constraints that can be used to refine a three-dimensional model for the SL1-SL2 domain in the final high affinity dimer state. In collaboration with Costin M. Gherghe, structures of the final dimer state were obtained using distance-constraints-based algorithms, analogous to those employed to refine structures using information from nuclear magnetic resonance experiments. Canonical hydrogen bonding and planarity constraints were imposed for base pairs predicted using a thermodynamic RNA folding algorithm, constrained by experimental SHAPE reactivities (Figs. 4.1b and 4.2). Base pairing was also enforced between C-G pairs in the GACG tetraloops. The site-directed hydroxyl radical cleavage experiments yielded 31 long-range intermolecular constraints between the two RNA strands that comprise the dimer (Fig. 4.5). Because the SL1-SL2 domain is a symmetrical dimer, the total number of constraints used to constrain the refinement is doubled, to 62 total.

The refined model represents the first experimentally constrained structure for any retroviral dimer component, larger than a single helix. Relative to the central-most solution, refined structure models converged to an average RMS deviation of ~4.1 Å over 128 phosphate positions (Fig. 4.8).

The SL1-SL2 domain in the final dimer forms a tightly packed tertiary structure in which SL1 is bent at the 338-341 bulge, such that each SL1-SL2 element in the dimer form a structure that is roughly U-shaped (see Fig. 4.8 and 4.9a). 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
Figure 4.8. Refined model of the SL1-SL2 domain in the final dimer state based on secondary structure and long-range site-directed cleavage restraints. (a) Stereo image of eight refined structures with lowest all-atom clash scores (£ 43). RNA backbones are shown as tubes; cross-strand G-C base pairs in the tetraloops are represented as black cylinders. (b, c) 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; and positions used as repelling constraints are shown in white next to gray arrowheads. RNAs derivatized with Fe(II)-EDTA groups are shown with gray backbones; attachment sites are emphasized with black spheres.
Figure 4.9. Refined model and two-step assembly of the SL1-SL2 domain. (a) 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. (b) Assembly of the high affinity SL1-SL2 dimer via a stringent loop-loop intermediate.
groove side of SL1 faces the minor groove of the SL2'; whereas, the major groove of SL1 fits snugly against the major groove side of SL1' in the second RNA (Fig. 4.8 and 4.9a).

Together, these experiments support a model for retroviral genomic dimerization in MuSV in which the SL1-SL2 domain plays a major role. Inter-genomic interactions form in two thermodynamic steps. In the first step, SL1-SL2 domain interactions in the monomer-like conformation are mediated primarily by simple loop-loop interactions involving canonical C-G pairs mediated by the GACG tetraloops (center panel, Fig. 4.9b). This complex has a relatively loose tertiary structure as judged by site-directed hydroxyl radical cleavage (Fig. 4.6). This interaction is highly stringent because only the native sequence or one closely related sequence forms stable interactions when the RNA is in this conformation (Fig. 4.4c).

During the dimerization reaction from the initial loop-loop pair, two PAL2 sequences eventually form an extended duplex (Fig. 4.1b) (refs. 10,11 and Chapter 3). Coincident with PAL2 duplex formation, the SL1-SL2 domain changes conformation to form a much higher affinity interaction (Fig. 4.4f). The change in secondary structure involves extension of SL1 by four base pairs and disruption of the 231-241/305-315 anchoring helix in the monomer-like conformation. Importantly, this conformational change allows extensive new tertiary interactions to form in the SL1-SL2 domain (right panel, Fig. 4.9b). Relative to the newly formed tertiary interactions, the initial SL1-SL2 loop-loop base pairs now comprise a smaller net fraction of the interactions in the dimer structure. In the final dimer state, RNAs with non-native sequences in the GACG tetraloops are also able to form high affinity complexes (see Fig. 4.4).
Dimerization of retroviral genomes appears to occur in the cytoplasm and prior to encapsidation into the immature viral particle\textsuperscript{30}. Dimerization must therefore exclude non-cognate interactions with the enormous background\textsuperscript{31,32} of nonviral RNAs in the cytoplasm. Simple formation of long, extended, PAL1 and PAL2 duplexes in the dimer state is unlikely to be sufficiently stringent to exclude many non-cognate interactions because duplex formation tends to allow mismatched and gapped helices to form with small energetic and kinetic penalties\textsuperscript{18}.

A suggested two-step mechanism (\textbf{Fig. 4.9b}) functions, in significant part, to resolve the specificity versus stringency problem in retroviral biology. At early stages of new viral RNA production in an infected cell, dimerization via loop-loop interactions alone is highly stringent, likely contributing to accurate dimerization and packaging. At late stages in the infectivity cycle, the interaction increases in affinity in the final dimer state, which may be important to prevent loss of one strand of the genomic RNA during packaging, reverse transcription or other viral processes. Analogous, multi-step formation of dimeric tertiary structures from simpler initial base pairing interactions may be broadly used to ensure stringent dimerization and packaging of retroviral genomes.
4.4 Materials and Methods

4.4.1 Retroviral RNA Constructs. DNA templates for in vitro transcription of the MiDAS, SL1-SL2 (final dimer), and SL1-SL2 (monomer) RNA constructs were generated by PCR from the pLNBS\textsuperscript{33} plasmid or synthetic oligonucleotide (Midland Certified) templates using same conditions discussed previously in Chapter 2. RNA constructs were generated by T7 RNA polymerase-mediated transcription [500 µl, 37 °C, 5 h; containing 80 mM HEPES (pH 7.4), 40 mM dithiothreitol (DTT), 0.01% (v/v) Triton X-100, 2 mM spermidine, 10 mM MgCl\textsubscript{2}, 2 mM each nucleoside triphosphate, ~25 µg PCR-generated template, 20 U SUPERase-In (Ambion) and 0.1 mg/ml polymerase]. Internally labeled RNAs were generated using 20 µCi α-[\textsuperscript{32}P]ATP and 0.5 mM unlabeled ATP. RNAs were purified by denaturing gel electrophoresis [5 and 15% (w/v) polyacrylamide, 7 M urea; for MiDAS and SL1-SL2 RNAs, respectively], excised from the gel, eluted overnight into 1/2× TBE (45 mM Tris-borate, 1 mM EDTA), and concentrated by ethanol precipitation. RNAs were stored in 10 mM HEPES (pH 7.5), 1 mM EDTA at -20 °C.

4.4.2 SHAPE Analysis of MiDAS and SL1-SL2 RNA. NMIA modification, primer extension and band quantification steps for the MiDAS (nts 175-381) were discussed previously in Chapter 2 and 3 \textsuperscript{17}. The isolated SL1-SL2 (nts 295-381) (10 pmol) was heated at 90 °C for 3 minutes in 7.2 µL water, cooled on ice, treated with 1.8 µL 5× dimerization buffer [250 mM Hepes (pH 8.0), 1 M potassium acetate (pH 7.5), 25 mM MgCl\textsubscript{2}], heated at 25 °C for 30 min to make dimer species and returned to ice. The RNA solution was then equilibrated at 37 °C for 5 min, treated with NMIA (1 µL, 180 mM in anhydrous DMSO), and allowed to react for 50 min (~ 5 half lives) at 37 °C. Control reactions contained DMSO without NMIA. The MiDAS and SL1-SL2 constructs used in these experiments also contain...
the nonviral RNA cassette at their 3' ends to facilitate analysis of the entire sequence by primer extension\textsuperscript{18}. SHAPE reactivity data were normalized to nt 393 in the UUCG tetraloop of the non-viral RNA cassette, because NMIA reactivity in this region is independent of viral RNA folding and dimerization.

4.4.3 Concentration-dependent Dimerization of SL1-SL2. Internally \([^{32}\text{P}]-labeled RNA (~ 0.1 \text{nM}) was combined with the same unlabeled RNA (at 0.15-640 \text{nM in 8 \text{\mu l}}). RNAs were denatured at 90 °C for 3 min, snap-cooled on ice for 30 sec, treated with 2 \text{\mu l} of 5x dimer buffer [250 mM HEPES (pH 7.5), 1 M potassium acetate (pH 7.5), 25 mM MgCl\textsubscript{2}; at 25 °C], incubated at 25 °C for 50 min, and placed on ice. Samples (3 \text{\mu l}) were mixed with 1 \text{\mu l} loading dye (30% glycerol, 5 mM MgCl\textsubscript{2}, 0.01% xylene cyanol and bromophenol blue) and resolved on non-denaturing gels (12% polyacrylamide, 5 mM MgCl\textsubscript{2}, 1x TBE) at 4 °C for 1 h using a running buffer that also contained 5 mM MgCl\textsubscript{2}. Gels were pre-run for 15 min prior to loading and the running buffer was re-equilibrated every 20 minutes to maintain a constant MgCl\textsubscript{2} concentration. Monomer and dimer species or visualized by phosphorimaging and the bimolecular dimerization dissociation constant \(K_{\text{dimer}}\) was obtained by fitting to the equation: fraction RNA dimer = \(\left(\frac{A}{4C_T}\right)^2 + \frac{8K_{\text{dimer}}^2 C_T}{K_{\text{dimer}}^2 + 8K_{\text{dimer}} C_T}\), where \(A\) is the fraction dimer at saturating RNA concentrations and \(C_T\) is the total concentration of RNA (see Appendix A.2.).

4.4.4 310-BABE and 336 ITE RNAs. A 65 nt RNA (nts 310-374) was transcribed \textit{in vitro} in the presence of 10 mM guanosine monophosphorothioate (GMPS, Dharmacon); GMPS is incorporated only at the initiating 5' G during transcription. 50 pmol of this RNA were treated with an excess (50 \mu mol) of bromoacetamidobenzyl-EDTA (BABE, Dojindo Labs) in 10 \mu l of 0.05 M potassium phosphate (pH 8.0) at 37 °C for 1 h, ethanol precipitated,
and resuspended in 10 µl of 0.15 M potassium phosphate (pH 8.0). This solution (8 µL) was incubated with 2 µl 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). The precursor RNA was synthesized chemically (Dharmacon) with a 2'-NH₂ nucleotide at position U336. 30 pmol of this RNA was 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 U SUPERase-In] at 37 °C overnight, ethanol precipitated, resuspended in 10 µl of 0.15 M potassium phosphate (pH 8.0), and allowed to form a complex with Fe²⁺ as described above.

4.4.5 Site-directed Hydroxyl Radical Cleavage. 0.1 pmol 5'-[³²P]-labeled SL1-SL2 RNA (Fig. 4d) was mixed with 1 pmol unlabeled 310-BABE or 336-ITE RNA in 5.25 µl water, heated for 3 min at 90 °C, snap-cooled on ice, mixed with 1.75 µl 4× dimer buffer [200 mM HEPES (pH 7.5), 800 mM potassium acetate (pH 7.5), 20 mM MgCl₂], and incubated for 30 minutes at 25 °C to form the heterodimer. Hydroxyl radical cleavage was initiated by adding hydrogen peroxide and ascorbic acid (in 3 µl) to final concentrations of 0.1% and 2.5 mM, respectively. Reactions (4 min at 25 °C) were quenched by adding 11 µl stop solution [0.1 M thiourea, 73% (v/v) formamide, 81 mM Tris-borate, 1.8 mM EDTA, and marker dyes]. Sites of cleavage were resolved by denaturing gel electrophoresis (15% polyacrylamide/7 M urea) and intensities for individual bands were obtained by integration²⁷.

4.4.6 Solvent-based Hydroxyl Radical Cleavage. 0.5 pmol of 5'-[³²P]-labeled SL1-SL2 RNA in 5.25 µl water was heated for 3 min at 90 °C, snap-cooled on ice, mixed with 1.75 µl 4× dimer buffer [200 mM HEPES (pH 7.5), 800 mM potassium acetate (pH 7.5), 20 mM MgCl₂], and incubated for 40 minutes at 25 °C to create the dimer species. Hydroxyl radical cleavage was then performed by adding (in 3 µl) final concentrations of 3 mM
(NH₄)₂Fe(SO₄)₂, 4.5 mM EDTA, 6 mM sodium ascorbate, and 15 mM DTT and reactions (1 hr at 25 °C) were quenched by addition of stop solution [0.1 M thiourea, 73% (v/v) formamide, 81 mM Tris-borate, 1.8 mM EDTA, and bromophenol blue and xylene cyanol dyes]. Cleavage products were resolved by denaturing electrophoresis (15% polyacrylamide/7 M urea) and net cleavage efficiency was quantified by subtracting the background observed in the absence of Fe²⁺.

4.4.7 Model Building and Refinement. Costin M. Gherghe refined three-dimensional models using simulated annealing and molecular mechanics computations, analogous to routines used to develop macromolecular structure models from NMR information²⁶,²⁷. The general refinement routine is based on scripts generously provided to us by J. Davis and S. Butcher²⁸ and additionally incorporates the optimized force field parameters developed by Moore and colleagues³⁵. Two classes of constraints were used during refinement. (1) Intramolecular constraints. Planarity and hydrogen-bonding constraints were imposed at all base paired positions; constraints involving idealized phosphate-phosphate and intra- and inter-ribose 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 nts 323 to 324 and 338 to 342 were constrained to be 5–7 and 16–18 Å, respectively. (2) Intermolecular constraints. Base pairing was enforced at the G-C pairs in the tetraloops; strong and medium site-directed cleavages (Fig. 5) 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, using square-well potentials. Regions that were both unreactive by site-directed cleavage and also solvent accessible as judged by cleavage using free Fe(II)-EDTA were used as repelling interactions and were
constrained to be $\geq 30\,\text{Å}$ from the tethered Fe(II)-EDTA group. Experiments with the 310-BABE construct yielded 3 strong, 4 moderate and 3 repelling (to positions 312, 340 and 358) constraints; 336-ITE yielded 6 strong, 11 moderate, and 4 repelling (to positions 313, 340, 353 and 374) constraints; total number of constraints for the dimer was therefore 62.

Refinement was performed in two steps. (1) Starting from an initial structure in which both RNA monomers were in an extended conformation\textsuperscript{28}, CNS 1.1 \textsuperscript{26} was used initially to generate 75 independent, globally folded sets of monomers using intramolecular constraints only; overall geometries were then improved using Xplor-NIH\textsuperscript{27}. (2) These 75 structures were then folded to candidate dimer structures in CNS, using both intramolecular and intermolecular constraints; local interactions were again optimized 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\textsuperscript{36,37}. Analysis is based on the eight refined structures with clash scores less than 43. These eight structures superimpose\textsuperscript{38} with RMS deviations of 3.5–5.0 (4.1 average) Å over 128 phosphate positions in the dimer. Structures were analyzed and figures were composed using PyMOL\textsuperscript{39}.
4.5 References


Appendix A
A.1. Using the Probability Drop Off Equation

\[ y = Ap^n + C, \] where \( A \) is maximum amplitude, \( p \) is probability, and \( n \) is the series # in kaleidagraph terms: \[ y = m2*(m1^m0)+340, \] where \( m1=0.99 \), \( m2=6300 \)

<table>
<thead>
<tr>
<th>fitting drop off</th>
<th>Monomer</th>
<th>Dimer</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p ) (m1)</td>
<td>0.99</td>
<td>0.993</td>
<td>0.992</td>
</tr>
<tr>
<td>( A ) (m2)</td>
<td>2500</td>
<td>3400</td>
<td>2950</td>
</tr>
<tr>
<td>( C )</td>
<td>340</td>
<td>340</td>
<td>340</td>
</tr>
</tbody>
</table>

keep in mind that at \( n=0 \), \( A=3290 \)
Next, I used equation below to adjust the SHAPE reactivity in monomer and dimer.

\[
CorrectedSHAPE\# = \frac{3290}{(2950(0.992^x) + 340)}
\]

Or in kaleidagraph terms:
\[\text{column0} = \text{column1} \times \frac{3290}{(2950(0.992^\text{seriescolumn}) + 340)}\]

Final Result:
A.2. Derivation of Fraction RNA Dimer Equation

\[ 2M \xrightarrow{\text{\hspace{1cm} \longleftrightarrow \hspace{1cm}}} D \]

\[ K_{\text{dimer}} = \frac{[M]^2}{[D]} \]

\[ [D] = \frac{F_D C_T}{2} \], where \( F_D \) is the fraction RNA dimer and \( C_T \) is the total concentration of RNA

\[ [M] = (1 - F_D) C_T \]

\[ K_{\text{dimer}} = \frac{2(1 - F_D)^2 C_T}{F_D} \]

\[ 0 = -K_{\text{dimer}} F_D + 2C_T (1 - 2F_D + F_D^2) \]

\[ 0 = 2C_T F_D^2 - (K_{\text{dimer}} + 4C_T) F_D + 2C_T \]

use quadratic equation, where \( a = 2C_T \), \( b = (K_{\text{dimer}} + 4C_T) \), and \( c = 2C_T \)

\[ F_D = \frac{(K_{\text{dimer}} + 4C_T) - \sqrt{K_{\text{dimer}}^2 + 8K_{\text{dimer}} C_T + 16C_T - 4(4C_T)}}{4C_T} \]

\[ F_D = \left( A / 4C_T \right) \left[ (K_{\text{dimer}} + 4C_T) - \sqrt{K_{\text{dimer}}^2 + 8K_{\text{dimer}} C_T} \right] \], where \( A \) is the fraction dimer at saturating RNA concentrations